# Developing universal models for the prediction of physical quality in Citrus fruits analyzed on-tree using portable NIRS sensors

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

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

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

[c] *Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14004 Cordoba, Spain*

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

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

## Abstract

The citrus sector seeks rapid, economical, environmentally-friendly and non-destructive technologies for monitoring the external and internal changes in physical quality taking place in fruit during on-tree ripening, thus allowing fruit quality to be evaluated at any stage in the ripening process. The use of portable NIRS sensors based on MEMS technology, in conjunction with chemometric data treatment models, has already been studied for quality-control purposes in two citrus species: oranges and mandarins. The critical challenge is to develop robust and accurate universal mathematic models based on hundreds of highly heterogeneous citrus samples in order to design quality prediction models applicable to all fruits belonging to the genus Citrus, rather than models that can only be applied successfully to a single citrus species. This study evaluated and compared the performance of MPLS and LOCAL regression algorithms for the prediction of major physical-quality parameters in all citrus fruits. Results showed that, while models developed using both linear (MPLS) and non-linear regression techniques (LOCAL) yielded promising results for the on-tree quality evaluation of citrus fruits, the LOCAL algorithm additionally increased the predictive capacity of models constructed for all the main parameters tested. These findings confirm that NIRS technology, used in conjunction with large databases and local regression strategies, increases the robustness of models for the on-tree prediction of citrus fruit quality; this will undoubtedly be of benefit to the citrus industry.

*Keywords:* NIRS; citrus; physical quality; universal models; MPLS regression; LOCAL algorithm

## 1. Introduction

Citrus fruits, which play a significant role in the human diet (Liu, Heying, & Tanumihardio, 2012), are among the world's major crops, and the highest value fruit crop in international trade (Magwaza et al., 2012b). The market price of harvested citrus fruit is at present based largely on external color, size and weight (Olmo, Nadas, & García, 2000; Nicolaï et al., 2007; Magwaza et al., 2012a); it would be useful to introduce, in the near future, quality-based pricing systems, using both external and internal quality indices.

In response to growing demand from producers, consumers and the industry, recent years have seen the development of rapid, accurate, economical and above all non-destructive technologies for determining food-produce quality. NIRS is one flexible and versatile technology, which has been successfully applied for the prediction of quality parameters in various citrus fruit species, and especially in oranges and mandarins. Numerous authors, including Fraser, Jordan, Künnemeyer, & McGlone (2003), Guthrie, Walsh, Reid, & Lienberg (2005a), Guthrie, Reid, & Walsh (2005b), Hernández-Gómez, He, & Pereira (2006), Sun, Zhang, & Liu (2009), Liu, Sun, Zhang, & Aiguo (2010b), Antonucci et al., (2011), Magwaza et al., (2012b, 2013b, 2014), Magwaza, Opara, Cronje, Landahl, & Terry (2013a) and Sánchez, De la Haba, & Pérez-Marín (2013a) have confirmed the potential of NIRS for predicting quality in mandarins; similar findings have been reported for oranges by Cayuela (2008), Cayuela & Weiland (2010), Liu, Sun, & Ouyang (2010a), Zheng et al., (2010), Magwaza et al., (2013c) and Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b).

At the same time, the citrus sector is increasingly demanding methods for the on-tree monitoring of fruit quality parameters throughout the ripening process, with a view to identifying the optimal harvesting time depending on the final destination of the product. Advances in NIRS instrumentation include the development of handheld and portable equipment, some of which has already been applied successfully for on-tree quality measurements in mandarins (Sánchez, De la Haba, & Pérez-Marín, 2013a) and, separately, in oranges (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b). However, the predictive capacity and robustness of the models thus developed could be improved by using larger and more varied sample sets. In this sense, universal models applicable to any citrus fruit species would be particularly useful, and would favor the uptake of this technology by the citrus sector. However, when using what might be termed "multi-product sample sets", the relationship to be modeled may not always be linear; as a result, classical regression methods are not always the most suitable (Pérez-Marín, Garrido-Varo, & Guerrero, 2007). Barton II, Shenk, Westerhaus, & Funk (2000) suggested that one option in these cases could be to use local approaches based on the development of specific calibrations for each sample to be predicted, enabling existing nonlinearity to be addressed through the production of "local" linear models.

The aim of this study was to evaluate the LOCAL algorithm using a citrus-fruit database for the development of models to predict physical quality parameters during on-tree ripening—regardless of species, growing-season and crop practices—using a handheld MEMS-NIRS spectrophotometer.

