

This is a pre print version of the following article:

Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese / Calvini, R.; Michelini, S.; Pizzamiglio, V.; Foca, G.; Ulrici, A.. - In: FOOD CONTROL. - ISSN 0956-7135. - 112:(2020), pp. 107111-107111. [10.1016/j.foodcont.2020.107111]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

02/05/2026 14:01

(Article begins on next page)

Manuscript Number: FOODCONT-D-19-03127

Title: Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese

Article Type: Research Paper

Keywords: Grated cheese; Rind percentage; NIR hyperspectral imaging; Multivariate calibration; Multivariate image analysis

Corresponding Author: Professor Alessandro Ulrici,

Corresponding Author's Institution: Università di Modena e Reggio Emilia

First Author: Rosalba Calvini

Order of Authors: Rosalba Calvini; Sara Michelini; Valentina Pizzamiglio; Giorgia Foca; Alessandro Ulrici

Abstract: Parmigiano Reggiano (P-R) is one of the most important Italian food products labelled with Protected Designation of Origin (PDO). The PDO denomination is applied also to grated P-R cheese products meeting the requirements regulated by the Specifications of Parmigiano Reggiano Cheese. Different quality parameters are monitored, including also the percentage of rind, which is edible and should not exceed the limit of 18% (w/w). The present study aims at evaluating the possibility of using near infrared hyperspectral imaging (NIR-HSI) to quantify the rind percentage in grated Parmigiano Reggiano cheese samples in a fast and non-destructive manner. Indeed, NIR-HSI allows the simultaneous acquisition of both spatial and spectral information from a sample, resulting more suitable than classical single-point spectroscopy for the analysis of heterogeneous samples like grated cheese. Hyperspectral images of grated P-R cheese samples containing increasing levels of rind were acquired in the 900-1700 nm spectral range. Each hyperspectral image was firstly converted into a one-dimensional signal, named hyperspectrogram, which codifies the relevant information contained in the image. Then, the matrix of hyperspectrograms was used to calculate a calibration model for the prediction of the rind percentage using Partial Least Squares (PLS) regression. The calibration model was validated considering two external test sets of samples, confirming the effectiveness of the proposed approach.

Suggested Reviewers: Jose Manuel Amigo PhD
Professor, Analytical Chemistry, Universidad del País Vasco
josemanuel.amigo@ehu.eus

Prof. Amigo is expert in Chemometrics and in Hyperspectral Imaging; many of his research works are focused on the application of hyperspectral imaging for the characterization of food samples.

Paolo Oliveri PhD
University researcher, Department of Pharmacy, University of Genova
oliveri@difar.unige.it

Dr. Oliveri is expert in the use of chemometrics for the characterization of food samples.

Paul Williams PhD

University researcher, Department of Food Science, Stellenbosch University
pauljw@sun.ac.za

Dr. Williams does research in Vibrational Spectroscopy, Hyperspectral Imaging and Multivariate Data Analysis.

*Reggio Emilia, October 31, 2019*

Dear Editor,
please find enclosed a copy of the manuscript:

Rosalba Calvini, Sara Michelini, Valentina Pizzamiglio, Giorgia Foca, Alessandro Ulrici

Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese

that we would like to be considered for publication in Food Control.

This article describes the development of a fast and non-destructive method based on near infrared hyperspectral imaging (NIR-HSI) for the quantification of rind percentage in grated Parmigiano Reggiano cheese. Although edible, in long-ripened cheeses such as Parmigiano Reggiano the rind has chemical and textural properties different from those of the inner part of the cheese. As a consequence, an excessive amount of rind in grated cheese can be perceived by consumers, negatively affecting the organoleptic properties of the product. For this reason, the percentage of rind in grated P-R cheese should not exceed the 18% (w/w) threshold value.

Since NIR-HSI allows to obtain both spatial and spectral information from a sample by detecting variations in the chemical composition over the sample surface, it can be effectively applied to the analysis of grated cheese, which is a heterogeneous food matrix composed by unevenly dispersed particles derived from both cheese rind and pulp.

In this work, 45 grated cheese samples containing different percentages of rind (from 0% to 40%) were considered. NIR hyperspectral images were acquired in the 900-1700 nm spectral range on three different aliquots of each sample, for a total of 135 images. In order to develop calibration models, the images were converted into a matrix of signals, named hyperspectrograms, which were obtained by merging in sequence the frequency distribution curves of quantities derived from a PCA model common to the whole dataset of images. The matrix of hyperspectrograms was then analysed by means of Partial Least Squares (PLS) algorithm for the determination of the amount of rind. The PLS calibration model was validated considering two external test sets of samples, confirming the effectiveness of the proposed approach. These results demonstrated the possibility to develop a fast and non-destructive method based on NIR-HSI to control the amount of rind in grated Parmigiano Reggiano cheese samples.

The article is original, unpublished and not being considered for publication elsewhere.

Best regards,

Alessandro Ulrici

Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese

R. Calvini ^{a, b}, S. Michelini^c, V. Pizzamiglio^c, G. Foca ^{a, b}, A. Ulrici ^{a, b, *}

^a *Department of Life Sciences, University of Modena and Reggio Emilia, Pad. Besta, Via Amendola, 2, Reggio Emilia, 42122, Italy*

^b *Interdepartmental Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia, Piazzale Europa, 1, Reggio Emilia, 42122, Italy*

^c *Parmigiano Reggiano Cheese Consortium, Via J.F. Kennedy, 18, Reggio Emilia, 42124, Italy*

* Corresponding author: alessandro.ulrici@unimore.it

Abstract

Parmigiano Reggiano (P-R) is one of the most important Italian food products labelled with Protected Designation of Origin (PDO). The PDO denomination is applied also to grated P-R cheese products meeting the requirements regulated by the Specifications of Parmigiano Reggiano Cheese. Different quality parameters are monitored, including also the percentage of rind, which is edible and should not exceed the limit of 18% (w/w). The present study aims at evaluating the possibility of using near infrared hyperspectral imaging (NIR-HSI) to quantify the rind percentage in grated Parmigiano Reggiano cheese samples in a fast and non-destructive manner. Indeed, NIR-HSI allows the simultaneous acquisition of both spatial and spectral information from a sample, resulting more suitable than classical single-point spectroscopy for the analysis of heterogeneous samples like grated cheese. Hyperspectral images of grated P-R cheese samples containing increasing levels of rind were acquired in the 900-1700 nm spectral range. Each hyperspectral image was firstly converted into a one-dimensional signal, named hyperspectrogram, which codifies the relevant information contained in the image. Then, the matrix of hyperspectrograms was used to calculate a calibration model for the prediction of the rind percentage using Partial Least Squares (PLS) regression. The calibration model was validated considering two external test sets of samples, confirming the effectiveness of the proposed approach.

Keywords

Grated cheese; Rind percentage; NIR hyperspectral imaging; Multivariate calibration; Multivariate image analysis.

36 **1. Introduction**

37 Parmigiano Reggiano (P-R) is a long-ripened, cooked, hard cheese produced in Italy and registered
38 with Protected Denomination of Origin (PDO). P-R represents one of the most important typical
39 Italian food products and it is exported worldwide. This cheese is manufactured from raw and
40 unheated bovine milk, and the whole production chain must take place in a restricted area in
41 Northern Italy ([Malacarne et al., 2008](#)).

42 The PDO is extended also to grated cheese obtained from Parmigiano Reggiano cheese wheels,
43 provided that the product is grated in the specific production area and packaged immediately
44 afterwards, in order to avoid modifications of its organoleptic properties.

45 Furthermore, grated cheese products designated as Parmigiano Reggiano should meet technical and
46 technological parameters ruled by the Specifications of Parmigiano Reggiano Cheese
47 (https://www.parmigianoreggiano.com/consortium/rules_regulation_2/default.aspx), that regulates
48 all stages of P-R production, including cow feeding, cheese manufacturing and ripening process.

49 One of the different quality parameters of grated P-R cheese regulated by the Specifications is the
50 amount of rind. The rind is the external part of cheese wheels; although edible, in long-ripened
51 cheeses it has chemical and physical properties different from those of the inner part of the cheese.
52 These differences are mainly caused by exposure to environmental conditions during ripening,
53 which determines a decrease in moisture content, proteolytic activity and a higher degree of
54 oxidation ([Cattaneo et al., 2008](#); [Karoui et al., 2007](#); [Malacarne et al., 2019](#)). Furthermore, the
55 different properties of rind and pulp also affect size and shape of grated particles. In fact, generally
56 rind particles are finer and less circular than those derived from the pulp ([Alinovi et al., 2019](#)). As a
57 consequence of its peculiar chemical and textural properties, an excessive amount of rind in grated
58 cheese can be perceived by consumers, negatively affecting the organoleptic properties of the
59 product ([Zannoni and Hunter, 2015](#)). For these reasons, the percentage of rind in grated P-R cheese
60 should not exceed the 18% (w/w) threshold value.

