

Peer Reviewed Paper **openaccess** [Paper Presented at IASIM 2018, June 2018, Seattle, WA, USA](#)

# Raman and Fourier transform infrared hyperspectral imaging to study dairy residues on different surfaces

V. Caponigro,<sup>a\*</sup> F. Marini,<sup>b</sup> R.M. Dorrepaal,<sup>c</sup> A. Herrero-Langreo,<sup>d</sup> A.G.M. Scannell<sup>e</sup> and A.A. Gowen<sup>f</sup>

<sup>a</sup>UCD School of Biosystems and Food Engineering and UCD Institute of Food and Health, Belfield, Dublin, Ireland.

E-mail: [vicky.caponigro@ucdconnect.ie](mailto:vicky.caponigro@ucdconnect.ie), <https://orcid.org/0000-0003-2298-0508>

<sup>b</sup>Dipartimento di Chimica, Sapienza Università di Roma, Piazzale Aldo Moro 5, Rome, Italy. <https://orcid.org/0000-0001-8266-1117>

<sup>c</sup>UCD School of Biosystems and Food Engineering, Belfield, Dublin, Ireland. <https://orcid.org/0000-0002-1429-5570>

<sup>d</sup>UCD School of Biosystems and Food Engineering and UCD Institute of Food and Health, Belfield, Dublin, Ireland.

<https://orcid.org/0000-0003-3258-6248>

<sup>e</sup>UCD Institute of Food and Health, UCD Centre for Food Safety and UCD School of Agriculture and Food Science, UCD, Belfield, Dublin, Ireland.

<https://orcid.org/0000-0001-5051-7436>

<sup>f</sup>UCD School of Biosystems and Food Engineering and UCD Institute of Food and Health, Belfield, Dublin, Ireland.

<https://orcid.org/0000-0002-9494-2204>

Milk is a complex emulsion of fat and water with proteins (such as caseins and whey), vitamins, minerals and lactose dissolved within. The purpose of this study is to automatically distinguish different dairy residues on substrates commonly used in the food industry using hyperspectral imaging. Fourier transform infrared (FT-IR) and Raman hyperspectral imaging were compared as candidate techniques to achieve this goal. Aluminium and stainless-steel, types 304-2B and 316-2B, were chosen as surfaces due to their widespread use in food production. Spectra of dried samples of whole, skimmed, protein, butter milk and butter were compared. The spectroscopic information collected was not only affected by the chemical signal of the milk composition, but also by surface signals, evident as baseline and multiplicative effects. In addition, the combination of the spectral information with spatial information can improve data interpretation in terms of characterising spatial variability of the selected surfaces.

**Keywords:** milk, Raman, FT-IR, hyperspectral imaging, aluminium, stainless steel, PCA, PLS-DA

## Introduction

Milk is a complex emulsion of fat and water in which different nutrients are dissolved: proteins (caseins and whey protein), vitamins, minerals and carbohydrates, such as lactose.<sup>1</sup> Its complex nature makes it difficult to completely and quickly characterise all components with a single analytical technique. Due to the importance of milk and its by-products in human nutrition, many research approaches are used. Two key research

approaches, relevant to this study are (1) characterisation of the different components of milk and (2) hygienic studies of the surfaces in contact with the different components.

Milk's composition can provide information about the production animal.<sup>2</sup> Focusing on cow's milk, the composition gives information about the quality of the products<sup>3</sup> as well as of adulteration,<sup>4</sup> effects of homogenisation<sup>5</sup>

### Correspondence

V. Caponigro ([vicky.caponigro@ucdconnect.ie](mailto:vicky.caponigro@ucdconnect.ie))

**Received:** 16 November 2018

**Revised:** 7 January 2019

**Accepted:** 7 January 2019

**Publication:** 14 January 2019

**doi:** 10.1255/jsi.2019.a3

**ISSN:** 2040-4565

### Citation

V. Caponigro, F. Marini, R.M. Dorrepaal, A. Herrero-Langreo, A.G.M. Scannell and A.A. Gowen, "Raman and Fourier transform infrared hyperspectral imaging to study dairy residues on different surfaces", *J. Spectral Imaging* 8, a3 (2019). <https://doi.org/10.1255/jsi.2019.a3>

© 2019 The Authors

This licence permits you to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper in this journal is given, the use is not for commercial purposes and the paper is not changed in any way.



or other industrial applications.<sup>6</sup> In addition, raw milk compositional information can be used to predict origin<sup>7</sup> and the lactation stage of the cow<sup>8,9</sup> which produced it.

One of the most commonly applied techniques to characterise milk composition is chromatography; however, in spite of its widespread use, it is difficult to determine common parameters, i.e. temperature, solvents, type of column and eluents to separate all milk components. Chromatography can analyse liquid milk or dried/powder milk, as often the samples are pre-treated or dissolved.<sup>10-13</sup> Another problem is that chromatography is a destructive technique: samples are split into their components during the analysis and are separated and eluted at different time points based on specific chemical properties, i.e. polarity, chirality etc.<sup>14</sup>

To circumvent some limitations experienced with chromatography, e.g. sample destruction, time and portability, different types of spectroscopy have been applied. These include Raman,<sup>15-17</sup> Fourier transform infrared (FT-IR)<sup>7,18,19</sup> and near infrared (NIR).<sup>19,20</sup> In addition, sampling techniques are also applied to analyse milk: attenuated total reflection (ATR),<sup>21</sup> micro attenuated total reflection ( $\mu$ ATR) and high throughput transmission (HTT) have also been compared.<sup>5</sup> Unlike chromatography, spectroscopy is a non-destructive method which allows for the simultaneous characterisation of various components with one analysis and samples can be analysed without being pre-treated.

Milk is a high nutrition food and supports a variety of microbial species.<sup>22,23</sup> Assessment of the hygienic level of the surfaces that are used in milk production or in everyday life is complicated by the complexity of milk's composition. Stainless steel is widely used in the food industry due to its corrosion resistance. It can be cleaned and sterilised. The most used stainless steel types in the food industry are AISI-304, which is made of 18% chromium and 10% of nickel and AISI-316 composed of 17% chromium, 12% nickel and 2.5% molybdenum. Stainless steel 316 is more resistant to chloride solutions, so it is also used in medical devices. This type of stainless steel is used most frequently in food production because it is a good compromise between price and hygiene. Aluminium should be avoided in the food chain, as it is not resistant to corrosion, though it is generally accepted in domestic kitchens.<sup>24</sup> In the literature, there have been some studies which investigate the interaction of surfaces and milk with respect to hygiene level<sup>25</sup> and their interactions with bacteria<sup>23,26</sup> and biofilms.<sup>22</sup>

From these studies, it appears that stainless steel has a long phase of adhesion for biofilm but produces a strong biofilm.<sup>22</sup> In addition, the hygiene status of stainless steel is based on composition but not on the finish.<sup>25</sup> Some milk protein absorption can inhibit bacteria adhesion.<sup>23</sup>

The purpose of this study is the characterisation of dried whole milk, skimmed milk, protein milk, butter milk and butter on stainless steel 316-2B, stainless steel 304-2B (2B defines stainless steel surfaces with a pickling finish) and aluminium with FT-IR and Raman hyperspectral imaging. The effect of the surfaces on the sample spectra is also investigated.

## Materials and methods

### Sample preparation

Rectangular slides (37 mm  $\times$  25 mm  $\times$  1 mm) made from aluminium, stainless steel types 304-2B finished and 316-2B finished [purchased from Rice Metals (Cornwall, United Kingdom)] were chosen as substrates in this study. The surfaces were supplied covered with a protective tape. To remove any traces of glue, the surfaces were washed with a solution of ethanol [99%, Absolute, Extra Pure, SLR purchased from Fisher Scientific (Dublin, Ireland), CAS Number: 64-17-5] and acetone [HPLC grade, obtained from Fisher Scientific (Dublin, Ireland), CAS Number: 67-64-1] (1:1 v/v). Each slide was immersed for 10 min in the acetone/ethanol solution and was then rinsed for 5 min with deionised (DI) water sourced from a Thermo Scientific™ Barnstead™ Smart2Pure™ water purification system (producing Type I ASTM water, with a resistance of 18.2 M $\Omega$ cm at 24.7°C).

An isolator (Grace Bio-Labs Press-To-Seal silicone isolator adhesive on one side. Round, 13 mm diameter, 2.5 mm depth, CAT Number: GBL665307) was applied on top of each support to contain the dairy products. Selected dairy substrates were whole milk, protein milk, skimmed milk, butter milk and butter produced by Avonmore Ltd (Avonmore) purchased at a local shop. Table 1 shows the nutrients as declared by Avonmore<sup>27</sup> for each type of dairy product. The isolators were filled with 0.5 mL of each milk sample and were dried at room temperature (about 20°C), for 48 h prior to analysis. In addition, a sample series of butter was prepared and analysed. In this case, the thickness of whole dried milk was reproduced with a layer of butter by applying a cover slip on top. The system was frozen to measure the thick-

Table 1. Nutrient content of Avonmore dairy products utilised in this work.<sup>27</sup>

Nutrient content of Avonmore dairy products	Whole milk (g/100 mL)	Skimmed milk (g/100 mL)	Protein milk (g/100 mL)	Butter milk (g/100 mL)	Butter (g/100 g)
Fat	3.5	1	1	3	81
of which saturates	2.2	0.6	0.6	1.6	53
Carbohydrate	4.7	4.8	4.8	5	0.8
of which sugars	4.7	4.8	4.8	5	0.8
Protein	3.4	3.5	5.1	3.4	0.5
Salt	0.11	0.11	0.11	0.1	1.8
Calcium (mg)	119	123	165	120	—
Vitamin B12 ( $\mu\text{g}$ )	0.4	0.4	0.4	—	—
Vitamin D ( $\mu\text{g}$ )	1	—	1	—	—

ness and to enable easy removal of the cover slip. The thickness of the butter plus glass slide was measured and compared with the untreated glass thickness to determine the smear thickness. The thickness (approximately 0.15 mm) of both whole milk and butter was measured by a Vogel electronic micrometer (range 0–25 mm and resolution 0.001 mm).

Two experimental repetitions, from two different batches, were prepared and analysed for each combination of substrate and support. The dairy products were transferred from their packaging into 30 2 mL vials and stored in a fridge at 6 °C, over the sample preparation period. To avoid sample contamination, before applying the dairy products on the surfaces, two vials for each sample were removed from the fridge, one to measure the temperature and the other to pour the milk on the surface. After about 10 min, when the dairy products reached room temperature (20 °C), they were poured into the spacer.

## Hyperspectral imaging systems and data collection

### Raman hyperspectral imaging system

In this study a Renishaw InVia micro-Raman spectroscopy system with a Leica DM2500 M microscope and NIR-enhanced deep depletion CCD array (1024 × 256 pixels) were used to analyse the samples. Two Raman shift ranges (range 1: 2410–190  $\text{cm}^{-1}$  and range 2: 4000–2360  $\text{cm}^{-1}$ ) were collected to cover a total range of 3900–190  $\text{cm}^{-1}$ . Preliminary tests were carried out to optimise the instrumental parameters. An edge 785 nm laser was chosen for the excitation in order to reduce the fluorescence caused by proteins present

in milk. The laser was set to 5% power (power measured at source was 306 mW) and 5 s acquisition time was selected for each spectrum. Images were obtained using a 10×, 0.25 NA objective lens. The average spectral resolution was 1.8778  $\text{cm}^{-1}$ . Each data cube contained 200 spectra, collected as a 1.90 × 0.90 mm map with a step size of 100  $\mu\text{m}$ . The data were encoded in Renishaw WDF format. The instrument was calibrated daily using an internal silicon reference.

### FT-IR hyperspectral imaging system

A Nicolet iN10 MX imaging microscope with LN<sub>2</sub>-cooled MCTA linear array detector was used to collect FT-IR spectra. The spectra were acquired point by point over the wavenumber range 7500–475  $\text{cm}^{-1}$ . 16 spectral acquisitions were collected per pixel with a total pixel acquisition time of 5 s. Spectra were collected in reflectance. An aperture size of 150 × 150  $\mu\text{m}$  and step size of 150  $\mu\text{m}$  was used to obtain a data cube of 300 spectra covering a spatial region of 3.00 × 2.25 mm. The spectral resolution was 1.928  $\text{cm}^{-1}$ . The data were saved in Nicolet MAP file format and exported to ENVI format files. A background was collected every 300 minutes using a gold reference sample.