## 2. Materials and methods

### *2.1. Fruit samples and reference data*

The initial sample set comprised 611 samples belonging to the genus Citrus: 378 oranges (*Citrus sinensis* L. cv. 'Powell Summer Navel') and 233 mandarins (*Citrus reticulata* Blanco cv. 'Clemevilla'), from two consecutive seasons, both grown on a commercial plantation near the village of La Campana (Seville, Spain). A total of 191 of the oranges were those used by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b), while all the mandarins were used by Sánchez, De la Haba, & Pérez-Marín (2013a).

Harvested oranges and mandarins were kept in refrigerated storage at 5ºC and 90% RH until the following day, when laboratory testing was performed. Prior to each test, samples were allowed to reach room temperature. All physical tests were performed at 20ºC.

For mandarins, external (weight, equatorial and axial diameters, color (L*, a*, b*, C* and h*)) and internal (firmness, pericarp thickness and juice weight) physical-quality parameters were measured following Sánchez, De la Haba, & Pérez-Marín (2013a); the same external and internal physical-quality parameters for oranges were measured following Sánchez, De la Haba, Serrano, & Pérez-Marín, (2013b).

### *2.2. NIR analysis*

NIR spectra of mandarins and oranges were collected in reflectance mode (log 1/R) using the Phazir 2400, an integrated handheld micro-electromechanical system (MEMS) spectrophotometer (Polychromix, Inc., Wilmington, MA, USA) that incorporates all the essential components to deliver on-tree applications. This instrument operates between 1600 and 2400 nm with an 8 nm non-constant sampling interval (pixel resolution 8 nm, optical resolution 12 nm). Four spectral measurements were made for each fruit on the tree, taking orientation (north, south, east and west) into account. The four spectra were averaged to provide a mean spectrum for each sample.

### *2.3. Definition of calibration and validation sets*

Prior to carrying out NIRS calibrations, the CENTER algorithm included in the WinISI II software package, version 1.50 (Infrasoft International, Port Matilda, PA, USA) was applied to ensure a structured population selection based solely on spectral information for the establishment of calibration and validation sets (Shenk & Westerhaus, 1991, 1995). This algorithm performs an initial principal component analysis (PCA) to calculate the centre of the population and the distance of samples (spectra) from that centre in an n-dimensional space, using the Mahalanobis distance (GH); samples with a statistical value greater than 4 were considered outliers or anomalous spectra.

The standard normal variate (SNV) and detrending (DT) methods were applied for scatter correction (Barnes, Dhanoa, & Lister, 1989), together with the mathematical derivation treatment 1,5,5,1; where the first digit is the number 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 & Westerhaus, 1995; ISI, 2000).

Once spectral outliers had been removed (i.e. 7 of the original 611 samples), a set consisting of 604 samples was used to develop calibration models. The set was divided into two: a training set containing about 75% of the samples (N = 457) and a test set containing the remaining 25% (N = 147).

Data were subjected to chemometric treatment using the WinISI II software package, version 1.50.

### *2.4. Construction of prediction models for major physical quality parameters in intact citrus fruits on-tree using the LOCAL algorithm*

The LOCAL algorithm operates by searching and selecting samples in large databases that have spectra similar to the sample being analyzed. The selected samples are then used to compute a specific calibration equation, based on Partial Least Squares (PLS) regression, for predicting the constituents of an unknown sample (Shenk, Westerhaus, & Berzaghi, 1997).

Different parameters have to be evaluated in order to optimize the LOCAL algorithm (Pérez-Marín, Garrido-Varo, & Guerrero, 2007). In the present study, an optimization design for the LOCAL algorithm was set up by varying the number of calibration samples (*k*) from 80 to 120 in steps of 20, and the number of factors (*l*) from 14 to 16 in steps of 1. This gave a factorial design of 3 x 3 or 9 runs. Finally, the number of PLS factors discarded was set at the first four.

For each analytical parameter, different mathematical treatments were evaluated for scatter correction, including SNV and DT methods (Barnes, Dhanoa, & Lister, 1989). Additionally, four derivative mathematical treatments were tested: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1 (Shenk & Westerhaus, 1995).

Global calibration using the same math pre-treatments used in LOCAL was performed (WinISI, II software package, version 1.50 (Infrasoft International, Port Matilda, PA, USA) in order to compare results obtained using the non-linear regression algorithm with those yielded by the classical prediction strategy based on MPLS regression. The same validation file for the genus Citrus was then predicted using both regression algorithms. The results provided by the models constructed using non-linear regression for mandarin + orange were also compared with those obtained for mandarin alone (Sánchez, De la Haba, & Pérez-Marín, 2013a) and for orange alone (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b), in both cases using MPLS regression. Standard errors of prediction (SEP) and coefficients of determination ($r^2$) using the LOCAL procedure and MPLS regression were compared.

