Fast and accurate quality assessment of Raf tomatoes using NIRS technology

Irina Torres[a], Dolores Pérez-Marín[b,*], María-José De la Haba[a], María-Teresa Sánchez[a,*]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Corresponding authors. Tel.: +34 957 212576; fax: 34 957 212000

E-mail addresses: teresa.sanchez@uco.es (M.T. Sánchez) or dcperez@uco.es (D. Pérez-Marín).

Abstract

Near infrared reflectance (NIR) spectroscopy was used as a fast and accurate technology for the simultaneous measurement of color, sugar and organic acid content in intact Raf tomatoes. The potential of this method coupled with chemometric techniques based on modified partial least squares regression was assessed by comparison with the currently-used traditional method for determining color, dry matter, soluble solid content, glucose, fructose, titratable acidity, malic acid and citric acid. At the same time, the performance of two spectrophotometers, differing primarily in terms of measurement principle and wavelength range, was evaluated. A total of 165 tomatoes (cv. "Raf") were used in the construction of calibration models, testing various spectral signal pretreatments. The technology was well suited to sorting Raf tomatoes on the basis of color parameters (a* and a*/b*), soluble solid content and titratable acidity, and useful, though less accurate, for the sorting of fruits by the rest of the color parameters tested (b*, L*), as well as by sugar content (glucose and fructose), dry matter and citric and malic content, particularly when the diode array instrument was used.

Keywords: near-infrared spectroscopy, tomato, external color, internal quality, MPLS regression.

1. Introduction

The tomato is the world's most widely-consumed vegetable, and thus a key product on the global agricultural market (Scibisz et al., 2011). In many countries, tomato production is largely aimed at the fresh-produce market, and therefore requires the comprehensive monitoring of external and internal quality parameters both during on-stem growth and ripening and during subsequent industrial handling (Costa and Heuvelimk, 2005; Alvés De Oliveira et al., 2014).

The tomato is composed mainly of water, soluble and insoluble solids, organic acids (principally citric acid) and micronutrients such as carotenoids and vitamins A and C (Pedro and Ferreira, 2007). Sugars and organic acids are responsible for sweetness and tartness, and also influence tomato flavor; as a result, they are the major factors affecting consumer acceptability (Baldwin et al., 2008, Kader, 2008; Causse et al., 2010). Color also has a marked influence on the initial purchasing decision by consumers, who tend to link fruit color to taste quality (Causse et al., 2010). López-Camelo and Gómez (2004) have suggested that the a*/b* ratio could be used for practical purposes as an objective ripening index, giving a realistic view of consumer perceptions.

However, since the measurement of external and internal quality parameters using traditional analytical methods is highly time-consuming, destructive, costly and contaminant, there is a clear need for fast, accurate and non-destructive analytical techniques that can be used both in the field and by the industry, and that enable individual classification of tomatoes by quality. NIRS technology meets these requirements and also offers other advantages, making it ideal for monitoring purposes and for ensuring traceability: low per-sample cost; little or no need for sample preparation; ability to analyze a wide range of products; and a high degree of

reproducibility and repeatability (Slaughter and Abbott, 2004; Garrigues and De la Guardia, 2013).

NIR technology is currently in widespread use for measuring chemical components and quality attributes in vegetable products (Saranwong and Kawano, 2007, Sánchez and Pérez-Marín, 2011). The few studies dealing with intact tomatoes, however, focus largely on measuring total soluble solid content (Slaughter et al., 1996; Flores et al., 2009), titratable acidity (Flores et al., 2009), dry matter (Khuriyati et al., 2004), color (Clement et al., 2008), and firmness (Shao et al., 2007). There are no reports in the literature regarding the use of NIRS spectroscopy for measuring glucose and fructose levels or citric and malic acid content in intact tomatoes, these being key factors for assessing ripeness and postharvest life, as well as exerting a crucial influence on the consumers' decision to purchase. This is particularly true of the Raf tomato which, though outwardly ugly due to its distinctive dark-green coloring and almost-black shoulder, boasts a salinity resistance guaranteeing an exquisite flavor rarely found in other varieties. No hitherto-published research has addressed the comparison of NIRS instruments differing in terms of cost, optical design, and suitability for on-site use for quality determination in tomatoes.

This study sought to assess the feasibility of using NIRS spectroscopy to measure external and internal quality attributes (color, total soluble solid content, fructose and glucose levels, titratable acidity, citric and malic acid levels and dry matter content) in intact Raf tomatoes. Data analysis included a comparison between two NIRS instruments with very different optical designs, one of which is highly suited to laboratory measurement (monochromator spectrophotometer) and the other better suited to on-line use in the packing house (diode-array spectrophotometer).

