1 **An innovative non-targeted control system based on NIR spectral**

2 **information for detecting non-compliant batches of sweet almonds.**

3

4

5 Miguel Vega-Castellote[a], María-Teresa Sánchez[a,*], Irina Torres[a], Dolores Pérez-Marín[b,*]

6

7

8

9 [a] *Department of Bromatology and Food Technology, University of Cordoba, Rabanales*

10 *Campus, 14071 Córdoba, Spain.*

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

12 *Córdoba, Spain.*

13

14

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

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

17 Marín).

18

19   **Abstract**

20

21   Nowadays, there is growing awareness about the need to develop new methodologies to

22   fight against deliberate fraud. This study explored the use of near infrared spectroscopy

23   (NIRS) as an instantaneous, non-targeted method for detecting non-compliant products;

24   in this case, when used to detect sweet almond batches adulterated with bitter almonds.

25   For this purpose, we simulated the adulteration of batches by preparing four different

26   types of mixed samples which contained 5%, 10%, 15% and 20% of bitter almonds,

27   respectively, using 90 samples of sweet almonds and 50 samples of bitter almonds. For

28   each of the adulteration percentages, 21 samples were produced. The samples were

29   analysed using the Aurora and the Matrix-F spectrophotometers. The procedure initially

30   constructed the desired standard or target using only the spectral information provided by

31   the sweet almond population (control population). To achieve this, after principal

32   components analysis, the spectral warning and action limits were calculated using the n-

33   dimensional statistic Mahalanobis global distance. Next, the spectral distances from the

34   product standard defined for those samples not belonging to the control population,

35   including the adulterated sweet almonds, were calculated and represented as Shewhart

36   control charts. The implementation of NIRS technology throughout the almond supply

37   chain enabled to identify 87 % (73/84) of the adulterated sweet almond batches. These

38   findings suggest that NIRS technology and the use of spectral distances could enable to

39   establish an innovative, non-targeted control system based only on spectral information

40   to assess almond batches. This system allows to carry out conformity tests both *in situ*

41   and online of the batches of almonds received and processed in the industry, as well as

42   establishing fast, cost-efficient anti-fraud alert systems, which would help to reduce the

43 number of batches to be analysed by expensive and time-consuming confirmatory

44 methods.

45 *Keywords*: NIR spectroscopy; Almond adulteration; Fraud detection; Non-targeted

46 method; Shewhart control chart

47

1. Introduction

Food fraud and deliberate adulteration is a perennial problem which, unfortunately, is still very prevalent these days, and causes significant health and economic impacts as well as giving rise to an understandably high level of consumer distrust in the food supply chain. Consumers, on their part, demand foods of high 'integrity', which is a comprehensive term referring to a nutritive, healthy, tasty, authentic, traceable, as well as ethically, safely, environment-friendly and sustainably food product [1].

To verify integrity in marketed products, the current analytical and sampling control systems need to be renewed, through the development of non-invasive, fast, massive and cost-effective analytical methods to monitor all the steps in the food supply chain. A new, innovative strategy which could be implemented is to adopt 'non-targeted' methods to provide information on quality, safety and authenticity [2-5]. In this case, the objective is to evaluate the product in an integrated way through patterns, i.e. analysing whole matrix characteristics and identifying differences in order to establish early alert systems. In contrast, the traditional targeted approach evaluates the products compound-by-compound.

The nature and versatility of near infrared spectroscopy (NIRS) sensors, combined with specific data processing techniques, fit perfectly with both targeted and non-targeted strategies, enabling rapid, non-destructive, accurate and cost-effective analysis of large volumes and numbers of samples and the measurement of multiple parameters in raw materials, products and processes [6-9].

One of the main advantages of NIRS technology is the large quantity of the product that can be inspected when it is used online, in continuous mode in the sorting

73   lines at an industrial level. NIRS provides a unique digital fingerprint of each product,

74   which is essential for meeting the current requirements of industry and consumers

75   regarding food integrity [9,10]. In the almond sector, the presence of bitter almonds in

76   different proportions in batches of sweet almonds can result in strange, unpleasant

77   flavours, due to the presence of cyanogenic compounds such as amygdalin, which is

78   present in high concentrations in bitter almonds, thus altering the sensory quality (taste

79   and aroma), safety and acceptability of the product [11]. It must also take into account

80   that the intake of high doses of amygdalin is harmful to the human body, although the

81   unpleasant taste of benzaldehyde – which is produced when it comes into contact with

82   saliva, acts as a warning sign, preventing the eater from swallowing the hydrocyanic acid

83   in amounts considered toxic for humans [12,13]. This may hinder the commercialization

84   of lots of sweet almonds in both national and international markets, and it is therefore of

85   maximum importance to eradicate this type of almond from the batches of sweet almonds

86   produced for the market.

87   Since it is extremely difficult to distinguish bitter almonds from sweet ones

88   visually in adulterated batches, it would be of great interest to the almond sector to be

89   able to use analytical tools with a high throughput which were suitable for continuous,

90   instantaneous discrimination. Thus, the implementation of NIRS sensors at receipt and in

91   the industrial sorting lines for detecting the adulteration of sweet almond batches with

92   bitter ones, could answer this demand.

93   The combination of the NIRS spectral data of the product, generated using tools

94   such as the Shewhart charts [14], allows to carry out conformity tests and product

95   deviations in comparison to the established standards, thus enabling to ensure product

96   integrity, monitor the production process and establish early warning systems. One of the

97   benefits of this approach is that it enables to reliably detect anomalies in the production

98   process. In this way, we can improve control and monitoring of the product quality

99   systems [15], using a tool which provides greater flexibility to deal with common non-

100 conformities in the product, since near infrared spectra provide comprehensive

101 information about the product encompassing highly diverse aspects related to its integrity

102 [5,16].

103   There are no articles in the scientific literature which explore the potential of NIRS

104 technology for detecting the presence of bitter almonds in lots of sweet almonds. Some

105 authors have used NIRS to classify sweet versus bitter almonds, by analysing them when

106 ground or as individual intact almond kernels and including only batches of sweet or bitter

107 almonds separately, not mixtures of both, which is the commonest way the fraud is

108 committed [17,18].

109   The aim of this research, therefore, was to analyse the viability of using NIRS

110 technology to detect the adulteration of batches of sweet almonds with bitter almonds,

111 establishing a non-targeted control procedure based exclusively on spectral information

112 to guarantee the integrity of product when received and processed in the industry, in order

113 to certify that the entire product is composed of sweet almonds. Different percentages of

114 adulterated samples were also assessed to establish the minimum limit that could be

115 detected with this methodology.