## 3. Results and discussion

### *3.1. Descriptive data for NIR calibration and validation*

After applying the CENTER algorithm to the overall set (N = 611), a total of 7 samples (2 oranges and 5 mandarins) were identified as anomalous spectra. Analysis showed that six of these displayed extreme values for the parameter a*, three being very green (2 mandarins and 1 orange at the start of harvesting), and three (mandarins) displaying a marked reddish hue at the end of harvesting. The other anomalous orange sample displayed an abnormally high value for pericarp thickness.

Values (range, mean, standard deviation and coefficient of variation, CV) obtained for each physical-quality parameter in the calibration and validations sets, after removing outliers, are shown in Table 1. Structured selection based on spectral information, using the CENTER algorithm proved suitable, in that the calibration and validation sets displayed similar values for range, mean and SD for all study parameters. Furthermore, the ranges of the validation set lay within those of the calibration set.

All physical parameters tested, except three of the color-related parameters (L*, b* and C* for the calibration and validation sets), displayed marked variability, with CV values of over 12% for both sets, covering a wide range of values. Other parameters also recorded CV values of over 40% in both sets, including weight, a*, firmness and juice weight.

Pérez-Marín, Garrido-Varo, & Guerrero (2005) have highlighted the importance of sample set and of sample distribution within the calibration set, noting that sample sets for calibration should ideally ensure uniform distribution of composition across the range of the study parameter in question.

### *3.2. Prediction of physical quality parameters in citrus fruits using the LOCAL algorithm*

Results for the prediction of citrus-fruit physical quality parameters using LOCAL algorithm are shown in Table 2.

It should be noted that for predicting the external validation set, the LOCAL algorithm used only between 80 and 100 samples to predict most of the parameters tested and only 120 samples for weight and L* prediction, rather than using all 457 samples in the calibration set (as was the case for MPLS regression); only those samples whose spectra were considered representative of the calibration set were used.

The results obtained using the LOCAL algorithm were better than those achieved with MPLS regression (Table 2) for universal citrus models; robustness was increased by minimizing prediction error and increasing the coefficient of determination for prediction.

The accuracy of the predictions obtained using the LOCAL algorithm was greater (i.e. SEP values were lower) than that of those obtained using the MPLS regression for all parameters tested in mandarin + orange. The greatest reduction in SEP using the LOCAL algorithm was recorded for the a* parameter (34.60%), followed by b* (23.91%). The smallest reductions in SEP using the LOCAL algorithm were recorded for equatorial diameter and juice weight (4.79% and 6.20%, respectively, with respect to MPLS regression).

An overall increase in the coefficient of determination was recorded for models obtained using the LOCAL algorithm with respect to those using MPLS. The most significant increases in value for $r^2$ were recorded for all color-related parameters ($r^2_{MPLS}$ = 0.30-0.44; $r^2_{LOCAL}$ = 0.50-0.63), firmness ($r^2_{MPLS}$ = 0.08; $r^2_{LOCAL}$ = 0.28), fruit weight ($r^2_{MPLS}$ = 0.65; $r^2_{LOCAL}$ = 0.73), and axial diameter ($r^2_{MPLS}$ = 0.74; $r^2_{LOCAL}$ = 0.82).

However, neither of the strategies yielded results for L*, a*, C*, h* and firmness that lay within the limits recommended by Windham, Mertens, & Barton (1989) for the coefficient of determination ($r^2 > 0.60$). Even so, the LOCAL algorithm improved the coefficient of determination by 29.55% for L*, 52.94% for a*, 69.70% for C*, 66.67% for h* and by 250% for firmness, compared to the MPLS regression.

### *3.2.1. Morphological parameters*

For morphological parameters (weight, equatorial and axial diameters) the citrus universal calibrations using MPLS performed worse in terms of accuracy and precision of prediction (Table 2), whilst the use of LOCAL reduced the SEP value by 13.39% for weight, by 4.79% for equatorial diameter, and by 17.80% for axial diameter. Moreover, the predictive models obtained for weight and equatorial diameter using the global strategy and MPLS regression only enabled fruit to be classified as high, medium or low, whereas the predictive capacity using the LOCAL algorithm may be considered good according to the limits defined by Shenk & Westerhaus (1996). For axial diameter, the LOCAL strategy yielded an $r^2$ value of 0.82 compared to 0.74 for MPLS, i.e. an increase of 10.81%.

Comparison of the results obtained using the LOCAL algorithm for universal models (i.e. mandarin + orange) with those yielded by MPLS for mandarins alone (Sánchez, De la Haba, & Pérez-Marín, 2013a) and for oranges alone (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b) showed that the $r^2$ values recorded for individual species were below the minimum recommended by Windham, Mertens, & Barton II (1989) for routine use of predictive models in the citrus sector, whereas models constructed using LOCAL regressions strategies for the three morphological parameters studied displayed $r^2$ values of over 0.70, and were therefore suitable for routine use. However, SEP values for the accuracy of predictive models developed using LOCAL strategies were slightly higher than those recorded using the linear regression models for the individual species tested, due to higher SD values in the universal equations.

