Rapid, simultaneous, and *in-situ* authentication and quality assessment of intact bell peppers using NIRS technology

María-Teresa Sánchez,[a]* Irina Torres,[a] María-José de la Haba,[a] Ana Chamorro,[a] Ana Garrido-Varo,[b] Dolores Pérez-Marín,[b]*

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

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

* *Correspondence to: María-Teresa Sánchez, Department of Bromathology and Food Technology, University of Córdoba, Campus Rabanales, 14071 Cordoba, Spain. E-mail teresa.sanchez@uco.es or Dolores Pérez-Marín, Department of Animal Production, University of Córdoba, Campus Rabanales, 14071 Cordoba, Spain. E-mail: dcperez@uco.es*

Abstract

BACKGROUND: The ability of near infrared (NIR) spectroscopy to authenticate individual bell peppers as a function of the growing system (outdoor or greenhouse) was tested using partial least squares discriminant analysis (PLS-DA). 394 bell peppers grown outdoors (130 samples) or in a greenhouse (264 samples) during the 2015 and 2016 seasons, were selected for this purpose and analysed using a portable, handheld MicroPhazir MEMS instrument (spectral range 1600-2400 nm), working in reflectance. Subsequently, the potential of NIRS as a non-destructive sensor for *in-situ* quality (dry matter and soluble solid content) measurements, was investigated.

RESULTS: The models correctly classified 89.73% and 88.00% of the samples by growing system, when trained with unbalanced and balanced sets, respectively, mainly due to the differences in physical-chemical attributes between bell peppers cultivated in both growing systems. Separate classification models for bell peppers grouped by ripeness (judged by the colour), allowed to classify 88.28%-91.37% of the samples correctly. The standard error of cross-validation (SECV) values for the quantitative models were 0.66% fw and 0.75 ºBrix for dry matter and soluble solid content, respectively.

CONCLUSIONS: The results showed that NIRS can be used successfully for predicting the growing systems used in bell pepper production, which is of particular value to guarantee the authentication of outdoor-grown peppers. Additionally, the results showed that NIRS can be used simultaneously as a rapid preliminary screening technique to measure quality.

Keywords: NIR spectroscopy; Bell pepper; *In situ* authentication; Quality; portable NIR device.

INTRODUCTION

In Spain, peppers are grown almost exclusively indoors - in greenhouses - although in some regions, as in the case of Andalusia, cultivation may take place outdoors.[1]

Bell peppers can be grown in greenhouses more compactly than outdoors. They are also pruned and trained differently: the former are cut back more severely in order to aerate the plants more inside the greenhouse, while in the latter system, the plants are allowed to grow more foliage to protect the peppers from the sun and chilly night-time temperatures.[2]

Likewise, peppers grown indoors have to be trained securely to support the fruits and prevent them from touching the ground or the branches from splitting, since the plants can reach up to 2 meters in height and the stems are far more tender than those grown outdoors. In contrast, peppers grown outdoors reach only one metre in height and do not need to be supported, as the stems are sturdier and do not grow high enough to bend or break.[3]

The variations in growing conditions between peppers grown outdoors and in a greenhouse can make an important difference to the quality of the product, especially in terms of the organoleptic characteristics linked to dry matter and sugar content. It should also be noted that consumer demand currently puts a high value on products which are local, seasonal and traditional, and peppers grown outdoors are favoured by these consumers.[4]

In general, consumers are interested in buying horticultural products obtained using this particular cultivation system and attribute higher quality standards to bell peppers grown in this specific way. It is therefore desirable for the horticultural production sector to have access to non-destructive technology which can carry out fast, highly accurate, *in-situ* analyses to guarantee the authenticity of the growing system. In this way, the consumer will receive accurate information about the differences in quality between vegetables produced using different agronomic techniques and about their origin. In this field, NIR spectroscopy has proved to be an ideal way of providing authentication/certification of raw horticultural materials produced using different types of agricultural methods, as well as for the authentication of varieties.[5-8]

However, no papers in the scientific literature are dealing with the authentication of peppers using NIRS technology based on their origin (outdoor or greenhouse cultivation) and the possible differences in quality between fruits from the different growing systems. In fact, there are very few articles which incorporate the use of this technology to measure quality parameters in pepper. Thus, Sánchez *et al.*[9] assessed the viability of NIRS to measure pesticide residues in intact, crushed, and dry extract system for infrared analysis (DESIR) peppers, while other authors[10-13] carried out the analysis of a number of quality parameters in different types of peppers.

The aim of this study was to evaluate the viability of NIR spectroscopy in providing non-destructive, *in situ* authentication for the growing system - outdoor or in a greenhouse - of bell peppers. In addition, quantitative models were developed to predict two of the main quality parameters (dry matter content and soluble solid content) in intact bell peppers, which could help to classify peppers more easily by their origin. Particular attention was paid to the robustness of the models.

MATERIALS AND METHODS

Sampling

394 bell peppers (*Capsucicum annum* L.) of different colours (green, red and yellow, depending on the degree of ripeness), grown outdoors (N = 130 bell peppers: green = 50, yellow = 41, red = 39) and in a greenhouse (N = 264 peppers: 88 of each colour), picked

in the 2015 and 2016 seasons, were analysed. The greenhouse samples were grown in the Region of Murcia (Spain), while the outdoor peppers were harvested in Santaella (Córdoba, Spain).

Spectral set

A handheld MEMS (micro-electro-mechanical system)-based NIR digital transform spectrometer (MEMS-NIRS) (MicroPhazir, Polychromix Inc., Wilmington, MA, USA), working in reflectance mode in the spectral range 1600-2400 nm with a non-constant interval of around 8 nm was used to collect the NIR spectra of all the samples in reflectance mode. Sensor integration time was 600 ms. The device was equipped with quartz protection to prevent dirt accumulation. Each spectrum was the mean of 5 scans with a lamp warm-up time of 45 seconds. To obtain the NIR spectra, four measurements were taken at the equatorial region of the fruits, which were then rotated 90° after each measurement. The four spectra were averaged to provide a mean spectrum for each fruit.

Measurement of physical-chemical quality parameters

Dry matter content was measured by desiccation at 105ºC for 24 h[14]; the final dry weight was calculated as a percentage of initial wet weight. Soluble solid content (SSC) in ºBrix was measured as the refractometer reading for the pepper juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany).

All the samples were analysed in duplicate and the standard error of laboratory (SEL) was estimated from these duplicates (Table 7). All the measurements were performed immediately after taking the NIRS measurements.

