FAST EXPLORATION AND CLASSIFICATION OF LARGE HYPERSPECTRAL IMAGE DATASETS FOR EARLY BRUISE DETECTION ON APPLES

Carlotta Ferrari, Giorgia Foca, Rosalba Calvini, Alessandro Ulrici*

Department of Life Sciences and Interdepartmental Research Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia

* Corresponding author: Alessandro Ulrici, Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola 2, 42122 Reggio Emilia, Italy. Phone: +39 0522 522043. Fax: +39 0522 522027. E-mail: alessandro.ulrici@unimore.it

Abstract

Hyperspectral imaging allows to easily acquire tens of thousands of spectra for a single sample in few seconds; though valuable, this data-richness poses many problems due to the difficulty of handling a representative amount of samples altogether. For this reason, we recently proposed an approach based on the idea of reducing each image into a one-dimensional signal, named *hyperspectrogram*, which accounts both for spatial and for spectral information. In this manner, a dataset of hyperspectral images can be easily and quickly converted into a set of signals (2D data matrix), which in turn can be analyzed using classical chemometric techniques. In this work, the hyperspectrograms obtained from a dataset of 800 NIR-hyperspectral images of two different apple varieties were used to discriminate bruised from sound apples using iPLS-DA as variable selection algorithm, which allowed to efficiently detect the presence of bruises. Moreover, the reconstruction as images of the selected variables confirmed that the automated procedure led to the exact identification of the spatial features related to the onset and to the subsequent evolution with time of the bruise defect.

Keywords

Hyperspectral Imaging; Data reduction; Fast Exploration; NIR spectroscopy; Apples; Early bruise detection.

1. Introduction

The increasing normative severity as well as market competitiveness have led the food industry to constantly ask for the improvement of products and process monitoring systems. In the context of fast, non-destructive and reliable techniques, image analysis-based methods have gained particular interest, thanks to their ability to spatially characterize heterogeneous samples. In particular, HyperSpectral Imaging (HSI), also known as chemical or spectroscopic imaging, represents an emerging technique that provides both spatial information of imaging systems and spectral information of spectroscopy [1]. HSI in fact allows to acquire hundreds of congruent images recorded at different wavelengths and stacked to each other, forming a three-dimensional (3D) matrix, or hypercube, which consists of one spectral (λ) and two spatial (x and y) dimensions. HSI maintains therefore the same advantages of spectroscopic methods, i.e., it is fast, non-destructive and it does not require the use of chemicals; additionally, it allows to collect spectral data not only from a single point, but at each pixel of an image, enabling the visualization of the chemical composition of the sample surface.

Notwithstanding the great potentialities offered by this technique, which have led to a constant increase of the number of applications in the field of food industry [2, 3], there is still an open issue of hyperspectral imaging which merits utmost attention: hypercubes with very large file sizes (frequently greater than 50 MB) and composed of an extremely high number of spectra (generally tens of thousands) can be easily acquired in few seconds. Even if on the one hand this *data-richness* represents the main advantage of HSI, on the other hand it also requires improved data handling strategies. In fact, large data sizes imply long computational times and high computational loads, complicating therefore the development of efficient and fast applications.

This issue, also referred to as *curse* of dimensionality [4, 5], often limits the possibility to easily evaluate a sufficiently large number of hyperspectral images altogether, i.e., to obtain a fast overview of the whole dataset, allowing for the simultaneous comparison of all the available hyperspectral images. For this reason, when it is necessary to consider a representative number of

samples so as to account for the large inter-sample variability that often occurs, e.g., in the case of food industry, usually a Region of Interest (ROI) is used to represent the whole image, from which the average spectrum or a limited number of representative spectra are extracted. In this case, however, the information extracted from the hyperspectral image is strictly dependent upon which ROI is considered, whose definition becomes therefore the critical step. The ROI selection step can be accomplished by manually defining the area of interest of each image, which results to be quite laborious, time consuming and affected by a high degree of subjectivity. Alternatively, ROI definition can be done by applying an automatic masking procedure, which needs to be developed ad-hoc on the basis of the *a priori* knowledge of the specific problem at hand.

A further drawback of limiting the analysis to a small region of the image is the impossibility to investigate the dataset of hyperspectral images as a whole, that is considering all the information conveyed by all the samples at the same time. The possibility to gain an overlook of the whole dataset of hyperspectral images could be in fact highly advantageous, for instance to detect possible system inconsistencies which are not strictly related to the specific problem of interest, as well as to easily identify particular patterns of samples or anomalous images.