2. Material and methods

2.1. Sampling

A total of 165 tomatoes (*Lycopersicum esculentum* Mill., cv. "Raf") were harvested at commercial maturity in greenhouses in Almería (Spain). On arrival at the laboratory, fruits were promptly placed in refrigerated storage at 10ºC and 95% relative humidity. Prior to each measurement, samples were left until the near-surface fruit temperature had risen to, and stabilized at, laboratory temperature.

2.2. Reference data

Skin or external color values (L*, a*, and b*) were individually measured at the equator, turning the fruit through 120° between measurements, using a Minolta Chroma Meter CR-400 (Minolta Corporation, Ramsay, NJ) (CIE, 2004). Illuminant D65 and the 2° standard observer were used for all measurements. The three measurements obtained per fruit for each of the color parameters tested were averaged.

After these non-destructive measurements, fruits were halved and tissue from each fruit was taken at the same positions as those for the NIRS measurements. Dry matter content was determined by desiccation at 105ºC for 24 h (AOAC, 2000) and results were calculated as a percentage of final dry weight of the initial wet weight. Soluble solid content (SSC, in %) was measured as the refractometer reading for tomato juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany). Titratable acidity (TA) was measured by titration with 0.1 NaOH to an end point of pH 8.1. An automatic titrator was used (Crison Micro TT 2050, Crison, Alella, Barcelona, Spain). Results were expressed as % citric acid. Sugars (glucose, fructose) and organic acids (citric and malic acids) were quantified by an enzymatic method using food-analysis kits (Boehringer Mannheim Co., Mannhein, Germany) and expressed as g 100 g^{-1} of fresh weight for sugars and mg 100 g^{-1} of fresh

116 weight for acids. These measurements were performed with a BM-704 automatic
117 analyzer (Hitachi, Tokyo, Japan).

118 Each sample was analyzed in duplicate. All measurements were performed
119 immediately after VIS/NIRS spectrum collection.

120 *2.3. NIR analysis*

121 NIRS analysis was performed using two instruments that differ considerably in
122 terms of both function and optical design: a Perten DA-7000, Flexi-Mode diode array
123 spectrophotometer (Perten Instruments North America, Inc., Springfield, IL, USA),
124 more suitable for "on site" measurements, and a FNS-6500 scanning monochromator
125 (FOSS NIRSystems, Silver Spring, MD, USA), traditionally used in a laboratory
126 setting. These instruments operate in the 400 to 1700 nm range with a 5 nm scanning
127 interval, and in the 400 to 2500 nm range with a 2 nm scanning interval, respectively.

128 Using the diode-array instrument, tomatoes were placed centrally on the fruit
129 holder, with the stem-calyx axis vertical, calyx up, and were irradiated from above by
130 the light source while they rotated. Three separate spectral measurements were made on
131 each intact tomato, after a 120^0 sample rotation each time. The three spectra were
132 averaged to provide a mean spectrum for each intact fruit.

133 The FNS-6500 instrument was interfaced to a remote reflectance fiber optic
134 probe (NR-6539-A) with a 43 x 43 mm window; a dark compartment (340 x 238 x 222
135 mm) was used to protect the detector assembly. Each fruit was hand-placed in the probe,
136 so that the desired fruit location was centered on, and in direct contact with, the probe.
137 The first measurement was made at a random location on the blossom of the fruit. The
138 next two measurements were taken on the blossom end at rotations of roughly 120º and
139 240º from the initial site. The three spectra were averaged to provide a mean spectrum
140 for each tomato.

141 *2.4. Spectral repeatability*

142 Before averaging the three spectra, the spectral repeatability of intact tomatoes
143 was evaluated using the Root Mean Squared (RMS) statistic to eliminate spectra
144 displaying considerable variations. One hundred and sixty-five samples were analyzed
145 for this purpose. Three spectra were collected from each sample in the FNS-6500 and
146 the DA-7000, in three different positions.

147 The RMS statistic is the averaged root mean square of differences between the
148 different subsamples scanned at n wavelengths (Shenk and Westerhaus, 1995a, 1996).
149 The RMS for an individual sample (j) is defined as:

$$RMS_J = 10^6 \sqrt{\frac{\sum_{i=1}^{n} D_{ij}^2}{n}};\ D_{ij} = y_{ij} - \bar{y}_i$$

150

151 where y_{ij} is log (1/R) at wavelength i for subsample j, and $\tilde{y}_i$ is log (1/R) at
152 wavelength i for the average spectrum of N subsamples of a sample; n is the number of
153 data points collected by the instrument (here, 1050 data points for the FNS instrument
154 and 228 data points for the Perten instrument).

155 In order to determine the admissible limit for spectrum quality and repeatability
156 for each instrument and sample presentation mode, the standard deviation (STD) limit
157 was used to obtain an RMS cut-off value (Martínez et al., 1998).

$$STD_{limit} = 1.036 \sqrt{\sum_{k=1}^{k=m} STD_k^2/m}$$

158

159 where STD is the standard deviation per sample and m is the number of samples.

$$STD = \sqrt{\sum_{j=1}^{N} (RMS_j)^2/(N-1)}$$

160

161 where N is the number of sub-samples.

