Routine NIRS analysis methodology to predict quality and safety indexes in spinach plants during their growing season in the field

Irina Torres[a], María-Teresa Sánchez[a,*], Miguel Vega-Castellote[a], Natividad Luqui-Muñoz[c], Dolores Pérez-Marín[b,*]

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

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

[c] *Agronomic Department, Gelagri Ibérica, SL, 31320 Milagro (Navarra), 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

Cultural practices and harvesting in spinach plants should be based not only on subjective indexes such as freshness and green colour, which are both related with the visual appearance of the plants, but also on objective indexes that can be quantified non-destructively. The aim of this research was to develop a methodology based on the use of near infrared spectroscopy to monitor routinely the growth process of the spinach plants in the field. Using the MicroNIR™ OnSite-W spectrophotometer, which is ideally suited for *in situ* analysis, 261 spinach plants were analysed. Initially, calibration models for dry matter, soluble solid and nitrate contents were developed using 1 spectrum per plant for dry matter content, and nine spectra per plant for the other two parameters. These were then validated using the same number of spectra per plant as for calibration purposes. After that, to establish a procedure more suitable to routine analysis in the field, the models were validated with only one spectrum per plant and the suitability of the predictions was measured considering the global and neighbourhood Mahalanobis distances, whose control limit values were defined as inferior to 4.0 and 1.0, respectively. The results showed that once the calibration models were developed, only one spectrum per plant was enough to predict dry matter and nitrate contents successfully. Therefore, the methodology developed will allow us to monitor in real time the complete growth process and to take decisions about spinach cultivation based on internal quality and safety indexes.

Keywords: Spinach plants; NIRS for in-field analysis; Monitoring vegetable quality and safety; Routine methodology

1. Introduction

In leafy vegetables, such as spinach plants, horticultural maturity and optimal harvest time are usually measured by checking the appearance of the plants. Thus, the main quality characteristics to be considered for their selection and harvest are size and the proportion of clean leaves in early- to mid-maturity, while older and yellowing leaves should be avoided [1]. Furthermore, other indexes such as freshness and characteristic green colour, which are closely related to nutritional quality, are attributes which also help to define the visual appearance of the spinach plants [2].

However, it is important to stress that this is a subjective evaluation and, as a result, the decisions taken as regards crop management and harvesting vary enormously, making the automation process difficult. For this reason, decisions regarding the quality of the spinach plants, their horticultural maturity and optimum harvest time should be based not only on visual appearance, but also on quality indexes, which involve measuring the physicochemical parameters. These types of attributes - among which dry matter content (DMC) and soluble solid content (SSC) are the foremost - allow us to establish clear and well-defined standards.

In this context, Conte et al. [3] showed the importance of DMC analysis in spinach plants for growers, since values of around 10–12 % ensure a good resistance to handling and allow a high visual quality to be maintained during postharvest storage. Likewise, Kramchote et al. [4] showed that SSC was a crucial parameter when choosing the optimum time for harvesting, for measuring the shelf life in leafy vegetables and for classifying the product at the industrial level.

In addition, in spinach plants, it is also essential to quantify parameters related to food safety and, in particular, nitrate content, since this determines the industrial destination of the product once it is harvested [5].

Not only should the quality of the vegetables be assessed at the time of their harvest, but it is also important to monitor the state of the plants during development, in order to decide on the most suitable crop management guidelines, mainly as regards the nitrogen fertilization and water needed at each stage [2, 6].

Therefore, non-destructive quantitative evaluations and monitoring of these internal parameters are absolutely vital in order to identify predictors of market quality and safety before harvest and at harvest time.

The combination of the new generation of near infrared spectral sensors and the advances in data processing offers the possibility of monitoring the growth of the horticultural products directly in the field, either on the fruit or on the plants. This provides cost-effective, value-added solutions to a wide range of fruit and vegetables, as well as providing, at the same time, opportunities to understand better the influence of the variety and the preharvest factors on the quality and safety of the products during their growth, at harvest and postharvest, thus facilitating real-time decision-making for the selection of varieties, crop management practices and harvest decisions.

There are no scientific reports on the implementation of near infrared spectroscopy (NIRS) in spinach plants during their growth in the field, although some authors have carried out feasibility studies or simulated harvest decisions at laboratory level based on quality and safety indexes measured by NIRS sensors [7-11].

The main aim of this study was to monitor the growing season of spinach plants in the field, developing and optimizing an NIRS analysis methodology based on a new-generation sensor that can be easily used by farmers to establish a harvest index, based

not only on the visual appearance of the plants, but also on the non-destructive readings of the quality and safety parameters. These indexes could be of great importance for the routine monitoring of crops and could also help us to take informed decisions about crop cultural practices and the selection of the varieties best adapted to specific conditions.

2. Material and methods

2.1. Sampling and reference analysis

A total of 261 spinach (*Spinacia oleracea* L. cv. 'Baboon', 'Bandicoot', 'Harmonica' and 'Solomon') plants grown outdoors on different farms in the province of Cordoba (Spain) were used in this study. The spinach plants were manually harvested during their growing period between December 2019 and March 2020.

SSC and nitrate content were measured using 9 leaves per plant and following Pérez-Marín et al. [10]. Thus, SSC (ºBrix) was measured as the refractometer reading for spinach juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany) while nitrate content (mg NO_3 kg^{-1}) was measured using an RQFlex reflectometer (Merck, Darmstadt, Germany). The reflectometer which measures the colour intensity of Reflectoquant ® test strips (Merck, Darmstadt, Germany) is based on a colorimetric method.

