



This document is a postprint version of an article published in Food Chemistry© Elsevier after peer review. To access the final edited and published work see <https://doi.org/10.1016/j.foodchem.2018.11.054>

Document downloaded from:



1 **Feasibility study of smartphone-based Near Infrared Spectroscopy (NIRS)**
2 **for salted minced meat composition diagnostics at different temperatures**

3 Andreas KARTAKOULLIS¹, Josep COMAPOSADA¹, Alvaro CRUZ-CARRIÓN^{1,2},
4 Xavier SERRA¹, Pere GOU^{1,3}

5 andreas.kartakoullis@irta.cat, josep.comaposada@irta.cat,
6 alvarojavier.cruz@urv.cat, xavier.serra@irta.cat, pere.gou@irta.cat

7 **Abstract**

8 This research work evaluates the feasibility of a smartphone-based spectrometer
9 (740-1070 nm) for salted minced meat composition diagnostics at industrial scale.
10 A commercially available smartphone-based spectrometer and a benchtop NIR
11 spectrometer (940-1700 nm) were used for acquiring 1312 spectra from minced
12 meat samples stored at four different temperatures ranging from -14 °C to 25 °C.
13 Thereafter, for each spectrometer, PLS and Random Forest regression models
14 specific for each temperature and global models were created to predict the fat,
15 moisture and protein contents. Fat and moisture can be estimated with the global
16 model in a wide range of temperatures by using the smartphone-based
17 spectrometer, which has an acceptable accuracy for quality control purposes
18 (RPD>7) and comparable to the accuracy of a benchtop spectrometer.

19

20 **Highlights:**

- 21 • Smartphone-based NIR was used to estimate salted meat composition
22 • Accuracy of fat and moisture estimates are acceptable for quality control
23 purposes
24 • Predictive ability of smartphone-based NIR is similar to a benchtop
25 spectrometer

¹ IRTA, Food Technology Program, Finca Camps i Armet 17121 Monells, Girona, Spain

² Universitat Rovira i Virgili, C/ Marcel·lí Domingo, 2-4-6. 43007 Tarragona, Spain

³ Corresponding author: +34-902789449 Ext. 1407

- 26 • Random Forest predictive models are robust at temperatures from -14 °C
27 to 25 °C

28

29 **1. Introduction**

30 Smartphone-based food diagnostic technologies have the potential to
31 revolutionize the food sector by allowing rapid, on-site and inexpensive analysis
32 of food and food products (Rateni, Dario, & Cavallo, 2017). This potential has been
33 recognized by the European Union (EU) and a Horizon Prize competition called
34 “Food Scanner” took place in 2016 for developing affordable and non-invasive
35 mobile solutions that will enable users to measure and analyze their food intake.
36 The winners were three European startups: Spectral Engines (winner), SCIO
37 (runner-up) and TellSpec (runner-up) (Horizon, 2016). These three companies
38 have developed novel smartphone-based food diagnostic applications by utilizing
39 the analytical power of Near Infrared Spectroscopy (NIRS). NIRS utilize the
40 spectral range from 800nm to 2500nm in order to provide chemical information
41 through the different response of molecular bonds and structural information of
42 the scanned sample (Burns & Ciurczak, 2007).

43 Several research studies have been published in the last years in the agrifood
44 sector that are taking advantage of the combined use of the rear smartphone
45 camera with augmented sensors. For example, bio-receptors (e.g. fluorescent
46 dyes) and colorimetric assays have been used for the on-site detection of
47 contaminants (Coskun, et al., 2017); (Chen, et al., 2014), pesticide residues (Levin,
48 Krishnan, Rajkumar, Halery, & Balkunde, 2016), (Wang, et al., 2016) and
49 foodborne pathogens (Zhu, Sikora, & Ozcan, 2012), (Seo, et al., 2016). As for
50 commercial applications, a great number of startups worldwide are offering
51 smartphone-based NIRS by exploding the technological advancements in the high-
52 volume fabrication of microelectromechanical systems (MEMS) and micro-mirror
53 arrays. These advancements have made possible the construction of miniaturized
54 and inexpensive NIR spectrometers while keeping the performance at similar
55 levels with the benchtop equipment (Ozcan, 2016). Device schematics and details

56 on how to design a low-cost smartphone-based spectrometers from off-the-shelf
57 components can be found in the literature (Das, Wahi, Kothari, & Raskar, 2016).

58 Spectral Engines solution uses a variety of Bluetooth-connected NIR sensors
59 together with advanced analytics, cloud-connectivity and spectral libraries to
60 reveal the fat, protein, sugar, and calories of food items (Antila, Kantojärvi, &
61 Mäkynen, 2016). TellSpec is using a pocket-sized NIR spectrometer based on the
62 Texas Instruments DLP module for scanning food through a mobile application in
63 order to identify calories, allergens, contaminants and food frauds (Beam Your
64 Health Up–TellSpec, 2018). Consumer Physics has the in-house-developed SCIO, a
65 pocket-sized NIR spectrometer for molecular analysis of different samples,
66 including food (Goldring & Sharon, 2011). In addition to the advancements in
67 miniaturizing spectrometers, a mobile phone-compatible hyperspectral imager
68 based on a tunable MEMS Fabry-Perot interferometer has been demonstrated for
69 food sensing applications among other applications (Rissanen, et al., 2016).

70 The abovementioned innovations have developed reliable food scanning mobile
71 applications but are mostly targeting the customer. There is a need to develop
72 more robust smartphone-based food diagnostic applications especially for the
73 meat industry in order to ensure food quality and safety. NIRS is a widely used
74 analytical technique in the food industry for the determination of the chemical
75 composition because of its non-destructive, fast and sensitive abilities, and
76 requires little or no sample preparation (Prieto, Roehe, Lavin, Batten, & Andres,
77 2009) (Van den Berg, Lyndgaard, Sørensen, & Engelsen, 2013). Furthermore, NIRS
78 coupled with multivariate analysis can determine the water and salt content at the
79 surface of fermented sausages during the drying process (Collell C. , Gou, Arnau,
80 Muñoz, & Comaposada, 2012), water and salt content at the surface of dry-cured
81 ham during the resting and drying processes (Collell C. , Gou, Arnau, &
82 Comaposada, 2011) and the authentication of meat products (Chiesa, et al., 2016).

83 A limitation for the adoption of NIRS-based innovations is the sample temperature
84 because it can cause significant variation to the acquired spectra. When the light
85 of wavelengths longer than 450 nm is absorbed by liquid water, the energy is
86 transferred to one or more of the vibrational modes of the O–H bond.