*3.2.2. Color-related parameters.*

As Table 2 shows, the precision of the models constructed for color parameters (L*, a*, b*, C*, h*) using the LOCAL algorithm may be considered acceptable for screening purposes ($0.50 \leq r^2 \leq 0.63$), enabling values for citrus fruits to be classified as high, medium and low; by contrast, the precision of the universal models developed using MPLS ($0.30 \leq r^2 \leq 0.44$) enabled only classification into high or low (Shenk &

Westerhaus, 1996). The LOCAL-based model enabled routine prediction of parameter b* (blue–yellow), while values for the other parameters came close to threshold values for this purpose. The ability to measure, using a single NIRS instrument, the changes in color from green-yellowish tones (negative a* and positive b*) to orange-reddish tones (positive a* and b*) typically occurring in the course of on-tree ripening, together with the non-destructive estimation of selected morphological parameters is undoubtedly of considerable interest in order to determine the optimal harvesting time.

For all study parameters, application of the LOCAL algorithm improved the accuracy of predictive models; reduction of the SEP for parameter a* was particularly noteworthy ($SEP_{LOCAL}$ = 7.24; $SEP_{MPLS}$ = 11.07).

Comparison of LOCAL results for mandarin + orange with those obtained for individual species using MPLS showed that precision was greater with LOCAL for all parameters except a* and h* in mandarins. SEP values for the universal equations were also better, except for a* and C* which were better in MPLS models for mandarin.

*3.2.3. Internal physical parameters*

Results obtained for the prediction of firmness using the LOCAL algorithm indicate that the predictive capacity of the model, though very low ($r^2$ = 0.28, SECV = 11.63 N), was higher than that obtained with MPLS; the standard error was reduced by 12.89% and the coefficient of determination increased by 250%. Though increased by the application of non-linear regression algorithms, this low predictive capacity underlines the difficulty in correlating destructive measurements made to a puncturing depth of 10 mm with non-destructive NIR measurements, particularly for thick-peel fruits such as this orange variety (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b). As Peirs, Scheerlinck, Touchant, & Nicolaï (2002) have noted, NIR light will only penetrate usefully down to a depth of between 1 and 5 mm, depending on the wavelength, the instrument and the fruit ripeness stage.

For pericarp thickness and juice weight, the robustness of universal models was enhanced by application of the LOCAL algorithm; SEP values were reduced by 8.91% for pericarp thickness and by 6.20% for juice weight, whilst $r^2$ was increased by 11.29% and 7.46%, respectively. Non-destructive prediction of both parameters is of particular interest to the citrus sector, which prizes fruit with reduced peel thickness and high juice content.

Comparison of the results obtained here with those reported by Sánchez, De la Haba, & Pérez-Marín (2013a) and by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) confirms the view expressed by Williams (2001) and Pérez-Marín, Garrido-Varo, & Guerrero (2005), among others, regarding the importance of using a sufficiently-large and sufficiently-varied calibration set for developing global calibration equations. Here, increased sample size and greater uniformity in terms of the number of samples available across the whole range of the test parameter improved the predictive capacity of the models.

The frequency histogram for juice weight is shown in Fig. 1. Juice weight is one of the parameters most affected by sample distribution over the entire range, especially when two different citrus species are tested together; here, the range for oranges (19.33-282.96 g) was much wider than for mandarins (2.62-90.69 g). The effect of combining the two species in calibration sets, in terms of increased range and improved distribution for the juice-weight parameter was evident when comparing the results obtained by Sánchez, De la Haba, & Pérez-Marín (2013a) and by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) for individual species (mandarin: $r^2 = 0.30$; SECV = 14.15 g; RPD = 1.19; orange: $r^2 = 0.33$; SECV = 22.62 g; RPD = 1.21) with those obtained here for combined orange-mandarin sets ($r^2$ = 0.71; SECV = 27.39 g; RPD = 1.85). Shenk, Westerhaus, & Berzaghi (1997) suggest that the samples selected for calibration should include all possible sources of variation encountered during prediction, in order to increase the robustness of the calibration, although this usually decreases the accuracy of prediction. However, the use of the LOCAL algorithm obviates the need to choose between accuracy and robustness of a calibration.

### *3.3. Matching calibration samples for the external prediction of physical quality parameters in citrus fruits using the LOCAL algorithm*

It was considered useful to determine the percentage of each fruit species in the training set used by the LOCAL algorithm to develop prediction models for that species in a combined validation set. Juice weight was the parameter selected for this purpose. Results are shown in Fig. 2.