DMC was determined gravimetrically by desiccation at 105 °C for 24 h, and the final dry weight was calculated as a percentage of the initial wet weight [12]. To measure this, only one leaf per plant was used, following Sánchez et al. [9]. All the

113 analytical measurements were performed in duplicate and the standard error of
114 laboratory (SEL) was calculated from these replicates.
115
116 *2.2. NIR spectrum acquisition*
117
118 NIR spectra of spinach plants were collected in-field using the MicroNIR™
119 OnSite-W spectrophotometer (VIAVI Solutions, Inc., San Jose, California, USA), a
120 portable miniature instrument adapted to *in situ* measurements. This instrument uses a
121 Linear Variable Filter (LVF) as the dispersing element and works in reflectance mode
122 (log 1/R) in the spectral range 908 to 1676 nm (taking data each 6.2 nm). It is a light (<
123 250 g) instrument, with an optical window of around 227 mm^2. The sensor integration
124 time was set at 11 ms and each spectrum was the mean of 200 scans. Among the key
125 innovations of this instrument are the Bluetooth and WiFi connections; it is also a fully-
126 integrated spectrophotometer, with no moving parts and IP65/IP67 waterproofing
127 and/or dust proofing.
128 Spectra acquisition was carried out using the VIAVI MicroNIR™ software Pro
129 version 3.1 (VIAVI Solutions, Inc., San Jose, California, USA). The instrument's
130 performance was checked every 10 minutes. For that purpose, a white reference
131 measurement was obtained using a NIR reflectance standard (Spectralon™) with a 99%
132 diffuse reflectance, while the dark reference was obtained by placing a black cover over
133 the analysis window.
134 The NIRS analysis of the spinach plants was carried out on the plants in the
135 field, during the growing period. Initially, 10 leaves were chosen per plant. Then, in 9 of
136 those leaves chosen for measuring SSC and nitrate content, 1 spectrum was taken per
137 leaf at one location of the leaf blade relative to the main vein and close to the petiole on

138 the adaxial side. Next, on the remaining leaf, which was used for the DMC
139 measurement, 2 spectra were taken, one on each side of the main vein. As 9 leaves per
140 plant were used for the chemical analyses of SSC and nitrates, a mean spectrum was
141 obtained from the 9 spectra taken for each plant to predict these parameters. To measure
142 DMC, as only one leaf was used for the wet measurement, one of the spectra taken on
143 the spinach leaf was randomly selected using the Matlab version 2015a (The
144 Mathworks, Inc., Natick, MA, USA) software package, thus providing a representative
145 spectrum for each plant.
146
147 *2.3. Definition of the calibration and validation sets and development of NIRS models*
148 *using MPLS algorithm*
149
150 Prior to the development of NIRS calibrations, data pre-processing and
151 chemometric treatments were performed using the Matlab version 2015a and WinISI II
152 version 1.50 (Infrasoft International LLC, Port Matilda, PA, USA) [13] software
153 packages.
154 Firstly, the optimum spectral range for the instrument was selected. To achieve
155 this, the 1,1,1,1 derivation treatment, where the first digit is the number of the
156 derivative, the second the gap over which the derivative is calculated, the third the
157 number of data points in a running average or smoothing, and the fourth the second
158 smoothing [14] without scatter correction, was applied, which allows to highlight the
159 areas of the spectrum which display a high level of noise [15].
160 To structure and compress the data matrix, the CENTER algorithm was applied
161 to the two available sets (set I for DMC and set II for SSC and nitrate content) following
162 the methodology proposed by Shenk and Westerhaus [14, 16]. Samples with

163 Mahalanobis distance (GH) values > 4 were considered as outliers or anomalous
164 spectra.

165 Once the spectral outliers were studied and removed from both sets and after
166 ordering the sample sets by spectral distances (from smallest to greatest distance from
167 the centre), four of every five samples were selected to form part of the calibration sets
168 (C1 for DMC and C2 for SSC and nitrate content), while the remainder constituted the
169 validation sets (V1 for DMC and V2 for SSC and nitrate content).

170 Modified partial least squares (MPLS) regression was used to obtain NIRS
171 predictive models for each parameter tested, using their specific sets. Four cross-
172 validation groups were included in order to avoid overfitting [17]. For each analytical
173 parameter, different mathematical pre-treatments were used. For scatter correction,
174 Standard Normal Variate (SNV) and Detrend (DT) methods were applied [18].
175 Additionally, a total of two mathematical derivation treatments were tested: 1,5,5,1 and
176 2,5,5,1 [13, 14].

177 The best calibration models for each parameter were selected by statistical
178 criteria, using the coefficient of determination for calibration (r^2_c), the standard error of
179 calibration (SEC), the coefficient of determination for cross validation (r^2_{cv}), the
180 standard error of cross validation (SECV) and the residual predictive deviation for cross
181 validation (RPD_{cv}) calculated as ratio of the standard deviation (SD) of the reference
182 data for calibration from the SECV.

183

184 *2.4. Routine analysis procedure*

185

186 Once the calibration equations were established, the feasibility of using this
187 technology as a method for monitoring crop development was studied, and an optimal

methodology was set up for routine analysis with portable NIRS sensors. Control reliability statistics based on spectral distances were established for the results obtained: the global Mahalanobis distance (GH), or the spectral distance from a sample to the centre of the calibration population, and the neighbourhood Mahalanobis distance (NH), which is the spectral distance between the plant and neighbouring or similar samples.

To do this, external validation of the best models developed was carried out using the protocol of Windham et al. [19]. The use of one spectrum calculated as the mean value of the 9 spectra available for SSC and nitrate content (in this case, optimising the quality of the spectral information collected) was compared to the use of a single spectrum per plant for the three parameters tested which is an adapted strategy that would facilitate the routine analysis in the field. For SSC and nitrate content, the spectrum was randomly selected from the 9 available for each of the plants included in the validation set (V2), while for DMC, the spectrum was also randomly chosen from the 2 available (V1 validation set), using Matlab v. 2015a software in both cases.