## Data pre-treatments and analysis

The two ranges collected with Raman were examined for cosmic rays and were corrected using an in-house function which detects cosmic rays and allows for manual inspection, correcting with the mean of five neighbouring wavenumber intensities.<sup>28</sup> The obtained Raman data from the two ranges were individually pre-treated (as described below) and the pre-treated spectra were

concatenated to form a single spectrum of range of 3900–190  $\text{cm}^{-1}$  for each pixel.

In the case of FT-IR data, both the near infrared (NIR) and mid infrared (MIR) ranges were collected simultaneously. They were split into two ranges (MIR: 3580–920  $\text{cm}^{-1}$  and NIR: 6200–3700  $\text{cm}^{-1}$ ) and pre-treated independently (as described below). Before re-concatenation, the spectra were normalised using the Euclidean norm.

The normalisation is necessary for FT-IR data because the collected spectra contain both NIR overtones and MIR molecular vibrations. Bands in the NIR range are 10–100 times weaker than the MIR range, so in order to merge the two ranges, normalisation is necessary. In the case of Raman spectra, these differences are not present because the ranges are collected from the same detector (as FT-IR) and do not exhibit any physical differences.

Raman and FT-IR data set were both pre-treated by obtaining the Savitzky–Golay derivative and using standard normal variate normalisation. Later, principal components analysis and partial least squares-discriminant analysis were applied (as described below).

*The Savitzky–Golay (SG) derivative* is a row pre-treatment method which has various applications. In the case of Raman data, it corrects the fluorescence baseline,<sup>29</sup> while it improves the signal-to-noise ratio by smoothing the noise and emphasises peak features in FT-IR data. In this study, a window size of 15 was used. The second-order polynomial was calculated, followed by calculation of the first derivative.<sup>30</sup> In addition, the Savitzky–Golay derivative was used to identify the saturated pixels in Raman spectra. The derivative spectra with a maximum value greater than 4000 units were identified as saturated and were subsequently masked.

*Standard Normal Variate (SNV)* was applied to the data in order to remove multiplicative and additive effects. The scattering between the spectra due to the irregular surface roughness was reduced using this technique.<sup>31</sup>

*Principal Component Analysis (PCA)* was applied to FT-IR and Raman mean centred data sets independently. In the case of Raman, the analysis was carried out on pre-treated data sets after the concatenation of the two ranges. Pixels with saturation were masked as described in the Savitzky–Golay section above. The FT-IR data sets were also analysed using PCA after pre-treatment, normalisation and the concatenation of the two ranges. For each sample, and for

each modality, both repetitions were included in the analysis.

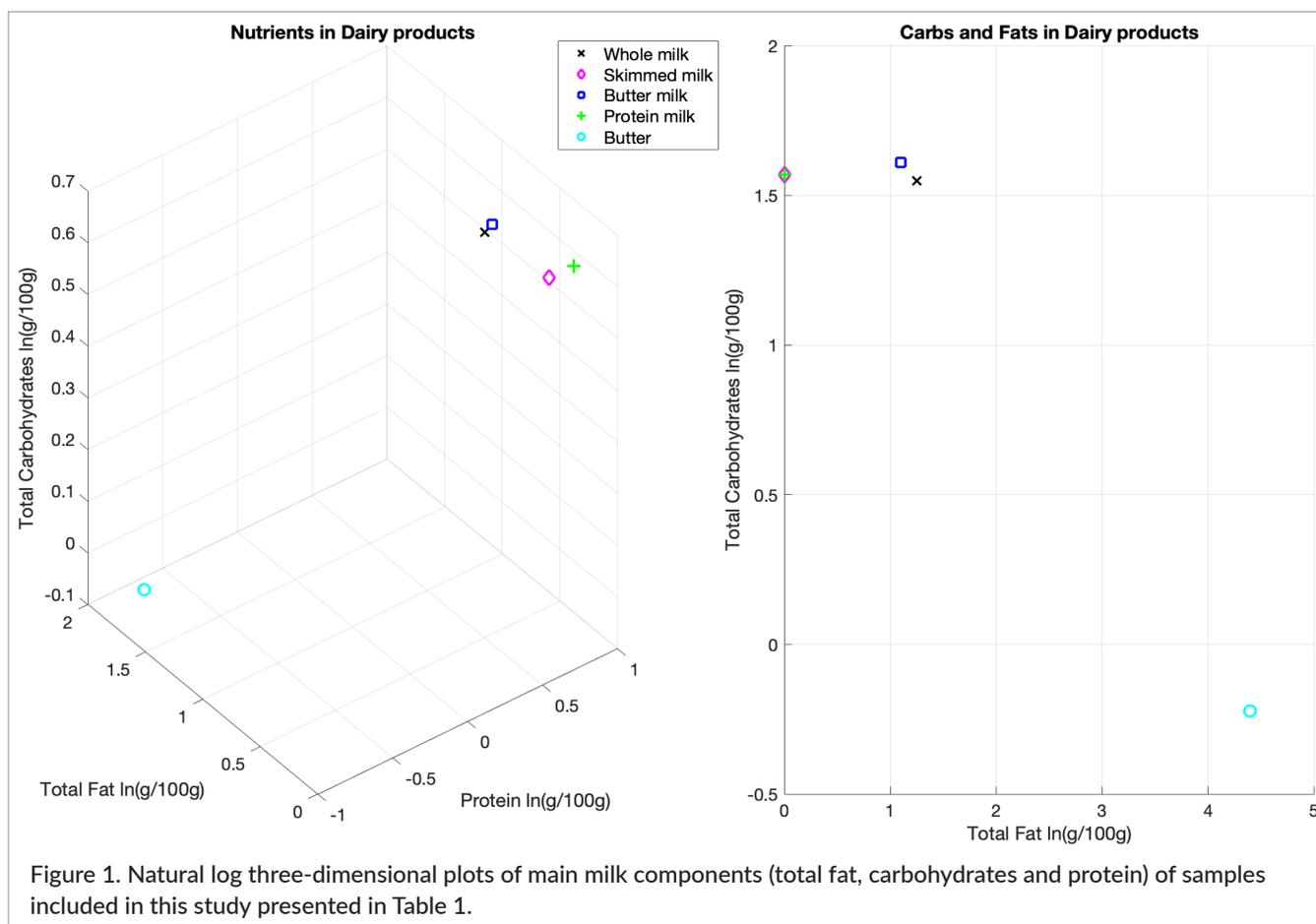
*Partial Least Squares-Discriminant Analysis (PLS-DA)* is a widely used chemometrics tool in both hyperspectral imaging and data analysis. As a supervised technique, PLS-DA performance metrics can be investigated in order to evaluate predictive ability. In this study, PLS-DA was applied on pre-treated data in order to classify the different dairy residues, regardless of substrate used.

One experimental repetition was used to train and calibrate the model while the second repetition was used for validation. The number of latent variables (LVs) was chosen to minimise the misclassification error on the calibration set. To evaluate the built model, both pixel and object interpretation were applied. For pixel interpretation, the percentage of correctly classified pixels (% CC) were calculated as the ratio between the sum of the correctly classified pixels and the total number of pixels of the validation set. Sensitivity and specificity were evaluated for each class. Sensitivity is defined as the ratio of the correctly identified pixels divided by the total number of pixels for that class. On the other hand, the percentage of pixels from other classes correctly predicted as not belonging to the category corresponds to the specificity. Object interpretation was carried out by considering the values of each predicted image that occurred most often (the mode).

As mentioned in the PCA section, FT-IR and Raman data were analysed independently. In the case of Raman data, the two ranges were pre-treated, concatenated and masked before carrying out PLS-DA analysis. The FT-IR data were pre-treated, normalised and concatenated before carrying out PLS-DA analysis.

## Results and discussion

An initial investigation of the samples included in the study was carried out based on the nutritional values presented in Table 1.<sup>27</sup> The most influential nutrients (i.e. proteins, carbohydrates and fats) of Table 1 are plotted in Figure 1. It is evident that butter is distinct from the other samples. Whole milk, skimmed milk and butter milk have similar values. Protein milk, of course, has a higher protein content than the other products but, when only considering the fat and carbohydrates ratio it completely overlaps with skimmed milk.



The infrared (IR) spectrum of milk is complex due to the combination of a large number of milk components and their three-dimensional structure (i.e., fat micelles interact with some lipoproteins). In general, the ranges  $4500\text{--}4200\text{cm}^{-1}$  and  $4300\text{--}4000\text{cm}^{-1}$  are assigned to CH aliphatic and aromatic groups, respectively. The range between  $4000\text{cm}^{-1}$  and  $3200\text{cm}^{-1}$  is assigned to amine (NH) and alcohol (OH) stretching signals which overlap.<sup>32</sup> Peaks between  $3100\text{cm}^{-1}$  and  $2800\text{cm}^{-1}$  are related to CH stretching vibrations due to both lipids and carbohydrates. The lower wavenumber region ( $1800\text{--}400\text{cm}^{-1}$ ) is the most rich in information but also the most difficult to interpret due to overlapping signals.<sup>17</sup> The range between  $1800\text{cm}^{-1}$  and  $1400\text{cm}^{-1}$  is used to study protein in milk.<sup>7</sup> In addition, the  $1200\text{--}800\text{cm}^{-1}$  range is generally related to the vibrational modes of carbohydrates, where the lactose signal also appears.<sup>17,21,33</sup> There are other typical bands used to identify different milk components as summarised in Table 2, which are used to interpret both Raman and FT-IR spectra and also the corresponding loadings obtained from PCA analysis.

## Raman analysis

Raman spectroscopy is widely used to characterise different types of milk because it has well-defined peaks related to different components. In this study, Raman spectra were collected in two different ranges and pre-processed as described previously in the Materials and methods section. Raman maps of dairy residue samples and clean substrates were both measured. Figure 2 shows the Raman spectra of both repetitions (first blue, second red) after cosmic ray correction and concatenation. Their mean is shown in Figure 3. Observing the collected data, it is difficult to recognise the typical liquid milk profile spectrum<sup>15</sup> or any obvious trend between the different samples. Milk components, i.e. fat micelles or protein, do not preserve the same three-dimensional structure in a dried form. Figure 4 represents the mean spectra of the pure aluminium, stainless steel 304-2B and 316-2B samples before and after pre-treatments. Comparing these spectra with those of Figure 2 and Figure 3, there are no evident influences of substrate on the spectra of the dairy residues.

To improve the interpretation, principal component analysis was carried out on the pre-treated, concatenated and masked map spectra as explained in materials and methods.

Score plot of PC1 vs PC2 vs PC3 and the loadings plot for these PCs are shown in Figure 5.

False colour red green and blue (RGB) image (Figure 6) was built with the same score values plotted in Figure 5,

i.e., image scores value of PC1, PC2 and PC3 are rescaled between 0 and 1 and concatenated to make an image in which each red, green and blue (RGB) index corresponds to PC1, PC2 and PC3, respectively. In Figure 6, white pixels are due to the applied mask described in the materials and methods section.

The PCA scores plot and false RGB image of both repetitions together indicate a clear pattern, where the

Table 2. Definition of bands over the IR range.