The LOCAL algorithm applied to the validation set (N = 147 samples; 88 oranges and 59 mandarins) used 80 samples to predict juice weight, rather than the 457 samples used in MPLS regression. In most cases, moreover, samples belonged to the species to be predicted. As Fig. 2a shows, 72 (81.82%) of the 88 oranges in the validation set were predicted with between 80% and 100% of oranges in the training set. In two cases, oranges in the validation set were predicted with less than 40% of orange samples, and between 62% and 71% of mandarin samples from the training set. As Fig. 2b shows, 36 of the 59 mandarins in the validation set (61.02%) were predicted using over 80% of mandarins in the training set.

No previously-published research has addressed the use of non-linear regression methods such as LOCAL to develop predictive models in other fruit species, but Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) used this algorithm to predict the same quality parameters tested here, in oranges, also reporting that LOCAL improved the predictive capacity of models for all parameters with respect to MPLS.

Sánchez, De la Haba, Guerrero, Garrido-Varo, & Pérez-Marín (2011) also found that the use of LOCAL rather than MPLS regression improved models for predicting quality parameters in nectarines using on-tree measurements.

## Conclusions

These findings confirm that NIRS technology using the LOCAL algorithm is a promising tool for the development of universal quality-prediction models for different fruit species belonging to the same genus, thus obviating the need to develop specific models for each species. The results also confirm the viability of NIRS technology, using latest-generation portable instruments, for the development of models enabling monitoring of the physical changes taking place during on-tree ripening. The LOCAL non-linear regression algorithm proved to be considerably more effective for this purpose than MPLS regression. To our knowledge, this is the first attempt to develop universal quality models using on-tree NIR spectroscopy for the genus Citrus. Over the coming years, however, recalibration may be required, increasing the number of samples in the calibration set by adding other species of this genus such as lemons, pomegranates, etc.

## Acknowledgements

Authors acknowledge Prof. Elías Fereres. Ms. Mª Carmen Fernández is also acknowledged for her technical support. This work was funded by the Spanish Ministry of Science and Innovation for the projects CONSOLIDER CSD2006-0067 and AGL2012-40053-C03-01, with participation of funds from FEDER (European Union) and by the Andalusian Regional Government under the Research Excellence Program, project no. P09-AGR-5129 MEMS and NIRS-image sensors for the in situ nondestructive analysis of food and feed. The authors have no conflict of interest to declare.

## References

Antonucci, F., Pallottino, F., Paglia, G., Palma, A., D'Aquino, S., & Menesatti, P. (2011). Non-destructive estimation of mandarin maturity status through portable VIS-NIRS spectrophotometer. *Food and Bioprocess Technology,* 4, 809–813.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43, 772–777.

Barton II., F. E., Shenk, J. S., Westerhaus, M. O., & Funk, D. B. (2000). The development of near infrared wheat quality models by locally weighted regressions. *Journal of Near Infrared Spectroscopy*, 8, 201–208.

Cayuela, J. A. (2008). Vis–NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance. *Postharvest Biology and Technology*, 47, 75–80.

Cayuela, J. S., & Weiland, C. (2010). Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biology and Technology*, 58, 113–120.

Fraser, D. G., Jordan, R. B., Künnemeyer, R., & McGlone, V. A. (2003). Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. *Postharvest Biology and Technology*, 27, 185–196.

Guthrie, J. A., Walsh, K. B., Reid, D. J., & Lienberg, C. J. (2005a). Assessment of internal quality attributes of mandarin fruit I. NIR calibration model development. *Australian Journal of Agriculture Research*, 56, 405–416.

Guthrie, J. A., Reid, D. J., & Walsh, K. B. (2005b). Assessment of internal quality attributes of mandarin fruit II. NIR calibration model robustness. *Australian Journal of Agriculture Research*, 56, 417–426.

Hernández-Gómez, A. H., He, Y., & Pereira, A. G. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIRspectroscopy techniques. *Jounal of Food Engineering*, 77, 313–319.

ISI. (2000). *The complete software solution using a single screen for routine analysis, robust calibrations and networking. Manual*. Silver Spring: Foss NIRSystems/Tecator, Infrasoft International.

Liu, Y., Sun, X., & Ouyang, A. (2010a). Non-destructive measurements of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLS and PCA-BPNN. *LWT-Food Science and Technology*, 43, 602–607.

Liu, Y., Sun, X., Zhang, H., & Aiguo, O. (2010b). Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy. *Computers and Electronics in Agriculture*, 71(S1), S10–S14.

Liu, Y., Heying, E., & Tanumihardjo, S. A. (2012). History, global distribution, and nutritional importance of citrus fruits. *Comprehensive Reviews in Food Science and Food Safety*, 11(6), 530-545.

Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje P. J. R., Saeys, W., & Nicolaï, B. M. (2012a). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food and Bioprocess Technology*, 5, 425-444.

Magwaza, L. S., Opara, U. L., Terry, L. A., Landahl, S., Cronje, P. J., Nieuwoudt, H., Mouazen, A. M., Saeys, W., & Nicolaï, B. M. (2012b). Prediction of “Nules Clementine” mandarin susceptibility to rind breakdown disorder using Vis/NIR spectroscopy. *Postharvest Biology and Technology*, 74, 1-10.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Landahl, S., & Terry, L. A. (2013a). Canopy position affects rind biochemical profile of “Nules Clementine” mandarin fruit during postharvest storage. *Postharvest Biology and Technology*, 86, 300-308.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Nieuwoudt, H. H., Landahl, S., & Terry, L. A. (2013b). Quantifying the effects of fruit position in the canopy on physical and biochemical properties and predicting susceptibility to rind breakdown disorder of "Nules Clementine" mandarin (*Citrus reticulate* Blanco) using Vis/NIR spectroscopy. *Acta Horticulturae* 1007, 83-91

Magwaza, L. S., Opara, U. L., Terry, L. A., Landahl, S., Cronje, P. J. R., Nieuwoudt, H. H., Hanssens, A., & Saeys, W. (2013c). Evaluation of Fourier transform-NIR spectroscopy for integrated external and internal quality assessment of Valencia oranges. *Journal of Food Composition and Analysis*, 31, 144-154.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Nieuwoudt, H. H., Mouazen, A. M., Nicolaï, B. M., & Terry, L. A. (2014). Assessment of rind quality of "Nules Clementine" mandarin using postharvest storage: 1. Vis/NIR PCA models and relationship with canopy position. *Scientia Horticulturae*, 165, 410-420.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., & Theron, I. K., & Lammertyn, J. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46, 99–118.

Olmo, M., Nadas, A., & García, J. M. (2000). Nondestructive methods to evaluate maturity level of oranges. *Journal of Food Science*, 65 (2), 365-369.

Peirs, A., Scheerlinck, N., Touchant, K., & Nicolaï, B. M. (2002). Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems Engineering*, 81, 305–311.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. (2005). Implementation of local algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs. *Applied Spectroscopy*, 59, 69–77.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. (2007). Non-linear methods in NIRS quantitative analysis. *Talanta*, 72, 28–42.

Sánchez, M. T., De la Haba, M. J., Guerrero, J. E., Garrido-Varo, A., & Pérez-Marín, D. (2011). Testing of a Local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. *Postharvest Biology and Technology*, 60, 130-135.

Sánchez, M. T., De la Haba, M. J.; & Pérez-Marín, D. (2013a). Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer. *Computers and Electronics in Agriculture*, 92, 66-74.

Sánchez, M. T., De La Haba, M. J., Serrano, I., & Pérez-Marín, D. (2013b). Application of NIRS for non-destructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest. *Food Analytical Methods*, *6*, 826-837.

Shenk, J. S., & Westerhaus, M. O. (1991). Population definition, sample selection and calibration procedures for near infrared spectra and modified partial least squares regression. *Crop Science*, 31, 469–474.

Shenk, J. S., & Westerhaus, M. O. (1995). *Routine operation, calibration, development and network system management manual*. Silver Spring: NIRSystem, Inc.

Shenk, J. S., & Westerhaus, M. O. (1996). Calibration the ISI way. In A. M. C. Davies & P. C. Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp. 198–202). Chichester: NIR Publications.

Shenk, J. S., Westerhaus, M. O., & Berzaghi, P. (1997). Investigation of a LOCAL calibration procedure for near infrared instruments. *Journal of Near Infrared Spectroscopy*, 5, 223–232.

Sun, X., Zhang, H., & Liu, Y. (2009). Nondestructive assessment of quality of 'Nanfeng' mandarin fruit by a portable near infrared spectroscopy. *International Journal of Agricultural and Biological Engineering*, 2, 65–71.

Williams, P. C. (2001). Implementation of near-infrared technology. In P. C. Williams & K. H. Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (pp. 145–171). St. Paul, MN: AACC, Inc.

Windham, W. R., Mertens, D. R., & Barton, F. E. II, (1989). Protocol for NIRS calibration: Sample selection and equation development and validation. In C. Martens, J. S. Shenk, & F. E. Barton, II (Eds.), *Near infrared spectroscopy (NIRS): Analysis of forage quality. Agriculture handbook, No. 643* (pp. 96–103). Washington, DC: USDA-ARS, US Government Printing Office.