Finally, the standard error of prediction (SEP) values for the models obtained for both validation strategies developed for the prediction of SSC and nitrate content - the mean of 9 spectra per plant and 1 spectrum per plant - were compared using Fisher's F test [20, 21]. Values for F were calculated as:

$$F = \frac{(SEP_2)^2}{(SEP_1)^2}$$

where SEP_1 and SEP_2 are the standard errors of prediction and $SEP_1 < SEP_2$. F is compared to $F_{critical}$ (1-P, n_1-1, n_2-1), as read from the table, with $P = 0.05$, and n_1 is the number of times the measurement is repeated with method 1, while n_2 is the number of times the measurement is repeated with method 2. If F is higher than $F_{critical}$, the two SEP values are significantly different.

Once the viability of using this working methodology (1 spectrum per plant) for crop monitoring was established, and in order to increase the robustness of the models by increasing the possible variability, the calibration and validation sets were merged and new global models were developed for the prediction of the three parameters tested, following the procedure mentioned above and using the same signal pre-treatments.

3. Results and discussion

3.1. Population characterization and development of the models for the prediction of quality and safety indexes in spinach plants

Before structuring the population by means of the CENTER algorithm, the spectral region affected by noise was studied, determining that the areas at the beginning (908–1001 nm) and at the end (1627–1676 nm) of the spectral range should be removed. Consequently, all the chemometric treatments were performed using a spectral range between 1001–1627 nm. Typical log(1/R), D_1 log (1/R) and D_2 log(1/R) spectra are shown in Fig. 1.

After using the CENTER algorithm to study the structure and spectral variability, anomalous samples were detected in the DMC (one sample) and SSC and nitrate content (one sample) sets respectively, which were removed.

The number of samples, range, mean, SD and coefficient of variation (CV) of the calibration (C1 and C2) and validation (V1 and V2) sets are shown in Table 1. The structured selection based on spectral information by means of the global Mahalanobis distance [16] proved to be useful, since the calibration and validation sets displayed

similar characteristics for all the study parameters and the validation set ranges lay within those of the calibration sets.

The frequency histogram for the nitrate content parameter is shown in Fig. 2. As can be seen in the figure, this parameter covers a wide range of values, from 70 to 3875 mg kg^{-1}, although approximately 41 % of the plants analysed (106 of 261 samples) had a nitrate content of below 500 mg kg^{-1}. The latter corresponded mainly to plants analysed at the beginning and end of the harvest period, as well as plants from plots of land which were subjected to lower doses of fertilizer. It is important to note that, although the nitrate content was not distributed evenly, it covered the entire range, representing the full variability of the parameter, which is essential for the subsequent development of the models [22].

Table 2 shows the statistics of the best prediction models obtained for the in-field prediction of quality and safety parameters in spinach plants using the sample sets C1 and C2 for calibration and the MicroNIR™ OnSite-W instrument.

When measuring the parameter DMC, for which a single spectrum taken from the leaf was used, which was subsequently analysed by the reference method, the predictive capacity of the model developed allowed to distinguish between high, medium, and low values of this parameter [23, 24]. Likewise, when measuring the SSC and nitrate content parameters, the predictive models obtained from the analysis of 9 leaves per plant also enabled to differentiate between high, medium and low values, as indicated by Shenk and Westerhaus [23] and Williams [24].

3.2. Implementation of a routine analysis procedure

Firstly, to analyse the predictive capacity of the models and their subsequent routine application**,** the best models selected were subjected to external validation using the V1 and V2 validation sets (Table 3). Following the protocol outlined by Windham et al. [19], the models developed for the three parameters analysed met the validation requirements in terms of r^2_p ($r^2_p > 0.6$), $SEP_{(c)}$ and bias. Additionally, the slope values for the DMC and the nitrate content fell within the recommended interval values (0.9-1.1), whereas for the SSC, the slope did not attain the limits, despite being close.

According to these results, these equations for DMC and nitrate content could be judged suitable for using in routine analysis, permitting the non-destructive measurement of quality and safety parameters, as well as facilitating decision-making about selection of varieties, fertilization requirements and deciding on the optimal time of harvest.

Once the feasibility of implementing NIRS technology for the in-field characterization of spinach was verified, the routine analysis process was established. To do this, the methodology recommended by Zamora-Rojas et al. [25] and Pérez-Marín et al. [26] was followed, which considers the global Mahalanobis (GH) and the neighbourhood Mahalanobis (NH) distances as control statistics for routine analysis. It was established that, during the analysis in routine in the field, those predicted samples which displayed a GH > 4 and/or NH > 1, had to be analysed again. Thus, for those samples which presented GH values > 4 and/or NH > 1 when predicted, another of the 9 spectra taken from that plant was randomly selected, in this case to simulate a second measurement. It was established in this procedure, therefore, that when the spectrum collected exceeded the established spectral limits, it should be repeated. If any of the samples once again presented prediction values higher than those established for the GH

and/or NH statistics, they would be considered outlier samples which should be analysed in the laboratory and incorporated into the next expansion of the equation.

The external validation statistics of the best models obtained to predict the three parameters analysed following the protocol established for the analysis in routine in the field are shown in Table 3. For each parameter, the results obtained prior to and after the repetition of the spectral selection procedure are displayed. Additionally, a graphical comparison between the reference and NIR predicted values obtained after repetition is shown in Fig. 3.

For DMC, only one sample had a slightly higher NH value than recommended (NH = 1.19), and the values of the validation statistics obtained before and after the repetition of this sample were practically identical. For this parameter and, after the sample which had an NH value greater than 1 was repeated, the model developed complied with the limits established by Windham et al. [19].