61 In order to ensure quality compliance of commercial products of grated P-R cheese and to avoid
62 counterfeits, it is essential to implement effective analytical methods to correctly estimate the rind
63 amount, meeting the requirements of low costs, limited sample preparation and short times of
64 analysis. In this context, Near Infrared (NIR) spectroscopy has been widely employed for fast and
65 non-destructive analysis and characterization of food products, thanks to its ability to easily provide
66 a spectral fingerprint codifying the chemical composition of the analysed sample ([Curda et al.,
67 2004](#); [Woodcok et al., 2008](#); [Foca et al., 2013](#); [Kraggerud et al., 2014](#)).

68 In particular, previous research studies evaluated the effectiveness of NIR spectroscopy in verifying
69 the authenticity of P-R grated cheese and in discriminating compliant from non-compliant samples
70 ([Cevoli et al., 2013a](#), [Cevoli et al., 2013b](#)).

71 However, grated cheese is an inhomogeneous food matrix composed by unevenly dispersed
72 particles derived from both cheese rind and pulp, which are characterized by different chemical
73 properties. When dealing with heterogeneous samples, classical NIR spectroscopy may lead to
74 inaccurate results since it is based on the acquisition of “average” spectra over a given sample area,
75 thus losing the information related to the compositional variability within the sample. The
76 importance of spatial information for the analysis of P-R grated cheese was recently demonstrated
77 by [Alinovi et al. \(2019\)](#), which found a relationship between rind percentage and parameters related
78 to particle size and distribution calculated from digital RGB images of the cheese samples.

79 The advantages of image-based methods and NIR spectroscopy can be coupled in NIR
80 Hyperspectral Imaging (NIR-HSI), an analytical technique based on the acquisition of particular
81 types of images, called hyperspectral images, where a whole NIR spectrum is registered for each
82 image pixel ([Gowen et al., 2007](#); [Amigo et al., 2013](#); [Calvini et al.; 2018](#)). More in detail, a
83 hyperspectral image, also called *hypercube*, is a three-dimensional data array with two spatial
84 dimensions (x pixel rows and y pixel columns) and one spectral dimension, corresponding to the λ
85 acquired wavelengths. Therefore, the hypercube can be seen as a stack of spectrally resolved
86 images, each one acquired at a given wavelength, or as a series of spatially resolved spectra, where
87 each spectrum characterizes one pixel of the image ([Wu and Sun, 2013a](#); [Amigo et al., 2015](#);
88 [Calvini et al.; 2016](#)).

89 Considering that each NIR spectrum acts like a fingerprint of the chemical properties of a specific
90 pixel, thanks to hyperspectral imaging it is possible to obtain both spatial and spectral information
91 from a sample by observing variations in the chemical composition over the sample surface. These
92 aspects resulted to be particularly useful for the analysis of heterogeneous matrices like food
93 products ([Wu and Sun, 2013b](#); [Dale et al., 2013](#); [Liu et al., 2017](#)), generally overcoming the
94 performances obtained with single-point spectroscopy ([Burger and Geladi, 2006](#); [Shonbichler et al.,](#)
95 [2013](#)).

96 However, the high amount of information contained in hyperspectral images can also become a
97 drawback, since each hypercube can contain up to tens of thousands of pixel spectra, resulting in
98 data handling and data storage issues. Indeed, in order to extract the relevant information from this
99 kind of data, the application of proper multivariate statistical methods is mandatory ([Burger and](#)
100 [Gowen, 2011](#)). This is known as Multivariate Image Analysis (MIA), which is based on the
101 application to images of common chemometric methods, like e.g. Principal Component Analysis

102 (PCA). This approach essentially consists in considering each pixel of the image as a separate
103 object and the main goal is to find similarities or differences among clusters of pixels based on their
104 spectral signatures (Prats-Montalban et al., 2011; Amigo et al., 2015).

105 When dealing with a large number of hyperspectral images that should be analysed altogether,
106 classical pixel-level MIA can become unfeasible due to the intensive computational loads, since it
107 would imply the simultaneous analysis of numerous images, each one with tens of thousands of
108 pixel spectra. In these situations, a possible solution is to move from a pixel-level approach to an
109 image-level approach, which consists in performing the analysis considering the image of each
110 sample as a single object and extracting a feature vector characterizing the whole image, and thus
111 the corresponding sample. In this manner, it is possible to analyse data matrices containing these
112 feature vectors, in order to gain a general overview of the image dataset, to identify images sharing
113 similar features or to quantify whole sample properties (Ulrici et al., 2012; Kucheryavskiy, 2013;
114 Girauda et al., 2018; Orlandi et al., 2018a; Orlandi et al., 2018b; Oliveri et al., 2019).

115 To this aim, a data dimensionality reduction method has been proposed, which consists in
116 converting each hyperspectral image of the dataset into a one-dimensional signal, named
117 hyperspectrogram, obtained by merging in sequence the frequency distribution curves of quantities
118 derived from a PCA model calculated on the images (Ferrari et al., 2013; Ferrari et al., 2015; Xu et
119 al., 2016; Calvini et al., 2016; Corti et al., 2017). In this manner, each hyperspectrogram
120 summarizes the relevant information contained in the corresponding hyperspectral image and a
121 large dataset of hyperspectral images is converted into a matrix of signals, which in turn can be
122 analysed by means of common chemometric methods.

123 In this context, the main goal of the present study consisted in evaluating the possibility of
124 exploiting the advantages of NIR-HSI coupled to data dimensionality reduction, in order to monitor
125 the rind percentage of grated P-R cheese products. In particular, hyperspectral images of grated P-R
126 cheese samples were analysed by means of the hyperspectrograms approach, and the resulting
127 hyperspectrograms were then used to predict the rind percentage by means of Partial Least Squares
128 (PLS) regression.

129

130 **2. Materials and Methods**

131 **2.1 Grated cheese samples**

132 Samples of grated Parmigiano Reggiano cheese containing varying rind percentages (RP) were
133 provided by Parmigiano Reggiano cheese Consortium. More in detail, the grated cheese samples
134 were prepared considering the following 15 percentages (w/w) of rind in pulp: 0%, 5%, 10%, 12%
135 14%, 16%, 18%, 20%, 22%, 24%, 26%, 28%, 30%, 35% and 40%.

136 In order to minimize possible effects of unwanted variations, the mixtures were prepared starting
137 from the same matrices of cheese pulp and rind, obtained by grating pulp and rind pieces derived
138 from different cheese wheels matured for a period of 12 months. The mixtures were replicated twice
139 (deliveries A and B) as reported in [Table 1](#), each time following a different random order. Firstly,
140 the matrices of grated pulp and rind were prepared, and a part of them was then used to obtain the
141 first set of 15 mixtures. The remaining part of the grated pulp and rind matrices was stored in the
142 dark at 4 °C, and after one week it was used to prepare the second set of 15 mixtures. For both the
143 replicate sets, the samples were stored in the dark at 4°C and the day after their preparation they
144 were delivered to the laboratory, where they were immediately analysed.
145 Furthermore, 15 additional samples with unknown rind percentage were provided by Parmigiano
146 Reggiano cheese Consortium (X1-X15). These samples were prepared, delivered and stored
147 considering the same procedure followed for the samples with known RP values.

148

149 **2.2 Image acquisition**

150 For each sample, three randomly sampled aliquots containing about 13 g of grated cheese were
151 collected and placed inside a plastic Petri dish of 5.5 cm diameter. Each hyperspectral image
152 included the three aliquots of two different samples. More in detail, the samples were positioned
153 according to a 3×2 chessboard scheme, as reported in [Figure 1](#).

154 The hyperspectral images were acquired using a line scanning system (NIR Spectral Scanner, DV
155 Optic) equipped with a Specim Inspector N17E imaging spectrometer coupled to a Xenics Xeva-
156 1.7-320 camera (320×256 pixels) embedding Specim Oles 31 f/2.0 optical lens. The hyperspectral
157 system covers the 900-1700 nm spectral range with a spectral resolution equal to 5 nm. Due to the
158 low signal-to-noise ratio at the edges of the spectral range, only 143 spectral channels between 960
159 and 1670 nm were considered for further analysis.