2.5. Population structuring and detection of spectral outliers prior to calibration

Principal Component Analysis (PCA) was performed on a set of N = 165 samples in order to decompose and compress the data matrix. After PCA, the center of the spectral population was determined in order to detect outlier samples. The Mahalanobis distance (GH) was calculated between each sample and the center; samples with a GH value greater than 3 were considered outliers (Shenk and Westerhaus, 1995a). As spectral pretreatments, the Standard Normal Variate (SNV) plus Detrending (DT) procedure (Barnes et al., 1989) was used to remove the multiplicative interferences of scatter, and one derivative mathematical treatment was performed (1,5,5,1), where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing (Shenk and Westerhaus, 1995b; ISI, 2000).

2.6. Construction and validation of prediction models by MPLS regression

Once spectral outliers had been removed (i.e. 7 of the original 165 samples), a set consisting of 158 samples was used to develop calibration models. The set was divided into two: a training set containing about 75% of the samples (N = 121) and a test set containing the remaining 25% (N = 37). These samples were selected following the method outlined by Shenk and Westerhaus (1991) using the CENTER algorithm included in the WinISI software package to calculate the Global Mahalanobis distance (GH). Samples were ordered based on the Mahalanobis distance to the center of the population, and three of every four were selected to be part of the calibration set.

Modified Partial Least Squares (MPLS) regression (Shenk and Westerhaus, 1995a) was used to obtain equations for predicting color, sugars, acids and dry matter content. Partial least squares (PLS) regression is similar to principal component

187 regression (PCR), but uses reference data (chemical, physical, etc.) and spectral
188 information to identify factors useful for fitting (Williams, 2001). MPLS is often more
189 stable and accurate than the standard PLS algorithm. In MPLS, the NIR residuals at
190 each wavelength, obtained after each factor is calculated, are standardized (divided by
191 the standard deviations of the residuals at a wavelength) before calculating the next
192 factor. When developing MPLS equations, cross-validation is recommended to select
193 the optimal number of factors and to avoid overfitting (Shenk and Westerhaus, 1995a).
194 For cross validation, the calibration set is partitioned in several groups; each group is
195 then validated using a calibration developed on the other samples; finally the validation
196 errors are combined to obtain a standard error of cross validation (SECV). In all cases,
197 cross validation was performed by splitting the population into six groups.

198 Signal noise at the beginning and end of the spectral range was eliminated for
199 both instruments: the resulting range for the DA-7000 spectrometer was from 515 to
200 1650 nm, while that of the FNS-6500 monochromator was from 516 to 2200 nm.

201 For each analytical parameter, different mathematical treatments were evaluated.
202 For scatter correction, the Standard Normal Variate (SNV) and Detrending (DT)
203 methods were tested (Barnes et al., 1989). Additionally, four derivative mathematical
204 treatments were tested in the development of NIRS calibrations: 1,5,5,1; 2,5,5,1;
205 1,10,5,1; 2,10,5,1 (Shenk and Westerhaus, 1995b).

206 The statistics used to select the best equations were: standard error of calibration
207 (SEC), coefficient of determination of calibration (R^2), standard error of cross-
208 validation (SECV), coefficient of determination for cross-validation (r^2), RPD or ratio
209 of the standard deviation of the original data (SD) to SECV, and coefficient of variation.
210 These latter two statistics enable SECV to be standardized, facilitating the comparison
211 of the results obtained with sets of different means (Williams, 2001).

The best models obtained for the calibration set, as selected by statistical criteria, were subjected to evaluation using samples not involved in the calibration procedure. A test set composed of 37 samples, not used previously in the model, was evaluated following the protocol outlined by Windham et al., (1989).

3. Results and discussion

3.1. Spectral repeatability

Optimization of spectrum quality and repeatability is crucial to the construction of models which are both accurate and robust. Statistical methods such as defined RMS cut-off limit can be useful for this purpose. The RMS cut-off was calculated for the two instruments as shown in section 2.4.

For the Perten DA-7000, the mean STD for the samples analyzed was 55,732 µlog (1/R), representing an RMS cut-off of 79,296 µlog (1/R). For the FNS-6500 instrument, mean STD and the RMS cut-off were 70,436 µlog (1/R) and 82,508 µlog (1/R), respectively. Any sample whose triplicated screening scans yielded an RMS above this value was eliminated and repeated until values fell below that limit, thus ensuring a high degree of spectrum repeatability.

No reference to the calculated RMS cut-off value for intact tomatoes has been found in the literature, although this statistic is essential to the generation of representative libraries.

The mean spectrum of the three replicates of each sample was used for further analysis.