87 Spectroscopic studies of liquid water have shown that the shift in frequency is
88 attributed to the weakness of intermolecular H-bonds due to the rise of
89 temperature, which strengthens the covalent O-H bonds and consequently causes
90 the water molecule to vibrate at higher frequencies (Segtnan, Sasic, Isaksson, &
91 Ozaki, 2001). Furthermore, the absorption of light from liquid water at the visible
92 and NIR region is getting stronger due to the rise of temperature above the
93 freezing point of water (Pegau, Deric, & Zaneveld, 1997). Unfortunately, the
94 temperature influence is usually ignored in NIRS predictive modeling because the
95 temperature control, especially in industrial conditions, is quite difficult. However,
96 recent studies have used compensation methods to mitigate the temperature
97 influence and to improve the accuracy of the prediction by employing
98 temperature-compensated PLSR global models (Yao, Chen, Xie, & Rao, 2013)
99 (Campos, Antolin, Deban, & Pardo, 2018).

100 The aim of this paper is to study the feasibility of smartphone-controlled near-
101 infrared spectrometers for salted minced meat composition prediction at different
102 temperatures in comparison to a benchtop spectrometer. Additionally, linear and
103 nonlinear regression methods have been used for constructing global models with
104 the smartphone-based spectrometer that are robust for a great range of
105 temperatures, from -14°C to 25°C. Furthermore, the effect of different data
106 partitioning methods on how they affect the construction of predictive models is
107 also investigated.

108 **2. Material and Methods**

109 **2.1 Experimental Design**

110 Green hams (n= 328) from different origins were purchased: 171 Serrano-type
111 hams from Jamonificio Subirats, S.L. (Els Hostalets de Balenyà, Barcelona, Spain),
112 72 Serrano-type hams from Càrniques de Juià, S.A. (Juià, Girona, Spain), 30 Duroc
113 Serrano-type hams from Càrnica Batallé, S.A. (Riudarenes, Girona, Spain) and 55
114 Iberian hams from Corporación Alimentaria Guissona, S.A., (Guissona, Lleida,
115 Spain).

116 Green hams were salted by covering them with a layer of salt of at least 10 cm for
117 different periods of time (7 days to 16 days) in order to obtain a high variability in

118 salt uptake, and therefore, in composition. Each green ham was carefully dissected
119 and the subcutaneous fat, lean and intermuscular fat were minced together in a
120 bowl chopper until a homogeneous meat paste was obtained. Four samples from
121 each ham were vacuum packed and stored at different temperatures (-14 °C, +5
122 °C, +12 °C and +25 °C) until data acquisition. The whole procedure resulted in
123 samples with fat content varying from 5% to 43%, protein content varying from
124 12% to 23% and moisture content varying from 35% to 69%.**2.2**

125 **Instrumentation**

126 Two NIR instruments were used. The first one is the smartphone-based SCiO
127 (Consumer Physics, Israel) that can acquire 331-points reflection spectra that are
128 ranging from 740 to 1070 nm. For controlling the SCiO spectrometer a Samsung
129 Galaxy Core Prime with an Android 5.1.1 operating system was used. The second
130 NIR instrument is a diode array Polychromix Spectral Probe (Polychromix Inc.,
131 Wilmington, USA) with InGaAs detector type that is covering the spectral region
132 between 940 and 1700 nm with a resolution of 8 nm. Polychromix was used to
133 compare the results obtained with SCiO. Polychromix reflectance data were stored
134 as the logarithm of reciprocal of reflectance (1/R). For both instruments, three
135 spectra per sample were collected and the mean spectrum per sample was used
136 for further analysis. All the equipment was set in reflectance mode and operated
137 at room temperature (20 ± 3 °C).

138 The moisture, protein and fat contents of the meat samples were determined by
139 using a FoodScan™ Lab (Foss Analytical, Hillerød, Denmark), which is an approved
140 method for the analysis of moisture, fat and protein in meat and meat products by
141 the Association of Analytical Communities, labeled as 2007.04 method (AOAC,
142 2007).

143 **2.3 NIRS Analysis**

144 **2.3.1 Data Acquisition**

145 Data acquisition was performed using the two different NIRS systems. Each
146 sample was removed from the storage rooms and immediately three SCiO scans
147 and three Polychromix scans were acquired at three different random spots and
148 subsequently averaged. Great attention was given to keep the total acquisition

149 time for both the instruments less than 2 minutes in order to ensure that the
150 temperature did not change significantly the sample composition during the
151 spectra acquisition.

152 **2.3.2 Data partitioning**

153 Dataset from each acquisition temperature was divided in two: a training dataset
154 with 80% of the initial data and a testing dataset with 20% of the initial data. The
155 general aim of performing the data partitioning is to prevent knowledge from the
156 training dataset to “leak” to the testing dataset, which will affect the predictive
157 power of the PLSR model (Wold, Sjöström, & Eriksson, 2001). Two different
158 methods for data partitioning have been selected, the first one is using only the
159 responses **Y** to construct the two datasets by separating the values of **Y** with
160 random stratification into groups. Each of the stratified groups contains
161 approximately the same percentage of each response as in the original **Y** dataset
162 (Molinaro, Simon, & Pfeiffer, 2005). For constructing the global model, the
163 datasets of each temperature were merged and the training and testing datasets
164 were obtained with the random stratification on the Y-dataset.

165 The second method is to use the properties of the NIR spectra from the predictors
166 **X** dataset. The Kennard-Stone (KS) algorithm was used for this partitioning, which
167 is a sequential method that covers the spectral space uniformly in order to build a
168 training dataset with flat distribution over the spectral space. It starts by
169 computing the geometric distances (e.g. Euclidean) between all the pairs of
170 spectra and the two spectra that are the farthest apart are assigned to the training
171 dataset. The selection procedure is repeated until to build the training dataset
172 according to the following maximin criterion, the next selected spectra must have
173 the least distance with respect to any other spectra already selected (Puzyn,
174 Mostrag-Szlichtyng, Gajewicz, Skrzyński, & Worth, 2011). Mahalanobis distance
175 can be used also for satisfying the maximin criterion by performing Principal
176 Component Analysis (PCA) on the spectra and computing the Euclidean distance
177 of the PCA scores according to the following definition of the Mahalanobis
178 distance:

$$179 \quad H_{ij}^2 = \sum_{n=1}^N (\hat{t}_{i,n} - \hat{t}_{j,n})^2 / \hat{\lambda}_n \quad (1)$$

180 where $\hat{t}_{i,n}$ is the n-th principal component score of point i , $\hat{t}_{j,n}$ the corresponding
181 principal component score for point j , $\hat{\lambda}_n$ is the eigenvalue of principal component
182 n and N is the number of the selected principal components.