Peaks ( $\text{cm}^{-1}$ )	Chemical groups	Components
1748	CO stretching ester	Fatty acids <sup>34</sup>
1650~1640	Combination of: CO stretching mode, CONH group, CC stretching	Amide I, proteins and fatty acids <sup>35,36,5</sup>
1500	Combination of: NH deformation and CN stretching	Amide II <sup>35,5</sup>
1450	$\text{CH}_2$ deformation	Fat and carbohydrates <sup>36,5</sup>
1340	CO stretching and COH deformation	Carbohydrates <sup>17</sup>
1304	$\text{CH}_2$ twisting	Lipids <sup>17</sup>
1260	$\text{CH}_2$ twisting mode	Carbohydrates <sup>17</sup>
1050–1150	CO stretching	Lipids <sup>32</sup>
1004	Ring breathing mode	Phenylalanine <sup>17</sup>

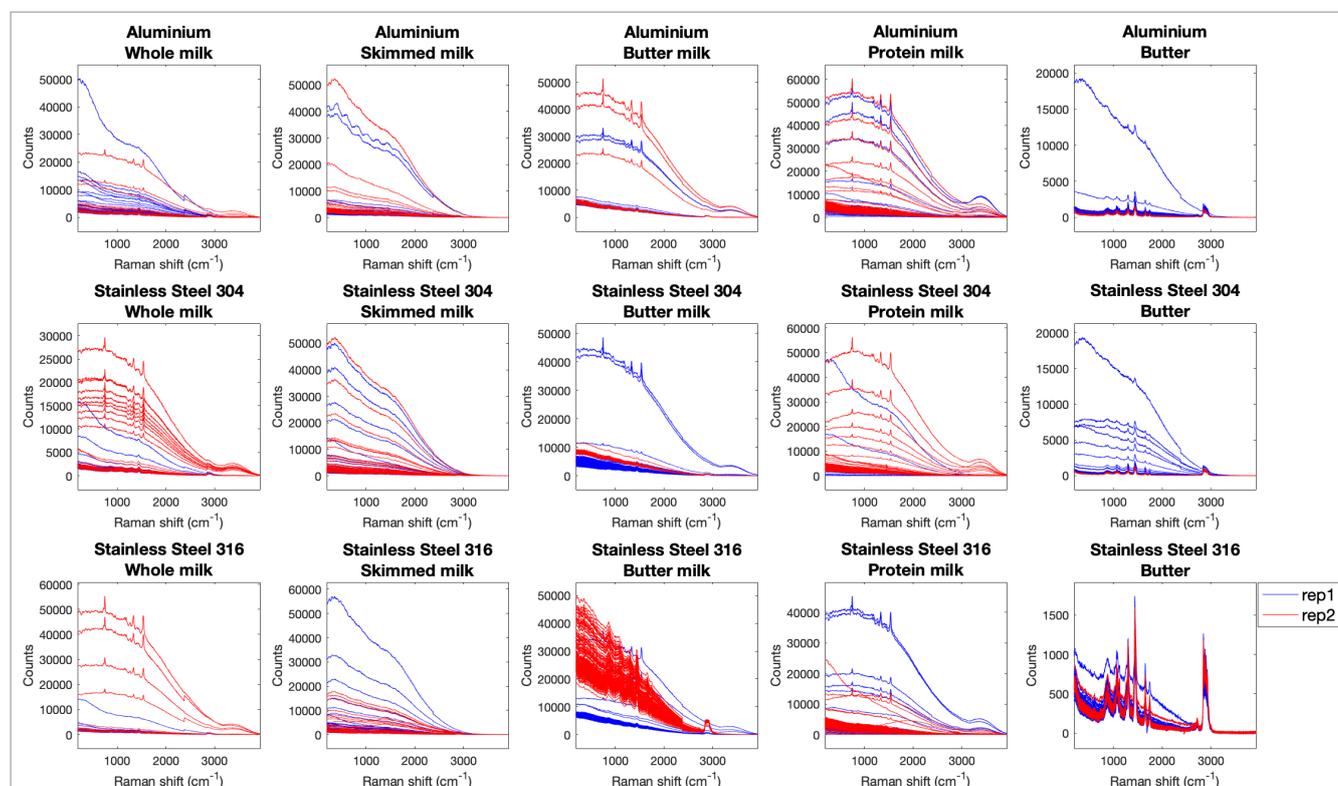
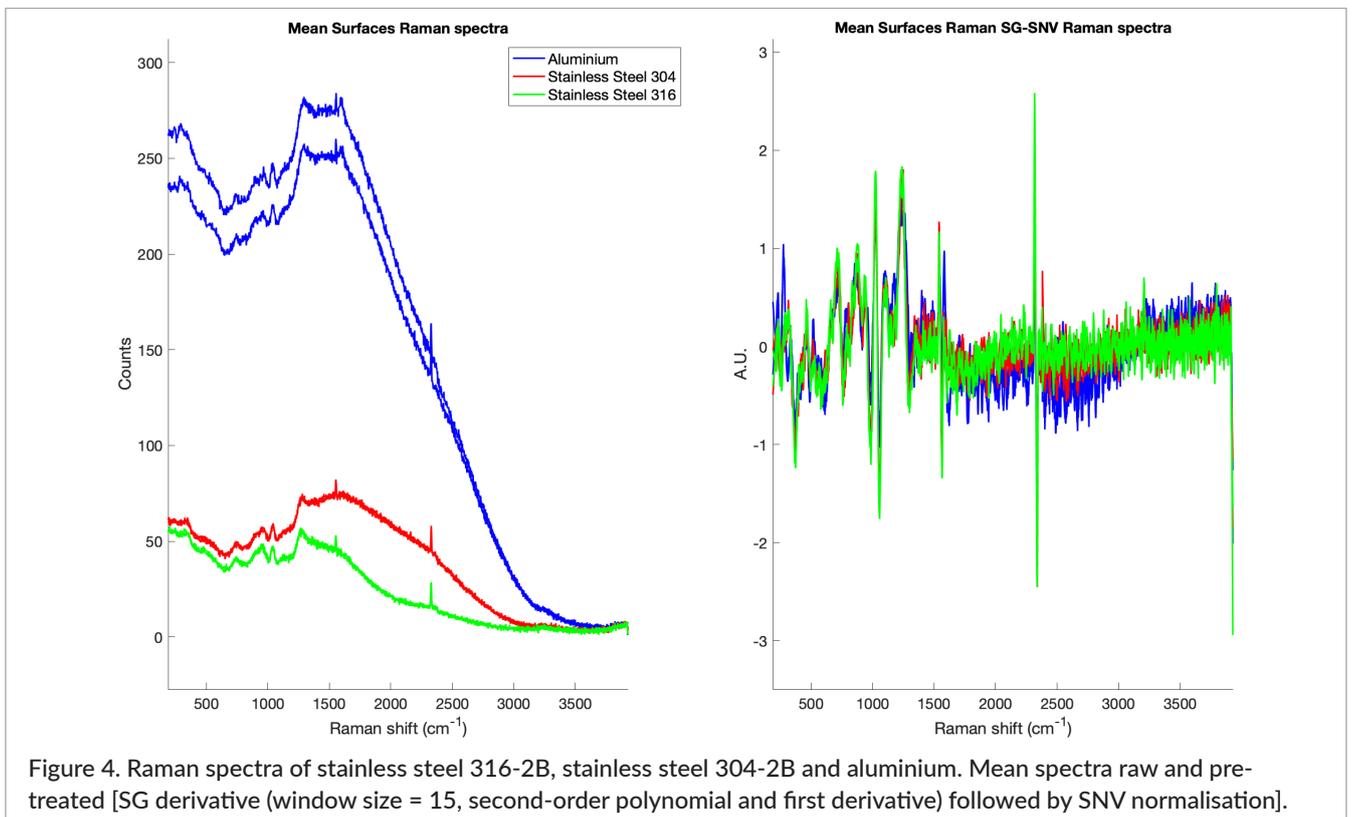
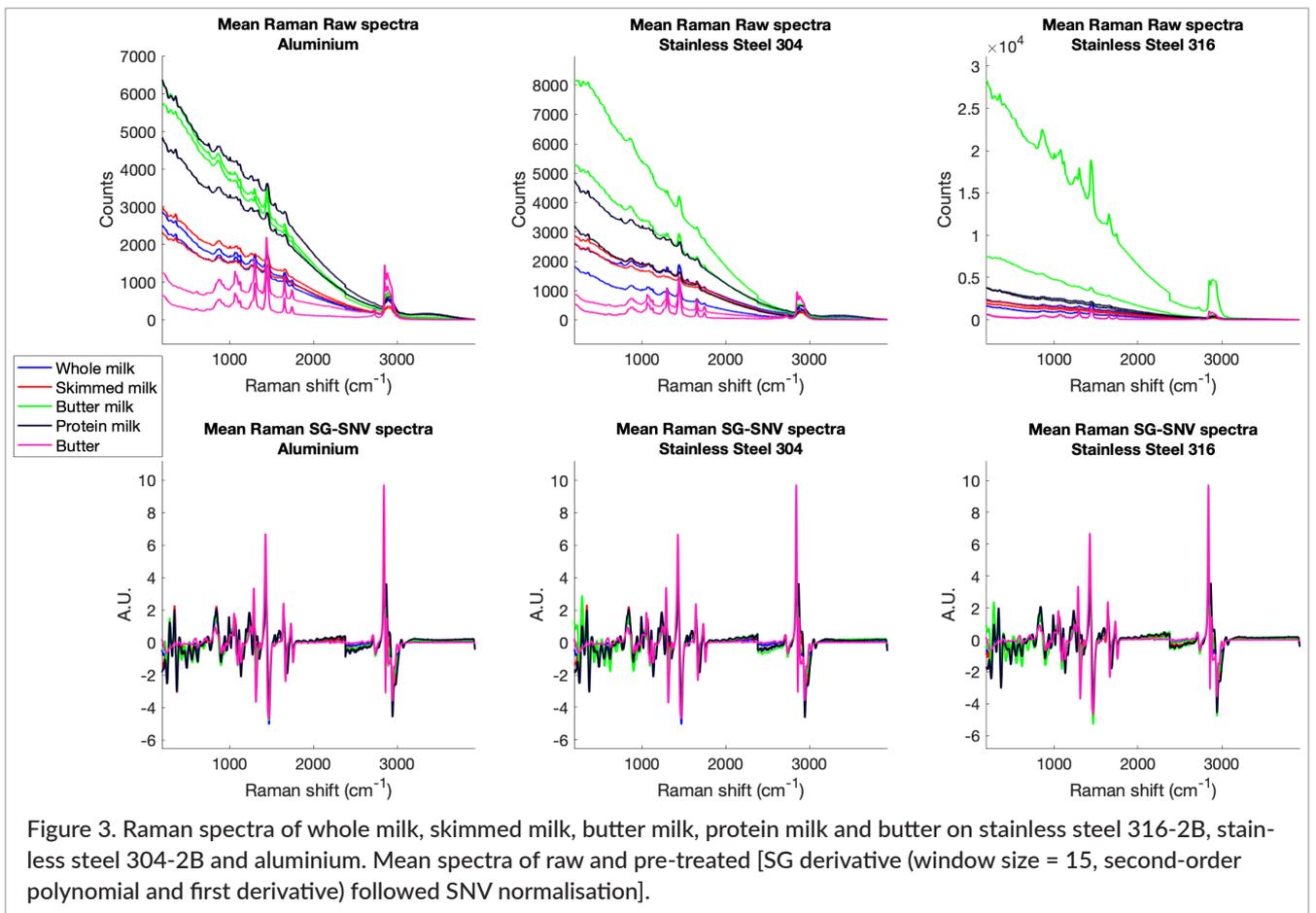
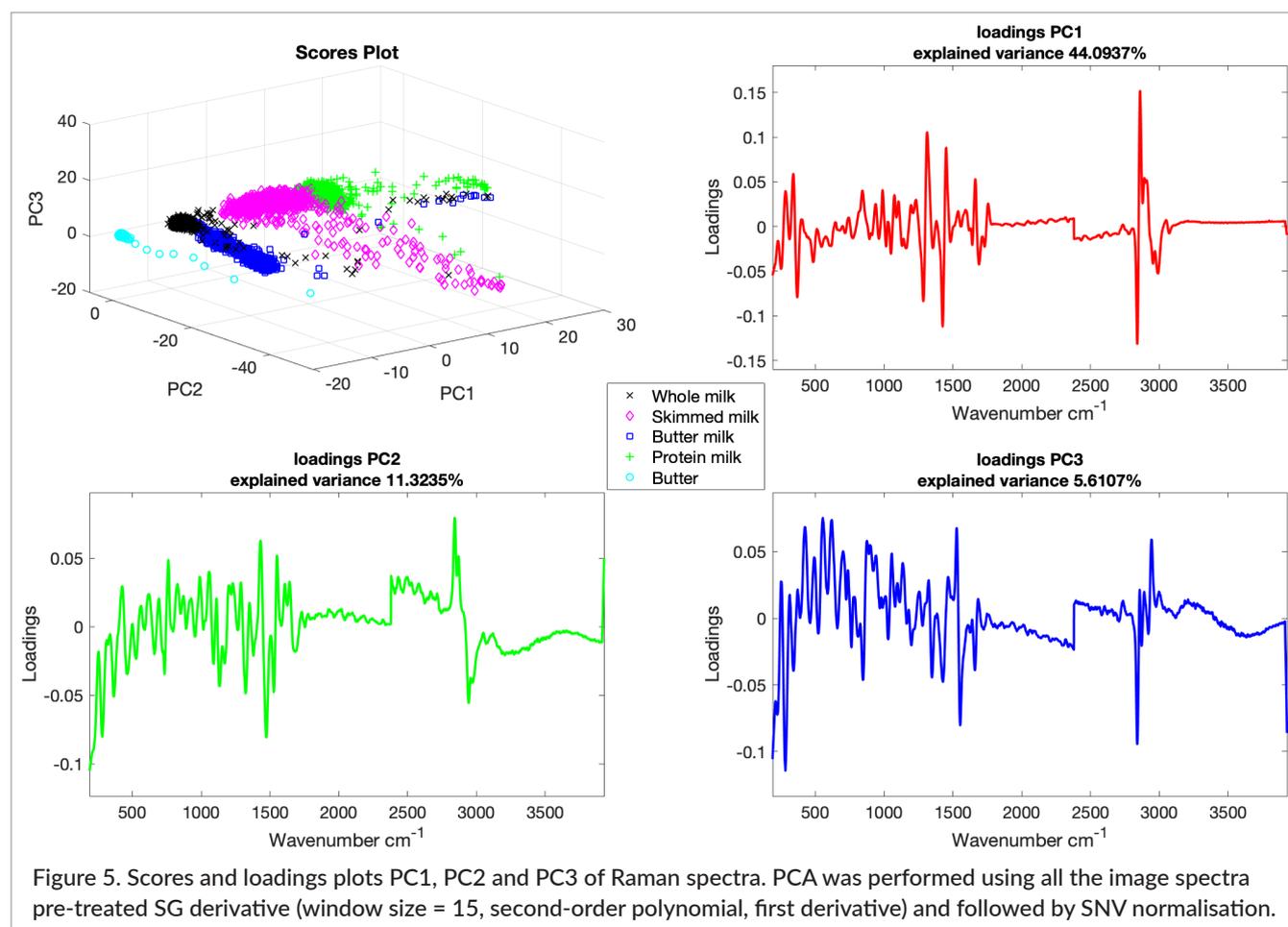


Figure 2. Raman spectra of both repetitions (first blue, second red) of whole milk, skimmed milk, butter milk, protein milk and butter on stainless steel 316-2B, stainless steel 304-2B and aluminium after cosmic ray correction.





different types of dairy residue form clusters in the PC1–3 space.