3.2. Descriptive data for NIR calibrations and validations sets

Values obtained for range, mean, SD and CV for each of the parameters measured (calibration and validation sets) are shown in Table 1. Structured selection based wholly on spectral information, using the CENTER algorithm, proved suitable, in

237 that the calibration and validation sets displayed similar values for range, mean and SD
238 for all study parameters; moreover, the ranges of the validation set lay within those of
239 the calibration set.

240 All parameters except color measurements L* and b* displayed marked
241 variability, with CV values of over 18% for both the calibration and validation sets.

242 Williams (2001) and Pérez-Marín et al., (2005) have highlighted the importance
243 both of sample set size and of sample distribution within the calibration set, noting that
244 sample sets for calibration should ideally ensure uniform distribution of composition
245 across the range of the study parameter in question.

246 *3.3. Prediction of color quality parameters using MPLS regression and NIR spectra*

247 The best equations for measuring color-related parameters (L*, a*, b* and a*/b*)
248 for the two instruments tested, using the combination of signal pretreatments that
249 yielded the best results in each case, are shown in Table 2.

250 Models obtained using the Perten DA-7000 instrument displayed greater
251 predictive ability, for all color parameters, than those obtained using the
252 monochromator. Models constructed for L* and b* using the diode-array instrument
253 enabled samples to be classified into high, medium and low values, whilst models for a*
254 and a*/b* displayed good predictive capacity within the limits established by Shenk and
255 Westerhaus (1996). Models obtained with the FNS-6500 monochromator only enabled
256 samples to be classified into high and low values (Shenk and Westerhaus, 1996).

257 No references have been found in the literature to color parameter prediction in
258 intact Raf tomatoes. However, the RPD values recorded here were lower than those of
259 between 2.81 and 7.22 reported by Clément et al. (2008) for color prediction (L*, a*, b*
260 and a*/b*) in Canadian tomatoes at varying degrees of ripeness using a Varian Cary 500
261 UV-VIS-NIR scanning spectrophotometer equipped with an integration sphere, working

262 in the spectral region 400-1000 nm. This highlights the difficulty of measuring color
263 parameters in Raf tomatoes, in which both form and color distribution are highly
264 irregular (Fig. 1).

265 Values for a*, like those of b* and a*/b*, increase significantly during ripening
266 due to higher carotenoid levels, and thus also provide a useful indicator of fruit ripeness
267 (Kader et al., 1978). It should also be stressed that the diode-array instrument enables
268 color parameters to be measured on-site, which is particularly useful for the tomato
269 handling industry. For Raf tomatoes, fruit color is regarded as synonymous with quality
270 and taste: darker-colored—almost bluish—fruits are likely to have the best taste
271 qualities; the green-black shoulder, while not an essential quality indicator, shows that
272 the fruit has received sufficient sunlight and is therefore sweeter, and is also an ideal
273 indicator for distinguishing Raf from similar tomatoes.

274 Validation statistics for the prediction of these parameters in intact tomatoes are
275 also shown in Table 2. In terms of the validation protocol recommended by Windham et
276 al., (1989) for the routine implementation of NIRS prediction models, the only models
277 yielding sufficiently accurate predictions were those constructed for parameters a* and
278 a*/b* using the DA-7000 spectrophotometer.

279 *3.4. Prediction of internal quality parameters using MPLS regression and NIR spectra*

280 Models obtained for all internal quality parameters using the Perten DA-7000
281 displayed greater predictive capacity than those constructed with the FNS-6500, with
282 the exception of dry matter content (Table 3).

283 For predicting dry matter, the model constructed using the monochromator and
284 D_2 log (1/R) (r^2 = 0.59; SECV = 0.26% fw) enabled samples to be classified into high,
285 medium and low values, whereas the model obtained with the diode-array instrument

286 and the same second derivative only enabled classification into high and low values (r^2
287 = 0.45; SECV = 0.29% fw).

288 Walsh et al., (2004) reported slightly better predictive capacity (r^2 = 0.64; SECV
289 = 0.20% fw) using a Carl Zeiss MMS1 NIR-enhanced spectrometer in the spectral
290 region from 300 nm to 1100 nm, noting that the low standard deviation value for the
291 sample set was probably the main cause of poor model performance. Increasing the
292 range for this parameter could improve the predictive capacity of the models. This
293 would be useful for the tomato packing industry, since non-destructive measurement of
294 tomato dry matter (DM) content is essential for fruit classification purposes, ensuring
295 that fruit batches are of similar DM levels. It may also have implications both for
296 consumer acceptability—fruits with higher dry matter content have a better flavor—and
297 for improving storage potential and ripe fruit quality.

298 Models for total soluble solid content obtained with D_2 log (1/R) using the
299 monochromator (r^2 = 0.77; SECV = 0.64%) and the diode-array instrument (r^2 = 0.79;
300 SECV = 0.59%) displayed good predictive capacity in terms of Shenk and
301 Westherhaus' recommendations (1996).