183 **2.3.3 Data pretreatment**

184 The scatter correction methods include Multiplicative Scatter Correction (MSC),
185 Inverse MSC, Extended MSC, Extended Inverse MSC, detrending, Standard Normal
186 Variate (SNV) and different kinds of normalization to unit-vector length (e.g.
187 scaling). The most common spectral derivatives methods are the gap-segment
188 (also known as Norris-Williams) derivatives and the Savitzky-Golay (SG)
189 polynomial derivative filters (Rinnan, van den Berg and Engelsen 2009).

190 The SG method is fitting polynomials to windows around each point in the
191 spectrum and these polynomials are then used to smooth the obtained data and
192 subsequently differentiate them. Finite difference method can be used instead of
193 the polynomial fitting for the computation of the first and second derivative by
194 considering two different points of the spectrum:

$$195 \quad x'_i = x_i - x_{i-1} \quad (2)$$

$$196 \quad x''_i = x_{i-1} - 2x_i + x_{i+1} \quad (3)$$

197 This numerical subtraction deemphasizes lower frequencies and emphasizes
198 higher frequencies (e.g. noise). Data smoothing methods are needed in order to
199 improve the signal-to-noise ratio of the obtained data. The gap-segment method
200 first performs a smoothing under a given segment size, followed by a gap
201 derivative:

$$202 \quad x'_i = x_{i+k} - x_{i-k} \quad (4)$$

$$203 \quad x''_i = x_{i-k} - 2x_i + x_{i+k} \quad (5)$$

204 where k is the gap size between the points.

205 **2.3.4 Predictive modelling**

206 Partial Least Squares (PLS) Regression (also known as Projection of Latent
207 Structures) is one of the most popular tools for multivariate analysis (Wold H. ,

208 1966). PLSR derives its popularity from the ability to analyze data with multiple,
 209 noisy, collinear (e.g. NIR spectra), and even incomplete variables. Its goal is to
 210 predict a set of dependent variables (denoted as \mathbf{Y}) from a set of predictors
 211 (denoted as \mathbf{X}). The prediction is achieved by extracting from the predictors a new
 212 set of orthogonal factors called Latent Variables (LVs) that have the best predictive
 213 power on unseen data. More specifically, PLSR tries to find two sets of weights \mathbf{w}
 214 and \mathbf{c} in order to extract two vectors from \mathbf{X} and \mathbf{Y} , $\mathbf{t}=\mathbf{X}\mathbf{w}$ and $\mathbf{u}=\mathbf{Y}\mathbf{c}$ such that
 215 maximize the covariation between \mathbf{X} and \mathbf{Y} . It can be achieved with the following
 216 cost function:

$$217 \quad \text{maximize cov}(\mathbf{X}\mathbf{w}, \mathbf{Y}) \text{ subject to } \mathbf{w}^T\mathbf{w}=1 \quad (6)$$

218 The number of LVs defines the complexity of the predictive model and selecting
 219 the optimal number is one the most important steps because the amount of
 220 variance explained by a LV indicates its importance in the prediction of \mathbf{Y} . For
 221 example, selecting too many LVs will result in an over-fitted model that takes into
 222 consideration not only the variance of the data but also noise. While selecting too
 223 few LVs implies an under-fitted model, which incorporates insufficient
 224 information of the data.

225 In the current study, the target \mathbf{Y} variables describe the concentration of certain
 226 components and \mathbf{X} are the spectral data. The PLSR analysis was applied to the
 227 training dataset for each acquisition temperature. We also constructed a global
 228 model for investigating further the temperature dependency on the predictive
 229 ability of the PLSR model. For each instrument, the four training datasets acquired
 230 at the four different temperatures were merged to create a global model for each
 231 instrument. The Root Mean Squared Error of Prediction (RMSEP), the coefficient
 232 of the determination (R^2) and the residual predictive deviation (RPD) were used
 233 to decide the number of LVs.

234 The RMSE of calibration was calculated:

$$235 \quad RMSEC = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N - LVs - 1}} \quad (7)$$

236 2.3.5 Validation of predictive models

237 Predictive models obtained after PLSR analysis of training dataset were applied to
238 the testing data set. The RMSEP, R^2 and RPD were calculated:

$$239 \quad RMSEP = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad (8)$$

$$240 \quad R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (9)$$

241 where N is the size of the dataset, Y_i is the experimentally measured reference
242 value for sample i , \hat{Y}_i is the predicted reference value for sample i and \bar{Y} is the
243 arithmetic mean of the corresponding dataset.

244 RPD is the ratio between the standard deviation of the reference values and the
245 error of prediction of the testing. It has been suggested that RPD values >3 are
246 considered good for screening purpose; RPD values >5 are good for quality control,
247 whereas RPD values >8 are considered excellent for all analytical tasks (Conzen,
248 2006).

249 **2.4 Software and Algorithms**

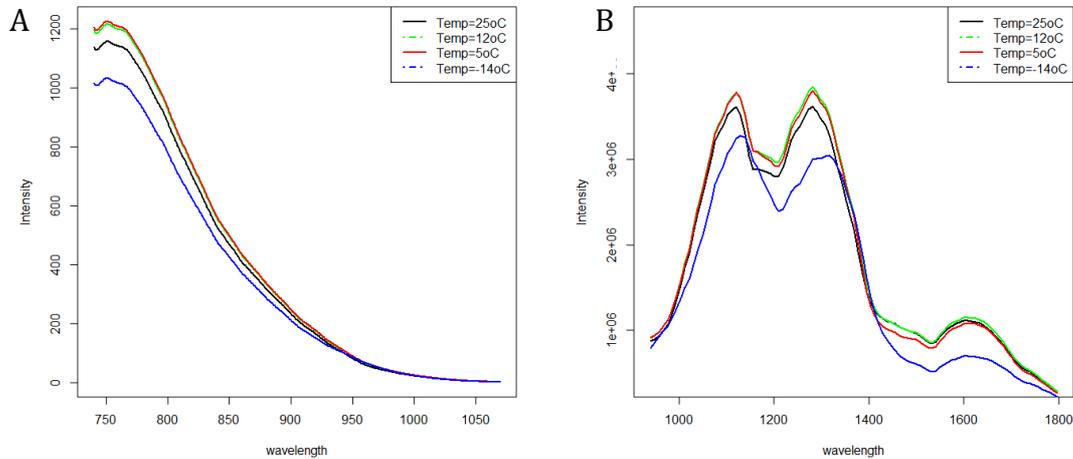
250 The open-source programming language R (version 3.4.0, codename: “You Stupid
251 Darkness”) (The R Core Team, 2018) was used to develop in-house codes based
252 on the PLS and randomForest packages (Mevik & Wehrens, 2007) (Liaw & Wiener,
253 2002).