Comparing the loadings plot in Figure 5 with the typical milk peaks reported above, it is evident that the first principal component (PC1), which explained 44.09% of the variance, is a combination of the contributions of proteins and carbohydrates. It is strongly connected with the lipid and carbohydrates signals around  $3100\text{--}2800\text{cm}^{-1}$ . The second principal component (PC2) (explained variance 11.32%) seems to be more influenced by the fat signal. PC2 has a very clear signal around  $2800\text{cm}^{-1}$  which may be due to both lipids and carbohydrates but, unlike PC1, it seems not to contain much information below  $1200\text{cm}^{-1}$ . Figure 6 supports this interpretation: butter is distinguishable from the other dairy products mainly due to the contribution of scores of PC2 (green component of the false RGB image). In spite of a considerable explained variance (5.61%), the third principal component (PC3) does not seem to bring any new chemical information related to the major nutrients. The peak

around  $2800\text{cm}^{-1}$  in PC3 may still be related to lipids and carbohydrates. However, Figure 6 seems to show that PC3 contributes only in combination with PC1, in the case of the purple pixels. These are outliers in the images, and they can be associated with less thick milk zones on the surface. In addition, from the same figure, it is evident that there is little contribution of the substrates: it is not possible to distinguish between aluminium, stainless steel 316-2B and 304-2B in the false colour RGB images. Indeed, it was not possible to identify any trend in the score scatter plot with respect to surface type (Appendix Figure A1).

In order to classify the samples based on their spectra, a PLS-DA model was built using the two independent repetitions, one as calibration set and one as validation set. 14 latent variables (LVs) were chosen based on the minimum misclassification value. Observing Figure 7, it is evident that there are some misidentifications at the pixel level. This model presented a value of total correct classification of 92.47%. The lowest sensitivity was for

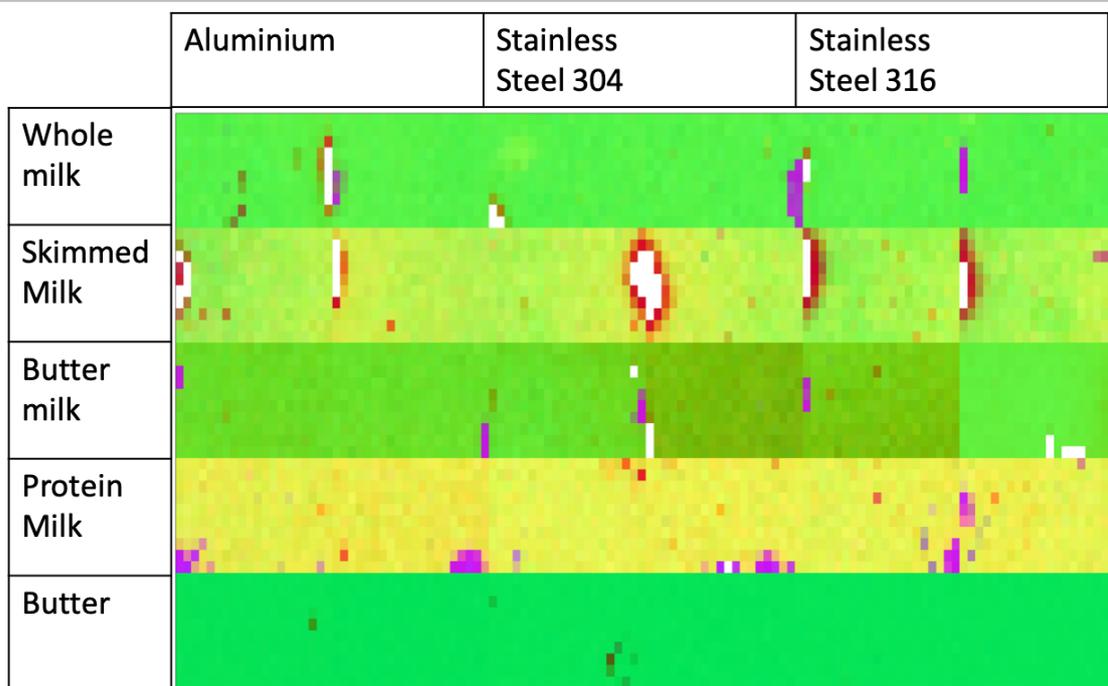


Figure 6. False RGB (red, green and blue) image built with score values of PC1, PC2 and PC3 (independently normalised between 0 and 1).

the butter milk (89.19%) where some pixels were classified as whole milk. This also affects the specificity of whole milk (96.67%), which was the lowest of all the samples. One possible explanation is that whole milk and butter milk are very similar in carbohydrates and fat composition. Whole milk has a slightly higher sensitivity than butter milk, 90%, and some pixels are classified as skimmed milk. In this case, it is necessary to consider that whole and skimmed milk have a comparable quantity of proteins and carbohydrates. The misclassification between skimmed and protein milk can be explained by the nutrients plot shown in Figure 1, where it is clear that they present the same carbohydrates/fats ratio. Butter is totally distinguishable from the other dairy products with a sensitivity of 100% and specificity of 99.96%. When representing each image as an object, the model seems to correctly predict all the dairy products (Table 4).

### FT-IR analysis

In Figure 8, the raw FT-IR spectra are plotted of both repetitions (first blue, second red), separated by type of surfaces and dairy product. Some plots (aluminium: butter and butter milk, stainless steel 304: whole milk and butter, stainless steel 304: whole milk) in Figure 8 present different baseline heights attributable to the height of the gold signal used for calibration. The spectra were split into two ranges (MIR,  $3580\text{--}920\text{ cm}^{-1}$ , and NIR,  $6200\text{--}3700\text{ cm}^{-1}$ , see black lines in the plots) and were pre-treated independently. As described in the materials and methods section, SG (window size = 15 points, second-order polynomial and first derivative and) was applied to improve the signal-to-noise ratio and to emphasise peak features. SNV was applied to remove the multiplicative and additive effects. Before the spectra were concatenated back together, they

Table 3. Values of PLS-DA analysis. It reports the values of sensitivity and specificity for each class based on a pixel-wise approach. Model built with Raman data pre-treated with a SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation. LVs = 14 and 92.47% CC.

Pixel level % CC = 92.47	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter
% Sensitivity	90.00	90.55	89.19	96.66	100
% Specificity	96.67	97.74	99.20	98.06	99.96

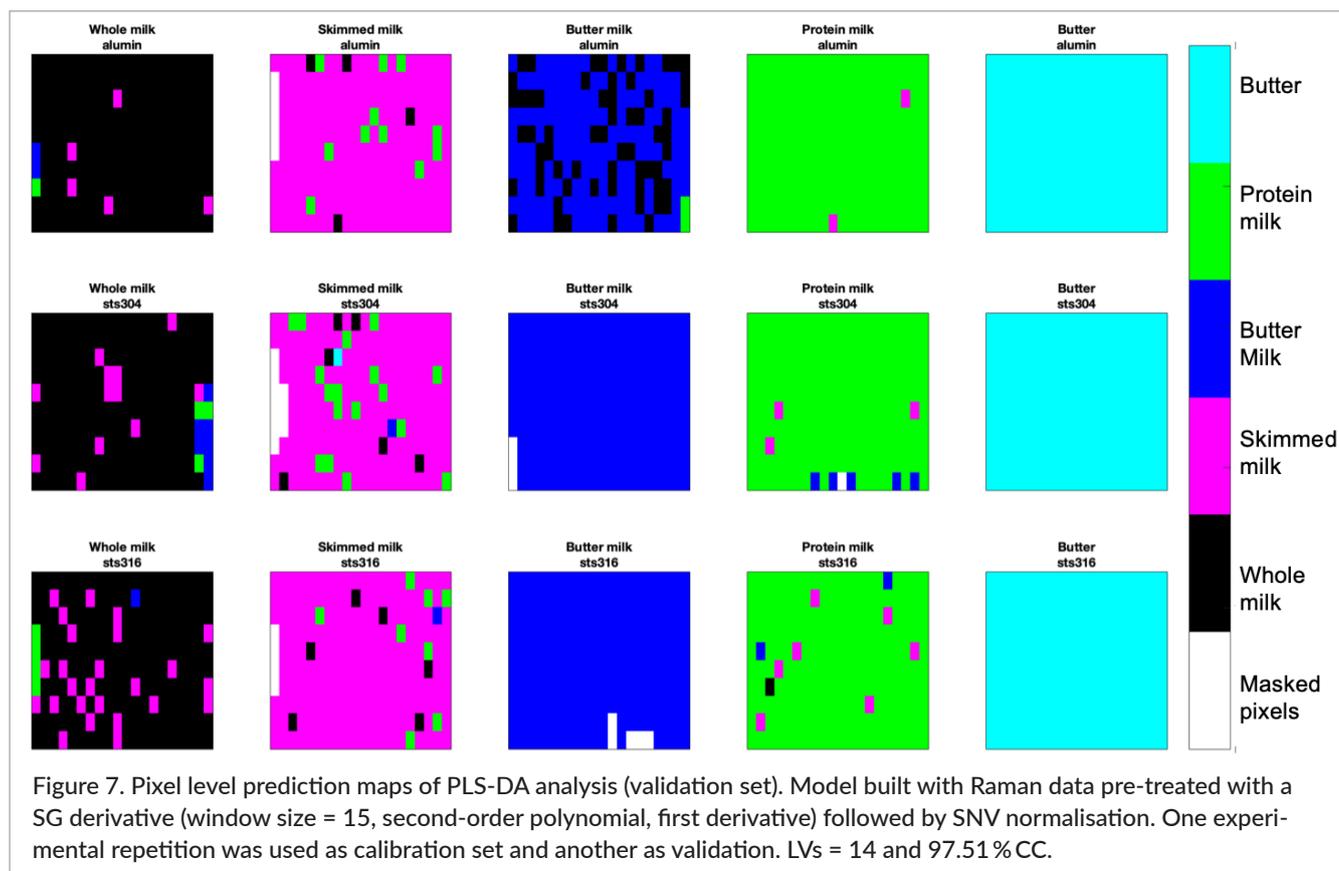


Table 4. Object identification of PLS-DA analysis. Model built with Raman data pre-treated with a SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation. LVs = 14 and 97.51% CC.

Object level	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter
Object identification	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter

were normalised using the Euclidean norm. As can be seen in Figure 9, this selection of pre-treatments eliminated the scattering between the two repetitions and showed a similar trend.

Observing the mean of pre-treated spectra separated by the different surfaces (Figure 10), it appears that in FT-IR the substrates also did not have a clear effect on the dairy residue spectra.