Data Processing

Before the spectral data was processed and using the WinISI II software package version 1.50 (Infrasoft International, Port Matilda, PA, USA), a study was conducted to select the most suitable spectral range for the instrument to carry out the authentication and quality control of peppers. To achieve this, the 1,1,1,1 derivation treatment was applied (the first digit being the number of the derivative, the second the gap over which the derivative is calculated, the third the number of data points in a running average or smoothing, and the fourth the second smoothing) without scatter correction, which allows to highlight the areas of the spectrum where the signal/noise ratio is degraded.[15]

Spectral repeatability

Spectrum quality was evaluated using the Root Mean Square (RMS).[16-17] This statistic indicates the similarity between different spectra of a single sample: in this case, between the four spectra collected per sample. To establish a threshold for this statistic, 36 bell peppers were selected, from which four spectra were taken in the equatorial region, rotating the fruit 90º after each measurement. An admissible limit for spectrum quality and repeatability was set following the procedure described by Martínez *et al.*[18] to calculate the standard deviation (STD) limit from the RMS statistic and obtain an RMS cut-off value.

Principal component analysis

In order to study the relationship between the quality (dry matter and SSC) of the bell pepper and the growing system used (outdoor or greenhouse), Principal Component Analysis (PCA) was carried out.

PCA is a mathematical procedure that reduces the dimensionality of the data to uncorrelated variables, including in each component the maximum residual variance of the data, and each component therefore contains a representation of the data variation.[19] The PCA scores represent the weighted sums of the original variables without significant loss of useful information, and loadings (weighting coefficients) were used to identify the major variables responsible for specific features appearing in the scores. Matlab software (version 2015a, The Mathworks, Inc., Natick, Massachusetts, USA) was used to conduct PCA, using the mean centre, by which the mean spectrum of the group is subtracted from each spectrum, as a pre-treatment.[20]

Authenticating bell peppers by growing method using NIR spectroscopy

To carry out the non-destructive authentication of bell peppers according to their growing system, discriminant models were designed to classify the peppers into two groups: bell peppers grown outdoors and bell peppers grown in a greenhouse.

Firstly, the spectral structure and variability of the sample population was studied to select the samples which would make up the training group. The CENTER algorithm was used for this, which is included in the WinISI II version 1.50 software. This algorithm was applied separately to each of the two training groups (130 outdoor-grown bell peppers and 264 greenhouse bell peppers). The mathematical treatments SNV (Standard Normal Variate) and DT (Detrend) for scatter correction were applied,[21] and the 1,5,5,1 derivate mathematical treatment.[22-23] 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 4.5 were considered outliers or anomalous spectra.[16] After discarding outliers, the sampling groups consisted of 128 samples of bell pepper grown outdoors and 259 samples of greenhouse-grown bell pepper.

After ordering the sample set by spectral distances (from smallest to greatest distance from the center), a structured selection of an external validation set (28 samples for each classification group) was performed following Shenk and Westerhaus.[24]

The difficulty involved in obtaining balanced learning groups in terms of the number of samples per class or classification category meant that its influence on the predictive capacity of the models had to be assessed. The results obtained were therefore contrasted with balanced and unbalanced classification models as regards the number of samples per class.

The samples of the balanced groups were selected using the SELECT algorithm included in the WinISI II software package version 1.50.[24] This algorithm enables spectral selection of a number of samples which are representative of the population as a whole, by calculating the 'NH' distance (Mahalanobis neighbour distance) between two spectra. An 'NH' of less than 0.6 implies that two spectra are too similar to each other ('neighbours'). After this algorithm was applied, 100 samples of the 'greenhouse-grown peppers' group were selected, thus leaving the number of samples of the training group for the two classes equal, and the classification models were then developed.

Discriminant models were constructed to authenticate bell peppers according to their growing system, using PLS-DA for supervised classification.[23] Briefly, PLS-DA uses a training set to develop a qualitative prediction or calibration model that may subsequently be applied for the classification of new unknown samples. This model seeks to correlate spectral variations (X) with defined classes (Y), attempting to maximise the covariance between the two types of variable. In this type of approach, the Y variables used are not continuous, as they are in quantitative analysis, but rather categorical 'dummy' variables created by assigning different values to the different classes to be distinguished. Specifically, the PLS2 algorithm was applied, using the ''Discriminant Equations'' option in the WINISI II version 1.50 software.

All the models were designed using four cross validation groups (i.e. the calibration set was divided into four groups, and each group was then predicted using a calibration obtained from the other samples), a spectral range from 1600 to 2168 nm, signal noise eliminated at the end of the spectral range, and combined SNV+DT treatment for scatter correction. First- and second-derivative treatments were tested by applying 1,5,5,1 and 2,5,5,1.[22]

The precision of the models obtained was evaluated using the percentage of correctly-classified samples, both for the model and for each class. In addition, the standard error of cross validation (SECV) was evaluated. Most of the papers use the value of 1.5 as discrimination limit, so that, if one sample obtain a variable value over the limit for a given class, it will be classified as belong to this class. However, in this paper it was also used the minimum difference (MD) value, calculated as the product of the value of the model's uncertainty factor (1.5) by the SECV, for the detection of uncertain samples when interpreting the results obtained. Samples with a MD higher than the MD value calculated should be considered as uncertain.[25] Regression coefficients were also used to discuss the contributions of individual wavelengths to the qualitative PLS models. [26]

Next, after selecting the best classification model for each of the established types (unbalanced and balanced models), these were validated. In this case, an external validation procedure was also carried out to determine the predictive capacity of the model using a sample group different to that used in the training of the model. In both models (unbalanced and balanced), 56 samples were selected in a structured way (28 samples for each of the culture systems: outdoor or greenhouse).[24]

Then, after analysing the results of the statistical tests which evaluated the influence of the cultivation systems and the state of ripeness (reflected by colouration) on the quality of the bell peppers harvested, new classification models were developed for the peppers according to the growing system, but also taking into account the colouration (green, yellow and red).

For each of the colours analysed, the structure and spectral variability of the sample population was studied for each of the growing systems, using the same methodology described above. A structured selection of the external validation set (5 samples for each classification group due to the low number of samples in the 'outdoor-grown' category, once the bell peppers were separated by colour), solely on the basis of spectral data, was performed following Shenk and Westerhaus.[24]

Quantitative models: sets, calibration development and validation procedure

Quantitative models were built to predict the parameters of dry matter and SSC, using all the available bell pepper samples, independently of the cultivation system used. The samples for the calibration and validation groups were selected by applying the CENTER algorithm in the 1600-2168 nm spectral range. The pre-treatments SNV and DT were used for scatter correction,[21] and one derivative mathematical treatment (Norris derivative) was performed (1,5,5,1).[22-23] Thus, having ordered the sample set by spectral distances (from smallest to greatest distance to the centre) and once outlier spectrum samples were eliminated (N = 2), the 130 samples forming the validation set were selected by taking one of every 3 samples in the initial 392-sample set; the calibration set thus comprised the remaining 262 samples.