254 **3. Results and Discussion**

255 **3.1 Spectral description of the samples**

256 Figure 1 shows the raw mean spectrum of all the collected spectra at each one of
257 the acquisition temperatures. As expected, there are differences between the
258 different temperatures even with a shift in frequency (frozen sample at -14 °C)
259 due to perturbations caused by the increase in temperature (Fig. 1B). The primary
260 vibrational modes of liquid water are in the mid-infrared region at 3049 nm (ν_1 ,
261 symmetric stretch), 6079 nm (ν_2 , bend), and 2865 nm (ν_3 , asymmetric stretch).
262 Several higher overtones of the vibrational modes can be found at the detection
263 area of our equipment between 750-1800nm and therefore the whole spectrum

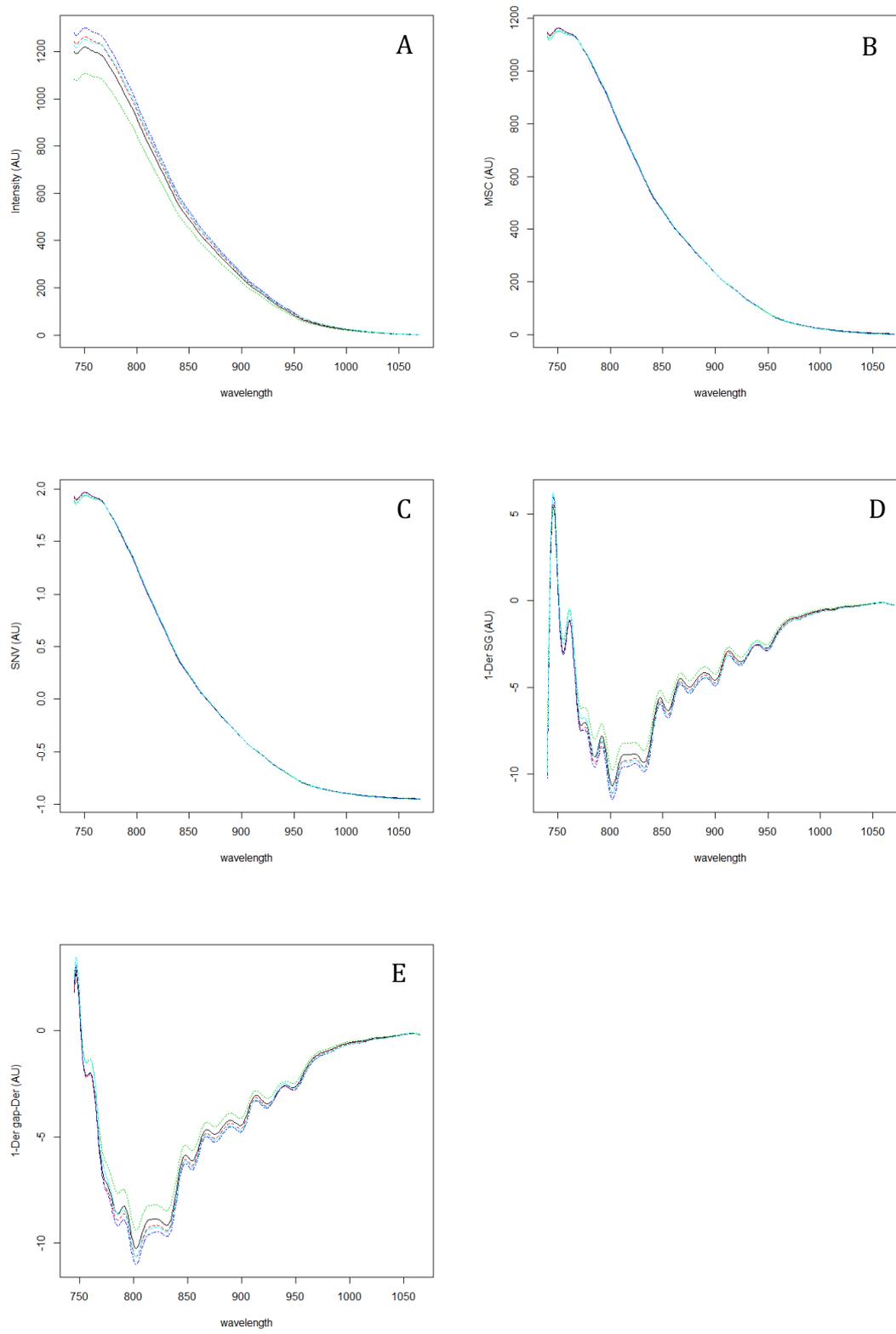
264 from each sample will be used for the construction of the predictive model in order
265 not to exclude any informative part of the spectrum.



266 *Figure 1: The mean spectrum for each temperature for SciO (A) and Polychromix (B).*

267 **3.2 Pretreatment of raw spectra**

268 Figure 2 shows the effect of the pretreating methods upon 5 different SciO spectra
269 collected from samples stored at 25 °C. As expected the SNV and MSC are returning
270 almost identical spectra with the only difference to be the offset of the intensity.
271 The spectra obtained from the two spectral derivatives methods shows distinct
272 differences between them, because of the different methods used for the
273 computation of the derivatives. The spectral derivatives need thoroughly tuning
274 because they have a number of parameters that are critical for building a
275 predictive model, such as the order and length of the polynomial for the SG, and
276 the size of the gap and the segment size of the smoothing filter for the gap-segment.



277 *Figure 2: SciO spectra without any pretreatment (A), with MSC pretreatment (B), with SNV pretreatment (C),*
 278 *with SG-1Der pretreatment (D), and with gap-segment derivative pretreatment (E).*

279 Several pretreatments have been used in order to build a PLSR model with the
 280 best overall predictive ability for the two different spectrometers. For scatter

281 correction, we have selected to use only SNV for practical reasons. The signal-
282 correction concept behind the SNV and MSC is similar except that a reference
283 spectrum is needed for the MSC correction. Usually, the reference spectrum is the
284 mean spectrum of all the measurements and that can lead to the so-called *data*
285 *leakage*. Data leakage is when unexpected information is being added to the
286 training dataset that can allow the model to learn something that it otherwise
287 would not know. This will hinder the performance of the predictive model under
288 construction (Nisbert, Elder, & Miner, 2009). Therefore, the selection of the MSC
289 reference spectrum is a great challenge and that gives to SNV a practical advantage
290 (Rinnan, van den Berg, & Engelsen, 2009). As for spectral derivatives, the gap-
291 segment method performs better with the SCiO data compared to the SG method
292 that performs better with the Polychromix data.

293 **3.3 Predictive ability of NIR devices with two different data partitioning** 294 **methods**

295 Table 1 and Table 2 show, for both NIR devices the predictive ability of the
296 different PLSR models created by using the random stratification method and the
297 KS method respectively for partitioning the data into training and testing datasets.