In the case of the SSC and nitrate content parameters, a total of 12 of the 52 samples available in the validation set (23.08%) had to be repeated; six of these samples displayed GH and NH values higher than the control limits established for both statistics, 2 samples had GH values greater than 4 and, the remaining 4 samples showed NH values above 1 (Table 4).

After carrying out a detailed study of these samples, it became clear that 8 of the 12 samples corresponded to 'Harmonica' plants analysed at the beginning of the season (December and the first weeks of January), which were poorly developed plants, with smaller leaves, size and thickness, and a lighter colour. Meanwhile, 3 of the samples (2 ‘Solomon’ and 1 ‘Baboon’ plants) corresponded to the last two weeks of the growing period (first fortnight of March), in which the plants were extremely thick, with a very intense green colour. The remaining sample, which belonged to the ‘Harmonica’ plants,

309 displayed no particular characteristics of interest, which means that a mistake could
310 have been made during the spectral acquisition process.

311 As can be seen in Table 3, for the SSC and nitrate parameters, the statistics
312 obtained after taking a new spectrum for these samples improved when compared with
313 the initial ones, with the SEP values decreasing by about 20% for both parameters,
314 which confirms the importance of carrying out the field measurement procedure as
315 rigorously as possible. It should be also mentioned that, after the spectrum was repeated
316 for these samples, none of them presented second GH and/or NH values above the
317 established limits. If, during the routine analysis in the field, any of these samples had
318 presented a second GH value greater than 4 and/or NH greater than 1, as previously
319 indicated, they would have been collected for analysis by the laboratory reference
320 method and subsequently incorporated into a future expansion of the equation.

321 The models developed for SSC and nitrate content after repeating the analysis of
322 those samples that displayed high values of GH and/or NH did not meet the validation
323 requirements established by Windham et al. [19] in terms of slope (0.90-1.10) and R^2_p
324 ($R^2_p > 0.6$), although in the case of SSC, this statistic is close to the minimum of 0.60.
325 For both parameters, the bias remained within the confidence limits, while the $SEP_{(c)}$
326 value obtained for SSC (1.4 ºBrix) was higher than the control limit (1.2 ºBrix).

327 In addition, it must be mentioned that the mean Mahalanobis distance between
328 each sample and the centre or the nearest neighbour after the repetition (GH = 1.57 and
329 NH = 0.41) was lower than the initial values (GH = 2.35 and NH = 0.66), which showed
330 the higher representativeness after repeating the spectra compared with the samples
331 included in the calibration set used to develop the models. Once the routine analysis
332 procedure was established, it was important to determine whether the reduction of the
333 number of acquired spectra affected the precision of the measurement.

Table 5 shows the comparison between the SEP values obtained considering the number of spectra (1 spectrum or the mean spectrum of the 9 taken) used for the external validation of the models for the SSC and nitrate content parameters.

According to Table 5, for the parameter SSC, the SEP value increased significantly when the number of spectra taken in routine monitoring is reduced from 9 to 1 (27.86 %), whereas for the nitrate content, the increase in terms of SEP was not significant.

Therefore, in light of the results obtained, it can be stated that taking a single spectrum would be sufficient to monitor the crop and determine the evolution of the nitrate content; in this way, both the behaviour of the different varieties, and the necessary management practices, principally those related to the dose and timing of the fertilizer, could be established.

For the SSC parameter, we studied the number of leaves that had to be analysed in the field, so that the difference between errors would not be significant, with 3 being the optimal number of leaves to analyse (SEP = 1.2 ºBrix). Although analysing a higher number of leaves for each plant was a key step for this parameter, it should bear in mind that the error obtained was an average uncertainty value, so the individual uncertainty of each sample may be lower; it must also be remembered that this increased error does not affect all the analysed samples equally and that a loss of precision of ± 0.5 ºBrix is not a determining factor in quality in this type of product. Therefore, given that the other two parameters (DMC and nitrate content), which are essential for monitoring field cultivation, can be predicted by measuring a single spectrum per plant without showing significant differences as regards the increase in the number of leaves analysed, we decided to incorporate the strategy of analysing a single leaf into the routine analysis, since this would enable us to speed up the measurement process in the

359 field and, therefore, to analyse many plants quickly and with low cost, which would
360 compensate for the loss of precision in the SSC parameter.
361
362 *3.3. New global model developments for the in situ quality and safety prediction of*
363 *spinach plants*
364
365 After the evaluation and establishment of the routine analysis procedure, to
366 increase the robustness of the models prior to being incorporated in routine monitoring,
367 the variability covered by these models for each of the parameters analysed was
368 increased. To achieve this, new calibration models were developed using the sample
369 sets obtained by merging the two groups, the calibration and validation sets (Table 1).
370 The results of the new global models developed using all the samples available are
371 shown in Table 6.
372 For the three parameters analysed, the new global models enabled to distinguish
373 between low, medium and high levels [23, 24]. By increasing the calibration set with
374 the validation samples, the predictive capacity of the models remains very similar, as
375 can be seen when comparing the RPD_{cv} values, with only a slight reduction (2.83%) in
376 the error, in the case of nitrate content.
377 However, with a view to predicting unknown samples in the future, it is
378 important to include all the possible sources of variation by expanding the set of
379 samples used to develop the models [27]. Furthermore, it must be highlighted that in
380 order to increase the robustness of the models that will be used in routine monitoring to
381 analyse samples from different seasons, regions or varieties, representative libraries of
382 the studied parameters are required [22].

Thus, these new global models would be implemented in routine monitoring for the *in situ* measurement of quality and safety parameters in spinach plants, allowing the incorporation of the NIRS technology as a fast method for the real-time decision-making for crop management practices and harvest decisions, considering that nowadays these decisions are mainly based on physical indexes, such as colour or size.