Modified partial least squares (MPLS) regression[16] was used to obtain equations for predicting dry matter and SSC. All the models were constructed using four cross-validation groups. The same signal pre-treatments and spectral region described earlier for authentication analysis were used for designing the quantitative models. The statistics used to select the best equations were: standard error of calibration (SEC), coefficient of determination of calibration (r^2_c), standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2_{cv}), RPD_{cv} or ratio of the standard deviation of the original data (SD) to SECV and the coefficient of variation (CV), defined as the percentage ratio of the SECV to the mean value of the reference data for the calibration set. These latter two statistics enable SECV to be standardized, facilitating the comparison of the results obtained with sets of different means.[27] Regression coefficients were also used to discuss the contributions of individual wavelengths to the quantitative models.[26]

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 130 samples, not used previously in the model, was evaluated following the protocol outlined by Windham *et al.*[28]

Statistical analysis

All the quantitative analyses were expressed as mean values ± standard deviation. The data for each attribute (dry matter and SSC) for outdoor and greenhouse bell peppers were analysed statistically by analysis of variance (ANOVA) using Statgraphics Centurion XV (StatPoint Inc., Warrenton, North Virgina, USA), and initially considering the origin of the pepper (outdoor or greenhouse cultivation) as a factor. Next, in order to study the influence of both the growing technique and the pepper colouring in the dry matter and soluble solids contents, a two-factor ANOVA variance analysis was carried out.

In both cases, the difference between the means were compared with the Fisher's Least Significant Difference (LSD) test, and differences at $P < 0.05$ were considered to be significant.

RESULTS AND DISCUSSION

Optimal spectral region and spectral repeatability

Prior to the model development, it was necessary to optimise the NIRS analysis by means of the spectrum quality and repeatability measurement.

The existence of noise in the spectrum was evaluated (spectral range 1600-2400 nm). To this end, the derivative treatment 1,1,1,1 was applied in order to determine the area of the spectral range affected by noise, given that it degrades the signal/noise relationship.[15] After this process, the spectral range between 2169–2400 nm was eliminated, and all the models were designed using the spectral range 1600–2168 nm.

Spectral repeatability is crucial to the construction of models that are both accurate and robust. Statistical methods such as a defined RMS cut-off limit can be useful for this purpose.

The mean STD for the samples analysed was 108,733 µlog (1/R), representing an RMS cut-off of 122,144 µlog (1/R).[18] Any sample whose quadruplicated screening scans yielded an RMS above this value was eliminated and the scan was 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 peppers has been found in the literature, although this statistic is essential for generating the representative libraries.

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

Principal Component Analysis

PCA was performed on the set comprising the spectra recorded for each growing system (outdoor or greenhouse) of intact bell peppers.

Fig. 1a displays scores of the second and third components of the PCA model. These two components were chosen because, although the first two principal components (PC1 and PC2) represented a high proportion of the explained variance 94.18% and 5.46%, respectively, they did not facilitate the grouping of the samples according to the growing system used; this grouping does however seem to become more evident when the latent variables PC2 and PC3 are used. Fig. 1b shows the PCA loadings for intact bell peppers in the spectral range 1600-2168 nm.

The graphic representation of the loadings for PC2 and PC3 (Fig. 1b) shows that the main absorption peaks for differentiating between the two growing systems of the bell peppers are those related to carbohydrates and water, respectively. The PC3 weighting coefficient exhibits a band of water around 1930 nm.[29] The peak points down so more water (less dry matter) means a more negative score on PC3, which is exactly what the greenhouse-grown peppers show (Table 1). PC2 exhibits a band that is characteristic of carbohydrates (~1680 nm)[29].

In the light of the PCA scores (Fig 1a) and bearing in mind the results of the ANOVA and LSD tests (Table 1) about the similarities or not in physical–chemical composition between bell peppers cultivated outdoors or in a greenhouse, it may be said that dry matter is indeed related to PC3 and significant differences ($P < 0.05$) were found for dry matter between both types of bell peppers. The positive PC3 scores are associated with fruits of higher dry matter content, while the negative PC3 scores are linked to fruits with lower dry matter values. As has already been mentioned, PC2 may be linked to carbohydrate content, and considering that no significant difference ($P > 0.05$) was found for SSC between outdoor and greenhouse bell peppers (Table 1), no grouping of samples by SSC was apparent using this component.

Authentication of bell peppers by NIRS

Values obtained for number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) for each of the quality parameters measured for the training and validation sets used in the discriminant models for the authentication of bell peppers by growing system are shown in Table 2.

Table 3 shows the results for the best classification models obtained, using PLS-DA, to authenticate the origin of the intact bell peppers analysed (grown outdoors or in a greenhouse).

The most accurate models were achieved using D_1 log (1/R), for both unbalanced and balanced sets. The total percentages of correctly classified samples were 89.73% and 88.00% for the unbalanced and balanced model, respectively. These results, regardless of the population size, confirm those obtained by Pérez-Marín *et al.*[30], who showed that PLS2 is less sensitive to the fact that the populations are unbalanced.

For the unbalanced model, 74 samples of the 100 forming the training group of outdoor-grown peppers were correctly classified, while for the greenhouse-grown peppers, 223 samples out of 231 were correctly classified. It is also important to note that of the 26 samples poorly classified in the outdoor-grown bell pepper category, 17 were within the 1.5 ± MD limit, while for peppers grown in the greenhouse, 7 out of the 8 poorly classified samples are also within this limit.

For the balanced model, 88 samples of the 100 contained in each of the two established training groups (outdoor and greenhouse) were correctly classified. In this case, 11 out of the 12 samples poorly classified in the ‘outdoor-grown bell pepper’ category were within the limits established by the uncertainty factor ± MD, while for the greenhouse pepper category, the 12 misclassified samples were also within this limit.

The models were then validated, using samples not included in their design. In the models created from the unbalanced populations, the percentage of correctly classified samples was 78.57% and 100.00%, for the outdoor and greenhouse cultivation systems, respectively (Fig. 2 and Table 3). Out of the 6 badly-classified samples in the ‘outdoor-grown’ bell pepper category, 4 were in the interval between the uncertainty factor ± MD.

In the case of the balanced populations, 85.71% of the peppers grown outdoors and 96.43% of greenhouse-grown bell peppers were correctly classified (Fig. 2 and Table 3). It is important to note how the 4 poorly classified samples in the ‘outdoor-grown bell pepper’ category were within the range of 1.5 ± MD, which was the same case as the single badly-classified sample from the ‘greenhouse-grown bell pepper’ category.

For the balanced model, the point clouds hardly change, but the threshold moves towards the outdoor-grown samples in that case (Fig. 2). The consequence of this is that for the smaller group the total accuracy is 78.57% when using the unbalanced set and increases to 85.71% when using the balanced set. Despite a low reduction in the accuracy of the larger set (100% *versus* 96.43%) when using the balanced set, the results for the smaller group set improve.

To examine more deeply the results of the classification models obtained, the results of the ANOVA (dry matter and SSC) tests and the LSD (dry matter) test (Table 1) were also considered, along with the results of the PCA (Fig. 1). Significant differences ($P < 0.05$) were detected in terms of the dry matter content between both types of bell peppers (the dry matter content was significantly higher in peppers grown outdoors), and the SSC content was higher - although not significantly ($P > 0.05$) - in the outdoor group.