298 The predictive models were constructed on the SNV+gap-seg and SNV+SG
299 pretreated datasets. Fat, protein and moisture content of the samples were
300 accurately predicted with the lowest value of the coefficient of the determination
301 to be $R^2=0.92$ and with 4 LVs for protein estimation at 25 °C. The RPD values for
302 all the predictions are above 3 and in some cases above 8, which means that our
303 PLSR modelling could be accurate enough for quality control of fat and moisture
304 contents and for screening purposes of protein content. It should be noted that
305 RMSEP, R^2 and RPD for fat and moisture prediction are worst in frozen samples,
306 especially with SCiO device. RMSEP values for fat and moisture prediction are
307 similar to the results reported in previous studies where benchtop spectrometers
308 have been used in minced mass of pork sausages, for example SEP: 0.94-1.41% for
309 fat and 0.76–1.01% for moisture (Ortiz-Somovilla, Espana-Espana, Gaitan-Jurado,
310 Perez-Aparicio, & De Pedro-Sanz, 2007), RMSEP= 0.622-0.675% for moisture
311 (Collell C. , Gou, Picouet, Arnau, & Comaposada, 2010), and in minced *Biceps*

312 *femoris* muscle of dry-cured ham RMSEP: 0.31-0.43% for fat and 0.44-0.51% for
 313 moisture (Prevolnik, et al., 2011).

314 The results obtained with datasets built with the KS method for partitioning the
 315 data (Table 2) are comparable to those obtained with datasets built with the
 316 stratified random partitioning based on Y values (Table 1). KS method is
 317 concentrating most of the diversity of the data at the training set that in some cases
 318 can lead to overoptimistic results (Wested & Marini, 2015), which is not the case
 319 for our study since the two data partitioning methods are returning similar results.
 320 Similar methods to KS, such as the Duplex algorithm (Daszykowski, Walczak, &
 321 Massart, 2002), have great potential for building new predictive models with
 322 smartphones because they require less computational effort compared to the
 323 random stratification.

324 *Table 1: Results of PLSR analysis after dataset partitioning based on random stratification sampling applied to*
 325 *response values (Y) for each temperature of spectra acquisition.*

Parameter	Pretreatment	LVs	RMSEC	RMSEP	R ²	RPD
SCiO, Temp=-14°C						
Fat	SNV+gap-seg	6	1.26%	1.47%	0.97	6.89
Moisture		6	1.23%	1.36%	0.96	5.48
Protein		6	0.43%	0.54%	0.93	3.89
Polychromix, Temp=-14°C						
Fat	SNV+SG (1D)	4	0.99%	0.92%	0.99	10.34
Moisture		5	0.78%	0.76%	0.99	7.46
Protein		5	0.49%	0.48%	0.94	3.91
SCiO, Temp=5°C						
Fat	SNV+gap-seg	5	0.97%	1.00%	0.99	6.99
Moisture		5	0.89%	1.06%	0.98	6.86
Protein		5	0.49%	0.43%	0.95	3.87
Polychromix, Temp=5°C						
Fat	SNV+SG (1D)	6	0.85%	0.74%	0.99	14.11
Moisture		6	0.82%	0.88%	0.98	7.99
Protein		5	0.42%	0.49%	0.93	4.76

SCiO, Temp=12oC						
Fat	SNV+gap-seg	4	1.05%	0.94%	0.99	12.05
Moisture		5	0.96%	0.83%	0.98	8.55
Protein		5	0.41%	0.45%	0.96	4.18
Polychromix, Temp=12°C						
Fat	SNV+SG (1D)	6	0.86%	0.92%	0.99	10.79
Moisture		4	0.79%	0.88%	0.98	9.76
Protein		4	0.46%	0.45%	0.95	5.04
SCiO, Temp=25°C						
Fat	SNV+gap-seg	6	0.97%	0.74%	0.99	12.65
Moisture		4	0.85%	0.71%	0.98	9.05
Protein		5	0.45%	0.47%	0.95	3.99
Polychromix, Temp=25°C						
Fat	SNV+SG (1D)	5	0.92%	0.86%	0.99	10.65
Moisture		5	0.76%	1.01%	0.98	9.56
Protein		6	0.38%	0.36%	0.96	4.09

326
327

Table 2: Results of PLSR after dataset partitioning based on the KS method applied to X descriptors (NIR spectra).

Parameter	Pretreatment	LVs	RMSEC	RMSEP	R ² _p	RPD
SCiO, Temp=-14°C						
Fat	SNV+gap-seg	6	1.23%	1.43%	0.98	7.12
Moisture		6	1.29%	1.02%	0.97	6.34
Protein		5	0.48%	0.49%	0.92	3.29
Polychromix, Temp=-14°C						
Fat	SNV+SG (1D)	4	1.13%	0.76%	0.99	7.51
Moisture		5	0.82%	0.72%	0.98	8.79
Protein		4	0.50%	0.45%	0.94	2.64
SCiO, Temp=5°C						
Fat	SNV+gap-seg	4	1.11%	0.83%	0.99	11.64
Moisture		4	0.99%	0.78%	0.99	8.57
Protein		6	0.46%	0.43%	0.94	3.87

Polychromix, Temp=5°C						
Fat	SNV+SG (1D)	4	0.95%	0.90%	0.99	9.38
Moisture		5	0.83%	0.90%	0.98	7.96
Protein		6	0.42%	0.47%	0.93	4.11
SCiO, Temp=12°C						
Fat	SNV+gap-seg	5	1.00%	1.19%	0.98	10.15
Moisture		5	0.93%	1.12%	0.97	7.95
Protein		6	0.46%	0.52%	0.93	3.21
Polychromix, Temp=12°C						
Fat	SNV+SG (1D)	4	1.02%	0.76%	0.99	8.54
Moisture		5	0.83%	0.79%	0.98	6.83
Protein		4	0.48%	0.43%	0.93	3.35
SCiO, Temp=25°C						
Fat	SNV+gap-seg	6	1.01%	0.80%	0.99	10.50
Moisture		6	0.87%	0.78%	0.98	7.54
Protein		5	0.49%	0.37%	0.95	3.37
Polychromix, Temp=25°C						
Fat	SNV+SG (1D)	4	0.98%	0.68%	0.99	9.46
Moisture		4	0.76%	0.84%	0.99	7.82
Protein		4	0.44%	0.47%	0.92	3.41

328 3.4 Global Model

329 For constructing the global model, the datasets of different temperatures were
330 merged and the training and testing datasets were obtained with the random
331 stratification on the Y-dataset. Table 3 shows that when a global temperature
332 model is applied, in most of the cases the SCiO has results similar to those obtained
333 with local temperature models. The predictive ability for the fat content of the
334 SCiO global model has slightly improved for the -14 °C samples from
335 RMSEP=1.47% to RMSEP=1.18% and for the 5 °C samples from RMSEP=1.00% to
336 RMSEP=0.88%. Predictive errors do not show any specific pattern related to the
337 temperature (Figure 3a-c).