160 The hyperspectral images were acquired using a black silicon carbide sandpaper sheet as
161 background and including in the image scene also a white ceramic tile with a 99 % reflectance
162 standard reference and two ceramic tiles with intermediate reflectance values corresponding to 89
163 % and 46 %, respectively. The raw data were converted into reflectance values using the
164 instrumental calibration based on the measure of the white high reflectance standard reference and
165 of the dark current ([Ulrici et al., 2013](#)).

166

167 **2.3 Image elaboration**

168 The first step of image elaboration consisted in the application of an additional internal calibration
169 procedure in order to reduce possible variations over time. This correction procedure is based on the

170 comparison of the average reflectance values of the white standard reference, of the two ceramic
171 tiles and of the black silicon carbide sandpaper between an image chosen as reference and all the
172 remainder images of the dataset. Further details about the image correction algorithm can be found
173 in [Ulrici et al. \(2013\)](#).

174 Subsequently, the corrected images were cropped in order to obtain a single hyperspectral image for
175 each aliquot of grated cheese sample. At the end of the cropping procedure, a total of 135
176 hyperspectral images were obtained (= 45 grated cheese samples \times 3 aliquots), as reported in the
177 last column of [Table 1](#).

178 After cropping, the pixels related to the black sandpaper background were removed using a
179 thresholding procedure carried out considering a wavelength equal to 1100 nm. Indeed, at 1100 nm
180 the pixels with reflectance value lower than 0.5 were ascribable to the background or to the plastic
181 Petri dish, thus they were not considered in the subsequent steps.

182 Finally, an additional morphological erosion procedure was performed using a disk-shaped
183 structuring element with radius equal to 8 pixels ([Van Den Boomgaard and Van Balen, 1992](#)).
184 Morphological erosion allowed to remove the pixels placed at the edges of the region of interest
185 obtained after background removal, since these pixels were mainly influenced by scattering
186 phenomena and specular reflections of the plastic Petri dish.

187 The image elaboration steps, which are summarized in Step 1 and Step 2 of [Figure 1](#), were
188 performed with routines written *ad hoc* in MATLAB language (ver. 9.3, The Mathworks Inc.,
189 USA).

190

191 **2.4 Data analysis**

192 **2.4.1. Exploratory analysis**

193 As a preliminary assessment, Principal Component Analysis (PCA) was performed at the pixel level
194 on some representative images ([Prats-Montalbán et al., 2011](#)). More in detail, three images
195 corresponding to RP values equal to 0%, 20% and 40% were merged together and analysed by
196 means of PCA after applying standard normal variate (SNV) and mean center as spectral
197 preprocessing methods. This preliminary analysis was carried out in order to obtain a qualitative
198 evaluation of the differences between samples containing an increasing amount of rind.

199

200 **2.4.2. Calibration models**

201 *Data organization*

202 Before calculating the calibration model to predict the rind percentage, the hyperspectral images of
203 the grated cheese samples were split into training images, for model computation, and test images

204 for external validation. The training images included the hyperspectral images of grated cheese
205 samples with RP values equal to 0%, 10%, 14%, 18%, 22%, 26%, 30% and 40%, for a total of 48
206 images (= 8 RP values \times 2 deliveries \times 3 aliquots). The remainder images were separated into two
207 different test sets:

- 208 - TS_{known} : including the images acquired on cheese samples with RP values equal to 5%, 12%,
209 16%, 20%, 24%, 28% and 35%, for a total of 42 images (= 7 RP values \times 2 deliveries \times 3
210 aliquots);
- 211 - TS_{unknown} : including the images acquired on the cheese samples of unknown composition,
212 for a total of 45 images (= 15 unknown samples \times 3 aliquots).

213

214 *Conversion into Common Space Hyperspectrograms*

215 The hyperspectral images were then converted into one-dimensional signals, named Common Space
216 Hyperspectrograms (CSH) (Calvini et al., 2016). The basic idea behind the hyperspectrograms
217 approach consists in converting each hyperspectral image of the dataset into a one-dimensional
218 signal, which acts like a feature vector retaining the useful spectral/spatial information of the
219 corresponding image (Ferrari et al., 2013; Ferrari et al., 2015). More in detail, in the case of CSH,
220 the signals are obtained by merging in sequence the frequency distribution curves of quantities
221 derived from a common PCA model, i.e. from a model calculated considering all the images of the
222 training set.

223 The first step in the computation of the CSH consisted in unfolding all the three-dimensional
224 hyperspectral images into two-dimensional matrices with as many rows as the pixels retained after
225 background removal and erosion, and as many columns as the number of spectral channels. Then,
226 the unfolded hypercubes were row-preprocessed using SNV and scaled according to the global
227 mean spectrum, obtained by averaging all the retained pixel spectra of the training images. After
228 unfolding and spectra preprocessing, for each training image the corresponding variance-covariance
229 matrix was calculated. Then, all the resulting variance-covariance matrices were summed in order
230 to obtain the kernel variance-covariance matrix of the whole training set (Geladi and Grahn, 1996).
231 The kernel-variance covariance matrix was then decomposed by singular value decomposition
232 (SVD) to obtain the loading vectors of the common PC space. In this case, the common PC space
233 was calculated considering 3 principal components, based on the results of the previous exploratory
234 data analysis described in Section 2.4.1.

235 Once calculated the PC space common to the training images, all the hyperspectral images
236 belonging to both training and test sets were projected onto the PC space to obtain the
237 corresponding scores, Q residuals and Hotelling's T^2 vectors. Finally, for each image, the

238 corresponding CSH signal was obtained by merging in sequence the frequency distribution curves
239 of the three score vectors, of the Q residuals vector and of the Hotelling's T^2 vector.

240 The frequency distribution curves were calculated considering a number of bins equal to 150 and
241 normalized according to the number of pixels retained after image elaboration, as described in
242 Section 2.3, to give the corresponding hyperspectrogram. Therefore, in this case each
243 hyperspectrogram was a 750 points long vector, resulting from 150 bins \times 5 quantities derived from
244 PCA (3 PCs + Q residuals + Hotelling's T^2). Further details about the algorithm used to calculate
245 the CSH can be found in [Calvini et al. \(2016\)](#). The conversion of the hyperspectral images into CSH
246 signals is schematically depicted in Step 3 of [Figure 1](#).

247 In this manner, at the end of the conversion procedure three different matrices of signals were
248 obtained: the training set (TR), the test set derived from the TS_{known} images and the test set of
249 unknown samples derived from the TS_{unknown} images.

250 [Figure 2](#) shows a plot of the CSH signals belonging to the training set, coloured according to the
251 rind percentage of the corresponding sample.

252

253 *Calibration model*

254 The training set matrix containing the CSH signals calculated from the training images was used to
255 calculate the calibration model to predict the RP value, using Partial Least Squares (PLS) algorithm
256 ([Geladi and Kowalski, 1986](#)). The signals were preprocessed using mean center and the optimal
257 number of latent variables (LVs) was chosen by minimizing the cross-validation error. In particular,
258 a custom cross-validation scheme was followed, considering 2 deletion groups, each one containing
259 the signals derived from samples belonging to the same delivery day.

260 The performances of the calibration models were evaluated both in terms of Root-Mean-Square
261 Error (RMSE) and of coefficient of determination (R^2). These parameters were calculated in
262 calibration (RMSEC and R^2 Cal), cross-validation (RMSECV and R^2 CV) and prediction of the
263 external test set (RMSEP and R^2 Pred).

264 The conversion of the hyperspectral images in CSH signals was done using a specific Graphical
265 User Interface (GUI), that was previously developed by some of the authors of the present work.
266 The GUI, which works under the MATLAB environment (ver. 9.3, The Mathworks, USA) and is
267 named *Hyperspectrograms GUI*, is freely downloadable from
268 www.chimslab.unimore.it/downloads. PCA and PLS models were calculated using PLS Toolbox
269 (ver. 8.5, Eigenvector Research Inc., USA) and MIA Toolbox (ver. 3.0.4, Eigenvector Research
270 Inc., USA).

271

272 **3. Results and discussion**

273 **3.1. PCA at the pixel-level**

274 For a first evaluation of the differences between grated cheese samples containing different amounts
275 of rind, three hyperspectral images corresponding to samples with RP values equal to 0%, 20% and
276 40% were merged together to obtain a unique hyperspectral image, which was analysed at the pixel-
277 level by PCA. [Figure 3.a](#) reports the resulting PC1-PC2 score plot, where each object represents a
278 single pixel and is coloured according to pixel density, i.e. a yellowish colour represents a region of
279 the PC1-PC2 score space with a high density of pixels, while blue corresponds to low pixel density.
280 From this score plot it is possible to observe the presence of three clusters of pixels, corresponding
281 to the imaged samples with different RP values. The separation between samples with different rind
282 levels is particularly evident along PC2. Indeed, the sample containing only cheese pulp is
283 characterized by higher PC2 score values, while samples with increasing percentages of rind have
284 decreasing PC2 score values, as shown in the PC2 score image reported in [Figure 3.b](#).