4. Conclusions

The results showed that once the calibration models were developed, the methodology proposed, based on taking a suitable spectrum (GH inferior to 4.0 or/and NH inferior to 1.0) per plant, allowed us to predict DMC and nitrate content in spinach plants successfully during their growing season in the field, without any loss of accuracy, thus making it possible for a greater number of plants to be analysed. This will enable to establish more precisely the influence of cultural practices such as irrigation and fertilization, mainly nitrogen, on crop development and its quality and safety, as well as establishing the optimal harvest time and classifying the different varieties with respect to the objective indexes studied. However, plants from a new growing season should be analysed to test *in situ* whether the proposed methodology has been applied correctly and, in turn, to extend the model with different sources of variation.

CRediT authorship contribution statement

Irina Torres: Data acquisition, Methodology, Formal analysis, Investigation, Software, Data curation, Validation, Writing - original draft, Writing - review & editing,

Visualization. **María-Teresa Sánchez**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. . **Miguel Vega-Castellote**: Data acquisition, Formal analysis, Software, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Natividad Luqui-Muñoz:** Data acquisition, Formal analysis, Investigation. **Dolores Pérez-Marín**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

This research was carried out as part of the research project P-12019005 'Quality determination of spinach grown in open-air fields using near infrared spectroscopy', funded by Gelagri Ibérica, S.L. The authors are grateful to Mrs. Mª Carmen Fernández of the Animal Production Department for her technical assistance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced in any way the work reported in this paper.

References

[1]. T.V. Suslow, M, Cantwell, Spinach: Recommendations for Maintaining Postharvest Quality, http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Vegetables_English/?uid=32&ds=799, 2002 (accessed 17 June 2020).

[2] M. Cantwell, R.F. Kasmire, Postharvest handling systems: flower, leafy and stem vegetables, in: A.A. Kader (Ed.), Postharvest Technology of Horticultural Crops, Division of Agriculture and Natural Resources, University of California, Oakland, CA, USA, 2002, pp. 423–433.

[3] A. Conte, G. Conversa, C. Scrocco, I. Brescia, J. Laverse, A. Eliba, M.A.D. Nobile, Influence of growing periods on the quality of baby spinach leaves at harvest and during storage as minimally processed produce. Postharvest Biol. Technol. 50 (2008) 190–196, https://doi.org/10.1016/j.postharvbio.2008.04.003.

[4] S. Kramchote, V. Srilaong, T. Wasusri, C. Wongs-Aree, S. Kanlayanarat, Low temperature storage maintains postharvest quality of cabbage (*Brassica oleracea* var. capitata L.) in supply chain, Int. Food Res. J. 19 (2012) 759–763.

[5] Official Journal of the European Union (OJEU), Commission Regulation (EU) No 1258/2011 of 2 December 2011 Amending Regulation (EC) No 1881/2006 as regards Maximum Levels for Nitrates in Foodstuffs, OJ L 320/15-17, 3.12.2011, 2011.

[6] E. Gutiérrez-Rodríguez, J.H. Lieth, J.A. Jernstedt, T.V. Suslow, Prediction of spinach quality based on pre- and postharvest conditions, Acta Hortic. 934 (2012) 1141–1148, https://doi.org/10.17660/ActaHortic.2012.934.153.

[7] H. Itoh, H. Tomita, Y. Uno, N. Shiraishi, Development of method for non-destructive measurement of nitrate concentration in vegetable leaves by near-

457 infrared spectroscopy, IFAC Proc. Vol. 44 (2011) 1773–1778,
458 https://doi.org/10.3182/20110828-6-IT-1002.00738.
459 [8] M.T. Sánchez, J.A. Entrenas, I. Torres, M. Vega, D. Pérez-Marín, Monitoring
460 texture and other quality parameters in spinach plants using NIR spectroscopy,
461 Comput. Electron. Agric. 155 (2018) 446–452,
462 https://doi.org/10.1016/j.compag.2018.11.004.
463 [9] D. Pérez-Marín, I. Torres, J.A. Entrenas, M. Vega, M.T. Sánchez, Pre-harvest
464 screening on-vine of spinach quality and safety using NIRS technology,
465 Spectrochim. Acta A Mol. Biomol. Spectrosc. 207 (2019) 242–250,
466 https://doi.org/10.1016/j.saa.2018.09.035.
467 [10] J.A. Entrenas, D. Pérez-Marín, I. Torres, A. Garrido-Varo, M.T. Sánchez,
468 Simultaneous detection of quality and safety in spinach plants using a new
469 generation of NIRS sensors, Postharvest Biol. Technol. 160, (2020) 111026, 1–8,
470 https://doi.org/10.1016/j.postharvbio.2019.111026.
471 [11] I. Torres, M.T. Sánchez, D. Pérez-Marín, 2020. Integrated soluble solid and nitrate
472 content assessment of spinach plants using portable NIRS sensors along the supply
473 chain, Postharvest Biol. Technol. 168, (2020) 111273, 1–7,
474 https://doi.org/10.1016/j.postharvbio.2020.111273.
475 [12] AOAC, Official Methods of Analysis of AOAC International, 17th ed., AOAC,
476 Gaithersburg, MD, USA, 2000.
477 [13] ISI, The Complete Software Solution Using a Single Screen for Routine Analysis,
478 Robust Calibrations, and Networking, Manual, FOSS NIRSystems/TECATOR,
479 Infrasoft International, LLC, Sylver Spring, MD, USA, 2000.

[14] J.S. Shenk, M.O. Westerhaus, Routine Operation, Calibration, Development and Network System Management Manual, NIRSystem, Inc., 12101 Tech Road, Silver Spring, MD, USA, 1995.