Similar to the approach carried out using the Raman data, PCA was applied to the pre-treated images (as described in the materials and methods section). A PCA score plot of PC1 vs PC2 vs PC3 and the corresponding loading plots for these PCs are shown in Figure 11. In addition, a false colour RGB image (Figure 12) was built as described previously for the Raman data. In this case no pixels were masked.

PCA analysis on the FT-IR data seems to give a better representation than the one carried out on Raman profiles, in that it allowed easy identification of the different types of dairy products from both the score plot in three-dimensional space and the false RGB image built with PC1, PC2 and PC3 (Figures 11–12).

In FT-IR PCA analysis, PC3 seems to contribute to the differentiation. However, it is difficult to attribute a specific constituent to each PC. The information in the NIR range seems to have more of an effect on PC1 and PC2 than on PC3. PC3 better represents the MIR range.

The range between  $1200\text{cm}^{-1}$  and  $920\text{cm}^{-1}$ , typical of carbohydrates, affects PC2 (explained variance 24.93%) and PC3 (explained variance 12.39%) more than PC1 (explained variance 42.29%). The protein bands between  $1800\text{cm}^{-1}$  and  $1400\text{cm}^{-1}$ , i.e. amide signals,

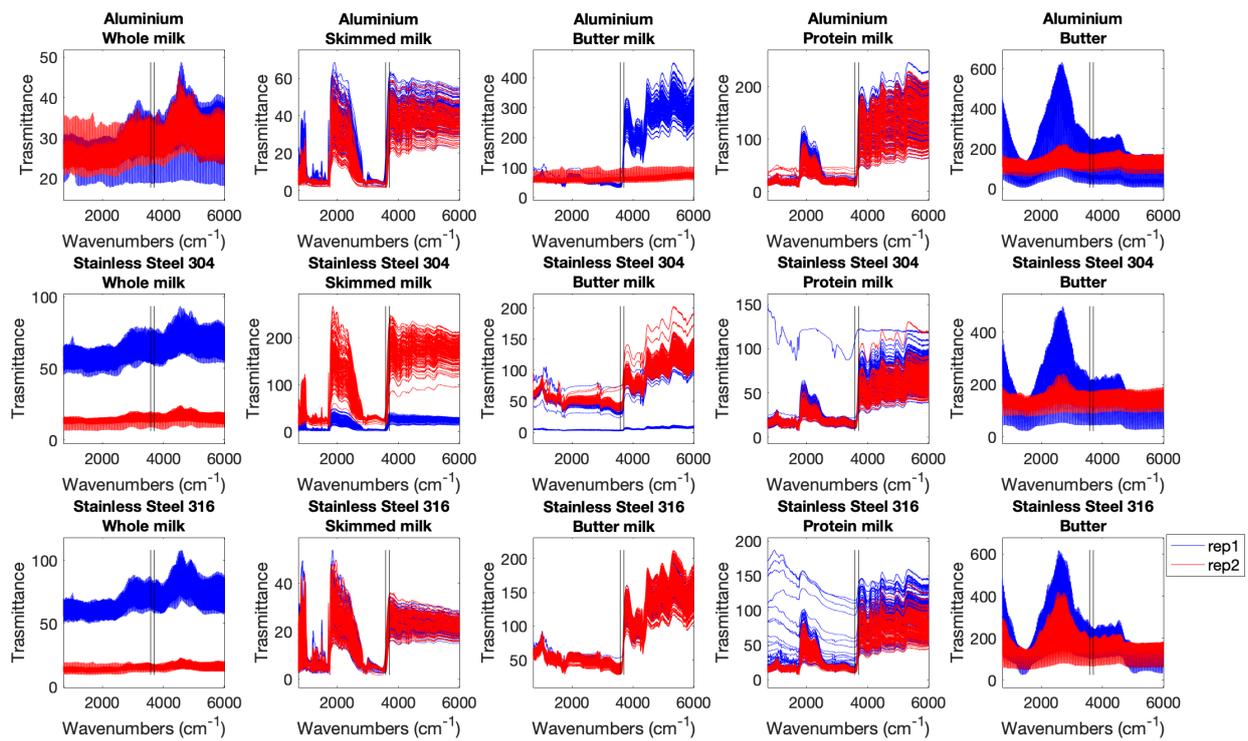


Figure 8. FT-IR raw spectra of both repetitions (first blue, second red) of whole milk, skimmed milk, butter milk, protein milk and butter on stainless steel 316-2B, stainless steel 304-2B and aluminium.

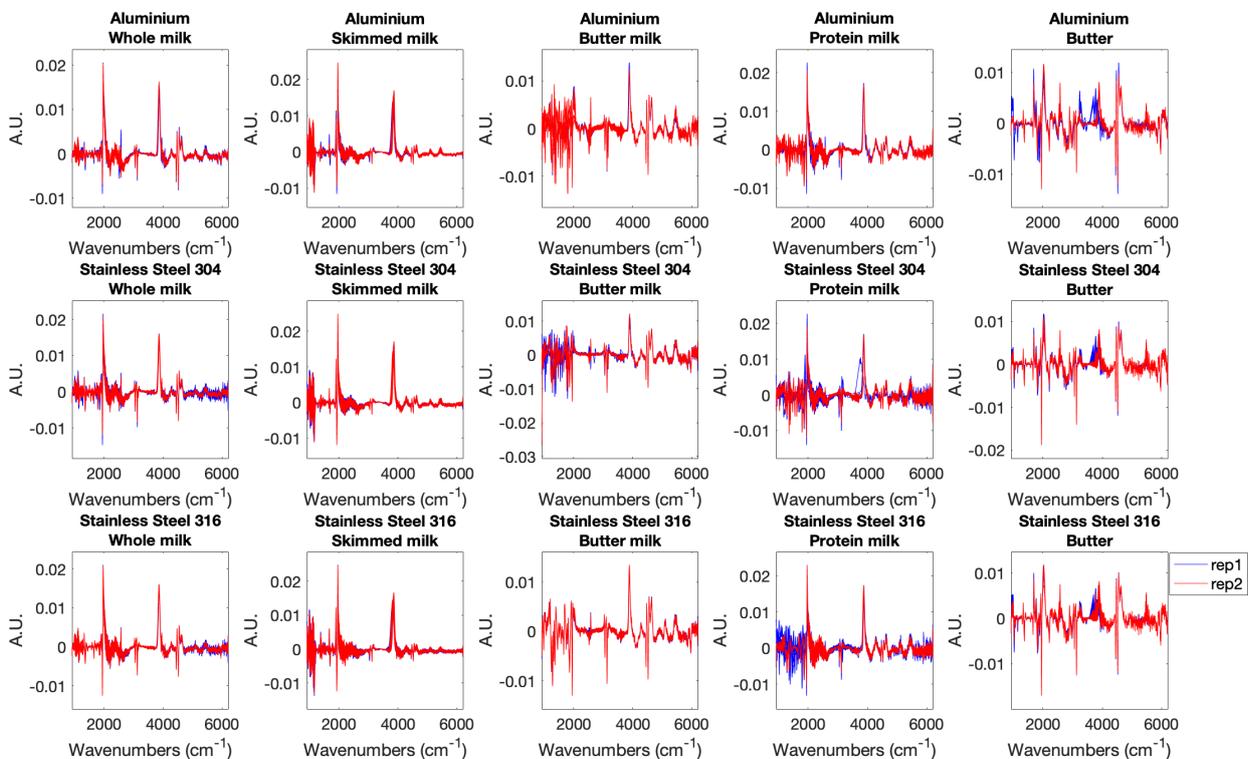
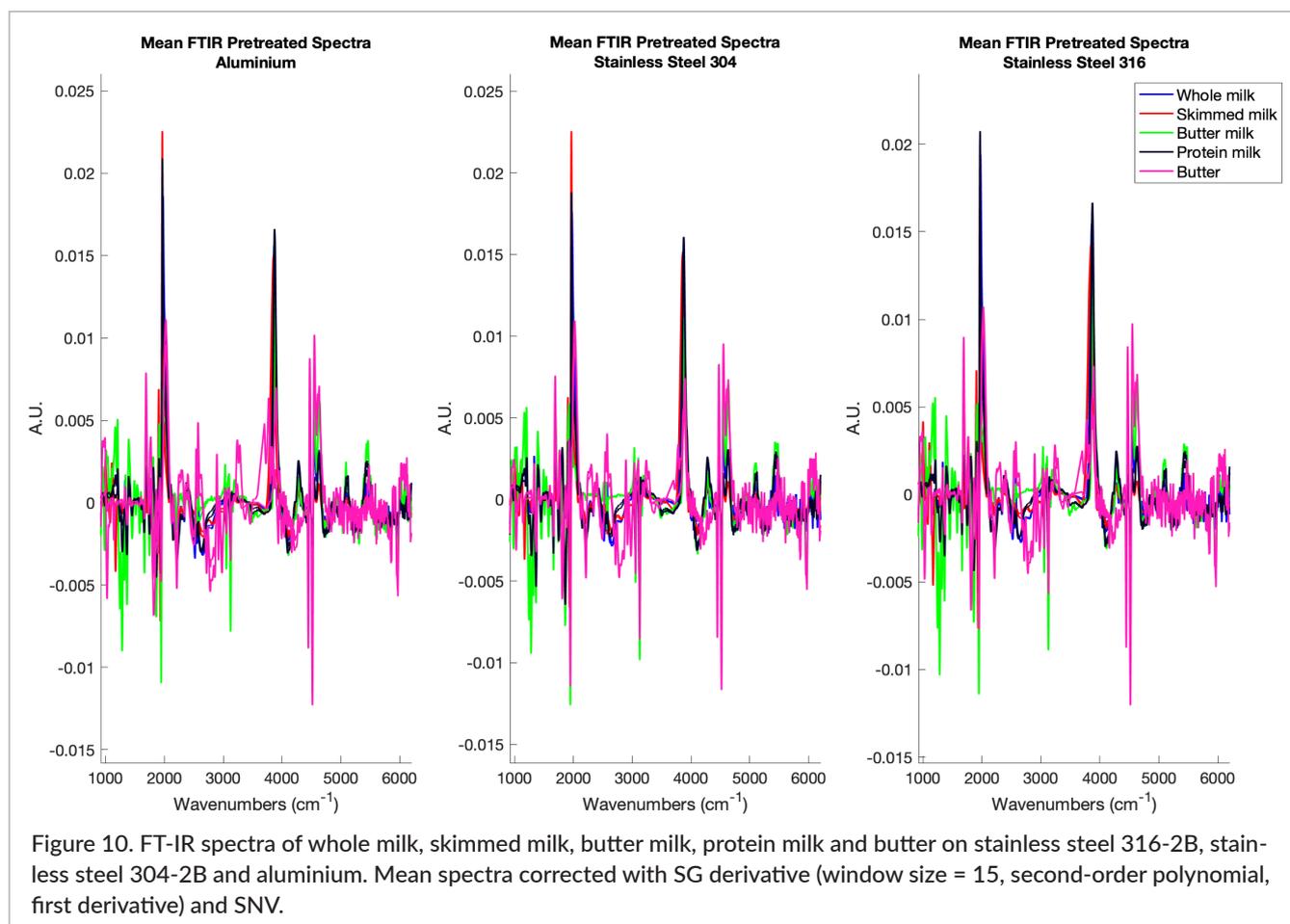


Figure 9. FT-IR spectra of both repetitions (first blue, second red) of whole milk, skimmed milk, butter milk, protein milk and butter on stainless steel 316-2B, stainless steel 304-2B and aluminium pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) and SNV normalisation.



are present in both PC1 and PC2. The clear signal of CH for both lipids and carbohydrates ( $3100\text{--}2800\text{cm}^{-1}$ ) observed in the Raman loadings (see Figure 5) has a lower influence here: just in PC1 and PC2. The band around  $3830\text{--}3840\text{cm}^{-1}$ , present in all three loadings, is in the range ( $4000\text{--}3200\text{cm}^{-1}$ ) assigned to both NH (amine) and OH (alcohol) groups. The signal of CH due to aromatic and aliphatic groups seems to affect only PC2 ( $4500\text{--}4000\text{cm}^{-1}$ ). From this analysis is evident that it is difficult to define the contribution of specific milk components in each PC. Indeed, all of the false RGB images in Figure 12 are obtained from the contribution of the three different PCs in combination. For example, butter is described by an equal contribution of PC3 (blue) and PC1 (red): magenta is obtained from the mixing of red and blue in equal parts.