116

117 **2. Material and methods**

118

119 *2.1. Sampling*

120

121   A total of 140 samples of shelled almonds, of which 90 belonged to sweet varieties

122 (*Prunus dulcis* Mill., cv. 'Antoñeta', 'Belona', 'Guara', 'Lauranne', 'Soleta' and

123 'Vairon') and 50 to various non-specific bitter varieties harvested during the 2018-2019

124 season, were analysed in this study. To conduct this work, the batch mixing process was

125 simulated by preparing four different types of mixtures: M5 (95 % sweet almonds and 5

126 % bitter almonds), M10 (90 % sweet almonds and 10 % bitter almonds), M15 (85 %

127 sweet almonds and 15 % bitter almonds) and M20 (80 % sweet almonds and 20 % bitter

128 almonds), ending up finally with 21 samples of each class M5, M10, M15 and M20. To

129 obtain the mixtures M5, M10, M15 and M20, samples were randomly chosen from the

130 90 samples of sweet almonds (class M0) and the 50 samples of bitter almonds (class

131 M100) available. The mixtures were prepared by weighing 400-500 g of sweet almonds

132 and 25-100 g of bitter almonds, depending on the percentage of the sample to be prepared,

133 using an electronic scale (model PB3002-S, Mettler Toledo, Barcelona, Spain). Once

134 weighed, a V mixer (Afau, Zaragoza, Spain) was used to mix the two types of almonds.

135

136 *2.2. Instrumentation and NIRS spectra acquisition*

137

138   NIR spectra of the shelled almonds were collected using two NIRS instruments,

139 the Aurora and the Matrix-F spectrophotometers, which were considered suitable for the

140 *in situ* and online analysis of the product, respectively.

141   The Aurora spectrophotometer (GraiNit S.r.l., Padova, Italia) is a handheld, robust

142 and compact instrument based on the diode array technology, which works in reflectance

143 mode in the spectral range 950-1650 nm (taking data every 2 nm), with an optical window

144 of 12.56 cm^2. The sensor integration time was 6.57 ms and each spectrum was the mean

145 of 50 scans. This instrument is equipped with an internal white reference, which was

146 collected after the analysis of each sample. The UCal 4^{TM} software (Unity Scientific LLC,

147 Milford, MA, USA) was used to acquire the spectral information. To acquire the spectra,

148   the samples were uniformly distributed on a white plastic tray covering the whole surface.

149   For the analysis, four spectra were taken per sample in dynamic mode, i.e. moving the

150   sensor along the tray with the almonds, covering all the area of the tray. The four spectra

151   were averaged to provide a mean spectrum per sample.

152   The Matrix-F spectrophotometer (Bruker Optik GmbH, Ettlingen, Germany) is a

153   Fourier Transform (FT)-near infrared (NIR) instrument interfaced to a fibre optic NIR

154   illumination and detection head. The light was collected and guided via fibre optic cable

155   (5 m length) to the spectrophotometer. Furthermore, the system was equipped with a

156   conveyor belt to move the sample. A distance of 10 cm between the instrument head and

157   the conveyor belt was established, which remained constant throughout the process of

158   taking spectra. The area illuminated by this instrument was around 154 cm^2. The spectra

159   were collected in reflectance mode in the spectral range from 4000 to 12000 cm^{-1} (834–

160   2502 nm), with a resolution of 16 cm^{-1} (1.07 nm). Each spectrum was the mean of 32

161   scans. An internal white reference was also collected every fifteen minutes. OPUS

162   7.0.122 software (Bruker Optics GmbH, Ettlingen, Germany) was used for spectra

163   acquisition. The samples were placed on the conveyor belt, covering the surface, and

164   illuminated by the instrument's own source of light and analysed in static (conveyor belt

165   stopped) and dynamic (conveyor belt in motion) modes. The static mode of analysis was

166   performed by keeping the samples in a fixed point, under the light source with the

167   conveyor belt stopped. To perform the dynamic mode, a conveyor belt speed of 3.5 cm/s

168   was set. For each analysis mode, two measurements per sample were taken and averaged.

169

170   *2.3. Data processing*

171

Data pre-processing and chemometric treatments were performed using the WinISI II software package version 1.50 (Infrasoft International LLC, Port Matilda, PA, USA) and Matlab software version 2019a (The Mathworks, Inc., Natick, MA, USA).

To assess whether some parts of the spectral range presented low signal quality levels – i.e. inappropriate levels of noise – the 1,1,1,1 derivation pre-treatment was applied in order to highlight those spectrum areas where the signal to noise ratio was degraded [19]. The first digit of the 1,1,1,1 derivation treatment refers to the order of the derivative, the second to the gap over which the derivative is calculated, the third to the number of data points in a running average or smoothing and the fourth to the second smoothing [20]. This procedure was applied to both instruments.

The next step was to study the structure and spectral variability of the population of sweet samples that would be used to fix the standard. To achieve this, we applied the CENTER algorithm [21] to class M0 (N = 90 samples) of the group of sweet almonds analysed, with both the Aurora instrument in dynamic mode and the Matrix-F in static and dynamic modes. This algorithm was applied using a combination of mathematical pre-treatments, standard normal variate (SNV) and de-trending (DT) for scatter correction [22], together with the 1,5,5,1 Norris derivative treatment, which enabled to classify the samples based on their distance from the centre of the population. The CENTER algorithm performed a principal component analysis (PCA), and the Mahalanobis global distance (GH) of each sample to this centre was then calculated. We then studied those samples considered as spectral anomalies (GH > 3.5) in order to demonstrate whether or not these samples could be justifiably removed from the M0 group. To compare results, the samples identified as outliers in any of the three NIRS assays carried out were removed at the same time from the three available groups (the samples were analysed with the Aurora (dynamic mode) and Matrix-F instruments in static and dynamic modes).

197

2.3.1. Constructing the Shewhart control charts

To detect the presence of bitter almonds in batches of sweet almonds, a methodology based on the Shewhart control charts was followed [5,23] using the values of the spectral distances (Mahalanobis global distance, GH) of each of the samples tested against a standard sample of sweet almonds.

Two different strategies, in terms of the number of samples included in the standard, were followed. Initially, approximately 75 % of the samples belonging to the M0 group were used to construct the standard and the remaining 25 % to validate the conformity test performed, i.e. to assess the quality of that standard (Strategy I). Strategy II consisted of using all the samples available of the M0 group to form the standard.

The set of samples analysed with the Aurora instrument in dynamic mode was used to select the samples that would define the product standard (unadulterated sweet almonds), following Strategy I. Once the spectral outliers were removed, and after ordering the sample sets by spectral distances, in Strategy I a set consisting of 89 samples was used to construct the $M0_{standard1}$ and $M0_{test}$ sets. To achieve this, approximately 75% of the samples from the M0 group were selected, choosing 3 out of every 4 samples, to make up the $M0_{standard1}$ set (N = 68), while the remaining samples (N = 21) were used to validate the standard ($M0_{test}$). Similarly, the same samples were selected to make up the $M0_{standard1}$ and ($M0_{test}$) sets, analysed in static and dynamic modes with the Matrix-F. Likewise, and once the spectral outliers were removed from the M0 set, all the samples (N = 89) were used to build the standard ($M0_{standard2}$) for Strategy II.