302 Although other studies of SSC prediction in tomatoes using NIRS technology
303 (Slaughter et al., 1996; Hong and Tsou, 1998; Walsh et al., 2004; He et al., 2005; Shao
304 et al., 2007) report models with r^2 values ranging from 0.49 to 0.97 and SEP values of
305 between 0.22 and 0.38ºBrix, the models constructed here for predicting SSC displayed
306 adequate predictive capacity, bearing in mind the irregular shape of this tomato variety,
307 which undoubtedly influences measurements.

308 For glucose, the model obtained using the DA-7000 (r^2 = 0.61; SECV = 0.38
309 g/100 g fw) displayed greater accuracy and precision than its counterpart constructed
310 using the FNS-6500 (r^2 = 0.50; SECV = 0.41 g/100 g fw), enabling values to be

311 classified into high, medium and low. For fructose, results obtained using both the
312 diode-array instrument (r^2 = 0.43; SECV = 0.37 g/100 g fw) and the monochromator (r^2
313 = 0.30; SECV = 0.37 g/100 g fw) only enabled classification into high and low values.
314 Fructose and glucose are components of the main sugars and carbohydrates in tomatoes.
315 MPLS regression showed that spectra could be employed to distinguish between sample
316 ripeness stages.

317 No published studies address the direct measurement of these sugars in intact
318 tomatoes. Pedro and Ferreira (2007) reported better predictive capacity both for glucose
319 (r^2 = 0.98; RMSEP = 0.54 %) and for fructose (r^2 = 0.94; RMSEP = 0.88 %) although
320 their results are not wholly comparable, since they used a set comprising samples of
321 tomato concentrate products with total solid content ranging from 6.9 to 35.9%, and
322 thus worked with a more varied calibration set.

323 Although measurement of acidity-related parameters in intact fruit is notoriously
324 difficult (Flores et al., 2009), the models obtained for predicting titratable acidity using
325 both instruments displayed good predictive capacity, with values of r^2 = 0.72 and 0.70
326 for the diode-array and monochromator, respectively, and SECV = 0.06% citric acid in
327 both cases.

328 Hong and Tsou (1998) recorded an r^2 value of 0.94, i.e. higher than that obtained
329 here, for measurements of titratable acidity, although they used chopped rather than
330 intact tomato; the residual error reported by these authors was similar to that recorded
331 here (0.06% citric acid).

332 Models constructed for citric and malic acid content using the diode-array
333 instrument yielded r^2 values of between 0.50 and 0.49, and SECV values in the range
334 80.82 - 22.43 mg/100 g fw, whilst with the monochromator r^2 value lay between 0.30
335 and 0.42 while SECV values ranged from 95.68 to 23.79 mg/100 g fw. These results

suggest that NIRS technology may be used for screening purposes, to distinguish between low and high levels of both acids.

There are no published reports on the measurement of malic and citric acid in intact tomatoes using NIRS technology. However, these parameters are linked to the behavior of the tomato during ripening, and may thus act as indicators of ripeness and thus of optimal harvesting time. Malic acid levels decreases significantly during the later stages of ripening, while citric acid content generally increases (Baldwin et al., 1991); non-destructive measurement of citric and malic acid content is therefore of considerable value.

Validation statistics for the prediction of internal parameters in intact Raf tomatoes using both instruments are shown in Table 3.

The models constructed for predicting SSC in intact tomatoes using both instruments tested, and for predicting TA using the diode array instrument, met the validation requirements in terms of r^2 ($r^2 > 0.6$) and both the SEP(c) and the bias were within confidence limits: the equations thus ensure accurate prediction, and can be applied routinely. For dry matter and glucose content, it should be stressed that SEP(c) and bias lay within confidence limits for both instruments, although r^2 results did not always attain recommended minimum values, indicating that the NIRS equations constructed should be regarded as a first step in the fine-tuning of NIRS technology for the on-site monitoring of internal quality parameters in this tomato.

Slight differences in accuracy were noted between models constructed using the two instruments tested, although better results were obtained with the diode-array instrument for all parameters except dry matter content.

The models predicted fructose content, citric and malic acid content in validation-set samples with low values for r^2, in neither case meeting the

361 recommendations of Windham et al., (1989). These models are thus not suitable for
362 routine applications.