In the score plot in Figure 10, high score values between PC1 and PC2 define butter milk. The score values of skimmed milk, butter and butter milk, whole milk and protein milk appear as separated clusters. Skimmed milk and protein milk maintained a similar

trend, but they are well defined in three-dimensional space.

In Figure 12, it is possible to see some intensity trends probably due to the surfaces, especially in the case of skimmed milk. However, when trying to plot the score values based on surfaces, no significant trend appeared (Appendix Figure A2). Therefore, despite some spatial trends which appear to be detectable, aluminium, stainless steel 304-2B and 316-2B do not affect the separation in PCA.

Three PLS-DA models were built and compared: independently pre-treated NIR and MIR ranges, and their combination. For the FT-IR model on the combined range, five LVs were selected and the total correct classification by pixel is 98.73% (Table 5). The combination of MIR and NIR provides better results than using the two ranges independently. In fact, for the MIR range five LVs (Table 7) were used, obtaining correct classification by pixel of 97.41%. In this case, only whole milk has a sensitivity of 100%, in fact it is the only dairy product always predicted correctly. In addition, only butter and butter

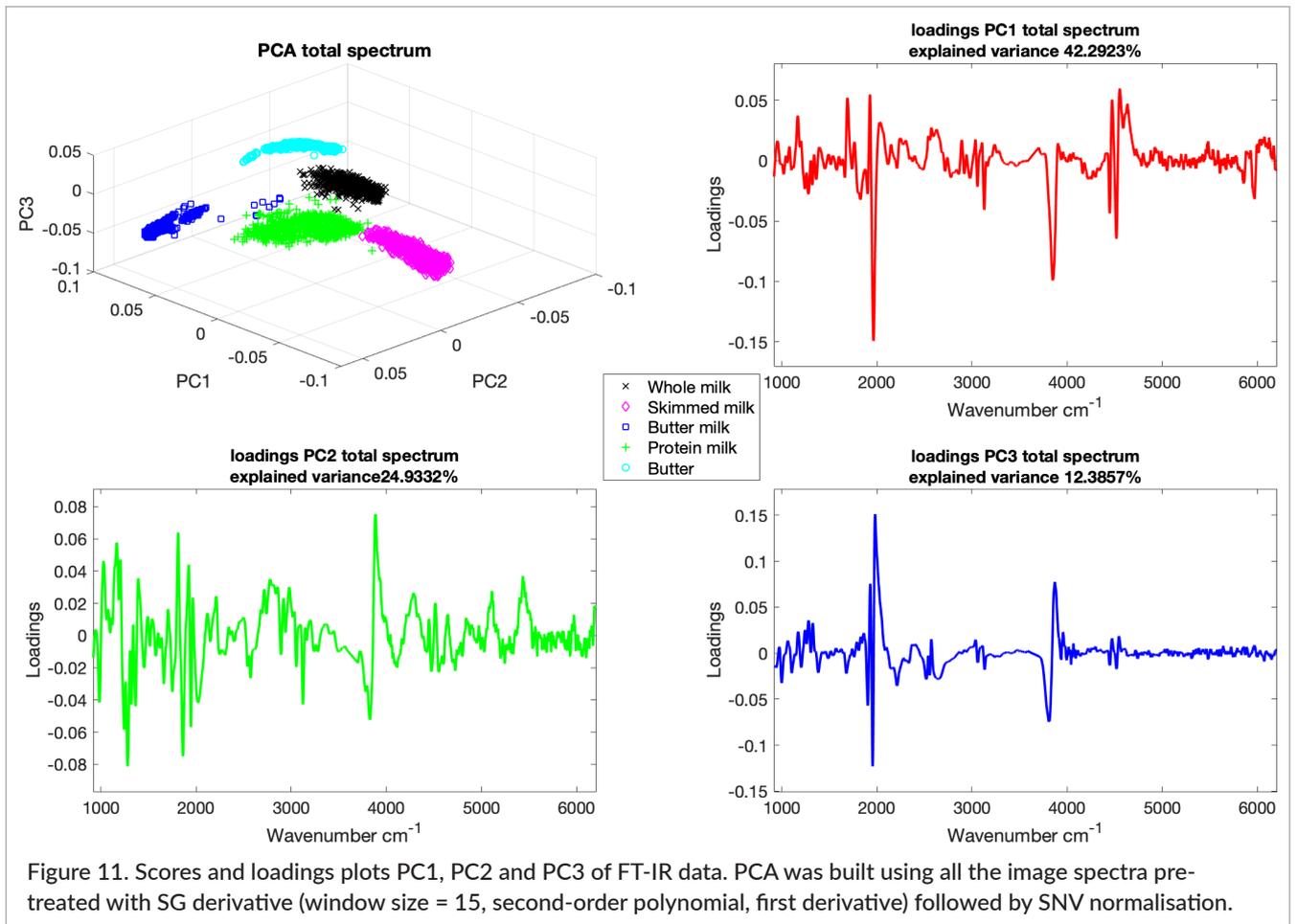


Figure 11. Scores and loadings plots PC1, PC2 and PC3 of FT-IR data. PCA was built using all the image spectra pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation.

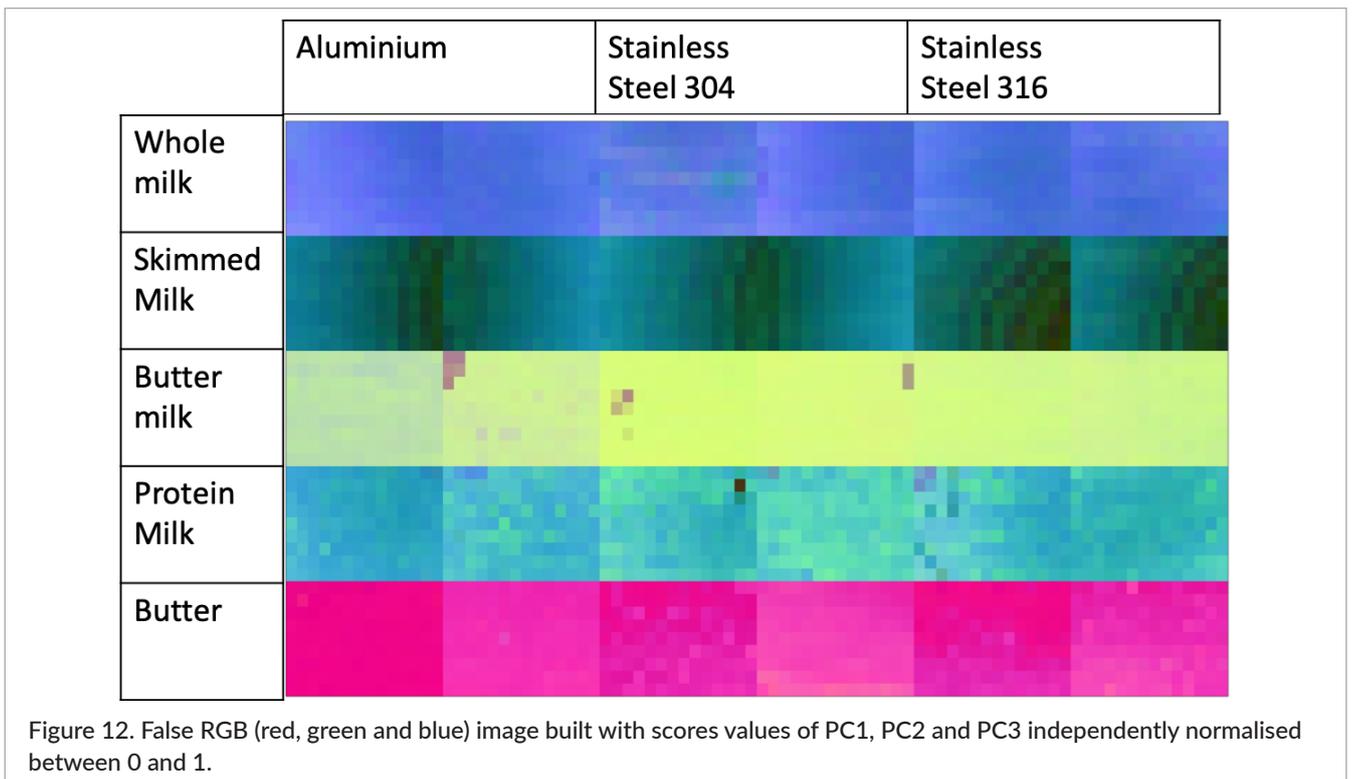


Figure 12. False RGB (red, green and blue) image built with scores values of PC1, PC2 and PC3 independently normalised between 0 and 1.

milk are never falsely predicted with a specificity of 100% (maps in Appendix, Figure A3). In the NIR range, six LVs were necessary to minimise the misclassification with an overall correct classification rate of 95.56% (Table 9). The corresponding model correctly predicts butter, protein milk and skimmed milk. Some pixels in whole milk are predicted as skimmed milk and some pixels in butter milk are predicted as whole milk (maps in appendix, Figure A4). The combination of the two ranges greatly improved the model: in fact, whole milk, protein milk and butter have

a sensitivity of 100% and whole milk, butter milk and butter have a specificity of 100% (Table 5). Following the object interpretation (Tables 6, 8 and 10), all the models correctly predict all the dairy products.

Comparing the total FT-IR range with the Raman model, the misclassifications between skimmed and protein milk are similar: due to having the same carbohydrates and fat ratio. Excluding these few pixels the total FT-IR range model seems more robust and has higher values in both sensitivity and specificity.

**Table 5.** Values of PLS-DA analysis. It reports the values of sensitivity and specificity for each class based on pixels approach. Model built with FT-IR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 5 and 98.73% CC.

<b>TOTAL RANGE: Pixel level % CC = 98.73</b>	<b>Whole milk</b>	<b>Skimmed milk</b>	<b>Butter milk</b>	<b>Protein milk</b>	<b>Butter</b>
% Sensitivity	100	94.18	99.47	100	100
% Specificity	100	99.87	100	98.54	100

**Table 6.** Object prediction of PLS-DA analysis. Model built with FT-IR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed with SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 5.

<b>TOTAL RANGE: Object level</b>	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter
<b>Object identification</b>	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter

**Table 7.** Values of PLS-DA analysis. It reports the values of sensitivity and specificity for each class using a pixel-wise approach. Model built with FT-IR-MIR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) and SNV. One repetition was used as calibration set and another as validation set. LVs = 5 and 97.41% CC.

<b>MIR RANGE: Pixel level % CC = 97.41</b>	<b>Whole milk</b>	<b>Skimmed milk</b>	<b>Butter milk</b>	<b>Protein milk</b>	<b>Butter</b>
% Sensitivity	83.07	100	94.71	100	100
% Specificity	98.68	95.77	100	100	100

**Table 8.** Object prediction of PLS-DA analysis. It reports the values of sensitivity and specificity for each class based on using a pixel-wise approach. Model built with FT-IR-MIR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 5.

<b>MIR RANGE: Object level</b>	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter
<b>Object identification (mode)</b>	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter

**Table 9.** Values of PLS-DA analysis. It reports the values of sensitivity and specificity for each class based on pixels approach. Model built with FT-IR-NIR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 5 and 95.56% CC.