[15] W.R. Hruschka, Data analysis: wavelength selection methods, in: P.C. Williams, K.H. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Inc., St. Paul, MN, USA, 2001, pp. 39–58.

[16] J.S. Shenk, M.O. Westerhaus, Population structuring of near infrared spectra and modified partial least squares regression, Crop Sci. 31 (1991) 1548–1555, https://doi.org/10.2135/cropsci1991.0011183X003100060034x

[17] J.S. Shenk, M.O. Westerhaus, Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy, Monograph, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD, USA, 1995.

[18] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777. https://doi.org/10.1366/0003702894202201.

[19] W.R. Windham, D.R. Mertens, F.E. Barton II, Protocol for NIRS calibration: sample selection and equation development and validation, in: G.C. Martens, J.S. Shenk, F.E. Barton II (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality, Agriculture Handbook, vol. 643, USDA-ARS, US Government Printing Office, Washington, DC, USA, 1989, pp. 96–103.

[20] D.L. Massart, B.G.M. Vandeginste, S.M. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Data Handling in Science and Technology 2, Elsevier Science, Amsterdam, The Netherlands, 1988.

[21] T. Naes, T. Isaksson, T. Fearn, A. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002.

[22] J.S. Shenk, J. Workman, M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. Ciurczac (Eds.), Handbook of Near Infrared Analysis, CRC Press, Francis & Taylor Group, Boca Raton, FL, USA, 2008, pp. 347–386.

[23] J.S. Shenk, M.O. Westerhaus, Calibration: the ISI way, in: A.M.C. Davies, P.C. Williams (Eds.), Near Infrared Spectroscopy: The Future Waves, NIR Publications, Chichester, UK, 1996, pp. 198–202.

[24] P.C. Williams, Implementation of near-infrared technology, in: P.C. Williams, K.H. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, AACC Inc., St. Paul, MN, USA, 2001, pp. 145–169.

[25] E. Zamora-Rojas, A. Garrido-Varo, E. De Pedro-Sanz, J.E. Guerero-Ginel, D. Pérez-Marín, Monitoring NIRS calibrations for use in routine meat analysis as part of Iberian pig-breeding programs, Food Chem. 129 (2011) 1889–1897. https://doi.org/10.1016/j.foodchem.2011.05.139.

[26] D. Pérez-Marín, A. Garrido-Varo, C. Riccioli, T. Fearn, Training guidelines and wider exploitations of NIRS, in: Food Integrity: Ensuring the Integrity of the European Food Chain, Deliverable: 19.7, https://secure.fera.defra.gov.uk/foodintegrity/secure/downloadFile.cfm?id=762. 2018 (accessed 17 June 2020).

[27] J.S. Shenk, M.O. Westerhaus, P. Berzaghi, Investigation of a LOCAL calibration procedure for near infrared instruments, J. Near Infrared Spectrosc. 5 (1997) 223–232. https://doi.org/10.1255/jnirs.115

529 **Table 1**

530 Descriptive statistics (number of samples (N), range, mean, standard deviation (SD) and

531 coefficient of variation (CV) for the calibration, validation and global sets.

Parameter	Statistics	Calibration set	Validation set	Global set
Dry matter	N	208	52	260
content (% fw)	Range	6.12–20.34	8.77–20.23	6.12–20.34
	Mean	13.29	13.39	13.31
	SD	2.31	2.19	2.29
	CV (%)	17.38	16.36	17.21
Soluble solid	N	208	52	260
content (ºBrix)	Range	5.2–15.2	6.4–14.3	5.2–15.2
	Mean	10.3	10.3	10.3
	SD	1.9	1.7	1.8
	CV (%)	18.16	16.73	17.87
Nitrate content	N	208	52	260
(mg kg^{-1})	Range	70–3875	95–3582	70–3875
	Mean	1225	1212	1223
	SD	1085	1122	1090
	CV (%)	88.52	92.61	89.16

532

533 **Table 2**

534 Calibration statistics for predicting quality and safety parameters in spinach plants in the
535 field. C1 and C2 calibration sets.

Parameter	Calibration statistics							
	N	Mean	SD	r^2_c	SEC	r^2_{cv}	SECV	RPD_{cv}
Dry matter content (% fw)	200	13.42	2.17	0.72	1.16	0.66	1.26	1.83
Soluble solid content (ºBrix)	203	10.3	1.8	0.73	0.9	0.68	1.0	1.84
Nitrate content (mg kg^{-1})	202	1171	1032	0.57	675	0.53	708	1.53

536 N: number of samples; SD: standard deviation of calibration set; r^2_c: coefficient of determination of
537 calibration; SEC: standard error of calibration; r^2_{cv}: coefficient of determination of cross validation;
538 SECV: standard error of cross validation; RPD_{cv}: residual predictive deviation for cross validation.

539

Table 3

External validation for prediction of quality and safety parameters in spinach plants following the procedures established for traditional measurement and routine monitoring in the field.

Parameter	Analysis procedure		r^2_p	SEP	Bias	$SEP_{(c)}$	Slope	SEP limit = 1.3·SEC	Bias limit = ± 0.6·SEC
Dry matter content (% fw)	Mean 9 spectra		0.68	1.27	-0.32	1.24	1.00	1.51	± 0.70
	Routine analysis	First analysis	0.68	1.27	-0.32	1.24	1.00		
		After repetition	0.68	1.27	-0.34	1.23	1.01		
Soluble solid content (ºBrix)	Mean 9 spectra		0.68	1.0	0.1	1.0	0.85	1.2	± 0.6
	Routine analysis	First analysis	0.47	1.7	0.1	1.8	0.49		
		After repetition	0.54	1.4	-0.3	1.4	0.65		
Nitrates (mg kg^{-1})	Mean 9 spectra		0.62	688	-22	695	0.95	889	± 410
	Routine analysis	First analysis	0.32	1052	-163	1049	0.56		
		After repetition	0.48	833	46	840	0.78		

r^2_p: coefficient of determination of prediction; SEP: standard error of prediction; $SEP_{(c)}$: standard error of prediction corrected for bias; SEC: standard error of calibration

Table 4

Mahalanobis (GH and NH) distances for repeated samples following established routine protocol for measurement in the field.