Likewise, as stated above, it is the PC3 related to water content and, therefore, to dry matter content, which facilitates the classification of bell peppers according to the cultivation system in which they are grown.

An ANOVA analysis was later carried out to study both the influence of the cultivation system used and the colouring of the pepper, which indicates its state of ripeness, on the dry matter and SSC in the bell peppers analysed. The results of the ANOVA test for the parameter dry matter content pointed to the existence of significant differences ($P < 0.05$) between the cultivation systems and colouration, as well as in the interaction between the cultivation system and the colouration. For SSC, no significant differences ($P > 0.05$) were detected between the cultivation systems and in the interaction between the cultivation system and the colouration. However, significant differences were detected ($P < 0.05$) between peppers of different colourations. The results of the Fisher's tests are shown in Table 4.

After analysing the results of the ANOVA and Fisher's tests, new models were designed to classify the peppers according to the cultivation system used and taking the colour into account.

Values obtained for number of samples (N), range, mean, SD and CV for each of the quality parameters measured for the training and validation sets used in the discriminant models for the authentication of green, yellow and red bell peppers by growing system are shown in Table 5.

The results obtained for the best classification models for bell peppers according to the cultivation system used and taking the colour into account are shown in Table 6.

For peppers with green colouration, 113 samples of the 128 available were correctly classified; of these, the model correctly classified 35 of the 45 samples in the outdoor-grown category and 78 of the 83 samples in the greenhouse category. When these models were externally validated, all the samples were correctly classified in the right category.

For yellow bell peppers, the percentage of samples correctly classified in the training group was 91.37% (106 out of 116), with percentages of 87.87% (29 of 33) and 92.77% (77 of 83) for peppers grown outdoors and in the greenhouse, respectively. When the models were validated, the 5 selected samples of outdoor-grown peppers were correctly classified, while 80% of the greenhouse peppers were correctly classified.

In the case of red bell peppers, 107 out of the 118 samples were correctly classified (90.67%). Category by category, 28 out of 35 samples were in the outdoor category and 79 out of 83 in the greenhouse category. When the models were validated, 80% and 100%, respectively, of the peppers from the outdoor and greenhouse categories were correctly classified.

The results of the classification models obtained (Tables 3 and 6) show that using NIR technology to predict the cultivation system of the intact bell peppers is a feasible option and it can be used to authenticate the origin of these vegetables.

Fig. 3 shows characteristic peaks and valleys that indicate which wavelength ranges are important for the balanced classification model of bell peppers by growing system. The figure indicates that the most relevant regression coefficients are located in the region 1660-1880 nm which is associated to the absorption band of a C-H stretching first overtone corresponding to sugars. [31-32] Other relevant coefficients appear in the regions 1930-1990 nm, related to water absorption [31] and 2064-2144 nm, also related with different types of sugars. [32]

In the scientific literature, no predictive models have been found based on NIRS to authenticate the origin of bell peppers depending on the culture system used. Only Sánchez *et al.*[9] assessed the feasibility of using NIR spectroscopy to classify peppers according to the presence of pesticide residues, confirming that NIRS technology may be used to provide swift, non-destructive preliminary screening for pesticide residues.

Predicting quality parameters in bell peppers using MPLS regression

Table 7 shows the characteristics of the calibration and validation sets used to develop the predictive models for dry matter and SSC.

Structured selection based wholly 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 the study parameters; moreover, the established ranges of the validation lay within those of the calibration set.

Table 8 shows the best calibration equations for the two quality parameters selected in bell peppers. For predicting dry matter and SSC in bell peppers, the models constructed allow to discriminate between high, medium and low values of these parameters.[17, 27]

As regards the dry matter parameter, Ignat *et al.*[12] reported predictive capacity (RPD_{cv} = 3.8) higher than those obtained here using a diode array instrument (spectral range: 477-950 nm), although these authors used a wider calibration set since they chose fruits picked during the growing season, from the 34th day after anthesis (DAA) until full ripening (88th DAA), and when fully grown.

For the SSC parameter, Penchaiya *et al.*[10] used a diode array spectrophotometer (spectral range 780-1690 nm) to obtain predictive capacity (RPD_{cv} = 2.08) superior to that of this research work, although the window for the spectrophotometer used (Corona Fiber VIS / NIR, Carl Zeiss Jena GmbH, Germany) was much wider than that of the instrument used here, and its measurement range was also different. In addition, these authors used a wide range of sample attribute in the calibration set, obtained by random harvesting at various stages of ripeness.

Also, for SSC, Ignat *et al.*[12] used the same instrument and spectral range and obtain predictive capacity (RPD_{cv} = 3.9); it is important to stress the greater variability of the fruits used, which also affected the 'dry matter' parameter, as aforementioned.

Toledo-Martín *et al.*[13], using an instrument based on MEMS technology with a 1000-1800 nm spectral range, obtained models for SSC with a predictive capacity (RPD_{cv} = 1.7) very similar to that obtained in this work.

When these results are compared with those of other authors, the importance of the spectrophotometer's measurement window can be seen for the robustness of the developed models. While MEMS instruments perform isolated readings on the product being studied with measurement windows of an area of only around 4 mm^2, the diode array instruments tested by the authors quoted above perform a scan of the whole sample, which is of vital importance in hollow, irregularly-shaped vegetables such as bell peppers.

Validations of the best calibration models obtained were performed using a set comprising 130 samples (Fig. 4).

For dry matter and SSC, it should be stressed that bias lay within confidence limits for both parameters, although SEP(c) and r^2_p results did not attain the recommended values for their routine use in equations,[28] indicating that the NIRS equations constructed should be regarded as a first step in the finetuning of NIRS technology for the *in-situ* monitoring of internal quality parameters in this type of pepper.

The SEL values for dry matter and SSC were 0.21% fw and 0.06 ºBrix, respectively (Table 7). Such a small SEL for SSC must be correctly interpreted when it is compared with the SEP value obtained for the prediction model. Firstly, it must be considered that sugar distribution is heterogenous in the fruit. It is for this reason that in the NIR analysis four spectra were taken in the equatorial region of the fruit. However, the reference value was obtained as the refractometer reading for the pepper juice. It means that the sampling error is included in the SEP value but not in the SEL value. Consequently, NIRS model developed for SSC was characterised by questionable performance, since SEP value obtained exceeded 5*SEL.[15] For dry matter, SEP fell between 3 and 4 SEL, indicating acceptable performance of NIRS model developed.

These findings must be considered for the correct interpretation of the statistic SEP in intact fruits and vegetables. Likewise, the use of handheld NIRS spectrometers is justified given the fact that they ensure in a short period of time, a more precise and accurate guarantee of internal quality of the horticultural product analysed, allowing increased sampling either on the surface of the product tested or in the batch produced.