<b>NIR RANGE: Pixel level % CC = 95.56</b>	<b>Whole milk</b>	<b>Skimmed milk</b>	<b>Butter milk</b>	<b>Protein milk</b>	<b>Butter</b>
% Sensitivity	100	92.86	98.15	99.74	96.30
% Specificity	99.07	99.47	100	98.21	100

Table 10. Object prediction of PLS-DA analysis. Model built with FT-IR-NIR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 5

NIR RANGE: Object level	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter
Object identification (mode)	Whole milk	Skimmed milk	Butter milk	Protein milk	Butter

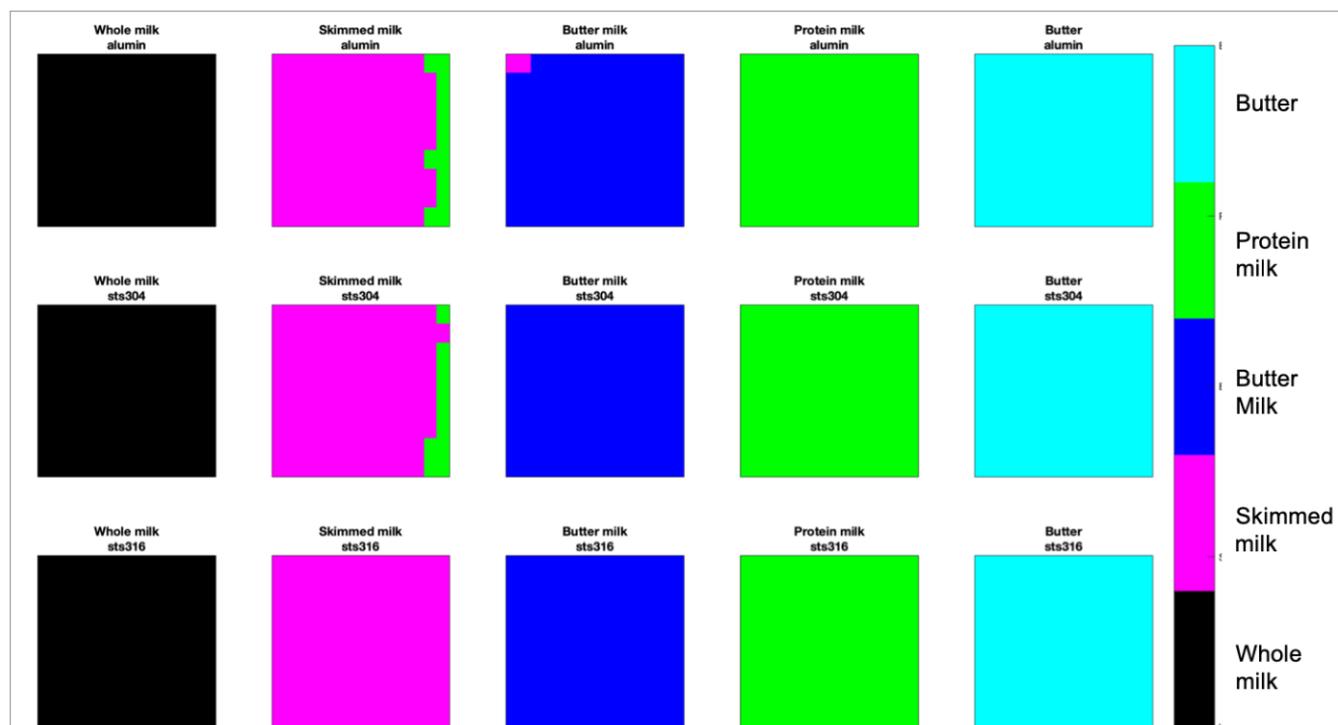


Figure 13. Prediction map of validation set of PLS-DA analysis. Model built with FT-IR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation. LVS = 5 and 99.58% CC.

## Conclusion

The objective of this study was to discriminate different types of dairy residues (whole milk, skimmed milk, butter milk, protein milk and butter) on surfaces (aluminium, stainless steel 304-2B and 316-2B) using FT-IR and Raman spectroscopy. The Raman spectra collected in two ranges were pre-treated independently. SG (second-order polynomial, first derivative) derivative was applied to remove the fluorescence signal that affected the spectra baseline. SNV removed the multiplicative and additive effects, and the spectra were successively concatenated. The FT-IR spectra were split into two ranges (MIR and NIR) and pre-treated independently. In this case, SG was applied to improve the signal-to-noise ratio and to emphasise peak features and SNV again to remove the multiplicative and additive effects. Before re-concatenating, they were normalised using the Euclidean norm. PCA and PLS-DA

were applied to these two data sets independently. PCA was applied to inspect for patterns in the data that might show separation between the different dairy products. For both Raman and FT-IR spectroscopy, PCA shows a good separation between the different dairy products. Despite that, observing the Raman spectra, it is easier to identify the signals of chemical groups and their effect on the loadings of PCA. In addition, PLS-DA was applied to classify each dairy product. The results of PLS-DA were evaluated from both a pixel-wise and object-wise perspective. Both spectroscopic techniques also give good results in prediction. However, Raman spectroscopy results are less precise at recognising different products. In fact, the PLS-DA model built using Raman data has 97.51% correct pixels in 14 LVs. On the other hand, FT-IR gives a better model with 98.73% of correctly classified pixels in 5 LVs.

In conclusion, from this study it is clear that both FT-IR and Raman spectroscopy, in combination with chemometrics approaches, are suitable to discriminate the various dairy product types.

One additional aim of this study was to also characterise the contributions due to the surfaces. In this respect, no influence due to the substrates was observable. This is probably due to the thickness of the sample (approximately 0.15 mm). Accordingly, testing different sample thicknesses to study the surface influence may be an interesting future investigation.

## Acknowledgement

Funding for this research was provided by Science Foundation Ireland (SFI) under the investigators programme Proposal ID 15/IA/2984—HyperMicroMacro.

## References

1. H.A. Mckenzie, "Milk proteins", *Adv. Protein Chem.* **22**, 55 (1967). [https://doi.org/10.1016/S0065-3233\(08\)60041-8](https://doi.org/10.1016/S0065-3233(08)60041-8)
2. S. Rafiq, N. Huma, I. Pasha, A. Sameen, O. Mukhtar and M.I. Khan, "Chemical composition, nitrogen fractions and amino acids profile of milk from different animal species", *Asian-Australasian J. Anim. Sci.* **29(7)**, 1022 (2016). <https://doi.org/10.5713/ajas.15.0452>
3. T.M.P. Cattaneo, G. Cabassi, M. Profaizer and R. Giangiacomo, "Contribution of light scattering to near infrared absorption in milk", *J. Near Infrared Spectrosc.* **17(6)**, 337 (2009). <https://doi.org/10.1255/jnirs.867>
4. P.M. Santos, E.R. Pereira-Filho and L.E. Rodriguez-Saona, "Application of hand-held and portable infrared spectrometers in bovine milk analysis", *J. Agric. Food Chem.* **61(6)**, 1205 (2013). <https://doi.org/10.1021/jf303814g>
5. B. Aernouts, E. Polshin, W. Saeys and J. Lammertyn, "Mid-infrared spectrometry of milk for dairy metabolomics: a comparison of two sampling techniques and effect of homogenization", *Anal. Chim. Acta* **705(1-2)**, 88 (2011). <https://doi.org/10.1016/j.aca.2011.04.018>
6. V. Raikos, "Effect of heat treatment on milk protein functionality at emulsion interfaces. a review", *Food Hydrocoll.* **24(4)**, 259 (2010). <https://doi.org/10.1016/j.foodhyd.2009.10.014>
7. A. Kher, P. Udabage, I. McKinnon, D. McNaughton and M.A. Augustin, "FTIR investigation of spray-dried milk protein concentrate powders", *Vibr. Spectrosc.* **44(2)**, 375 (2007). <https://doi.org/10.1016/j.vibspec.2007.03.006>
8. J. Król, Z. Litwińczuk, A. Brodziak and J. Barłowska, "Lactoferrin, lysozyme and immunoglobulin G content in milk of four breeds of cows managed under intensive production system.", *Pol. J. Vet. Sci.* **13(2)**, 357 (2010). <https://www.ncbi.nlm.nih.gov/pub-med/20731193>
9. M.L. Bainbridge, L.M. Cersosimo, A.-D.G. Wright and J. Kraft, "Content and composition of branched-chain fatty acids in bovine milk are affected by lactation stage and breed of dairy cow", *PLoS One* **11(3)**, e0150386 (2016). <https://doi.org/10.1371/journal.pone.0150386>
10. W.M. Stoop, H. Bovenhuis, J.M.L. Heck and J.A.M. van Arendonk, "Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows", *J. Dairy Sci.* **92(4)**, 1469 (2009). <https://doi.org/10.3168/jds.2008-1468>
11. L. Pellegrino and A. Tirelli, "A sensitive HPLC method to detect hen's egg white lysozyme in milk and dairy products", *Int. Dairy J.* **10(7)**, 435 (2000). [https://doi.org/10.1016/S0958-6946\(00\)00065-0](https://doi.org/10.1016/S0958-6946(00)00065-0)
12. V. Bonfatti, L. Grigoletto, A. Cecchinato, L. Gallo and P. Carnier, "Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants", *J. Chromatogr. A* **1195(1-2)**, 101 (2008). <https://doi.org/10.1016/j.chroma.2008.04.075>
13. G. Bordin, F.C. Raposo, B. de La Calle and A.R. Rodriguez, "Identification and quantification of major bovine milk proteins by liquid chromatography", *J. Chromatogr. A* **928(1)**, 63 (2001). [https://doi.org/10.1016/S0021-9673\(01\)01097-4](https://doi.org/10.1016/S0021-9673(01)01097-4)
14. E. Heftmann, *Chromatography: Fundamentals and Applications of Chromatography and Related Differential Migration Methods. Part A, Fundamentals and Techniques*. Elsevier (2004).
15. S. Mazurek, R. Szostak, T. Czaja and A. Zachwieja, "Analysis of milk by FT-Raman spectroscopy", *Talanta*

- 138, 285 (2015). <https://doi.org/10.1016/j.talanta.2015.03.024>
16. C.M. McGoverin, A.S.S. Clark, S.E. Holroyd and K.C. Gordon, "Raman spectroscopic quantification of milk powder constituents", *Anal. Chim. Acta* **673(1)**, 26 (2010). <https://doi.org/10.1016/j.aca.2010.05.014>
17. M.R. Almeida, K.D.S. Oliveira, R. Stephani and L.F.C. De Oliveira, "Fourier-transform Raman analysis of milk powder: a potential method for rapid quality screening", *J. Raman Spectrosc.* **42(7)**, 1548 (2011). <https://doi.org/10.1002/jrs.2893>
18. F.A. Iñón, S. Garrigues and M. De La Guardia, "Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques", *Anal. Chim. Acta* **513(2)**, 401 (2004). <https://doi.org/10.1016/j.aca.2004.03.014>
19. D. Lefier, R. Grappin and S. Pochet, "Determination of fat, protein, and lactose in raw milk by Fourier transform infrared spectroscopy and by analysis with a conventional filter-based milk analyzer", *J. AOAC Int.* **79(3)**, 711 (1996).
20. P.M. Santos, E.R. Pereira-Filho and L.E. Rodriguez-Saona, "Application of hand-held and portable infrared spectrometers in bovine milk analysis", *J. Agric. Food Chem.* **61(6)**, 1205 (2013). <https://doi.org/10.1021/jf303814g>
21. S. Šašić and Y. Ozaki, "Short-wave near-infrared spectroscopy of biological fluids. 1. quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment", *Anal. Chem.* **73(1)**, 64 (2001). <https://doi.org/10.1021/ac000469c>
22. C. Malegori, L. Franzetti, R. Guidetti, E. Casiraghi and R. Rossi, "GLCM, an image analysis technique for early detection of biofilm", *J. Food Eng.* **185**, 48 (2016). <https://doi.org/10.1016/j.jfoodeng.2016.04.001>
23. L.M. Barnes, M.F. Lo, M.R. Adams and A.H. Chamberlain, "Effect of milk proteins on adhesion of bacteria to stainless steel surfaces.", *Appl. Environ. Microbiol.* **65(10)**, 4543 (1999).
24. H.L.M. Lelieveld, J.T. Holah and D. Napper (Eds), *Hygiene in Food Processing: Principles and Practice*. Woodhead Publishing/Elsevier (2014).
25. C. Jullien, T. Bénézech, B. Carpentier, V. Lebret and C. Faille, "Identification of surface characteristics relevant to the hygienic status of stainless steel for the food industry", *J. Food Eng.* **56(1)**, 77 (2003). [https://doi.org/10.1016/S0260-8774\(02\)00150-4](https://doi.org/10.1016/S0260-8774(02)00150-4)
26. J.J. Ojeda, M.E. Romero-González and S.A. Banwart, "Analysis of bacteria on steel surfaces using reflectance micro-Fourier transform infrared spectroscopy", *Anal. Chem.* **81(15)**, 6467 (2009). <https://doi.org/10.1021/ac900841c>
27. "Products | avonmore". [Online at <https://www.avonmore.ie/products>]. [Accessed: 23 October 2018].
28. S. Piqueras, L. Duponchel, R. Tauler and A. de Juan, "Monitoring polymorphic transformations by using *in situ* Raman hyperspectral imaging and image multi-set analysis", *Anal. Chim. Acta* **819**, 15 (2014). <https://doi.org/10.1016/J.ACA.2014.02.027>
29. S. André, L. Saint Cristau, S. Gaillard, O. Devos, É. Calvosa and L. Duponchel, "In-line and real-time prediction of recombinant antibody titer by *in situ* Raman spectroscopy", *Anal. Chim. Acta* **892**, 148 (2015). <https://doi.org/10.1016/J.ACA.2015.08.050>
30. A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.", *Anal. Chem.* **36(8)**, 1627 (1964). <https://doi.org/10.1021/ac60214a047>
31. R.J. Barnes, M.S. Dhanoa and S.J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", *Appl. Spectrosc.* **43(5)**, 772 (1989). <https://doi.org/10.1366/0003702894202201>
32. G. Socrates, *Infrared and Raman Characteristic Group Frequencies: Tables and Charts*. John Wiley & Sons (2007).
33. P. Di Mascio, M.E. Murphy and H. Sies, "Antioxidant defense systems: the role of carotenoids, tocopherols, and thiols", *Am. J. Clin. Nutr.* **53(1)**, 194S (1991). <https://doi.org/10.1093/ajcn/53.1.194S>
34. Q. Zhou, S.-Q. Sun, L. Yu, C.-H. Xu and X.-R. Zhang, "Sequential changes of main components in different kinds of milk powders using two-dimensional infrared correlation analysis", *J. Mol. Struct.* **799(1-3)**, 77 (2006). <https://doi.org/10.1016/J.MOLSTRUC.2006.03.025>
35. P.F. Fox, T. Uniacke-Lowe, P.L.H. McSweeney and J.A. O'Mahony, *Dairy Chemistry and Biochemistry*. Springer International Publishing (2015).
36. H. Schulz, M. Baranska and R. Baranski, "Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis", *Biopolymers* **77(4)**, 212 (2005). <https://doi.org/10.1002/bip.20215>