338 However, Figure 3b shows that the predicted values of moisture are always lower
 339 than the measured values at the midrange (53-59%). Figure 3a for fat predictions
 340 has a similar trend at the midrange but to a lesser extent. There is a non-linear
 341 relationship between moisture or fat content and spectral response over large
 342 ranges of fat or moisture. PLS, as a linear regression method, cannot handle this
 343 non-linear relationship and some part of the explanatory information is going to
 344 the residuals of the model. Therefore, for estimating correctly the predicted values
 345 of moisture and fat, non-linear regression methods are needed such as Random
 346 Forest (RF) regression that has the ability of fully utilizing the explanatory
 347 information from various types of data (Hastie, Tibshirani, & Friedman, 2009).

348 RF is an ensemble learning algorithm that can be used for both classification and
 349 regression. It is using a large number of decision trees (ntree) to split the data into
 350 an equal number of bootstrap samples that will be randomly sampled by a number
 351 of predictors (mtry). This adds randomness to the model while growing the
 352 number of the decision trees, and the algorithm chooses the best split between the
 353 sampled variables leading to a wide diversity that generally results in a better
 354 model (Breiman, 2001).

355 Figure 3e-f depicts the new prediction plots and it is obvious that the RF
 356 regression has corrected the non-linear relationship between moisture and fat
 357 content. The RMSEC values at Table 2 have significantly improved because the
 358 model captures all the explanatory information from the data, but the RMSEP
 359 values are at the same levels as PLS results.

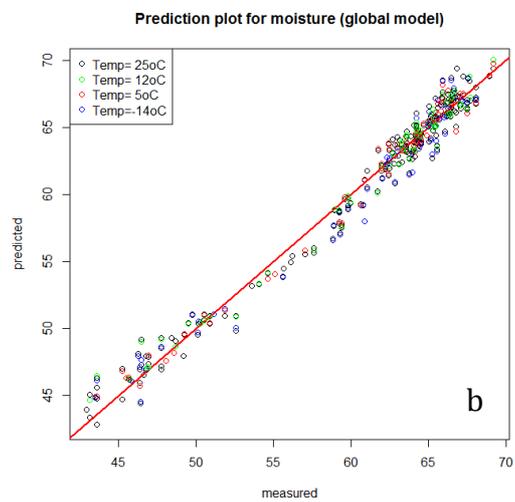
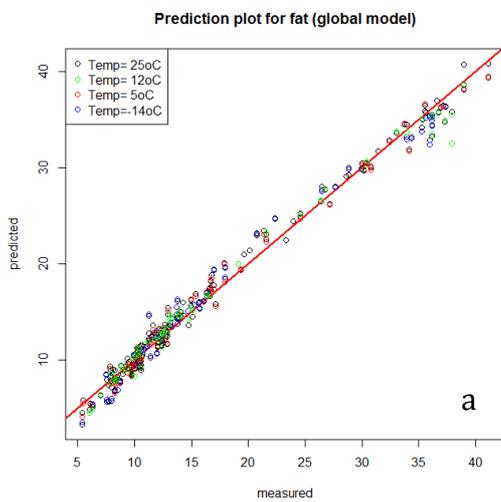
360 Therefore, the RF regression global temperature model build with smartphone-
 361 based NIR spectrometers has the potential to predict fat and moisture contents at
 362 different temperatures, even in frozen samples, with high accuracy and protein
 363 content with good accuracy.

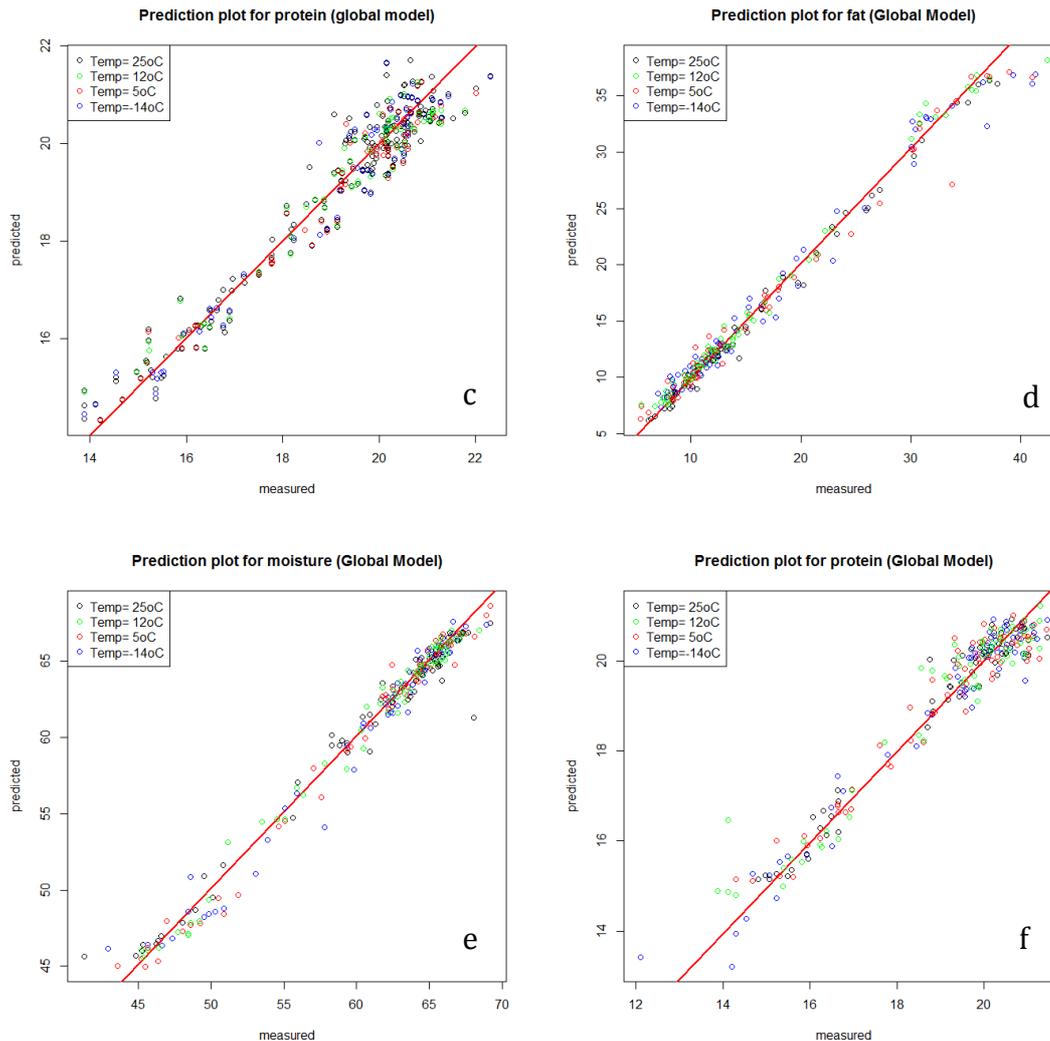
364 *Table 3: Results of the global PLS and RF regression after datasets selection based on random stratification*
 365 *sampling applied to response values (Y) for all temperatures of spectra acquisition.*