Finally, the regression coefficients for the best predictive models for dry matter and SSC are illustrated in Fig. 5. These regression coefficients show significant importance for the region around 1650–1850 nm which correspond to the first overtone of the C-H stretching bonds and at around 1920-1960 nm due to O-H group contribution. The absorbance region at 2040–2100 nm could be attributed to NH and OH stretching modes besides C=O vibration bands. [29]

It is also important to point out that the most relevant peaks and valleys coincide in Fig. 3 (qualitative model) and in Fig. 5 (quantitative models). These results reinforce the idea that the discrimination between outdoor-grown and greenhouse-grown bell peppers has a scientific explanation based on the differences in dry matter and SSC between both type of bell peppers. Nicolaï *et al.* [33] indicate that the water absorption bands dominate the spectrum of fruit and vegetables, and it is not likely that minor constituents can be measured well. The authors also state that evidently, when the concentration of such a minor constituent is correlated to, e.g., sugar content, the calibration results may seem reasonable. From the observation of Fig. 5, it can be detected that while peaks and valleys at 1680-1690 nm and 1800 nm are relevant for both, dry matter and SSC, the valley at 1776 nm only dominates in the SSC spectrum. The scientific literature concerning absorptions bands in fruits and vegetables are dominated by papers which use a more limited range of wavelengths that the considered in this study, due to many of them use transmittance instruments. Therefore, further studies are needed to confirm the bands indicated as the most relevant ones in the spectral region analysed.

CONCLUSIONS

The results confirm that NIR spectroscopy using a portable manual instrument based on MEMS technology can be used at any time in the food chain (from the field to the dinner table) to authenticate intact bell peppers depending on the type of cultivation (outdoor *versus* greenhouse) used for growing the crop. Also, NIRS technology could be used as a fast and *in-situ* preliminary screening technique for the classification of bell peppers by dry matter and SSC. However, further research is needed to make the quantification of these parameters more robust.

ACKNOWLEDGEMENTS

This research was carried out under the Research Project 'Quality determination of pepper grown in Santaella (Córdoba)', funded by Gelagri Ibérica, S.L. The authors are grateful to Mrs. María Carmen Fernández of the Animal Production Department for her technical assistance. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres in the form of the Training Programme for Academic Staff (FPU).

REFERENCES

1 López-Marín J, Porras I, Ros C and Brotons-Martínez JN, Study of the profitability of pepper cultivation (*Capsicum annum*) in greenhouse with the use of shading. *ITEA-Inf Tec Econ Ag* **112**:57 – 71 (2016) [in Spanish].

2. Maroto JV, *Special Herbaceous Horticulture*. Mundi-Prensa, Madrid (2002) [in Spanish].

3. Serrano Z, *Handbook of Pepper Cultivation*. Consejería de Agricultura y Pesca, Junta de Andalucía, Seville (2011) [in Spanish].

4. Meca D and Céspedes A, Varietal trends in protected horticulture. *Vida Rural* **406**:54 – 58 (2015) [in Spanish].

5. Pérez D, Sánchez MT, Cano G and Garrido A, Authentication of green asparagus varieties by Near-Infrared Reflectance Spectroscopy. *J Food Sci* **6**:323 – 237 (2001).

6. Sánchez MT, De La Haba MJ, Benítez-López M, Fernández-Novales J, Garrido-Varo A and Pérez-Marín D, Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *J Food Eng* **110**:102 – 108 (2012).

7. Sánchez MT, Garrido-Varo A, Guerrero JE and Pérez-Marín D, NIRS technology for fast authentication of green asparagus grown under organic and conventional production systems. *Postharvest Biol Technol* **85**:116 – 123 (2013).

8. Torres I, Pérez-Marín D, De la Haba MJ and Sánchez MT, Fast and accurate quality assessment of Raf tomatoes using NIRS technology. *Postharvest Biol Technol* **107**:9 – 15 (2015).

9. Sánchez MT, Flores-Rojas K, Guerrero JE, Garrido-Varo A and Pérez-Marín D, Measurement of pesticide residues in peppers by near infrared reflectance spectroscopy. *Pest Manag Sci* **66**:580 – 586 (2010).

10. Penchaiya P, Bobelyn E, Verlinden BE, Nicolaï BM and Saeys W, Non-destructive measurement of firmness and soluble solids content in bell pepper using NIR spectroscopy. *J Food Eng* **94**:267 – 273 (2009).

11. Ignat T, Schmilovitch Z, Feföldi J, Bernstein N, Steiner B, Egozi H *et al.,* Nonlinear methods for estimation of maturity stage, total chlorophyll, and carotenoid content in intact bell peppers. *Biosyst Eng* **114**:414 – 425 (2013).

12. Ignat T, Alchanatis V and Schmilovitch Z, Maturity prediction of intact bell peppers by sensor fusion. *Comput Electron Agr* **104**:9 – 17 (2014).

13. Toledo-Martín EM, García-García MC, Gómez P, Moreno-Rojas JM, González A, Moya M *et al.,* Application of visible/near-infrared reflectance spectroscopy for predicting internal and external quality in pepper. *J Sci Food Agric* **96**:3114 – 3125 (2016).

14. AOAC, *Official Methods of Analysis of AOAC International*, 17th ed. AOAC, Gaithersburg, MD (2000).

15. Hruschka WR, Data analysis: Wavelength selection methods, in *Near-Infrared Technology in the Agricultural and Food Industries*, ed. by Williams PC and Norris KH. American Association of Cereal Chemists, Inc., St. Paul, MN, pp. 39 – 58 (2001).

16. Shenk JS and Westerhaus MO, *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy*. NIRSystems, Silver Spring, MD (1995).

17. Shenk JS and Westerhaus MO, Calibration: the ISI way, in *Near Infrared Spectroscopy: The Future Waves*, ed. by Davies AMC and Williams PC. NIR Publications, Chichester, pp. 198 – 202 (1996).

18. Martínez ML, Garrido A, De Pedro EJ and Sánchez L, Effect of sample heterogeneity on NIR meat analysis: the use of the RMS statistic. *J Near Infrared Spectrosc* **6**:313 – 320 (1998).

19. Mark H, Data analysis: Multilinear regression and Principal Component Analysis, in *Handbook of Near-Infrared Analysis* ed. by Burns DA and Ciurczak EW. CRC Press, Florida, Vol. 8, pp. 151 – 188 (2001).

20. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W and Koch RS, *PLS_ToolBox 4.0. Manual for Use with MATLAB (TM) [Computer software]*. Eigenvector Research, Inc., Wenatchee, WA (2006).

21. Barnes RJ, Dhanoa MS and Lister SJ, Standard Normal Variate transformation and De-trending of near infrared diffuse reflectance spectra. *Appl Spectrosc* **43**:772 – 777 (1989).

22. Shenk JS and Westerhaus MO, *Routine Operation, Calibration, Development and Network System Management Manual*. NIRSystems, Silver Spring, MD (1995).

23. ISI, *The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibration, and Networking. Manual, FOSS NIRSystems/TECATOR*. Infrasoft International, LLC, Silver Spring, MD (2000).