285 In order to investigate the spectral features involved in the definition of the PC space, the
286 corresponding PC1-PC2 loading vectors are reported in [Figure 3.c](#). The highest absolute values of
287 the PC2 loading vector can be found in the 1195-1225 nm wavelength range, corresponding to the
288 C-H bond second overtone ascribable to lipids ([Burns and Ciurzak, 2008](#); [Karoui et al., 2006](#)), in
289 the 1330-1340 nm spectral range corresponding to asymmetric stretching vibration of water ([Ozaky,](#)
290 [2002](#)), and in the region centred at 1400 nm ascribable to the O-H bond first overtone of free water
291 ([Burns and Ciurzak, 2008](#)).

292 Therefore, the amount of rind of the cheese samples can be somehow described by the distribution
293 of the corresponding pixel spectra along the principal components. A convenient way to summarize
294 this pixel distribution consists in using the frequency distribution curves of the score vectors of each
295 sample, as reported in [Figure 3.a](#) that shows the frequency distribution curves of both PC1 and PC2
296 score vectors for each image. From this figure it is possible to observe the presence of a clear shift
297 of the frequency distribution curves of PC2, which is related to the rind percentage of the
298 corresponding samples. Although less marked, a variation with RP can be observed also for the
299 frequency distribution curves of PC1 scores, which tend to become sharper and with a maximum
300 located at lower PC1 values with increasing values of rind percentage.

301 Since the hyperspectrograms approach is based on the use of frequency distribution curves of score
302 vectors calculated from a PCA model in order to summarize the relevant information contained in
303 the images, this preliminary analysis suggests the effectiveness of this approach for the
304 determination of the rind amount in hyperspectral images of grated cheese samples.

305

306 **3.2 PLS calibration model with the CSH approach**

307 The training set of CSH signals reported in [Figure 2](#) was then used to calculate a PLS regression
308 model for the quantification of rind percentage in the samples of grated Parmigiano Reggiano
309 cheese, leading to the results reported in [Table 2](#).

310 The optimal model dimensionality was found to be equal to 8 LVs, leading to a RMSECV value
311 equal to 1.70 RP units, corresponding to a R^2 CV value of 0.979.

312 The calibration model was then used to predict the RP of the samples belonging to the TS_{known} test
313 set. In this case, the prediction results were calculated considering both the whole range of rind
314 levels (0% - 40%) and only the interval of rind percentages ranging from 10% to 30%, which better
315 reflects RP values that generally may occur in real situations.

316 The prediction results confirm the effectiveness of the calibration model in quantifying the rind
317 percentage, leading to RMSEP values equal to 1.91 and 1.85 RP units in the 0-40% and 10-30%
318 ranges, respectively.

319 [Figure 4](#) shows the plot of the rind percentage values predicted for the TS_{known} test set versus the
320 corresponding experimental values, where the samples are coloured according to the delivery day.

321 Generally, all the objects are close to the bisector, indicating the good prediction performances of
322 the model. In addition, from [Figure 4](#) it is also possible to observe that there are not evident
323 systematic variations in the prediction results caused by the different delivery days, which further
324 confirms the robustness of the calibration model toward replicated measurements.

325 Considering that compliant Parmigiano Reggiano grated cheese samples should have an RP value
326 lower or equal than 18%, all the test set samples with a rind percentage falling outside this limit are
327 correctly identified by the model.

328 The calibration model was also used to predict the samples with unknown composition belonging to
329 the TS_{unknown} test set. The predicted rind percentages of the unknown samples were communicated
330 to the Parmigiano Reggiano Cheese Consortium, which then revealed the corresponding
331 experimental values. In this manner, it was possible to perform a further external validation of the
332 calibration model.

333 It has to be considered that, for each unknown sample, three different aliquots were imaged and,
334 therefore, three RP values were obtained in prediction by the model. In order to have a single
335 estimate of the RP value for each unknown sample, the three RP predicted values corresponding to
336 the three aliquots were averaged for each sample. The results are reported in [Table 3](#), together with
337 the corresponding experimental values, further confirming the good prediction ability of the
338 calibration model. More in detail, an RMSEP value equal to 2.50 RP units was obtained,
339 corresponding to an R^2 value equal to 0.955. The highest difference between predicted and

340 experimental rind percentage, equal to 5 RP units, was observed for sample X15, while samples X2
341 and X4 were exactly predicted.

342 In order to have a comprehensive evaluation of the hyperspectrogram regions most relevant to the
343 calibration model, [Figure 5.a](#) reports the Variable Importance in Projection (VIP) scores: variables
344 with VIP score values higher than one are considered significant for the model. This figure shows
345 that all the frequency distribution curves of the PCA quantities included in the CSH signals have
346 regions with significant variables. In particular, among the three score vectors included in the
347 signals, the regions related to the frequency distribution curve of PC2 reach the highest VIP score
348 values, together with the signal regions related to Hotelling T^2 vales.

349 As an example, [Figure 5.b](#) reports the PC2 score images of some representative test set samples
350 with increasing percentage of rind. Similarly to what was previously reported in Section 3.1, images
351 of samples with a lower RP value are characterized by higher PC2 score values, and the increase of
352 the RP value in the grated cheese samples causes a shift of the corresponding pixels toward lower
353 PC2 score values. Actually, this is due to the fact that the PC2 loading vector of the common PCA
354 model used to calculate the CSH signals is very similar to the PC2 loading vector calculated on the
355 three sample hyperspectral images, reported in [Figure 3c](#).

356

357 **Conclusions**

358 The present study demonstrated the possibility of using NIR-HSI as a tool for the quantification of
359 the amount of rind in grated Parmigiano Reggiano cheese samples. The combined use of a data
360 dimensionality reduction approach, namely the Common Space Hyperspectrograms approach, with
361 PLS regression allowed to obtain satisfactory prediction performances. The calibration model was
362 validated using two different test sets: the first test set consisted of cheese samples with known RP
363 values, while for the second test set the experimental RP values of the analysed samples were
364 revealed by the operators of Parmigiano Reggiano Cheese Consortium only after providing them
365 with the RP values predicted by the model. The RMSEP values obtained for both test sets,
366 corresponding to 1.91 RP units and 2.50 RP units, respectively, confirm the advantages of coupling
367 spatial and spectral information of a sample brought by NIR-HSI in the analysis of heterogeneous
368 products, like grated cheese.

369 However, it has to be considered that commercial samples of grated Parmigiano Reggiano cheese
370 may be affected by different variability factors, such as months of ripening, fat content and rind
371 processing methods, among others. Therefore, in order to obtain a more robust calibration model
372 the influence of these factors should be properly evaluated and included in the model. As a

373 consequence, the increasing complexity of the model should be faced by including further steps in
374 the data analysis workflow, like e.g. spectral and/or spatial feature selection.

375

376 **Acknowledgements**

377 The authors wish to thank Consorzio del Formaggio Parmigiano Reggiano for providing technical
378 and financial support.