## Appendix

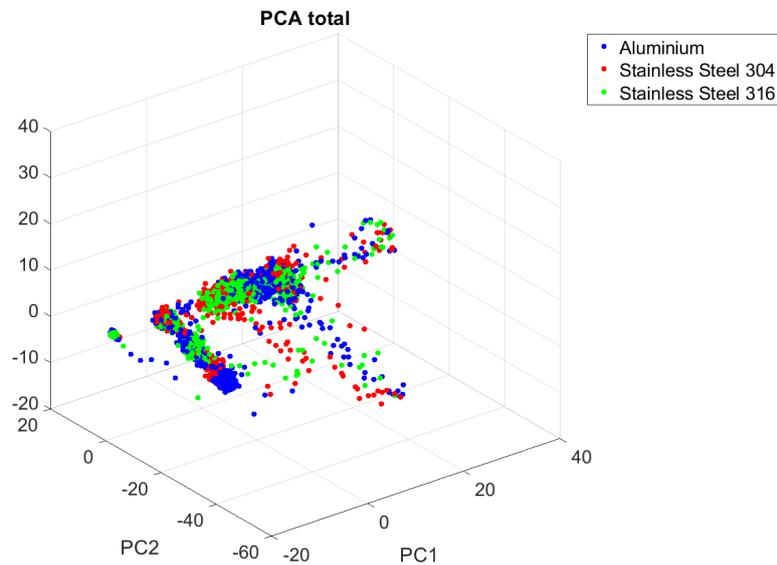


Figure A1. Scores plots PC1, PC2 and PC3 of Raman data based on surfaces. PCA was built using all the image spectra pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation.

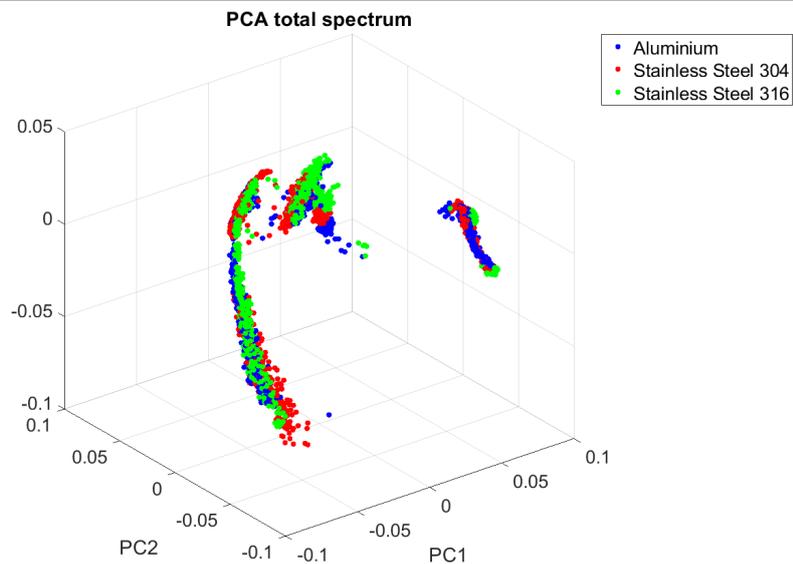
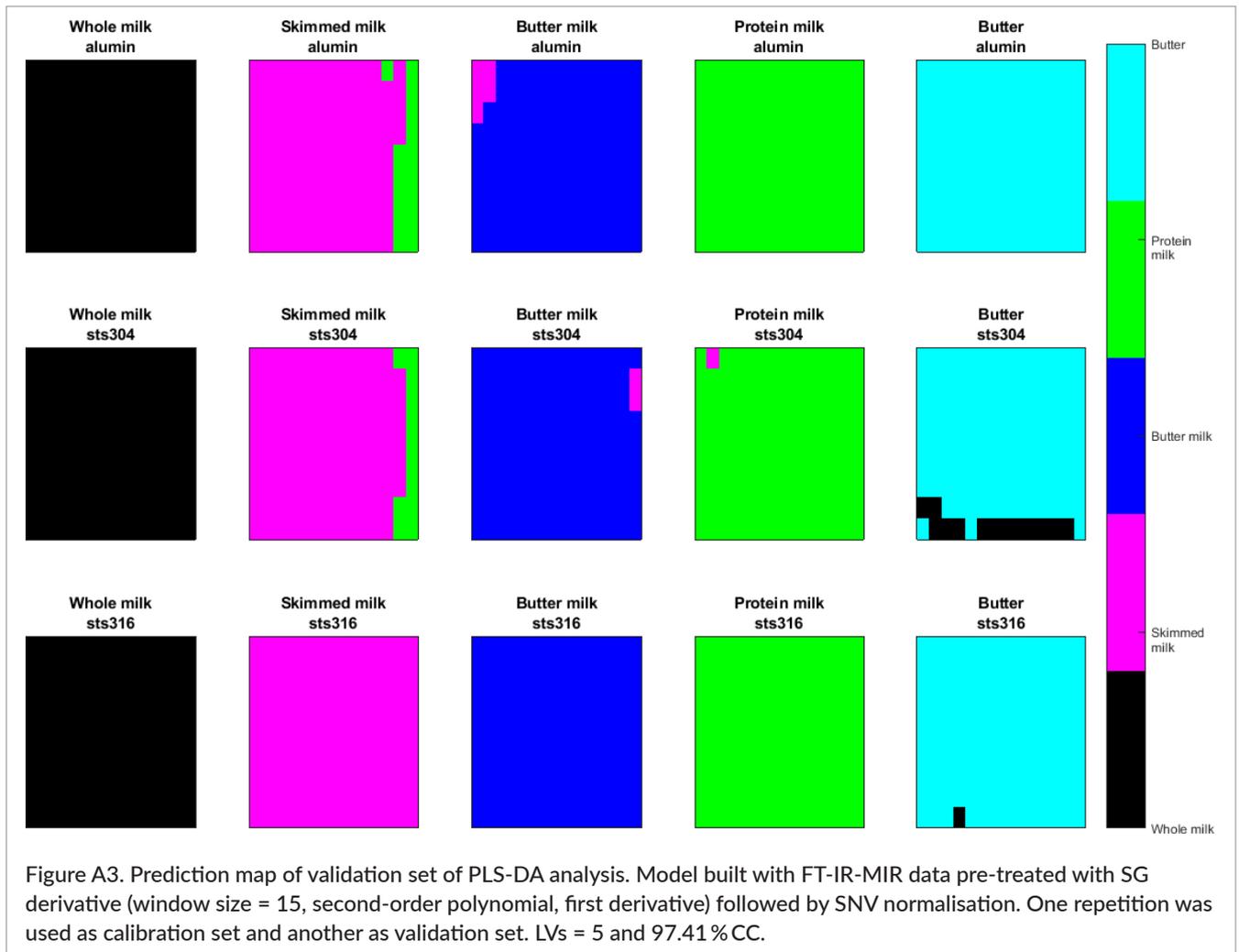


Figure A2. Scores plots PC1, PC2 and PC3 of FT-IR data based on surfaces. PCA was built using all the image spectra pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation.



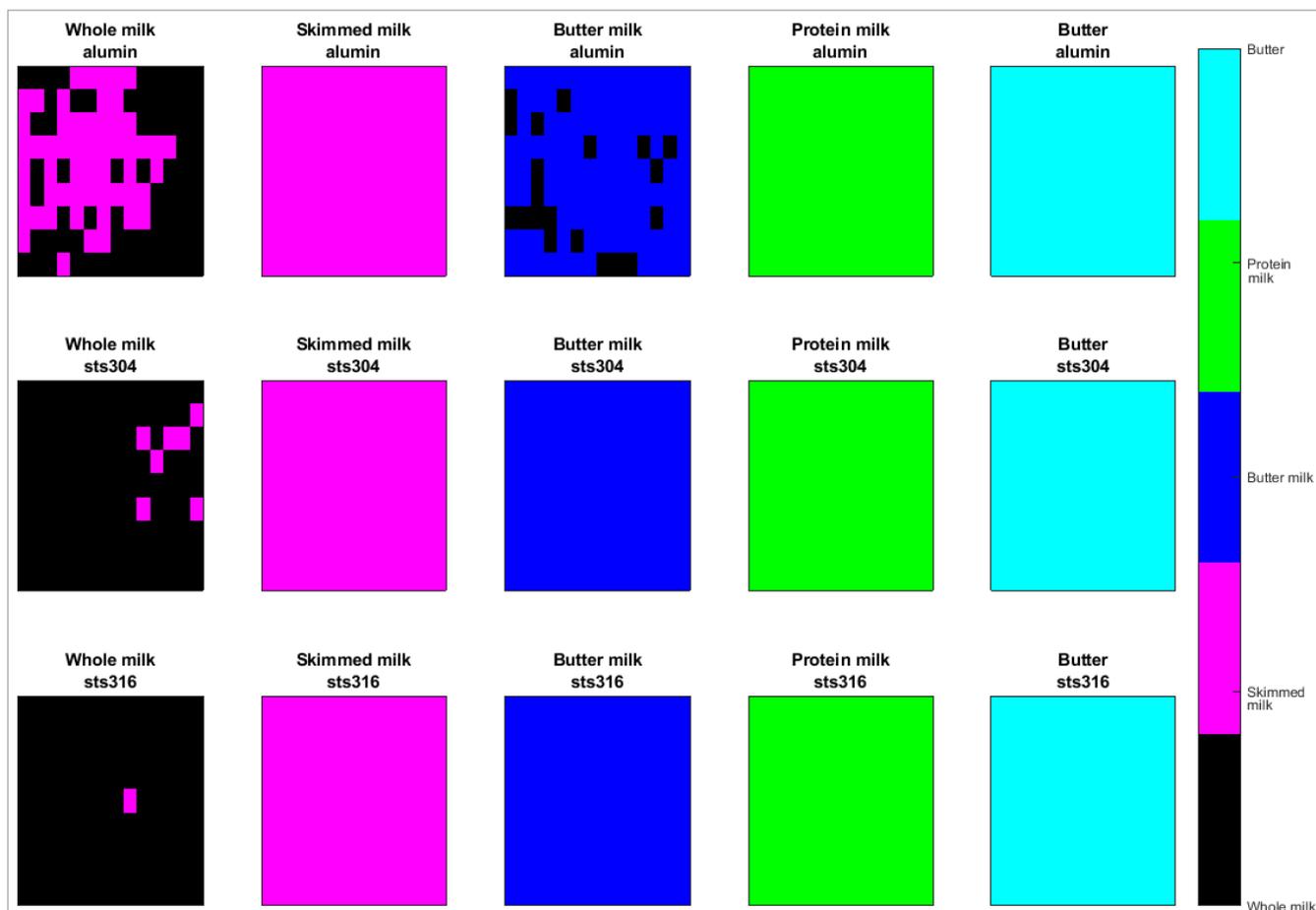


Figure A4. Prediction map of validation set of PLS-DA analysis. Model built with FT-IR-NIR data pre-treated with SG derivative (window size = 15, second-order polynomial, first derivative) followed by SNV normalisation. One repetition was used as calibration set and another as validation set. LVs = 6 and 95.56% CC.