For the spectral definition of the two standards (strategies I and II), a new PCA was conducted using the sample sets $M0_{standard1}$ and $M0_{standard2}$, respectively. Next, the $M0_{standard1}$ was compared independently with each of the 6 classes of analysed samples

222   (M0$_{test}$, M5, M10, M15, M20 and M100), each consisting of 21 samples, while the

223   M0$_{standard2}$ was compared with classes M5, M10, M15, M20 and M100. To do this, each

224   one of these samples was projected in the new n-dimensional space obtained with the

225   PCA defined with the product standard, in order to set up a compliance test and an early

226   warning system to control the integrity of the analysed product. This system was based

227   exclusively on spectral information obtained from the GH spectral distances of each of

228   these samples compared with the standard initially established.

229   The Shewhart chart warning and action limits were defined as the extreme

230   percentiles of the in-control distribution of the normalised Mahalanobis distance or GH

231   statistic. As this statistic is non-normally distributed, a program was developed in Matlab

232   software version 2019a (The Mathworks, Inc., Natick, MA, USA) to calculate these limits

233   for GH, following the methodology proposed by Pérez-Marín et al. [5]. In the WinISI II

234   software, GH is defined as D/p, where 'D' is the Mahalanobis distance and 'p' the number

235   of principal component factor scores utilised to calculate 'D'. For data originating from a

236   normal distribution, the distribution of D is χ^2 with p degrees of freedom. As this

237   distribution has a mean of p, GH = D/p has a mean of 1. In the Shewhart control chart,

238   the mean line was plotted as a straight line with a constant value of 1 and the action and

239   warning limits were positioned at the levels corresponding to the 97.5% and 99.5%

240   percentiles of χ^2_p divided by p. Lower limits were not considered, since small GH values

241   were not indicative of a problem. Next, the calculated GH values of the samples, which

242   were compared with the standard, were represented in the Shewhart control chart with the

243   previously calculated warning and action limits, in order to identify any samples

244   containing bitter almond kernels which would not comply with the industry's aim of

245   eliminating the presence of bitter almonds in batches of sweet almonds.

246

3. Results and discussion

3.1. Selection of optimal spectral work region and identification of outlier samples

The first derivative pre-treatment applied to the spectra of those samples analysed using the Aurora and the Matrix-F instruments in dynamic mode (Fig. 1A and Fig. 1B, respectively) showed that in the case of Matrix-F, it can be seen that at the beginning and the end, the sides of the spectral signal were degraded. With this instrument, the spectral signal is transmitted by optical fibre, which commonly produces a loss of signal quality on extreme wavelengths [24]. Consequently, the regions between 834-1165 nm and 2370-2502 nm were removed to define the optimal spectral region for study. In the case of the Aurora instrument, the whole spectral range 950-1650 nm was used.

After selecting the optimal spectral range for each instrument, the samples that presented a GH > 3.5 were studied. No samples were identified as spectral outliers in the groups of samples analysed using the Aurora instrument and the Matrix-F instrument in static mode. However, when the samples were analysed using the Matrix-F in dynamic mode, one sample presented a GH = 7.31. This sample was removed from the three groups of sweet almond samples obtained in the three tests carried out.

3.2. Definition of the quality standard

In Strategy I, the limits for the Shewhart charts were calculated using 7 principal components (PCs) in the test carried out using the Aurora instrument in dynamic mode; 9 and 8 PCs were used when the test was carried out using the Matrix-F instrument in static and dynamic mode, respectively. When Strategy II was followed, 9, 10 and 9 PCs

were used, respectively. The number of PCs were selected using the CENTER algorithm which recommends the number of PCs that make the differences in explained variance non-significant. The values obtained for the warning and action limits (Table 1) were 2.56, 2.44, 2.34, 2.26 and 3.15, 2.97, 2.83, 2.71, when the standards were calculated using 7, 8, 9 and 10 PCs, respectively [5]. More complex models, i.e. models in which a larger of principal components are used, would involve more restrictive limits.

After calculating the warning and action limits and the GH statistic values of each of the samples in the principal components space defined by the standard (control populations: $M0_{standard1}$ and $M0_{standard2}$), we identified those samples which did not meet the established criteria, with the aim of ensuring product integrity.

In Strategy I, when comparing the $M0_{test}$ group (unadulterated samples used to validate the compliance test) with the standard ($M0_{standard1}$), no samples presented a GH value higher than the action limit in the three NIRS assays carried out (Fig. 2A, Fig. 3A and Fig. 4A). These results confirm that the standard we constructed was suitable, since when samples with spectral characteristics similar to the group of $M0_{standard1}$ sweet almonds were projected in the principal components space defined by the target group, they were below the established action limit.

3.3. Identification of adulterated sweet almond batches using NIRS technology

3.3.1. In situ analysis of conformity using the handheld diode array NIRS instrument

The samples corresponding to mixtures of sweet and bitter almonds with different percentages of adulteration (M5, M10, M15 and M20) and the bitter samples (M100) were projected against the standard $M0_{standard1}$ (Fig. 2A) and $M0_{standard2}$ (Fig. 2B).

13

When Strategy I was followed, 9/21 (43 %) of the bitter almond samples (M100) were identified as 'non-compliant produce' (GH value over the action limit); this figure rose to 18/21 (86 %) when Strategy II was followed. In addition, it must be noted that the average GH value for the M100 class following Strategy I was lower compared to the one obtained using strategy II (Table 1). The differences between Strategy I and II confirm that a key aspect in this methodology is to define the standard carefully, since it is clear that when the standard covers a wider variability, the later discrimination is more accurate. Consequently, in view of the results presented above, the detection of samples of mixtures (M5, M10, M15 and M20) has certain limitations when Strategy I is followed.

After analysing the samples of mixtures, we observed that the total number of samples identified as 'non-compliant produce' following Strategy I and II were 31/84 (37 %) and 44/84 (52 %), respectively. In particular, for Strategy I, 6/21 (29 %), 8/21 (38 %), 5/21 (24 %) and 12/21 (57 %) samples, analysed in each of the 4 groups of mixtures M5, M10, M15 and M20, respectively, presented GH values above the action limit. For Strategy II, the number of samples that showed a GH value higher than this limit (Table 2) improved the percentages of adulteration detected —compared with those provided by Strategy I— for all the groups except for M10, which remained exactly the same. We also observed a large number of samples which exceeded the warning limit both with Strategy I – (8/21 (38 %), 12/21 (57 %), 7/21 (33 %) and 12/21 (57 %) samples of those analysed for groups M5, M10, M15 and M20 respectively – but especially with Strategy II (Table 2). In addition, when Strategy II was followed, we noticed that a large number of samples belonging to the M5 group – in which the amount of bitter almond in the mix was very low (5%) – were identified as 'non -compliant produce'.