379

380 **References**

- 381 Alinovi, M., Mucchetti, G., & Tidona, F. (2019). Application of NIR spectroscopy and image
382 analysis for the characterisation of grated Parmigiano-Reggiano cheese. *International dairy journal*,
383 92, 50-58.
- 384 Amigo, J. M., Martí, I., & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect
385 combination for the analysis of food structure, composition and quality. In *Data handling in science*
386 *and technology* (Vol. 28, pp. 343-370). Elsevier.
- 387 Amigo, J. M., Babamoradi, H., & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A
388 tutorial. *Analytica chimica acta*, 896, 34-51.
- 389 Burger, J., & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a
390 comparison between image and spectrometer data for studying organic and biological samples.
391 *Analyst*, 131(10), 1152-1160.
- 392 Burger, J., & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and*
393 *Intelligent Laboratory Systems*, 108(1), 13-22.
- 394 Burns, D. A., Ciurczak, E. W. (2008) Handbook of near infrared analysis, CRC Press, Boca Raton,
395 Florida, USA, (Chapter 17, pp. 347-386). ISBN 978-0-8493-7393-0
- 396 Calvini, R., Foca, G., & Ulrici, A. (2016). Data dimensionality reduction and data fusion for fast
397 characterization of green coffee samples using hyperspectral sensors. *Analytical and bioanalytical*
398 *chemistry*, 408(26), 7351-7366.
- 399 Calvini, R., Orlandi, G., Foca, G., Ulrici, A. (2018). Development of a classification algorithm for
400 efficient handling of multiple classes in sorting systems based on hyperspectral imaging. *Journal of*
401 *Spectral Imaging*, 7, a13.
- 402 Cattaneo, S., Hogenboom, J. A., Masotti, F., Rosi, V., Pellegrino, L., & Resmini, P. (2008). Grated
403 Grana Padano cheese: new hints on how to control quality and recognize imitations. *Dairy Science*
404 *and Technology*, 88(4-5), 595-605.
- 405 Cevoli, C., Gori, A., Nocetti, M., Cuibus, L., Caboni, M. F., & Fabbri, A. (2013). FT-NIR and FT-
406 MIR spectroscopy to discriminate competitors, non compliance and compliance grated Parmigiano
407 Reggiano cheese. *Food research international*, 52(1), 214-220.
- 408 Cevoli, C., Fabbri, A., Gori, A., Caboni, M. F., & Guarnieri, A. (2013). Screening of grated cheese
409 authenticity by nir spectroscopy. *Journal of Agricultural Engineering*, 44, 264-267.
- 410 Corti, M., Gallina, P. M., Cavalli, D., & Cabassi, G. (2017). Hyperspectral imaging of spinach
411 canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content.
412 *Biosystems engineering*, 158, 38-50.
- 413 Čurda, L., & Kukačková, O. (2004). NIR spectroscopy: a useful tool for rapid monitoring of
414 processed cheeses manufacture. *Journal of Food Engineering*, 61(4), 557-560.
- 415 Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F. (2013).
416 Hyperspectral imaging applications in agriculture and agro-food product quality and safety control:
417 a review. *Applied Spectroscopy Reviews*, 48(2), 142-159.
- 418 Ferrari, C., Foca, G., & Ulrici, A. (2013). Handling large datasets of hyperspectral images:
419 Reducing data size without loss of useful information. *Analytica chimica acta*, 802, 29-39.
- 420 Ferrari, C., Foca, G., Calvini, R., & Ulrici, A. (2015). Fast exploration and classification of large
421 hyperspectral image datasets for early bruise detection on apples. *Chemometrics and Intelligent*
422 *Laboratory Systems*, 146, 108-119.

423 Foca, G., Salvo, D., Cino, A., Ferrari, C., Fiego, D. P. L., Minelli, G., & Ulrici, A. (2013).
424 Classification of pig fat samples from different subcutaneous layers by means of fast and non-
425 destructive analytical techniques. *Food research international*, 52(1), 185-197.

426 Geladi, P., & Grahn, H. F. (1996). Multivariate image analysis. *Encyclopedia of Analytical*
427 *Chemistry*.

428 Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica*
429 *acta*, 185, 1-17.

430 Giraud, A., Calvini, R., Orlandi, G., Ulrici, A., Geobaldo, F., Savorani, F. (2018). Development of
431 an automated method for the identification of defective hazelnuts based on RGB image analysis and
432 colourgrams. *Food Control*, 94, 233-240.

433 Gowen, A. A., O'Donnell, C., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral
434 imaging—an emerging process analytical tool for food quality and safety control. *Trends in food*
435 *science & technology*, 18(12), 590-598.

436 Karoui, R., Mouazen, A. M., Dufour, É., Pillonel, L., Schaller, E., De Baerdemaeker, J., & Bosset,
437 J. O. (2006). Chemical characterisation of European Emmental cheeses by near infrared
438 spectroscopy using chemometric tools. *International Dairy Journal*, 16(10), 1211-1217.

439 Karoui, R., Dufour, É., & De Baerdemaeker, J. (2007). Front face fluorescence spectroscopy
440 coupled with chemometric tools for monitoring the oxidation of semi-hard cheeses throughout
441 ripening. *Food Chemistry*, 101(3), 1305-1314.

442 Kraggerud, H., Næs, T., & Abrahamsen, R. K. (2014). Prediction of sensory quality of cheese
443 during ripening from chemical and spectroscopy measurements. *International Dairy Journal*, 34(1),
444 6-18.

445 Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images.
446 *Chemometrics and Intelligent Laboratory Systems*, 120, 126-135.

447 Liu, Y., Pu, H., & Sun, D. W. (2017). Hyperspectral imaging technique for evaluating food quality
448 and safety during various processes: A review of recent applications. *Trends in food science &*
449 *technology*, 69, 25-35.

450 Malacarne, M., Summer, A., Formaggioni, P., Franceschi, P., Sandri, S., Pecorari, M., ... &
451 Mariani, P. (2008). Dairy maturation of milk used in the manufacture of Parmigiano-Reggiano
452 cheese: effects on physico-chemical characteristics, rennet-coagulation aptitude and rheological
453 properties. *Journal of dairy research*, 75(2), 218-224.

454 Malacarne, M., Summer, A., Franceschi, P., Formaggioni, P., Pecorari, M., Panari, G., & Mariani,
455 P. (2009). Free fatty acid profile of Parmigiano-Reggiano cheese throughout ripening: Comparison
456 between the inner and outer regions of the wheel. *International dairy journal*, 19(10), 637-641.

457 Oliveri, P., Malegori, C., Casale, M., Tartacca, E., & Salvatori, G. (2019). An innovative
458 multivariate strategy for HSI-NIR images to automatically detect defects in green coffee. *Talanta*,
459 199, 270-276.

460 Orlandi, G., Calvini, R., Foca, G., Ulrici, A. (2018). Automated quantification of defective maize
461 kernels by means of Multivariate Image Analysis. *Food Control*, 85, 259-268.

462 Orlandi, G., Calvini, R., Pigani, L., Foca, G., Vasile Simone, G., Antonelli, A., Ulrici, A. (2018).
463 Electronic eye for the prediction of parameters related to grape ripening. *Talanta*, 186, 381-388.

464 Ozaki Y. (2002) Two dimensional near-infrared correlation spectroscopy. In: Siesler H.W., Ozaki
465 Y., Kawata S. and Heise M. (Eds). *Applications in chemistry in near-infrared spectroscopy:*
466 *principles, instruments and applications*, Wiley-VCH, Weinheim, Germany.

467 Prats-Montalbán, J. M., De Juan, A., & Ferrer, A. (2011). Multivariate image analysis: a review
468 with applications. *Chemometrics and intelligent laboratory systems*, 107(1), 1-23.

469 Schönbichler, S. A., Bittner, L. K. H., Weiss, A. K. H., Griesser, U. J., Pallua, J. D., & Huck, C. W.
470 (2013). Comparison of NIR chemical imaging with conventional NIR, Raman and ATR-IR
471 spectroscopy for quantification of furosemide crystal polymorphs in ternary powder mixtures.
472 *European Journal of Pharmaceutics and Biopharmaceutics*, 84(3), 616-625.

473 Specifications of Parmigiano Reggiano Cheese.
474 https://www.parmigianoreggiano.com/consortium/rules_regulation_2/default.aspx. Accessed 2
475 September 2019.

476 Ulrici, A., Foca, G., Ielo, M. C., Volpelli, L. A., & Fiego, D. P. L. (2012). Automated identification
477 and visualization of food defects using RGB imaging: Application to the detection of red skin
478 defect of raw hams. *Innovative Food Science & Emerging Technologies*, 16, 417-426.

479 Ulrici, A., Serranti, S., Ferrari, C., Cesare, D., Foca, G., & Bonifazi, G. (2013). Efficient
480 chemometric strategies for PET-PLA discrimination in recycling plants using hyperspectral
481 imaging. *Chemometrics and Intelligent Laboratory Systems*, 122, 31-39.

482 Van Den Boomgaard, R., & Van Balen, R. (1992). Methods for fast morphological image
483 transforms using bitmapped binary images. *CVGIP: Graphical Models and Image Processing*,
484 54(3), 252-258.

485 Woodcock, T., Fagan, C. C., O'Donnell, C. P., & Downey, G. (2008). Application of near and mid-
486 infrared spectroscopy to determine cheese quality and authenticity. *Food and Bioprocess
487 Technology*, 1(2), 117-129.

488 Wu, D., & Sun, D. W. (2013). Advanced applications of hyperspectral imaging technology for food
489 quality and safety analysis and assessment: A review—Part I: Fundamentals. *Innovative Food
490 Science & Emerging Technologies*, 19, 1-14.