363 **4. Conclusions**

364 Near infrared reflectance spectroscopy combined with multivariate analysis is a very
365 promising tool for determining the overall composition of intact Raf tomatoes, allowing
366 ripeness to be monitored not only in terms of visual appearance but also in terms of
367 taste, within one minute. The results of external validation indicate that parameters such
368 as color (a* and a*/b*), SSC and TA can be routinely predicted using the diode array
369 instrument, thus considerably reducing analysis time and enabling incorporation of
370 these models into on-line NIR grading systems for measuring the ripeness of individual
371 fruits in lines of harvested tomatoes. This could in turn lead to improved taste
372 acceptability for this product. By contrast, the models constructed were unable to
373 accurately predict citric and malic acid levels in tomatoes. It should be stressed that the
374 results obtained here using a diode-array sensor should be regarded as the first step in
375 the fine-tuning of NIRS for on- site quality monitoring of the Raf tomato, a complex
376 vegetable with an irregular form. Over the coming years, recalibrations may be required
377 in order to enhance the robustness of the models obtained; the variability observed in
378 this type of tomato could be reflected by including fruits harvested in different years and
379 from different orchards, since these factors influence the chemical composition of
380 tomatoes.

381 **Acknowledgements**

382 This research was funded by the Andalusian Regional Government under the Research
383 Excellence Program (Project No. 3713 'Safety and Traceability in the Food Chain using
384 NIRS'). The authors thank Ms. Katherine Flores-Rojas and Ms. Mª Carmen Fernández
385 for their technical assistance.

386 **References**

387 Alvés De Oliveira, G.A., Bureau, S., Renard, C.M.-G.C., Pereira-Netto, A.B., de
388 Castilhos, F., 2014. Comparison of NIRS approach for prediction of internal
389 quality traits in three fruit species. Food Chem. 143, 223-230.

390 AOAC, 2000. Official Methods of Analysis of AOAC International. 17th ed. AOAC,
391 Gaithersburg, MD, USA.

392 Baldwin, E.A., Goodner, K.L., Plotto, A., 2008. Interaction of volatiles, sugars, and
393 acids on perception of tomato aroma and flavor descriptors. J. Food Sci. 73,
394 S294-S307.

395 Baldwin, E.A., Nisperos-Carriedo, M.O., Moshonas, M.G., 1991. Quantitative analysis
396 of flavor and other volatiles and for certain constituents of two tomato cultivars
397 during ripening. J. Am. Soc. Hortic. Sci. 116, 265–269.

398 Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate Transformation
399 and De-trending of near infrared diffuse reflectance spectra. App. Spectrosc. 43,
400 772-777.

401 Causse, M., Friguet, C., Coiret, C., Moneta, E., Grandillo, S. 2010. Consumer
402 preferences for fresh tomato at the European scale: a common segmentation on
403 taste and firmness. J. Food Sci. 75, 531-541.

404 CIE, 2004. Colorimetry, 3rd ed. Commission Internationale De L'eclairage, Vienna,
405 Austria.

406 Clément, A., Dorais, M., Vernon, M., 2008. Nondestructive measurement of fresh
407 tomato lycopene content and other physicochemical characteristics using
408 Visible-NIR Spectroscopy. J. Agr. Food Chem. 56, 9813–9818.

409 Costa, J.M., Heuvelimk, E., 2005. Introduction: the tomato crop and industry. In:
410 Heuvelink, E. (Ed.), Tomatoes. CABI Publishing, Oxfordshire, UK. pp. 1–19.

411 Flores, K., Sánchez, M.T., Pérez-Marín, D., Guerrero, J.E., Garrido-Varo, A., 2009.
412 Feasibility in NIRS instruments for predicting internal quality in intact tomato. J.
413 Food Eng. 91, 311-318.
414 Garrigues, S., De la Guardia, M., 2013. Non-invasive analysis of solid samples. TrAC-
415 Trend Anal. Chem. 43, 161–173.
416 He, Y., Zhang, Y., Pereira, A. G., Gómez, A. H., Wang, J., 2005. Nondestructive
417 determination of tomato fruit quality characteristics using Vis/NIR spectroscopy
418 technique. Int. J. Inform. Technol. 11, 97–108
419 Hong, T. L., Tsou, S.C.S., 1998. Determination of tomato quality by near infrared
420 spectroscopy. J. Near Infrared Spectrosc. 6, A321-A324.
421 ISI, 2000. The Complete Software Solution Using a Single Screen for Routine Analysis,
422 Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR.
423 Infrasoft International, LLC, Silver Spring, MD.
424 Kader, A.A. 2008. Flavor quality of fruits and vegetables. J. Sci. Food Agric. 88, 1863–
425 1868.
426 Kader, A.A., Morris, L.L., Stevens, M.A., Albright-Holton, M., 1978. Composition and
427 flavor quality of fresh market tomatoes as influenced by some postharvest
428 handling. J. Am. Soc. Hortic. Sci. 103, 6-11.
429 Khuriyati, N., Matsuoka, T., Kawano, S., 2004. Precise near infrared spectral
430 acquisition of intact tomatoes in interactance mode. J. Near Infrared Spectrosc.
431 12, 391–395.
432 López-Camelo, A. F., Gómez, P. A. 2004. Comparison of color indexes for tomato
433 ripening. Hortic. Bras. 22, 534-537.