The results obtained are of particular interest, since they show that this portable manual instrument could be used at the product reception points in the industry to carry

14

321 out an initial check, aimed at identifying any batches of sweet almonds which may have

322 been adulterated with bitter almonds. This would prevent parts of the batches which may

323 have been adulterated with bitter almonds being received by the industry and then being

324 processed, thus saving time and money. It is vital, however, to stress the difficulty of

325 carrying out a dynamic analysis of products with an irregular surface when the instrument

326 used is a contact instrument, as is the case with the Aurora device; on the other hand, the

327 spectrum obtained is more representative of the sample than when using one-off

328 measurements.

329

330 *3.3.2. Online analysis of conformity using the FT-NIR instrument*

331 The number of samples of mixtures (M5, M10, M15 and M20) identified as 'non-

332 compliant produce' in the test carried out in static mode (Fig. 3) following Strategies I

333 and II was 35/84 (42 %) and 29/84 (35 %), respectively. In comparison, when the test

334 was carried out in dynamic mode (Fig. 4), a total of 41/84 (49 %) and 65/84 (77 %) of

335 'non-compliant produce' when Strategies I and II were followed. The detailed study of the

336 number of samples classified as 'non-compliant produce' for each of the groups M5, M10,

337 M15 and M20 reveals that, for the M10 and M15 groups when Strategy I was followed

338 and in all cases in which Strategy II was followed, the number of samples identified as

339 'non-compliant produce' was higher in the test carried out in dynamic mode than when

340 the analyses were carried out in static mode.

341 Furthermore, studying the average GH value per category (Table 1) shows that,

342 for the test carried out in dynamic following both strategies, a greater number of bitter

343 almonds present in the sample mixture led to a sharp rise in the GH value, although the

344 increase was not so noticeable when the test was carried out in static mode following both

345 strategies. These results highlight the great importance of both sampling and acquiring

346   spectral information which is representative of the sample as a whole, which according

347   to Kuiper and Paoletti [25] and Adame-Siles et al. [26] are just as important as the

348   analytical methodology itself to achieve reliable results. In the analysis carried out in

349   dynamic mode, an average spectrum is obtained which allows to define the whole sample

350   more accurately compared to the spectrum obtained when the analysis is carried out in

351   static mode.

352   The results obtained in the test carried out in dynamic mode (Fig. 4) show that the

353   number of samples identified as 'non-compliant produce' was always higher when

354   Strategy II was followed, i.e., when the standard consisted of all the available sweet

355   samples. In fact, the greatest difference in terms of the number of samples showing a GH

356   value over the action limit when comparing Strategies I and II was obtained for the M5

357   and M10 groups, respectively, with 11 and 8 more samples above the limit in the NIRS

358   assay carried out with the Matrix-F instrument in dynamic mode, following Strategy II.

359   The close similarity between the standard group (pure sweet almonds) and the M5

360   and M10 groups (sweet almonds adulterated with only 5 % and 10 % of bitter almond

361   kernels) highlights the great importance of an accurate definition of the target when

362   identifying adulterated samples with small amounts of unwanted product. Consequently,

363   collecting a sufficient number of samples to build the standard is a key factor when

364   working with non-targeted systems, in order to cover all the possible variations inherent

365   in the target product [4]. Furthermore, it is also essential to define accurately the product

366   to be analysed and the quality of the standard, as can be seen in terms of the number of

367   samples belonging to the M100 group identified as 'non-compliant produce' and the

368   average GH values (Table 1) of the M100 group when the analyses were carried out in

369   dynamic *versus* static modes and when the Strategy I rather than Strategy II was followed.

370   The number of samples which presented GH values above the action limit when Strategy

16

371   I was followed (Fig. 3A and Fig. 4A) was 8/21 (38 %) and 21/21 (100 %) when the tests

372   were carried out in static and dynamic modes, respectively, which in turn presented

373   average GH values for the M100 group of 2.84 and 5.74. When the Strategy II was

374   followed (Fig. 3B and Fig. 4B), a total of 10/21 (48 %) and 21/21 (100 %) were identified

375   as 'non-compliant produce' in the tests carried out in static and dynamic mode, with mean

376   GH values of 2.97 and 6.44, respectively.

377   In view of these results, we can confirm the suitability of using the Matrix-F

378   instrument in dynamic mode for online detection in the sorting lines of batches of

379   adulterated sweet almonds which have not been detected in the reception controls when

380   the raw material is received in the industry, thus enabling us to discard those batches from

381   the production process.

382

383   *3.4. Implementation of NIRS technology throughout the almond supply chain to detect*

384   *bitter almonds in sweet almond batches*

385

386   The fact that the two instruments used in this study can be used in a

387   complementary way throughout the almond supply chain allows to identify in the process

388   lines any batches of sweet almonds which include bitter almonds which may not have

389   been detected using the portable manual instrument at the product reception points in the

390   industry. We therefore proceeded to study the results of all the conformity analyses for

391   Strategy II together, which were obtained when the product was analysed dynamically *in*

392   *situ* and online. Thus, the number of samples analysed dynamically which exceeded the

393   warning and action limits with the Aurora portable manual instruments and the online

394   Matrix-F for groups M5, M10, M15 and M20 following strategy II (Table 2) shows that

17

395   the largest number of samples of sweet almonds adulterated with bitter almonds that were

396   not identified by either of the two NIRS assays belonged to class M5.

397   We can therefore state that, although the percentage of samples identified as 'non-

398   compliant produce' in the M5 group was high – 16/21 (76 %) –, it is more difficult to

399   identify adulterated samples when they have a percentage of bitter almonds of 5 % or

400   less, and so the detection capabilities of the system developed in this study for this type

401   of mixture need to be improved.

402   In addition, we should note that from classes M10, M15 and M20, only 10%, 5%

403   and 14% of the samples adulterated with bitter almonds, respectively, did not exceed the

404   value of the action limit in any of the two tests carried out in dynamic mode with the

405   handheld and online NIR instruments. Admittedly, the heterogeneity of the mixture can

406   make it difficult to obtain a representative measure of the sample in those cases in which

407   a layer of sweet almonds covers the surface to be analysed and the bitter almond kernels

408   lie below that layer. Next, a detailed study of those samples belonging to the M5, M10,

409   M15 and M20 groups which were not identified as 'non-compliant produce' by any of

410   the instruments working in dynamic mode was made. This study revealed that 7 out of

411   the 11 samples that were not identified were prepared using the sweet variety 'Belona' –

412   3 belonging to M5 group, 2 to M10 and 2 to M20 – which has a large, flat kernel which

413   tends to completely cover the testing surface and hide the bitter almond kernels.