491 Wu, D., & Sun, D. W. (2013). Advanced applications of hyperspectral imaging technology for food
492 quality and safety analysis and assessment: A review—Part II: Applications. *Innovative Food
493 Science & Emerging Technologies*, 19, 15-28.

494 Xu, J. L., Riccioli, C., & Sun, D. W. (2016). Efficient integration of particle analysis in
495 hyperspectral imaging for rapid assessment of oxidative degradation in salmon fillet. *Journal of
496 Food Engineering*, 169, 259-271.

497 Zannoni, M., & Hunter, E. A. (2015). Relationship between sensory results and compliance scores
498 in grated Parmigiano-Reggiano cheese. *Italian Journal of Food Science*, 27(4), 487-494.

499

500 **Captions to Tables and Figures**

501 **Table 1.** Summary information about the grated cheese samples considered in the present study

502 **Table 2.** Results of PLS regression for the determination of rind percentage.

503 **Table 3.** Prediction results of the unknown test samples and corresponding experimental RP values.

504

505 **Figure 1.** Key steps involved in image elaboration and analysis

506 **Figure 2.** Hyperspectrograms of the training set images; the signals are coloured according to the
507 rind percentage of the corresponding grated cheese sample

508 **Figure 3.** In (a) PC1-PC2 score plot of the image containing grated cheese sample with 0%, 20%
509 and 40% percentages of rind, and corresponding frequency distribution curves of PC1 and PC2
510 score vectors calculated separately for each sample. In (b) PC2 score image. In (c) PC1 and PC2
511 loading vectors.

512 **Figure 4.** Results of the PLS model: TSkknown test set predicted rind percentage (Y Predicted) vs
513 experimental rind percentage (Y measured).

514 **Figure 5.** VIP scores of the PLS model (a) and PC2 score images of samples with increasing rind
515 percentage values(b).

516

Highlights

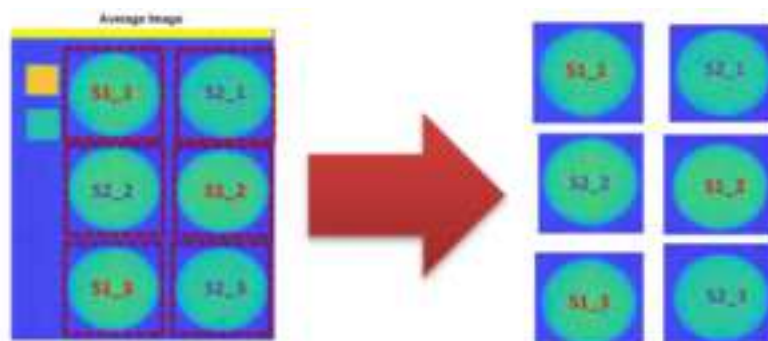
- Rind percentage in grated Parmigiano Reggiano cheese should not exceed 18% (w/w)
- Hyperspectral imaging (HSI) was used to estimate rind percentage in grated cheese
- The images were converted in one-dimensional signals named hyperspectrograms
- The hyperspectrograms were used to build PLS calibration models
- Validation with test samples confirmed the effectiveness of the proposed approach

Conflict of interest

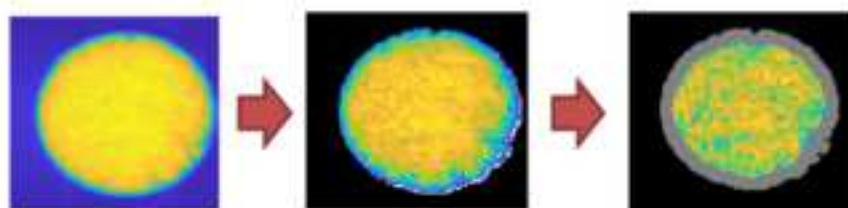
The authors declare that they have no conflict of interest.