Sample number	First analysis		After repetition	
	GH	NH	GH	NH
1	4.965	2.135	0.721	0.113
2	2.786	1.009	2.471	0.793
3	3.635	1.458	0.783	0.164
4	4.189	1.429	1.748	0.737
5	4.182	0.831	1.281	0.074
6	5.507	2.681	1.619	0.379
7	3.701	1.286	0.961	0.178
8	4.856	1.100	1.397	0.536
9	6.577	0.538	0.450	0.071
10	9.800	2.559	1.052	0.204
11	3.704	1.230	3.071	0.820
12	4.038	1.084	1.842	0.298

549 **Table 5**

550 Comparison between SEP values of the two validation strategies for *in situ* prediction of

551 the soluble solid and nitrate contents using MicroNIR^{TM} OnSite-W.

Parameter	Spectra per plant	r^2_p	SEP	Bias	$SEP_{(c)}$	F	$F_{critical}$
Soluble solid content	1	0.54	1.4	-0.3	1.4	1.92	1.60*
(ºBrix)	9	0.68	1.0	0.1	1.0		
Nitrate content (mg kg^{-1})	1	0.47	833	46	840	1.52	1.60
	9	0.63	676	-70	679		

552 *: Significant differences ($P < 0.05$).

553 r^2_p: coefficient of determination of prediction; SEP: standard error of prediction; $SEP_{(c)}$: standard error of

554 prediction corrected for bias.

555
556

557

Table 6

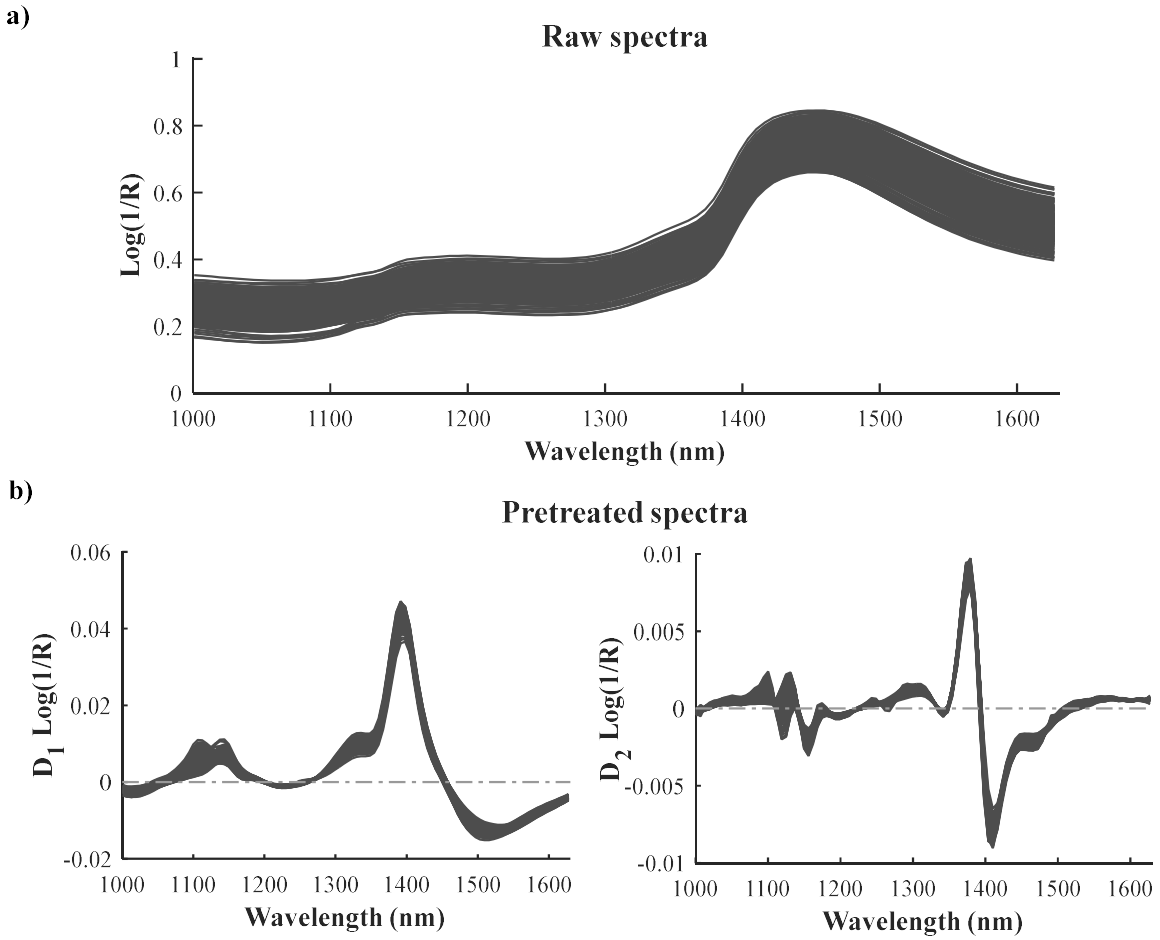
Calibration statistics for predicting quality and safety parameters in spinach plants in the field using the global sets.