414   These results are extremely promising as regards the use of this non-targeted fraud

415   identification approach as a suitable way of carrying out both *in situ* and online screening

416   of the product when it is received in batches and processed in the industry. In addition,

417   the results obtained confirm the great utility of the non-targeted system used in this study,

418   since it allows to reduce the number of analyses conducted by a confirmatory system by

419   employing a fast, economical method using spectral information, which could be limited

420   exclusively to carrying out an analysis of those samples identified as 'disconformities' by

421   the non-targeted system [27,28].

422   This constitutes a major benefit, since some of the confirmatory systems used can

423   be expensive, complex, slow and destructive, such as the traditional method for

424   measuring cyanogenic compounds in almonds using high performance liquid

425   chromatography [18].

426

4. Conclusions

428

429   The results obtained illustrate that spectral NIR analysis combined with the

430   Shewhart control charts derived from the spectral information acquired with the Aurora

431   and Matrix-F instruments provide an extremely useful tool for detecting adulterated

432   batches of sweet almonds in the processing industry, both on receipt and on the sorting

433   lines. This approach to non-targeted fraud identification enables to detect cases of non-

434   compliance with the standards for sweet almonds established by the industry. The results

435   confirm the importance of accurately defining the standard, in terms of setting the

436   objectives and the variability of the population: here, it is important to highlight that

437   larger, more comprehensive databases would allow to define in a more universal way the

438   desired target of the produce, which would provide a more robust approach to detecting

439   non-compliant batches. In future research, the number of samples detected as 'non-

440   compliant product' when the percentage of bitter almonds in the sweet almond batches is

441   less than 5 % should be increased, with readjusted action and warning limits to take into

442   consideration not only GH statistical distribution but also the population characteristics

443   of the samples used to set the standard, thereby ensuring a more robust model for

444   detecting non-compliant batches.

445

CRediT authorship contribution statement

447

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

458

Declaration of Competing Interest

460

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

464

473 **References**

474

475 [1] C. Elliott, Elliott Review into the Integrity and Assurance of Food Supply Networks

476 –Final Report, A National Food Crime Prevention Framework, HM Government,

477 (2014)

478 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/atta

479 chment_data/file/350726/elliotreview-final-report-july2014.pdf. (accessed 5

480 November 2020).

481 [2] J. Riedl, S. Esslinger, C. Fauhl-Hassek, Review of validation and reporting of non-

482 targeted fingerprinting approaches for food authentication, Anal. Chim. 885

483 (2015) 17–32.

484 [3] G.M. Fiorino, C. Garino, M. Arlorio, A. F. Logrieco, I. Losito, L. Monaci, Review

485 article: overview on untargeted methods to combat food frauds: a focus on fishery

486 products, J. Food Qual. (2018) 1581746.

487 [4] T.F. McGrath, S.A. Haughhey, J. Patterson, C. Fauhl-Hassek, J. Donarski, M.

488 Alewjin, S. van Ruth, C.T. Elliott, What are the scientific challenges in moving

489 from target to non-targeted methods for food fraud resting and how can they be

490 addressed?- Spectroscopy case of study, Trends Food Sci. Technol. 76 (2018) 38–

491 55.

492 [5] D. Pérez-Marín, L. Calero, T. Fearn, I. Torres, A. Garrido-Varo, M.T. Sánchez, A

493 system using *in situ* NIRS sensors for the detection of product failing to meet

494   quality standards and the prediction of optimal postharvest shelf-life in the case

495   of oranges kept in cold storage, Postharvest Biol. Technol. 147 (2019) 48–53.

496   [6] J.U. Porep, D.R. Kammerer, R. Carle, Online application of near infrared (NIR)

497   spectroscopy in food production, Trends Food Sci. Technol. 46 (2015) 211–230.

498   [7] V. Cortés, J. Blasco, N. Aleixos, S. Cubero, P. Talens, Monitoring strategies for

499   quality control of agricultural products using visible and near-infrared

500   spectroscopy: A review, Trends Food Sci. Technol. 85 (2019) 138–148.

501   [8] K. B. Beć, J. Grabska, H. W. Siesler, C. W. Huck, Handheld near-infrared

502   spectrometer: Where are we heading?, NIR news 31 (2020) 28–35.

503   [9] D. Pérez-Marín, A. Garrido, NIR sensors for the in-situ assessment of Iberian ham,

504   in: A. Cifuentes (Ed.), Comprehensive foodomics, Elsevier in press, Kidlington,

505   Oxford, UK, 2020.

506   [10] D. Pérez-Marín, New perspectives and challenges to face food integrity issues using

507   NIRS: unlocking its value as a smart sensing device. Keynote Lecture 19th

508   International Council for NIR Spectroscopy Meting, NIR2019, 15-20 September

509   2019, Gold Coast, Australia.

510   [11] R. Sánchez-Pérez, K. Jørgensen, C.E. Olsen, F. Dicenta, B.L. Møller, Bitterness in

511   almonds, Plant Physiol. 146 (2008) 1040e1052.

512   [12] J. Vetter, Plant cyanogenic glycosides, Toxicon 38 (2000) 11–36.

513   [13] A.V. Morant, K. Jørgensen, C. Jørgensen, S.M. Paquette, R. Sánchez-Pérez, B.L.

514   Møller, S. Bak, β-Glucosidases as detonators of plant chemical defense,

515   Phytochemistry 69 (2008) 1795–1813.

516   [14] W.A. Shewhart, Economic control of quality of manufactured product, Van Nostrand

517   Reinhold Co., Princeton, NJ, USA, 1931.

518   [15] P. Gejdoš, Continuous quality improvement by statistical process control, Procedia

519   Econ. Financ. 34 (2015) 565–572.

520   [16] M. Montoya, J. Laxalde, M. Veleva, J.G. Rosas, F. Soulas, Control of raw materials

521   with near infrared spectroscopy: a qualitative approach, NIR news 24 (2013) 4–6.

522   [17] E. Borrás, J.M. Amigo, F. Van den Berg, R. Boqué, O. Busto, Fast and robust

523   discrimination of almond (*Prunus amygdalus*) with respect to their bitterness by

524   using near infrared and partial least squares-discriminant analysis, Food Chem.

525   153 (2014) 15–19.

526   [18] V. Cortés, P. Talens, J.M. Barat, M.J. Lerma-García, Potential of NIR spectroscopy

527   to predict amygdalin content established by HPLC in intact almonds and

528   classification based on almond bitterness, Food Control 91 (2018) 68–75.

529   [19] W. R. Hruschka, Data analysis: wavelength selection methods, in: P. C. Williams,

530   K. H. Norris (Eds.), Near-infrared technology in the agricultural and food

531   industries, American Association of Cereal Chemists Inc., St. Paul, MN. 2001, pp.

532   39–58.