Zheng, Y., He, S., Yi, S., Zhou, Z., Mao, S., Zhao, X., & Deng, L. (2010). Predicting oleocellosis sensitivity in citrus using vis–NIR reflectance spectroscopy. *Scientia Hoticulturae*, 125, 401–405.

Table 1 - Number of samples (N), range, mean, standard deviation (SD) and coefficient of variation (CV) in calibration and validation sets.

| Parameter | Set | N | Range | Mean | SD | CV (%) |
|---|---|---|---|---|---|---|
| Weight (g) | Calibration | 457 | 44.20-598.30 | 243.93 | 108.35 | 44.42 |
| | Validation | 147 | 54.93-561.00 | 239.46 | 113.27 | 47.30 |
| Equatorial diameter (mm) | Calibration | 457 | 38.05-108.34 | 76.52 | 12.60 | 16.47 |
| | Validation | 147 | 41.37-107.18 | 76.10 | 12.93 | 16.99 |
| Axial diameter (mm) | Calibration | 457 | 42.10-113.92 | 73.71 | 16.24 | 22.03 |
| | Validation | 147 | 45.47-107.20 | 72.76 | 16.24 | 22.32 |
| L* | Calibration | 457 | 46.12-79.52 | 65.53 | 4.23 | 6.46 |
| | Validation | 147 | 48.61-70.81 | 65.23 | 3.81 | 5.84 |
| a* | Calibration | 457 | -16.34-42.41 | 21.40 | 12.44 | 58.13 |
| | Validation | 147 | -15.43-41.35 | 21.63 | 13.17 | 60.89 |
| b* | Calibration | 457 | 34.89-78.14 | 64.77 | 7.30 | 11.27 |
| | Validation | 147 | 36.94-76.49 | 64.42 | 6.84 | 10.62 |
| C* | Calibration | 457 | 37.61-81.98 | 69.24 | 8.11 | 11.71 |
| | Validation | 147 | 38.92-80.42 | 69.10 | 7.79 | 11.27 |
| h* | Calibration | 457 | 51.74-112.40 | 72.71 | 10.95 | 15.06 |
| | Validation | 147 | 52.77-108.36 | 72.43 | 11.42 | 15.77 |
| Firmness (N) | Calibration | 457 | 2.07-79.88 | 19.21 | 14.60 | 76.00 |
| | Validation | 147 | 2.65-62.18 | 16.92 | 12.63 | 74.65 |
| Pericarp thickness (mm) | Calibration | 457 | 1.59-10.27 | 5.02 | 1.64 | 32.67 |
| | Validation | 147 | 2.25-9.19 | 5.01 | 1.63 | 32.53 |
| Juice weight (g) | Calibration | 456 | 2.62-282.96 | 102.79 | 52.21 | 50.79 |
| | Validation | 147 | 17.79-260.67 | 101.72 | 52.38 | 51.49 |

Table 2 - Statistics for validation of citrus samples using LOCAL and MPLS regression strategies

| Parameter | Best of LOCAL. Citrus genus | | | MPLS regression | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Citrus genus | | Mandarin[b] | | Orange[c] | |
| | SEP | $r^2$ | Settings[a] | SEP | $r^2$ | SEP | $r^2$ | SEP | $r^2$ |
| Morphological parameters | | | | | | | | | |
| Weight (g) | 58.86 | 0.73 | 120, 14 | 67.96 | 0.65 | 26.12 | 0.39 | 50.32 | 0.38 |
| Equatorial diameter (mm) | 6.96 | 0.71 | 80, 14 | 7.31 | 0.68 | 5.83 | 0.39 | 5.03 | 0.49 |
| Axial diameter (mm) | 6.88 | 0.82 | 80, 15 | 8.37 | 0.74 | 4.36 | 0.31 | 5.18 | 0.51 |
| Color parameters | | | | | | | | | |
| L* | 2.52 | 0.57 | 120, 14 | 2.96 | 0.44 | 2.26 | 0.47 | 1.00 | 0.43 |
| a* | 7.24 | 0.52 | 80, 16 | 11.07 | 0.34 | 8.41 | 0.65 | 1.53 | 0.39 |
| b* | 4.17 | 0.63 | 100, 16 | 5.48 | 0.39 | 3.03 | 0.42 | 1.86 | 0.15 |
| C* | 5.19 | 0.56 | 100, 16 | 6.55 | 0.33 | 5.92 | 0.35 | 1.66 | 0.26 |
| h* | 8.09 | 0.50 | 80, 16 | 9.81 | 0.30 | 6.55 | 0.64 | 1.38 | 0.21 |
| Physical internal parameters | | | | | | | | | |
| Firmness (N) | 11.02 | 0.28 | 80, 14 | 12.65 | 0.08 | 3.03 | 0.15 | 15.05 | 0.30 |
| Pericarp thickness (mm) | 0.92 | 0.69 | 100, 16 | 1.01 | 0.62 | 0.54 | 0.51 | 1.76 | 0.43 |
| Juice weight (g) | 28.13 | 0.72 | 80, 15 | 29.99 | 0.67 | 14.71 | 0.28 | 24.07 | 0.28 |

[a]LOCAL settings: number of selected samples, number of PLS factors.