Parameter	N	Mean	SD	r^2_{cv}	SECV	RPD_{cv}
Dry matter content (% fw)	254	13.39	2.17	0.65	1.29	1.77
Soluble solid content (ºBrix)	254	10.3	1.8	0.69	1.0	1.87
Nitrate content (mg kg^{-1})	252	1187	1050	0.57	688	1.58

N: Number of samples; SD: standard deviation of calibration set; r^2_{cv}: coefficient of determination of cross validation; SECV: standard error of cross validation; RPD_{cv}: residual predictive deviation for cross validation.

565 **Fig 1.** Raw (a) and pretreated (first and second derivative) spectra (b) of the spinach

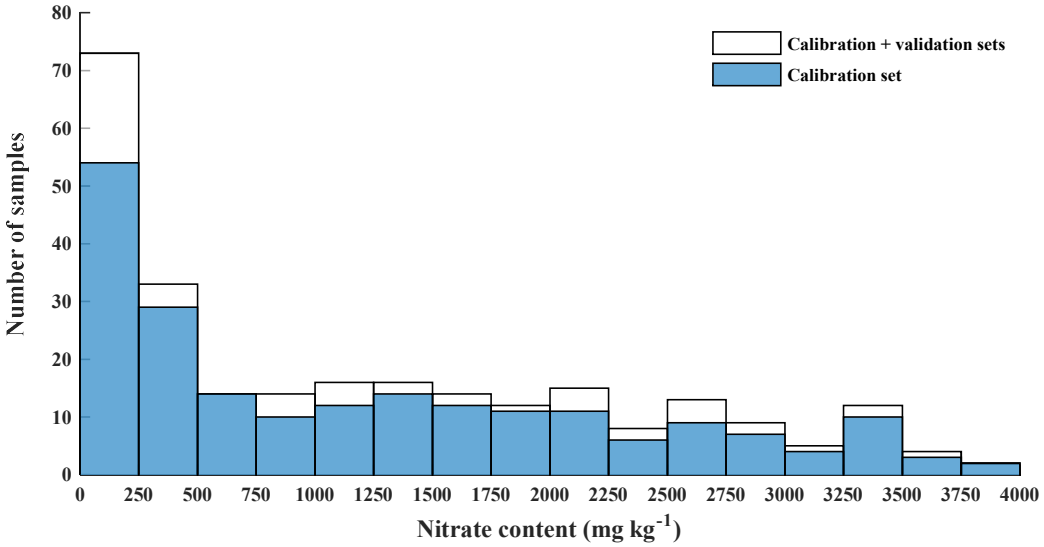
566 plants analysed in the field.



567

568

569 **Fig. 2.** Sample distribution for the nitrate content parameter.



570

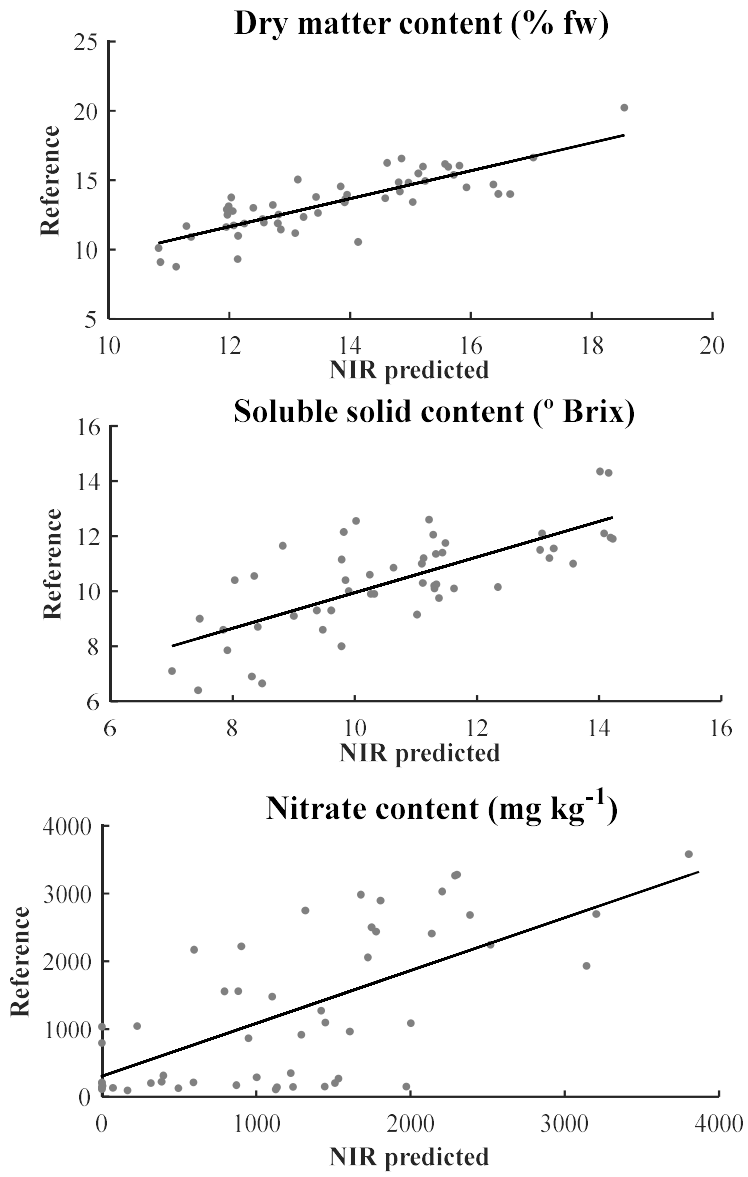
571

572

573 **Fig. 3.** Reference *versus* NIR predicted data for the validation procedure following the

574 established routine protocol for measurement in the field.

575



576

577