533   [20] J.S. Shenk, M.O. Westerhaus, Routine operation, calibration, development and

534   network system management manual, NIRSystems Inc., Silver Spring, MD, USA,

535   1995b.

536   [21] J.S. Shenk, M.O. Westerhaus, Analysis of agriculture and food products by near

537   infrared reflectance spectroscopy, NIRSystems Inc., Silver Spring, MD, USA,

538   1995a.

539   [22] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard Normal Variate transformation and

540   De-trending of near infrared diffuse reflectance spectra, Appl. Spectrosc. 43

541   (1989) 772–777.

542 [23] R.A. Sanusi, M. Riaz, N. Abbas, Combined Shewhart CUSUM charts using auxiliary

543   variable, Comput. Ind. Eng. 105 (2017) 329–337.

544 [24] A. Garrido-Varo, A. Sánchez-Bonilla, F. Maroto-Molina, C. Riccioli, D. Pérez-

545   Marín, Long-length fiber optic near-infrared (NIR) spectroscopy probes for online

546   quality control of processed land animal proteins, Appl. Spectrosc. 72 (2018)

547   1170–1182.

548 [25] H. A. Kuiper, C. Paoletti, Food and feed safety assessment: The importance of proper

549   sampling, J. AOAC Int. 98 (2) (2015) 252–258.

550 [26] J. A. Adame-Siles, J. E. Guerrero-Ginel, T. Fearn, A. Garrido-Varo, D. Pérez-Marín,

551   Performance comparison of sampling designs for quality and safety control of raw

552   materials in bulk: A simulation study based on NIR spectral data and geostatistical

553   analysis, Chemometr. Intell. Lab. Syst. 198 (2020) 103940.

554 [27] C. Black, S.A. Haughey, O.P. Chevallier, P. Galvin-King, C.T. Elliott, A

555   comprehensive strategy to detect the fraudulent adulteration of herbs: The oregano

556   approach, Food Chem. 210 (2016) 551–557.

557 [28] E. Wielogorska, O. Chevallier, C. Black, P. Galvin-King, M. Delêtre, C.T. Kelleher,

558   S.A. Haughey, C.T. Elliott, Development of a comprehensive analytical platform

559   for the detection and quantification of food fraud using a biomarker approach. The

560   oregano adulteration case study, Food Chem. 239 (2018) 32–39.

561 **Table 1**

562 Number of principal components (PCs) used to calculate the control limits, values of the control limits and average GH value for the groups of

563 mixtures analysed using the handheld and online NIRS instruments.

| Instrument	Analysis mode	Standard strategy	PCs	Control limits		Average GH value				
				Warning	Action	M5 (95 % sweet + 5 % bitter)	M10 (90 % sweet + 10 % bitter)	M15 (85 % sweet + 15 % bitter)	M20 (80 % sweet + 20 % bitter)	M100 (100% bitter)
Aurora	Dynamic	I	7	2.56	3.15	2.56	4.17	2.79	3.42	3.84
		II	9	2.34	2.83	3.05	4.70	3.67	3.69	7.78
Matrix-F	Static	I	9	2.34	2.83	2.01	2.19	3.39	3.33	2.84
		II	10	2.26	2.71	1.72	1.85	2.92	2.93	2.97
	Dynamic	I	8	2.44	2.97	2.06	3.12	5.57	3.99	5.74
		II	9	2.34	2.83	3.74	5.08	7.99	6.89	6.44

564

565 **Table 2**

566 Number of samples showing GH values over the warning and action limits, analysed in dynamic mode in different steps of the almond supply

567 chain for Strategy II.

| Industrial step	Instrument	Limits	Mixture				
			M5 (95 % sweet + 5 % bitter)	M10 (90 % sweet + 10 % bitter)	M15 (85 % sweet + 15 % bitter)	M20 (80 % sweet + 20 % bitter)	M100 (100% bitter)
Reception	Aurora	Warning	14/21 (67 %)	11/21 (52 %)	12/21 (57 %)	15/21 (71 %)	19/21 (90 %)
		Action	12/21 (57 %)	8/21 (38 %)	11/21 (52 %)	13/21 (62 %)	18/21 (86 %)
Processing lines	Matrix-F	Warning	12/21 (57 %)	17/21 (81 %)	20/21 (95 %)	18/21 (86 %)	21/21 (100 %)
		Action	12/21 (57 %)	16/21 (76 %)	20/21 (95 %)	17/21 (81 %)	21/21 (100 %)
Reception +	Aurora +	Warning	17/21 (81 %)	19/21 (90 %)	20/21 (95 %)	18/21 (86 %)	21/21 (100 %)
Processing lines	Matrix-F	Action	16/21 (76 %)	19/21 (90 %)	20/21 (95 %)	18/21 (86 %)	21/21 (100 %)

568

569 **Fig. 1.** First derivative spectra for the different mixtures of almond samples analysed in