Martínez, M.L., Garrido, A., De Pedro, E.J., Sánchez, L. 1998. Effect of sample heterogeneity on NIR meat analysis: The use of the RMS statistic. J. Near Infrared Spectrosc. 6, 313–320.

Pedro, A. M. K., Ferreira, M. M. C. 2007. Simultaneously Calibrating Solids, Sugars and Acidity of Tomato Products Using PLS2 and NIR Spectroscopy. Anal. Chim. Acta 595, 221–227.

Pérez-Marín, D.C., Garrido-Varo, A., Guerrero-Ginel, J.E., Gómez-Cabrera, A. 2005. Implementation of local algorithm with NIRS for compliance assurance in compound feedingstuffs. Appl. Spectrosc. Rev. 59, 69-77.

Sánchez, M.T., Pérez-Marín, D., 2011. Nondestructive measurement of fruit quality by NIR spectroscopy. In: Vázquez, M., Ramírez, J.A. (Eds.), Advances in Post-Harvest Treatments and Fruit Quality and Safety. Nova Science Publishers Inc., Hauppauge, NY, USA, pp. 101-163

Saranwong, S., Kawano, S., 2007. Applications to agricultural and marine products: fruits and vegetables. In: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc., New Jersey, Hoboken, pp. 219–242.

Scibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, M., Bertrand, D., Renard, C., 2011. Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. Food Chem. 125, 1390-1397.

Shao, Y., He, Y., Gómez, A.H., Pereira, A.G., Qiu, Z., Zhang, Y., 2007. Visible/near infrared spectrometric technique for nondestructive assessment of tomato "Heatwave" (*Lycopersicum esculentum*) quality characteristics. J. Food Eng. 81, 672-678.

458 Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near infrared spectra and
459 modified partial least squares regression. Crop Sci. 31, 1548-1555.
460 Shenk, J.S., Westerhaus, M.O., 1995a. Analysis of Agriculture and Food Products by
461 Near Infrared Reflectance Spectroscopy. Monograph, NIRSystems, Inc., 12101
462 Tech Road, Silver Spring, MD 20904, USA.
463 Shenk, J.S., Westerhaus, M.O., 1995b. Routine Operation, Calibration, Development
464 and Network System Management Manual. NIRSystems, Inc., 12101 Tech Road,
465 Silver Spring, MD 20904, USA.
466 Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davies, A.M.C.,
467 Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR
468 Publications, Chichester, UK, pp.198-202.
469 Slaughter, D.C., Abbott, J.A., 2004. Analysis of fruits and vegetables. In: Roberts, C.A.,
470 Workman, J., Reeves III, J.B. (Eds.), Near-Infrared Spectroscopy in Agriculture.
471 ASA, CSSA and SSSA, Madison, WI, pp. 377–398.
472 Slaughter, D.C., Barrett, D., Boersig, M., 1996. Nondestructive determination of soluble
473 solids in tomatoes using near infrared spectroscopy. J. Food Sci. 61, 695-697.
474 Walsh, K.B., Golic, M., Greensill, C.V., 2004. Sorting of fruit using near infrared
475 spectroscopy: application to a range of fruit and vegetables for soluble solids and
476 dry matter content. J. Near Infrared Spectrosc. 12, 141–148.
477 Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C.,
478 Norris, K.H. (Eds.), Near-Infrared Technology in the Agricultural and Food
479 Industries. AACC, Inc., St. Paul, Minnesota, pp. 145-169.
480 Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration:
481 sample selection and equation development and validation. In: Martens, G.C.,
482 Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis

of Forage Quality. Agriculture Handbook, vol. 643. Government Printing Office, Washington, DC, pp. 96–103.

Table 1

Statistical analysis of calibration and validation sets: data range, mean, standard deviation (SD), and coefficient of variation (CV).

Parameter	Set	Range	Mean	SD	CV (%)
L*	Calibration	28.92-58.18	47.73	3.30	6.93
	Validation	43.28-54.32	48.03	2.65	5.53
a*	Calibration	-18.14-10.42	-8.46	4.77	56.36
	Validation	-14.96-7.36	-8.62	5.18	60.06
b*	Calibration	20.96-47.15	28.88	3.50	12.15
	Validation	23.99-35.19	28.69	2.60	9.07
a*/b*	Calibration	-0.53-0.37	-0.29	0.16	57.05
	Validation	-0.53-0.26	-0.29	0.17	59.19
Dry matter (% fw)	Calibration	0.60-3.03	1.55	0.42	27.58
	Validation	0.69-2.67	1.64	0.47	28.93
SSC (%)	Calibration	2.50-9.00	5.29	1.32	25.11
	Validation	2.75-8.00	5.36	1.25	23.47
Glucose (g 100 g^{-1} fw)	Calibration	0.86-4.39	2.17	0.63	29.14
	Validation	1.03-3.89	2.18	0.64	29.62
Fructose (g 100 g^{-1} fw)	Calibration	0.91-4.35	1.88	0.53	28.59
	Validation	0.91-3.00	1.89	0.50	26.77
Titratable acidity (% citric acid)	Calibration	0.16-0.67	0.36	0.11	30.90
	Validation	0.20-0.58	0.35	0.09	27.09
Citric acid (mg 100 g^{-1} fw)	Calibration	187.46-895.68	449.57	122.96	27.35
	Validation	250.75-647.71	446.55	97.06	21.74
Malic acid (mg 100 g^{-1} fw)	Calibration	51.84-282.01	134.82	34.35	25.48
	Validation	62.90-180.20	132.60	24.54	18.51