Figure 1

1. Image correction and cropping



2. Background removal and erosion



3. Conversion in CSH signals

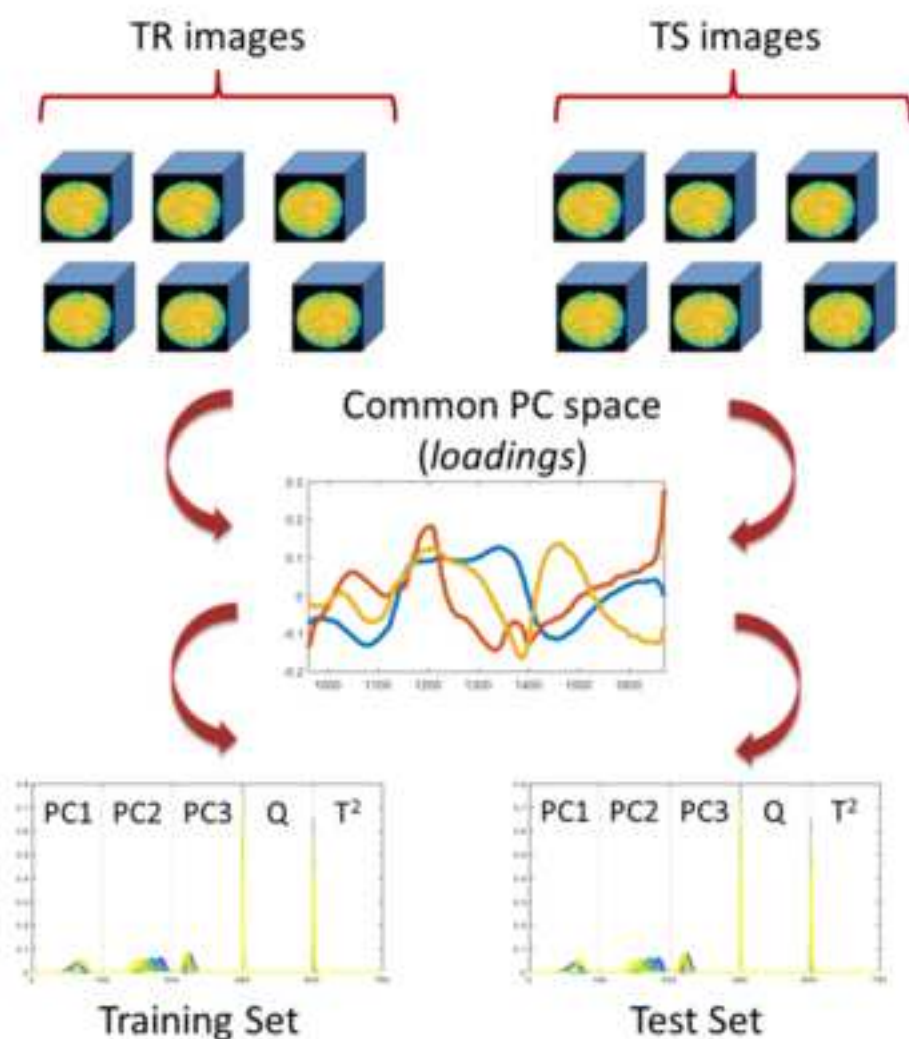


Figure 2

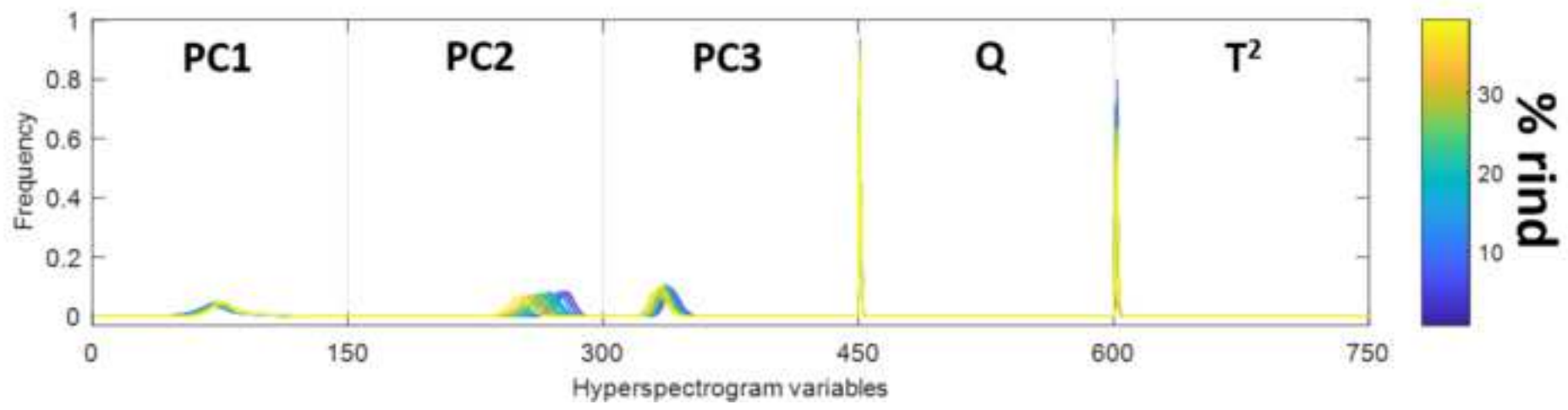


Figure 3

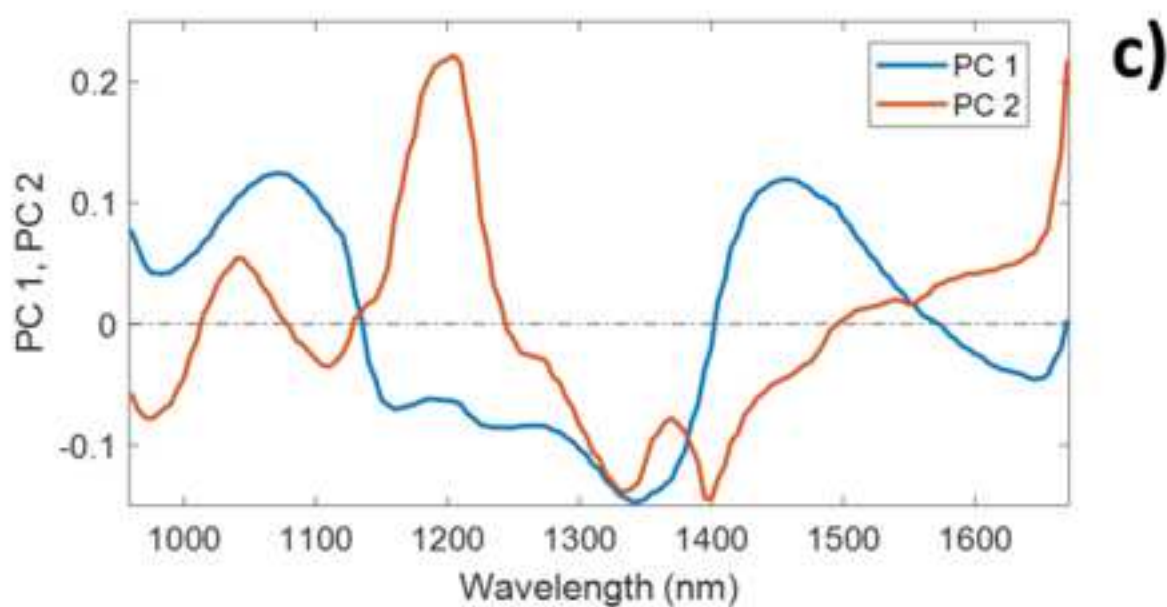
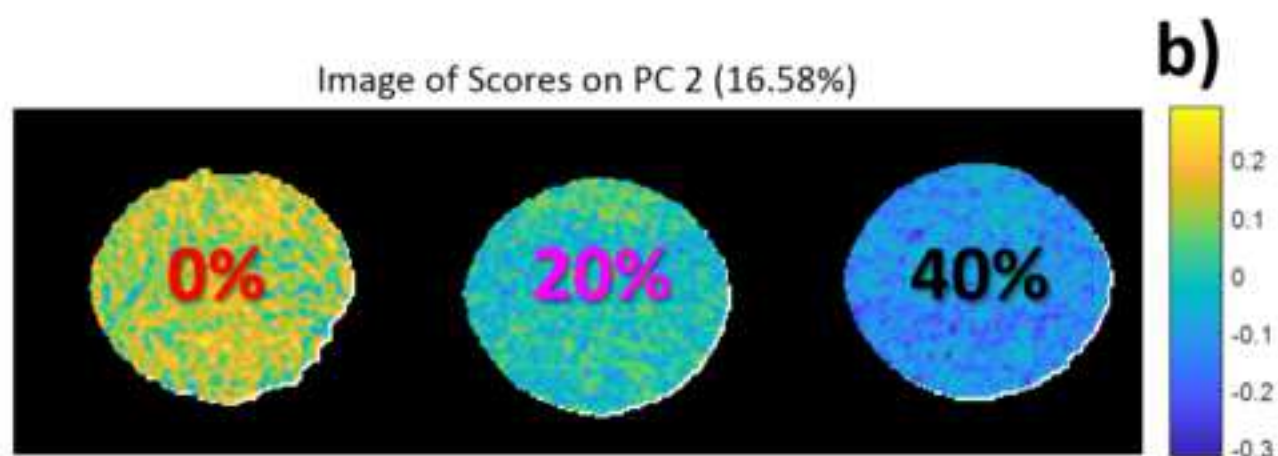
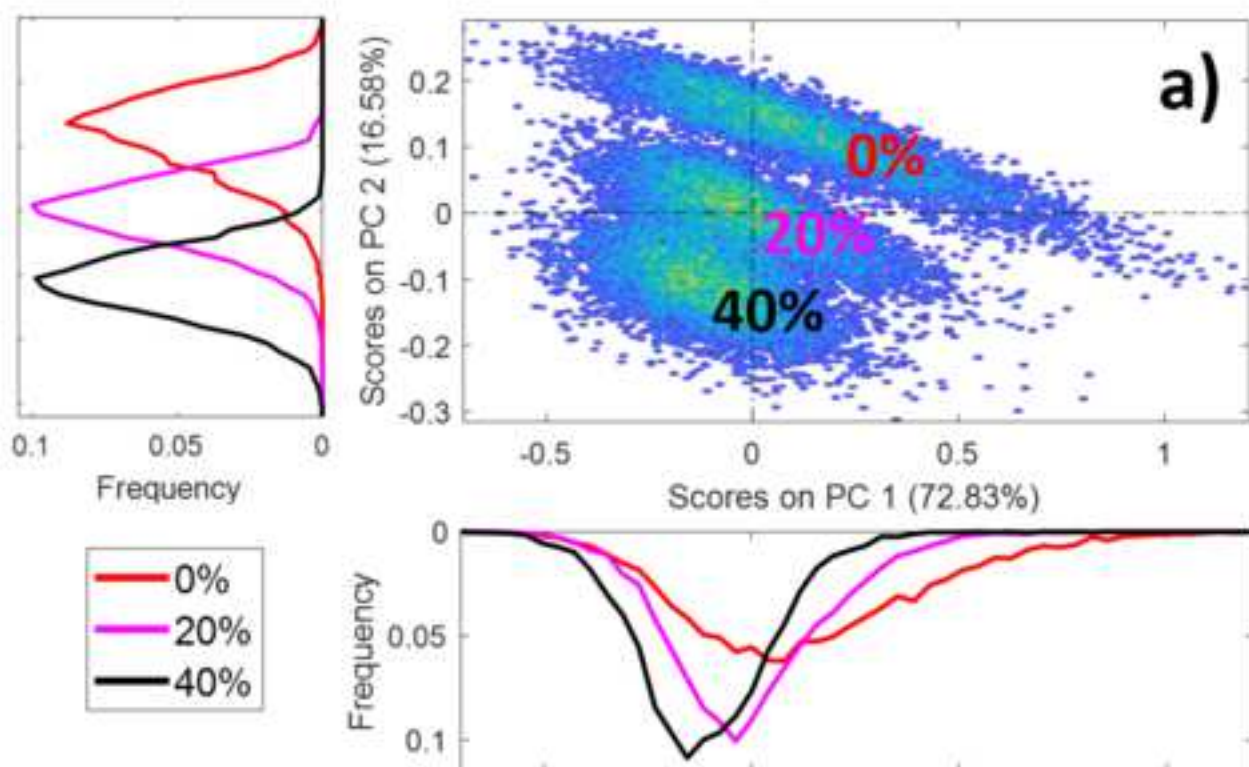