Parameter	Pretreatment	Parameter	RMSEC	RMSEP	R ² _p	RPD
PLS SCiO Global						
Fat	SNV+gap-seg	8 LVs	1.21%	1.07%	0.99	8.89

Moisture		7 LVs	1.01%	1.03%	0.98	7.16
Protein		7 LVs	0.46%	0.49%	0.94	3.94
RF SCiO Global						
Fat	SNV+gap-seg	ntree=500 mtry=104	0.58%	1.18%	0.98	7.91
Moisture		ntree=500 mtry=106	0.47%	0.98%	0.98	6.88
Protein		ntree=500 mtry=105	0.21%	0.48%	0.94	4.21

366





367 *Figure 3: Prediction plots for the SCiO global model using PLS regression (a-c) and RF regression (d-f).*

368 **3.5 Prediction error**

369 The whole procedure of this study has resulted in a predictive model with high
 370 RPD values. However, NIRS-based predictions were used for obtaining the
 371 reference values of **Y** instead of using wet chemistry methods and this introduces
 372 an additional error to our final predictive model. The Foss FoodScan has the
 373 following Standard Errors of Prediction (SEP): for fat 1.01%, for moisture 0.72%
 374 and for protein 0.62%. By following simple error propagation, our global RF
 375 predictive model has the following SEPs: for fat 1.56%, for moisture 1.21% and for
 376 protein 0.78%.

377 **4. Conclusion**

378 Robust, efficient and reliable methods for predicting minced meat composition at
379 four temperatures (-14°C, 5°C, 12°C and 25°C) have been demonstrated by using
380 a smartphone-based spectrometer. The predictive ability of the smartphone-
381 based spectrometer is comparable with the one obtained from a benchtop
382 spectrometer with in-house developed predictive analytics. Two different data
383 partition techniques were applied to construct a PLSR predictive model, which
384 perform well for moisture and fat contents even for frozen meat samples (RPD>7).
385 Furthermore, a global model with all the temperatures was constructed for
386 further examining the abilities of the smartphone-based spectrometer. In this case,
387 the PLS modelling has shown problems to capture all the explanatory information
388 from the global model, which was leaked to the residuals of the model. Random
389 Forest regression was used instead of PLS, which improved the predictive
390 modelling.

391 It is expected that as the technological improvements will continue, the
392 smartphones will include greater computational power, high-end sensing
393 technologies and higher connectivity. This together with the massive volume of
394 smartphone users, which is over than 7 billion, will provide a paradigm shift in
395 how measurement science for food quality and safety will be practiced globally.

396 **Acknowledgement**

397 The research leading to these results has received funding from the project “D1-
398 17 NIR low cost to control quality parameters of dry fermented meat products” of
399 the Innovative Business Groups of the Spanish Ministry of Economy, Industry and
400 from the CERCA programme from Generalitat de Catalunya. Andreas Kartakoullis
401 acknowledges financial support from the Marie Skłodowska-Curie COFUND P-
402 SPHERE project, which has received funding from the European Union’s Horizon
403 2020 research and innovation programme under the Marie Skłodowska-Curie
404 grant agreement No 665919.

405 **Conflicts of Interest Statement**

406 The authors whose names are listed on the first page of this manuscript certify
407 that they have NO affiliations with or involvement in any organization or entity
408 with any financial interest (such as honoraria; educational grants; participation in

409 speakers' bureaus; membership, employment, consultancies, stock ownership, or
410 other equity interest; and expert testimony or patent-licensing arrangements), or
411 non-financial interest (such as personal or professional relationships, affiliations,
412 knowledge or beliefs) in the subject matter or materials discussed in this
413 manuscript.

414 **References**

- 415 Antila, J., Kantojärvi, U., & Mäkynen, J. (2016). *USA Patent No. WO2016071572A1*.
- 416 AOAC. (2007). Determination of fat, moisture and protein in meat and meat
417 products by using the Foss Foodscan Near-Infrared Spectrophotometer
418 with Foss Artificial Neural Network calibration model and associated
419 database. *Journal of AOAC International*, 1073-1083.
- 420 *Beam Your Health Up–TellSpec*. (2018). Retrieved from Tellspec:
421 <http://tellspec.com/en/>
- 422 Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- 423 Burns, D., & Ciurczak, E. (2007). *Handbook of Near-Infrared Analysis, Third*
424 *Edition*. CRC Press.
- 425 Campos, M., Antolin, G., Deban, L., & Pardo, R. (2018). Assessing the influence of
426 temperature on NIRS prediction models for the determination of sodium
427 content in dry-cured ham slices. *Food Chemistry*, 257, 237-242.
- 428 Chen, A., Wang, R., Bever, C., Xing, S., Hammock, B., & Pan, T. (2014). Smartphone-
429 interfaced lab-on-a-chip devices for field-deployable enzyme-linked
430 immunosorbent assay. *Biomicrofluidics*, 8, 064101.
- 431 Chiesa, L., Panseri, S., Bonacci, S., Procopio, A., Zecconi, A., Arioli, F., . . . Moreno-
432 Rojas, J. (2016). Authentication of Italian PDO lard using NIR
433 spectroscopy, volatile profile and fatty acid composition combined with
434 chemometrics. *Food Chemistry*, 212, 296-304.
- 435 Collel, C., Gou, P., Arnau, J., Muñoz, I., & Comaposada, J. (2012). NIR technology for
436 on-line determination of superficial aw and moisture content during the
437 drying process of fermented sausages. *Food Chemistry*, 135, 1750–1755.
- 438 Collell, C., Gou, P., Arnau, J., & Comaposada, J. (2011). Non-destructive estimation
439 of moisture, water activity and NaCl at ham surface during resting and
440 drying using NIR spectroscopy. *Food Chemistry*, 129, 601-607.
- 441 Collell, C., Gou, P., Picouet, P., Arnau, J., & Comaposada, J. (2010). Feasibility of
442 near-infrared spectroscopy to predict aw and moisture and NaCl contents
443 of fermented pork sausages. *Meat Science*, 85, 325-330.