24. Shenk JS and Westerhaus MO, Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci* **31**:1548 – 1555 (1991).

25. Murray I, Aucott LS and Pike HI, Use of discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fishmeal with meat and bone meal. *J Near Infrared Spectrosc* **9**:297 – 311 (2001).

26. Martens H and Naes T, *Multivariate Calibration.* John Wiley & Sons, Chichester (1989).

27. Williams PC, Implementation of near-infrared technology, in *Near Infrared Technology in the Agricultural and Food Industries*, ed. by Williams PC and Norris KH. American Association of Cereal Chemists, Inc., St Paul, MN, pp. 145 – 171 (2001).

28. Windham WR, Mertens DR and Barton II FE, Protocol for NIRS calibration: sample selection and equation development and validation, in *Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality (USDA-ARS Agriculture Handbook No. 643)*, ed. by Martens GC, Shenk JS and Barton II FE. US Government Printing Office, Washington, DC, pp. 96 – 103 (1989).

29. Shenk J, Workman JJJr and Westerhaus M, Application of NIR spectroscopy to agricultural products, in *Handbook of Near-Infrared Analysis*, ed. by Burns DA and Ciurczak E. Marcel Dekker, Basel, pp. 419 – 474 (2008).

30. Pérez-Marín D, Garrido-Varo A and Guerrero JE, Optimization of discriminant partial least squares regression models for the detection of animal by-product meals in compound feeding stuffs by near-infrared spectroscopy. *Appl Spectrosc* **60**:1432 – 1437 (2006).

31. Osborne BG, Fearn T and Hindle P, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific and Technical, London (1993).

32. Downey G, Fouratier V and Kelly JD, Detection of honey adulteration by addition of fructose and glucose using near infrared transflectance spectroscopy. *J Near Infrared Spectrosc* **11**:447 – 456 (2003).

33. Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI and Lammertyn J, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* **46**:99 – 118 (2007).

Table 1. Dry matter and SSC in outdoor-grown and greenhouse-grown bell peppers

Growing system	Parameter	
	Dry matter (% fw)	SSC (ºBrix)
Outdoor	7.02 (1.30) ^{a}	6.38 (1.38) ^{a}
Greenhouse	6.63 (1.10) ^{b}	6.37 (1.30) ^{a}

[1] Standard deviations in brackets

[2] Different letters in the same column indicate statistical significance ($P < 0.05$)

Table 2. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) of the quality parameters for the different training and validation sets used in the discriminant models for the authentication of bell peppers by growing system.

	Training set								Validation set			
	Unbalanced model				Balanced model							
	Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse	
	DM [1]	SSC [2]	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC
N	100	100	231	231	100	100	100	100	28	28	28	28
Range	4.48-11.37	3.85-9.50	4.52-9.94	3.85-10.05	4.48-11.73	3.85-9.50	4.74-9.24	3.90-10.05	4.74-9.49	4.20-9.15	5.05-8.67	4.50-8.50
Mean	7.03	6.44	6.65	6.41	7.03	6.44	6.51	6.23	7.07	6.19	6.48	6.11
SD	1.27	1.38	1.13	1.34	1.27	1.38	1.07	1.35	1.42	1.41	0.93	1.07
CV (%)	18.07	21.43	16.99	20.90	18.07	21.43	16.44	21.67	20.08	22.78	14.35	17.51

[1] DM: Dry matter (% fw)

[2] SSC: Soluble solid content (ºBrix)

Table 3. Discriminant models for the authentication of bell peppers by growing system. PLS-DA

Qualitative Group	Unbalanced model		Balanced model	
	Percentage of correctly-classified samples: 89.73% (297/331) Model SECV: 0.32 Number of synthetic variables: 11 Mathematical treatment: 1,5,5,1-SNV+DT		Percentage of correctly-classified samples: 88.00% (176/200) Model SECV: 0.35 Number of synthetic variables: 9 Mathematical treatment: 1,5,5,1-SNV+DT	
Growing system	Training set	Validation set	Training set	Validation set
Outdoor	74.00% (74/100)	78.57% (22/28)	88.00% (88/100)	85.71% (24/28)
Greenhouse	96.54% (223/231)	100.00% (28/28)	88.00% (88/100)	96.43% (27/28)

Table 4. Dry matter and SSC in outdoor-grown and greenhouse-grown bell peppers

Growing system	Parameter					
	Dry matter (% fw)			SSC (ºBrix)		
	Green	Yellow	Red	Green	Yellow	Red
Outdoor	5.91 (0.70) ^{b}	7.46 (0.96) ^{d}	7.96 (1.16) ^{e}	5.06 (0.63) ^{a}	6.77 (0.81) ^{b}	7.61 (1.00) ^{c}
Greenhouse	5.62 (0.59) ^{a}	6.58 (0.67) ^{c}	7.69 (0.83) ^{d,e}	4.92 (0.49) ^{a}	6.62 (0.65) ^{b}	7.59 (0.88) ^{c}

[1] Standard deviations in brackets

[2] Means with different superscripts differ significantly ($P < 0.05$)

Table 5. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) of the quality parameters for the different training and validation sets used in the discriminant models for the authentication of bell peppers of different colours by growing system.

	Green bell peppers								Yellow bell peppers								Red bell peppers							
	Training set				Validation set				Training set				Validation set				Training set				Validation set			
	Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse	
	DM[1]	SSC[2]	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC
N	45	45	83	83	5	5	5	5	33	33	83	83	5	5	5	5	35	35	83	83	5	5	5	5
Range	4.48-7.68	3.85-7.90	4.52-7.23	3.85-6.10	5.26-7.33	4.60-6.00	5.33-6.21	4.65-5.60	4.97-9.64	4.80-8.35	5.14-8.33	5.30-8.40	7.04-8.63	6.40-7.40	5.53-7.69	5.90-8.25	5.11-11.73	5.05-9.20	5.66-9.50	5.60-10.05	8.09-8.82	7.30-9.50	6.80-9.94	6.70-8.80
Mean	5.90	5.06	5.61	4.92	6.07	5.10	5.76	4.95	7.43	6.80	6.56	6.60	8.05	6.93	6.80	6.91	7.87	7.47	7.67	7.57	8.28	8.38	7.98	7.80
SD	0.69	0.75	0.60	0.50	0.83	0.58	0.35	0.38	0.93	0.82	0.66	0.64	0.64	0.42	0.85	0.87	1.22	0.98	0.81	0.88	0.30	0.88	1.23	0.87
CV (%)	11.68	14.93	10.74	10.21	13.75	11.43	6.15	7.69	12.56	12.01	10.12	9.71	8.00	6.03	12.56	12.60	15.50	13.10	10.52	11.67	3.66	10.49	15.37	11.18

[1] DM: Dry matter (% fw)

[2] SSC: Soluble solid content (ºBrix)

Table 6. Discriminant models for the authentication of bell peppers of different colours by growing system. PLS-DA