[b]Mandarins: results in Sánchez et al.. 2013a.

[c]Oranges: results in Sánchez et al.. 2013b.

**Fig. 1. - Distribution of juice weight (g) for mandarins, oranges and mandarins + oranges during on-tree ripening**.

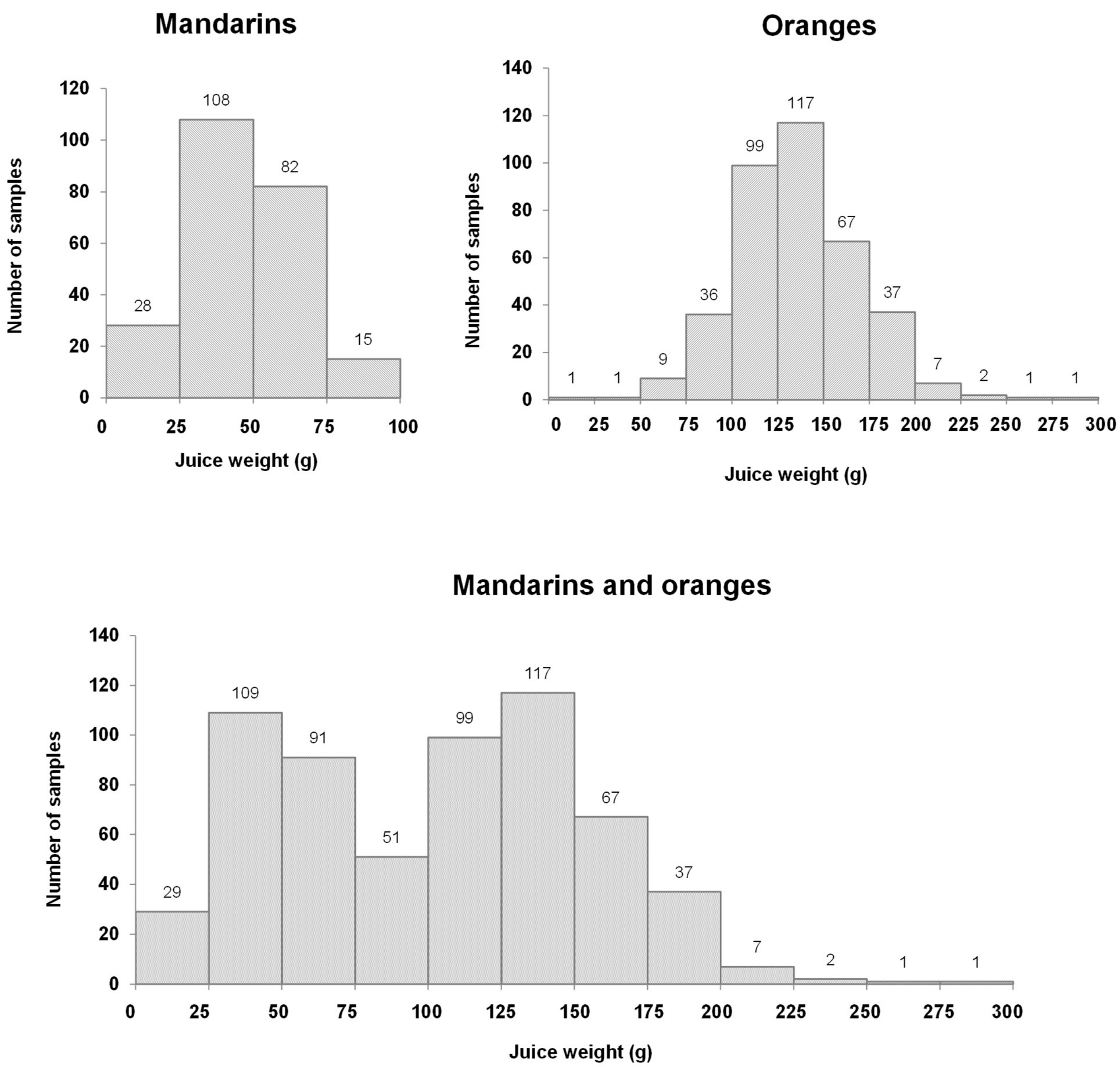

**Fig. 2. - Prediction of the validation set for juice weight using LOCAL algorithm.**