Table 2

MPLS regression statistics for NIR-based models for predicting external quality parameters in Raf tomatoes.

Parameter	Instrument	Spectral range (nm)	Mathematic treatment	Calibration					Validation				
				N	SECV	r^2	RPD	CV	N	r^2	SEP	SEP (c)	Bias
L*	FNS-6500	516-2200	2,10,5,1	116	2.04	0.48	1.37	5.83	36	0.31	2.06	2.09	0.12
	DA-7000	515-1650	2,10,5,1	111	1.56	0.59	1.56	5.07	35	0.50	1.85	1.84	-0.32
a*	FNS-6500	516-2200	1,10,5,1	116	3.16	0.47	1.36	48.99	37	0.37	4.15	4.21	-0.09
	DA-7000	515-1650	2,5,5,1	113	2.19	0.74	1.97	49.59	35	0.76	2.58	2.60	0.32
b*	FNS-6500	516-2200	2,5,5,1	120	2.56	0.34	1.21	10.78	37	0.16	2.46	2.50	-0.07
	DA-7000	515-1650	2,5,5,1	111	1.81	0.67	1.71	10.71	36	0.23	2.53	2.55	-0.24
a*/b*	FNS-6500	516-2200	2,10,5,1	118	0.12	0.38	1.26	52.75	37	0.44	0.13	0.13	-0.01
	DA-7000	515-1650	2,10,5,1	110	0.07	0.80	2.23	51.46	34	0.75	0.09	0.09	-0.01

Table 3

MPLS regression statistics for NIR-based models for predicting internal quality parameters in Raf tomatoes.

Parameter	Instrument	Spectral range (nm)	Mathematic treatment	Calibration					Validation				
				N	SECV	r^2	RPD	CV	N	r^2	SEP	SEP (c)	Bias
Dry matter (% fw)	FNS-6500	516-2200	2,10,5,1	116	0.26	0.59	1.54	26.39	34	0.49	0.32	0.30	0.12
	DA-7000	515-1650	2,10,5,1	116	0.29	0.45	1.30	24.76	35	0.39	0.33	0.33	0.01
SSC (%)	FNS-6500	516-2200	2,5,5,1	119	0.64	0.77	2.08	25.00	36	0.60	0.83	0.84	0.01
	DA-7000	515-1650	2,5,5,1	113	0.59	0.79	2.13	24.23	36	0.75	0.65	0.65	-0.05
Glucose (g 100 g^{-1} fw)	FNS-6500	516-2200	1,10,5,1	113	0.41	0.50	1.40	26.71	35	0.52	0.44	0.45	0.02
	DA-7000	515-1650	1,10,5,1	111	0.38	0.61	1.57	27.75	34	0.53	0.42	0.42	-0.07
Fructose (g 100 g^{-1} fw)	FNS-6500	516-2200	1,5,5,1	113	0.37	0.30	1.20	24.33	36	0.35	0.38	0.39	0.02
	DA-7000	515-1650	2,10,5,1	116	0.37	0.43	1.29	25.57	37	0.36	0.40	0.41	-0.02
Titratable acidity (% citric acid)	FNS-6500	516-2200	1,5,5,1	118	0.06	0.70	1.83	30.15	37	0.56	0.07	0.07	-0.01
	DA-7000	515-1650	2,5,5,1	115	0.06	0.72	1.87	29.92	35	0.69	0.06	0.06	-0.01
Citric acid (mg 100 g^{-1} fw)	FNS-6500	516-2200	2,10,5,1	119	95.68	0.30	1.18	25.57	35	0.31	85.93	86.58	10.14
	DA-7000	515-1650	1,10,5,1	117	80.82	0.50	1.39	25.27	37	0.38	81.18	82.09	5.86
Malic acid (mg 100 g^{-1} fw)	FNS-6500	516-2200	2,10,5,1	115	23.79	0.42	1.28	22.84	37	0.27	22.07	22.30	-1.75
	DA-7000	515-1650	1,10,5,1	117	22.43	0.49	1.40	23.60	37	0.34	21.76	21.87	-2.80

Fig.1. Raf tomato