- 444 Conzen, J. P. (2006). *Multivariate calibration. A practical guide for developing*
445 *methods in the quantitative analytical chemistry*. Ettlingen, Germany:
446 Bruker Optik GmbH.
- 447 Coskun, A., Wong, J., Khodadadi, D., Nagi, R., Tey, A., & Ozcan, A. (2017). A
448 personalized food allergen testing platform on a cellphone., *Lab Chip*, *13*,
449 636–640.
- 450 Das, A. J., Wahi, A., Kothari, I., & Raskar, R. (2016). Ultra-portable, wireless
451 smartphone spectrometer for rapid, non-destructive testing of fruit
452 ripeness. *Scientific Reports*, *6*, 32504 .
- 453 Daszykowski, M., Walczak, B., & Massart, D. L. (2002). Representative subset
454 selection. *Analytica Chimica Acta*, *468*, 91-103.
- 455 Goldring, D., & Sharon, D. (2011). *USA Patent No. WO2013065035A1*.
- 456 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical*
457 *Learning*. New York: Springer-Verlag.
- 458 Horizon. (2016). Retrieved from The Horizon Prize for a food scanner:
459 [http://ec.europa.eu/research/horizonprize/index.cfm?prize=food-](http://ec.europa.eu/research/horizonprize/index.cfm?prize=food-scanner)
460 [scanner](http://ec.europa.eu/research/horizonprize/index.cfm?prize=food-scanner)
- 461 Levin, S., Krishnan, S., Rajkumar, S., Halery, N., & Balkunde, P. (2016). Monitoring
462 of fluoride in water samples using a smartphone. *Sci. Total Environ.*, *551–*
463 *552*, 101–107.
- 464 Liaw, A., & Wiener, M. (2002). Classification and Regression by RandomForest. *R*
465 *News*, *2/3*, 18-22.
- 466 Mevik, B. H., & Wehrens, R. (2007). The pls Package: Principal Component and
467 Partial Least Squares Regression in R. *Journal of Statistical Software*, *18*, 1-
468 24.
- 469 Molinaro, A., Simon, R., & Pfeiffer, R. (2005). Prediction error estimation: a
470 comparison of resampling methods. *Bioinformatics*, *21*(15), 3301-3307.
- 471 Nisbert, R., Elder, J., & Miner, G. (2009). *Handbook of Statistical Analysis and Data*
472 *Mining Applications*. Academic Press.
- 473 Ortiz-Somovilla, V., Espana-Espana, F., Gaitan-Jurado, A., Perez-Aparicio, J., & De
474 Pedro-Sanz, E. (2007). Proximate analysis of homogenized and minced
475 mass of pork sausages by NIRS. *Food Chemistry*, *101*, 1031-1040.
- 476 Ozcan, A. (2016). Mobile phones democratize and cultivate next-generation
477 imaging, diagnostics and measurement tools. *Lab Chip*, *14*, 3187.
- 478 Pegau, S., Deric, G., & Zaneveld, V. (1997). Absorption and attenuation of visible
479 and near-infrared light in water: dependence on temperature and salinity.
480 *Applied Optics*, *36*, 6035-6046.

- 481 Prevolnik, M., Skrlep, M., Janes, L., Velikonja-Bolta, S., Skorjanc, D., & Candek-
482 Potokar, M. (2011). Accuracy of near infrared spectroscopy for prediction
483 of chemical composition, salt content and free amino acids in dry-cured
484 ham. *Meat Science*, *88*, 299-304.
- 485 Prieto, N., Roehe, R., Lavin, P., Batten, G., & Andres, S. (2009). Application of near
486 infrared reflectance spectroscopy to predict meat and meat products
487 quality: A review. *Meat Science*, *83*, 175–186.
- 488 Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. (2011).
489 Investigating the influence of data splitting on the predictive ability of
490 QSAR/QSPR models. *Struct Chem*, *22*(4), 795-804.
- 491 Rateni, G., Dario, P., & Cavallo, F. (2017). Smartphone-Based Food Diagnostic
492 Technologies: A Review. *Sensors*, *17*, 1453.
- 493 Rinnan, A., van den Berg, F., & Engelsen, S. (2009). Review of the most common
494 pre-processing techniques for near-infrared spectra. *Trends in Analytical
495 Chemistry*, *28*, 1201-1222.
- 496 Rissanen, A., Saari, H., Rainio, K., Stuns, I., Viherkanto, K., Holmlund, C., . . . Ojanen,
497 H. (2016). MEMS FPI-based smartphone hyperspectral imager. In
498 *Proceedings of the SPIE Next-Generation Spectroscopic Technologies IX* (p.
499 985507). Baltimore, MD, USA: SPIE.
- 500 Segtnan, V., Sasic, S., Isaksson, T., & Ozaki, Y. (2001). Studies on the Structure of
501 Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy
502 and Principal Component Analysis. *Analytical Chemistry*, *73*, 3153–3161.
- 503 Seo, S., Kim, S., Jeon, J., Kim, J., Kim, H., Cho, J., . . . Paek, S. (2016). Food
504 contamination monitoring via internet of things, exemplified by using
505 pocket-sized immunosensor as terminal unit. *Sens. Actuators B Chem.*, *233*,
506 148–156.
- 507 The R Core Team. (2018). R: A Language and Environment for Statistical
508 Computing. R Foundation for Statistical Computing. Retrieved from R-
509 project: <https://www.r-project.org/>
- 510 Van den Berg, F., Lyndgaard, C., Sørensen, K., & Engelsen, S. (2013). Process
511 analytical technology in the food industry. *Trends Food Sci. Technol.*, *31*,
512 27–35.
- 513 Wang, Y., Li, Y., Bao, X., Han, J., Xia, J., Tian, X., & Ni, L. (2016). A smartphone-
514 based colorimetric reader coupled with a remote server for rapid on-site
515 catechols analysis. *Talanta*, *160*, 194–204.
- 516 Wested, F., & Marini, F. (2015). Validation of chemometric models e A tutorial.
517 *Analytica Chimica Acta*, *893*, 14-24.
- 518 Wold, H. (1966). Estimation of principal components and related models by
519 iterative least squares. In H. Wold, *In Proceedings s of international
520 symposium on multivariate analysis* (Vol. 2, pp. 391-420). New York:
521 Academic Press.

- 522 Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of
523 chemometrics. *Chemom. Intell. Lab. Syst.*, 58, 109 – 130.
- 524 Yao, Y., Chen, H., Xie, L., & Rao, X. (2013). Assessing the temperature influence on
525 the soluble solids content of watermelon juice as measured by visible and
526 near-infrared spectroscopy and chemometrics. *Journal of Food*
527 *Engineering*, 119, 22-27.
- 528 Zhu, H., Sikora, U., & Ozcan, A. (2012). Quantum dot enabled detection of
529 *Escherichia coli* using a cell-phone. *Analyst*, 137, 2541–2544.
- 530