Qualitative Group	Green bell peppers		Yellow bell peppers		Red bell peppers	
	Percentage of correctly-classified samples: 88.28% (113/128) Model SECV: 0.34 Number of synthetic variables: 3 Mathematical treatment: 2,5,5,1-SNV+DT		Percentage of correctly-classified samples: 91.37% (106/116) Model SECV: 0.35 Number of synthetic variables: 10 Mathematical treatment: 1,5,5,1-SNV+DT		Percentage of correctly-classified samples: 90.67% (107/118) Model SECV: 0.33 Number of synthetic variables: 3 Mathematical treatment: 1,5,5,1-SNV+DT	
Growing system	Training set	Validation set	Training set	Validation set	Training set	Validation set
Outdoor	77.77% (35/45)	100.00% (5/5)	87.87% (29/33)	100.00% (5/5)	77.77% (28/35)	80.00% (4/5)
Greenhouse	93.97% (78/83)	100.00% (5/5)	92.77% (77/83)	80.00% (4/5)	93.97% (79/83)	100.00% (5/5)

Table 7. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) for the calibration and validation sets and standard error of laboratory (SEL)

Parameter	Calibration set					Validation set					SEL
	N	Range	Mean	SD	CV	N	Range	Mean	SD	CV	
Dry matter (% fw)	262	4.48-11.73	6.78	1.16	17.10	130	4.52-9.64	6.68	1.20	17.96	0.21
SSC (ºBrix)	262	3.85-10.05	6.39	1.29	20.18	130	3.85-10.05	6.33	1.39	21.95	0.06

Table 8. Calibration statistics for NIR-based models for predicting quality parameters in intact bell peppers

Parameter	**Math treatment**	**Mean**	**SD**	**SECV**	r^2_{cv}	**RPD$_{cv}$**	**CV (%)**
Dry matter (% fw)	1,5,5,1-SNV+DT	6.72	1.08	0.66	0.62	1.64	9.82
SSC (ºBrix)	1,5,5,1-SNV+DT	6.31	1.24	0.75	0.63	1.65	11.88

Figure 1. Scores plot (a) and loadings weight (b) for the second (PC2) and third (PC3) principal components for intact bell peppers.

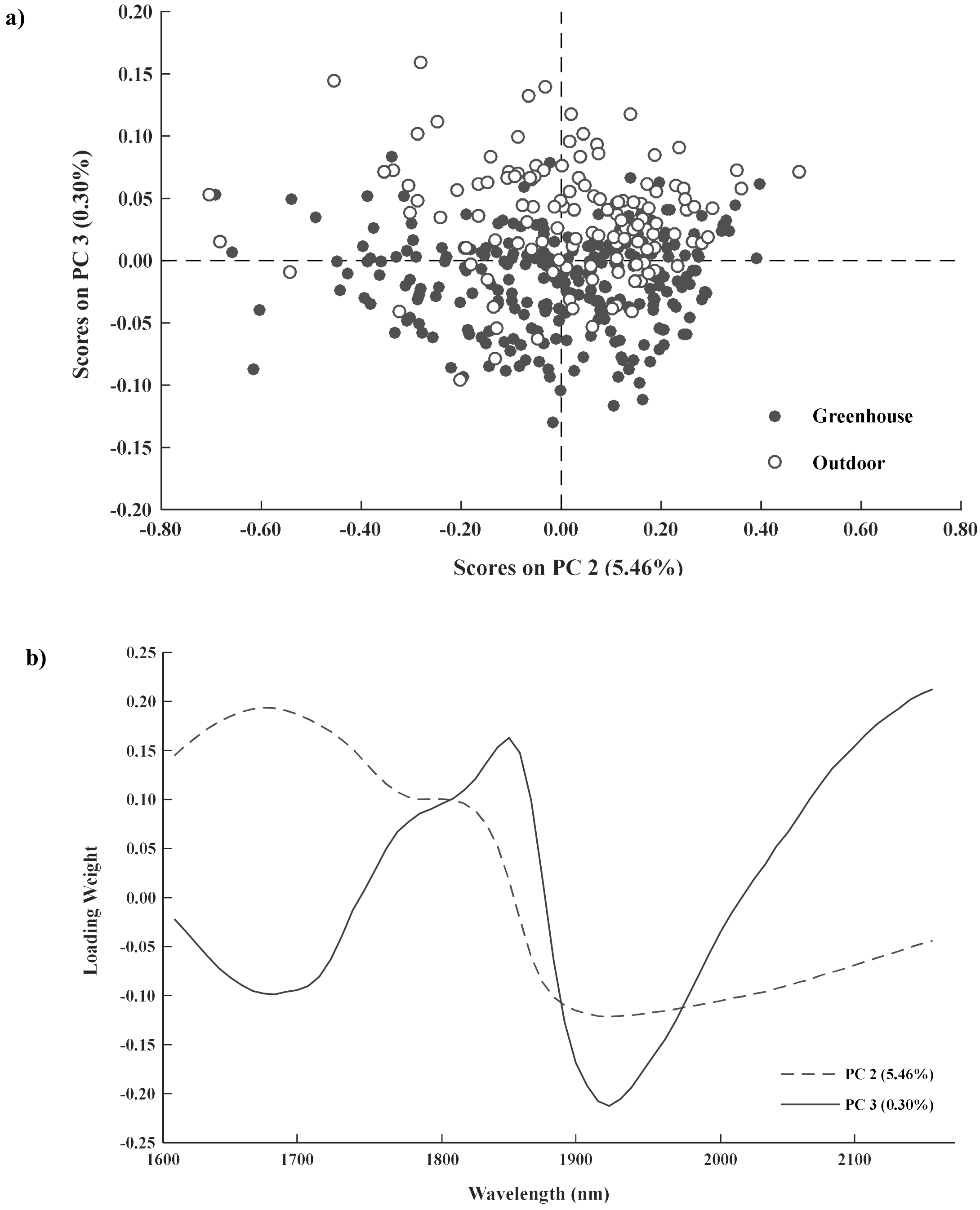


Figure 2. Values of the discriminatory variable obtained for the different validation groups. Unbalanced and balanced models.