570 dynamic mode using the Aurora and the Matrix-F spectrophotometers.
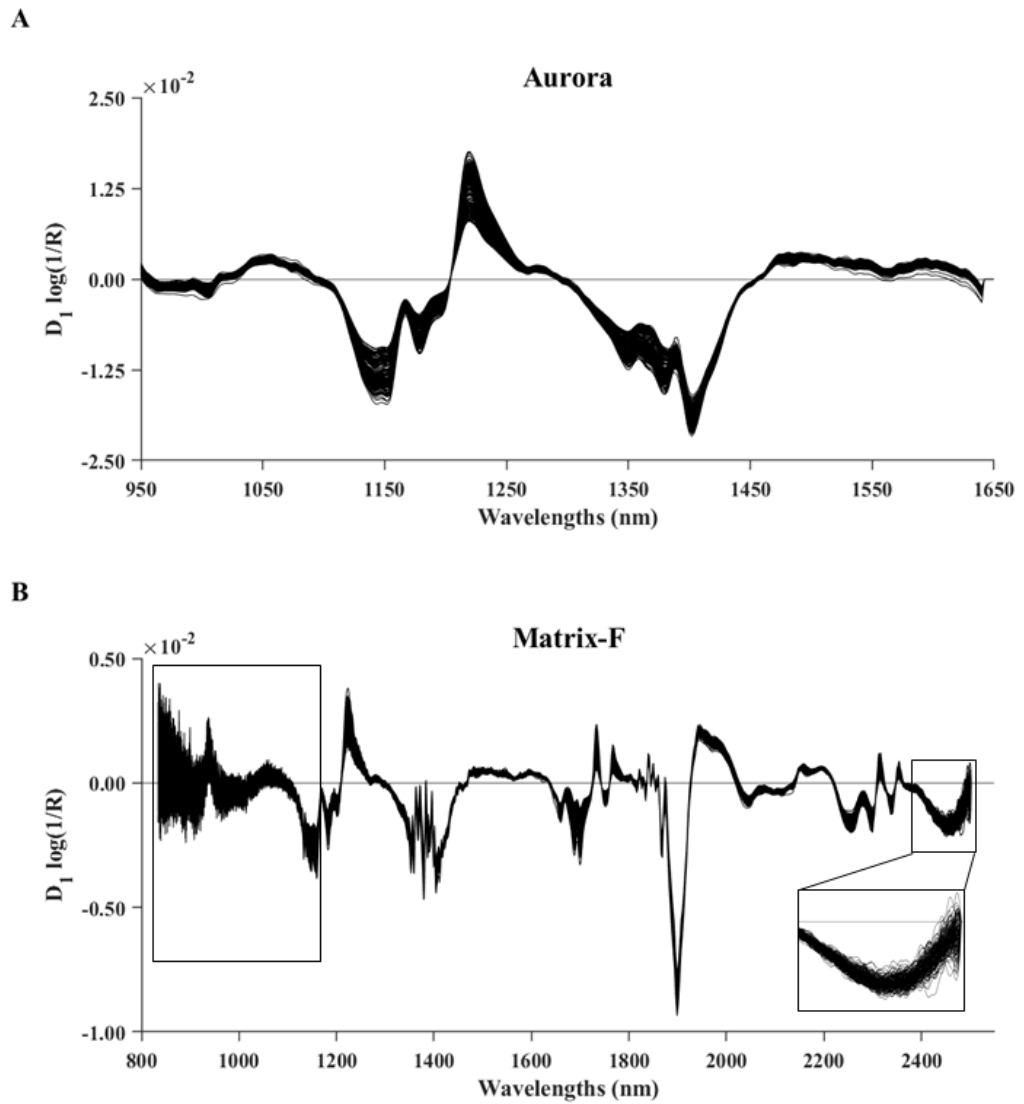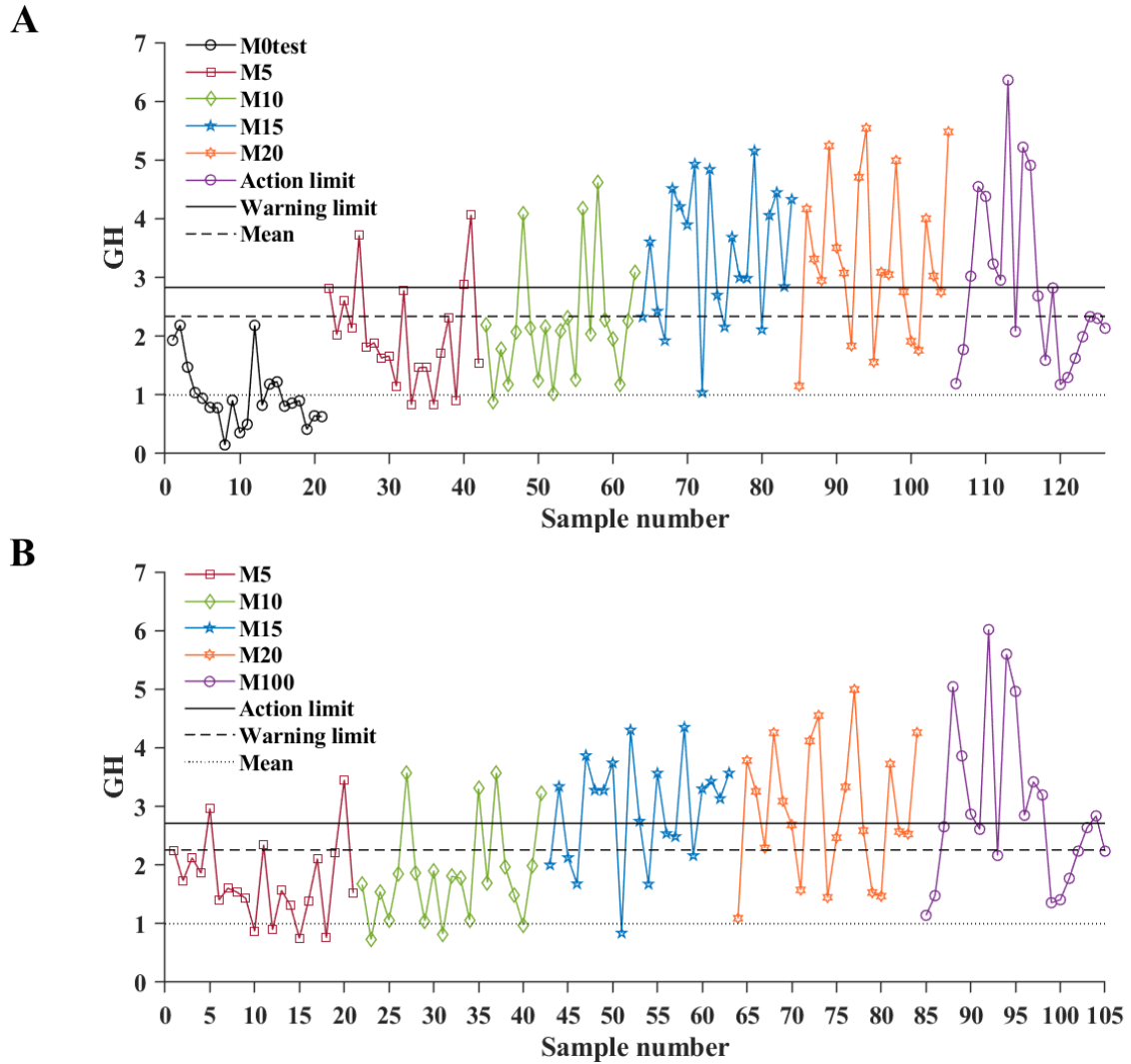

571 .

572 **Fig. 2.** Shewhart control chart based on the GH values derived from the Principal