Figure 4

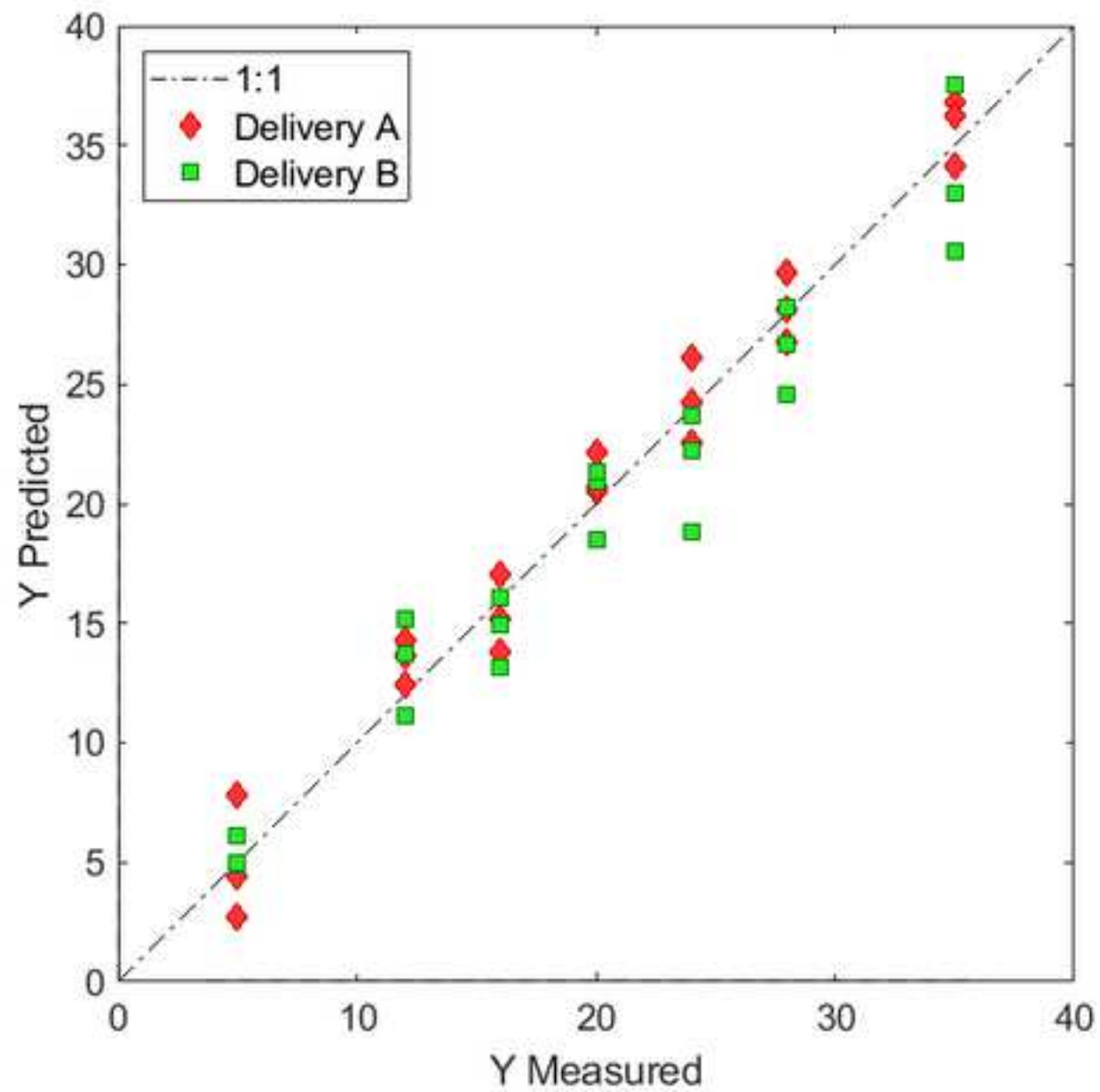


Figure 5

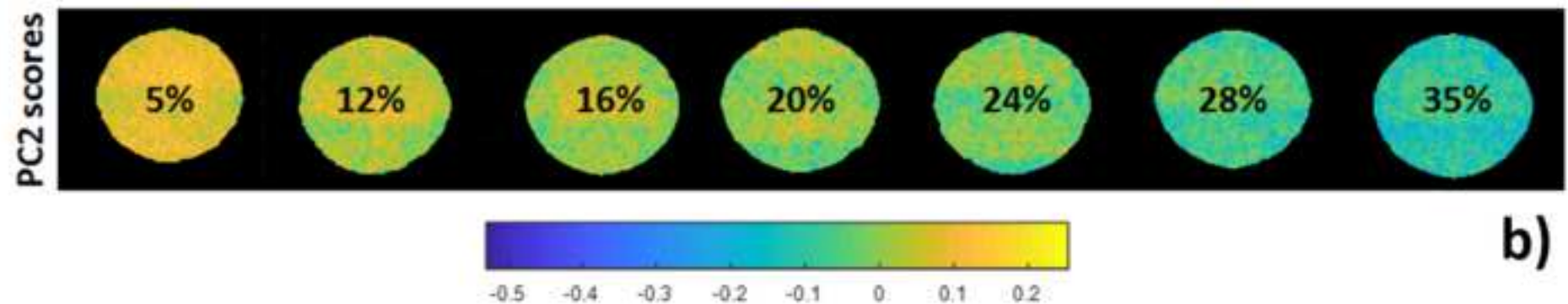
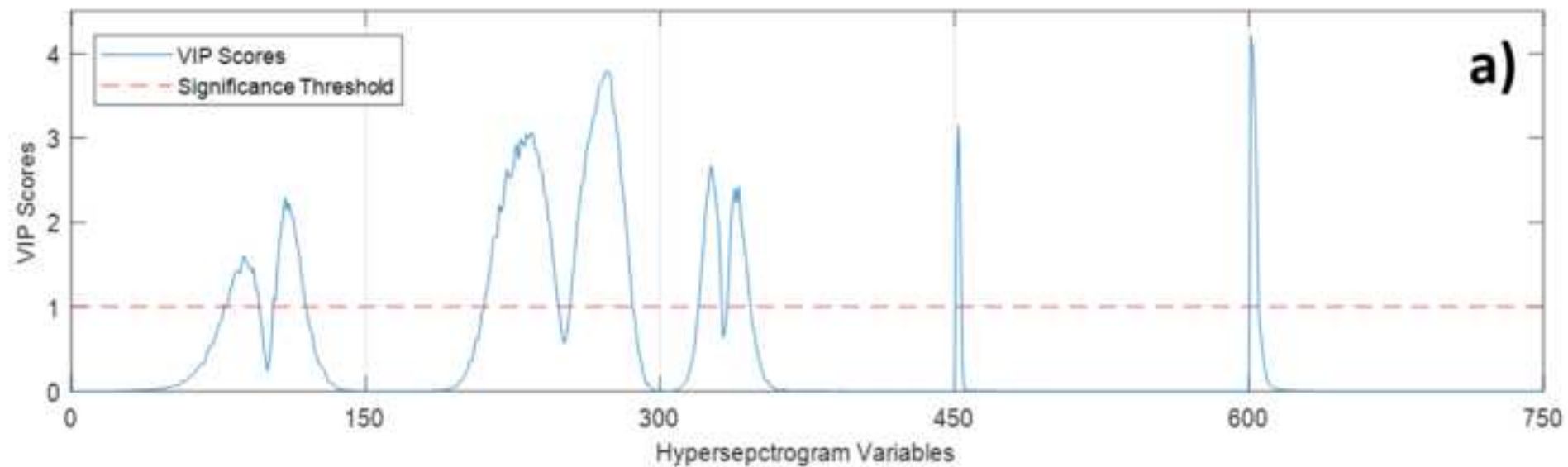


Table 1. Summary information about the grated cheese samples considered in the present study

	Delivery Date	RP (w/w)	Number of samples	Number of images
Delivery A	12/04/2018	0%, 5%, 10%, 12%, 14%, 16%, 18%, 20%, 22%, 24%, 26%, 28%, 30%, 35%, 40%	15	45 (=15×3)
		X1-X7	7	21 (=7×3)
Delivery B	12/11/2018	0%, 5%, 10%, 12%, 14%, 16%, 18%, 20%, 22%, 24%, 26%, 28%, 30%, 35%, 40%	15	45 (=15×3)
		X8-X15	8	24 (=8×3)

Table 2. Results of PLS regression for the determination of rind percentage.

LVs	8
RMSEC	0.98
RMSECV	1.70
RMSEP_{0-40%}	1.91
RMSEP_{10-30%}	1.85
R² Cal	0.993
R² CV	0.979
R² Pred_{0-40%}	0.958
R² Pred_{10-30%}	0.893

Table 3. Prediction results of the unknown test samples and corresponding experimental RP values.

Sample name	Predicted RP	Experimental RP
X1	30	32
X2	23	23
X3	20	18
X4	0	0
X5	7	5
X6	24	23
X7	13	10
X8	43	40
X9	14	12
X10	18	16
X11	30	28
X12	9	5
X13	35	32
X14	1	0
X15	23	18