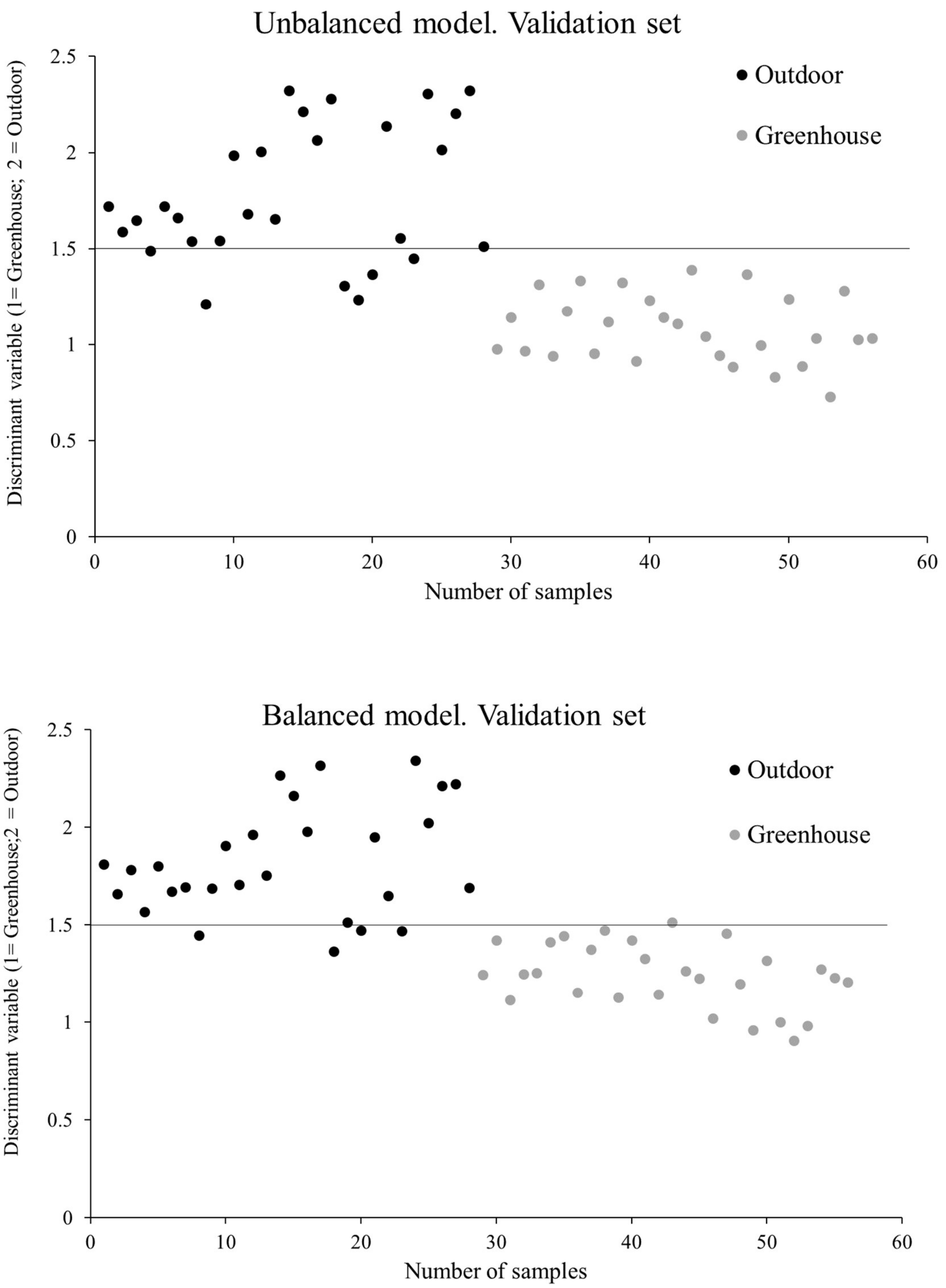


Figure 3. Regression coefficients for the bell pepper discriminant analysis. Balanced model.

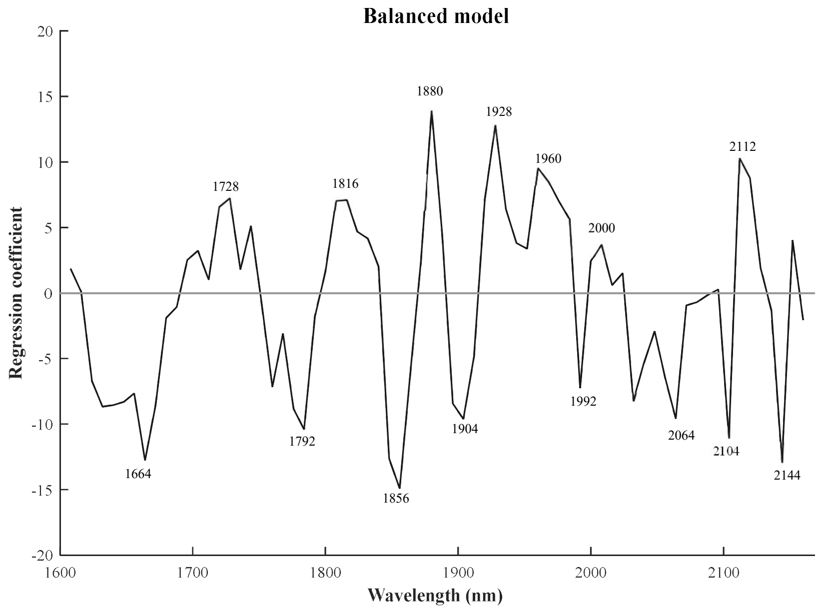
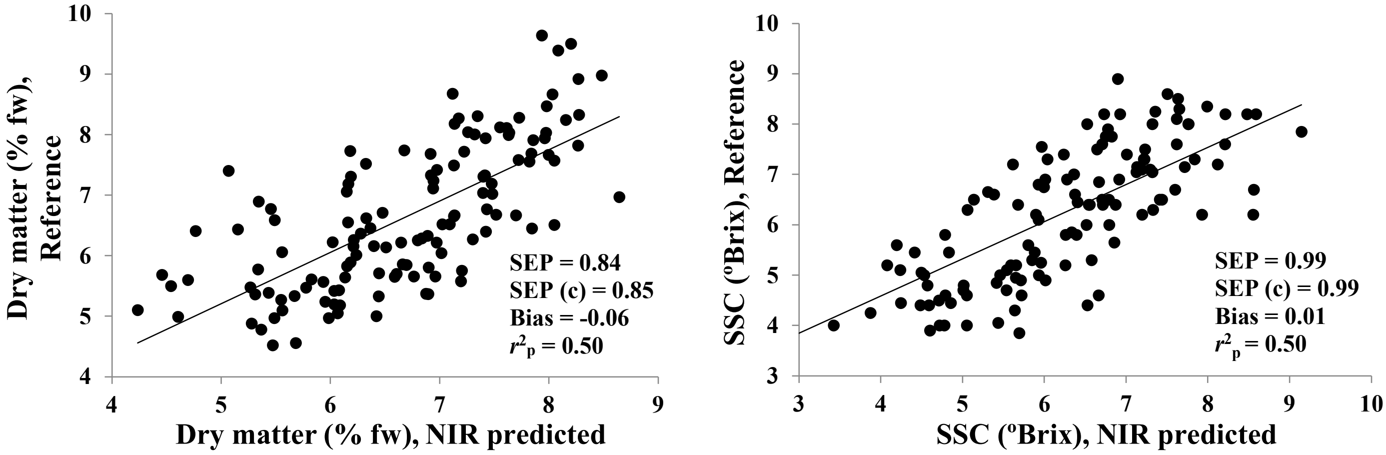


Figure 4. Reference *versus* NIR predicted concentration of dry matter (a) and SSC (b) in bell pepper.

a) **b)**



712 **Figure 5**. Regression coefficients for bell pepper dry matter and soluble solid content