573 Component Analysis following Strategy I (A) and II (B) for the samples analysed using

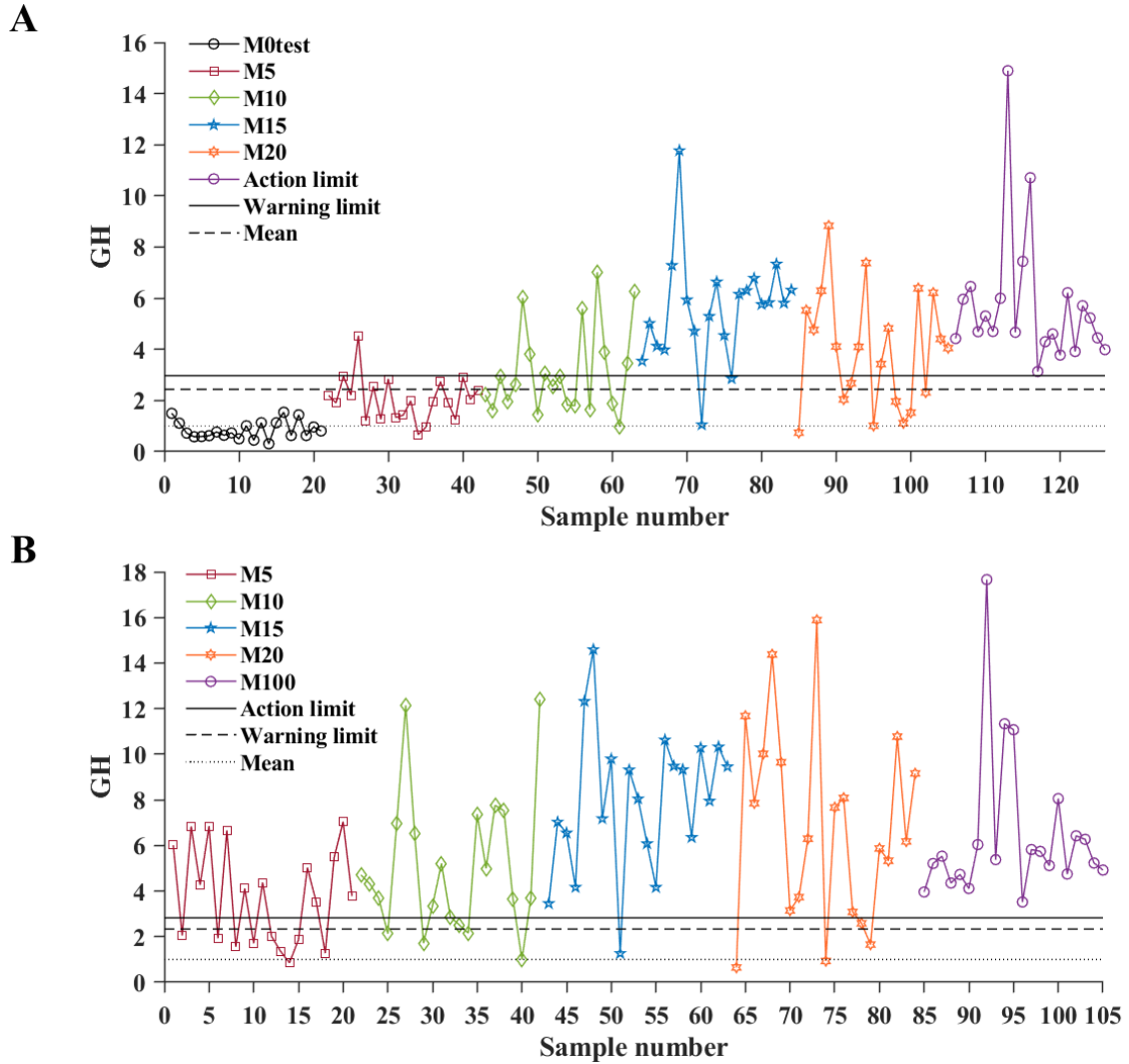574 the Aurora instrument in dynamic mode.
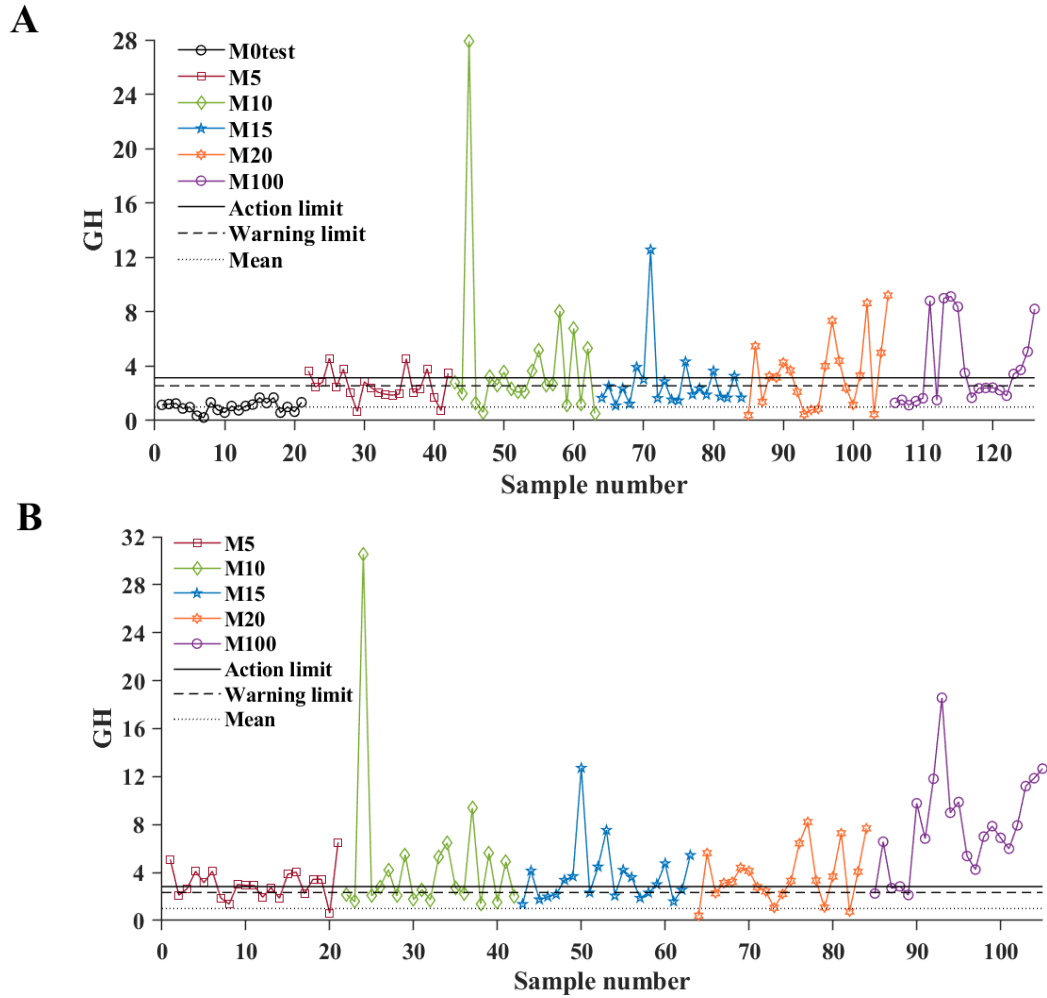
575

576

577   **Fig. 3.** Shewhart control chart based on the GH values derived from the Principal

578   Component Analysis following Strategy I (A) and II (B) for the samples analysed using

579   the Matrix-F instrument in static mode.

580

581

582   **Fig. 4.** Shewhart control chart based on the GH values derived from the Principal

583   Component Analysis following Strategy I (A) and II (B) for the samples analysed using

584   the Matrix-F instrument in dynamic mode.

585