In order to deal with datasets composed of a large number of hyperspectral images, our research group recently proposed an alternative approach [6], which is based on the idea of reducing each hyperspectral image into a signal, named *hyperspectrogram*, built in a way to convey both spatial and spectral information. This procedure is derived from the colourgrams approach, which was developed for the elaboration of RGB images [7–9]. By means of the hyperspectrograms approach, datasets composed of tens to hundreds of hypercubes can be therefore transformed into a matrix of signals, which in turn can be analyzed simply using multivariate analysis methods such as Principal Component Analysis (PCA), Partial Least Squares (PLS) and Partial Least Squares–Discriminant Analysis (PLS-DA). The main characteristic of this approach, whose main steps are depicted in Scheme 1, is that the relevant information brought by each hypercube is codified into a signal, which can be considered as a fingerprint of the relevant image data, independently of the

specific analyzed sample and/or of the problem to be faced. For this reason, hyperspectrograms can be used to easily investigate large datasets in their entirety, providing a complete overview of their structure and of the main sources of variations.

Moreover, the application to the hyperspectrograms of proper variable selection algorithms (like, for example, interval-PLS or Genetic Algorithms) allows one to identify the features which are most relevant for the solution of a specific calibration/classification problem. It must be underlined that in this case the feature selection step is performed automatically by the algorithm in a blind way, with no a priori assumptions on the basis of the nature of the considered food matrix and without the need to perform any (long and subjective) ROI definition step. The selected features can then be evaluated both as images and in the spectral domain, enabling a critical interpretation of the features which have been automatically identified by the multivariate analysis algorithms, only on the basis of their statistical significance.

The main aim of hyperspectrograms is therefore to allow the fast and simultaneous comparison of all the hyperspectral images of a newly acquired dataset, even when their number is very high. Generally, this is done using average (or median) spectra, which are very effective in the case of homogeneous samples, but do not perform properly when there is the need to identify spatially localized features within the image scene, since average spectra and median spectra do not account for spatial variability. Hyperspectrograms provide the user with an additional tool for fast exploration of datasets of hyperspectral images: the overview of the whole dataset, achievable e.g., by a PCA of the hyperspectrograms matrix, is made at the image level, and not a pixel level, but taking into account also spatial variability-related aspects of each image. The hyperspectrogram approach is therefore mainly meant as a tool that can be employed before performing more in-depth and ad-hoc analyses of selected groups of images at the pixel level by means of more refined pixel-oriented techniques such as Multivariate Image Analysis (MIA) or Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [10-14].

In the present work, the hyperspectrogram approach is used for the early detection of bruises of two different apple varieties. Mechanical damages which can occur during fruit harvest, handling and transport represent a major cause of economic losses, and bruising in particular is the most common type of postharvest mechanical injury [15]. In order to reduce these losses, automatic systems able to early detect bruised fruits are highly valuable, since visual assessment is always affected by operator subjectivity, and the human eye could have not a sufficient sensibility or consistency. Moreover, in the early stage bruises are frequently not visible with the naked eye, while they can be quite clearly detected by means of systems operating in the Near Infrared (NIR) spectral range. For these reasons, in the last decades spectroscopy has gained particular interest as a non-destructive and fast tool to discriminate bruised apple tissues [16-18]. However, the use of conventional NIR spectroscopic techniques can pose limitations to the effectiveness of defect detection, where spatial localization of the defect-related information represents a crucial point. Conversely, the use of HSI allows to solve successfully this issue, as it was already shown in the literature. Lu [19] developed a hyperspectral system working in the 900-1700 nm region, combining Principal Component transform (PC) and Minimum Noise Fraction transform (MNF) to enhance bruise features and to reduce data dimensionality. A number of works also considered the 400-1000 nm spectral range to develop multispectral systems able to detect bruises on apples through the selection of a limited number of effective wavelengths that were employed to build models based on PCA [20, 21], PLS-DA [21], PLS and stepwise discriminant analysis [22]. An extensive work has been recently presented by Baranowski et al. [23], where a wide dataset composed of 580 apples belonging to five different varieties was considered. This dataset was imaged in the visible/near-infrared (400–1000 nm) and in the short wavelength infrared (1000–2500 nm) ranges. Eight supervised classification methods (including support vector machines, linear logistic regression, neural networks and decision trees) were used and compared to check their effectiveness to estimate the number of days after bruising.

In this context, we present the possible advantages offered by the use of hyperspectrograms for apple bruise detection. In particular, the hyperspectrogram approach allowed us to analyze a very high number of hyperspectral images (800 on the whole) using for the first time to our knowledge a completely automated procedure, i.e., without the need to define any ROI and to design ad-hoc masking procedures, which are quite subjective, laborious and time consuming steps. Moreover, beyond the achievement of high classification performances in the detection of apple bruises also when the defect was not appreciable with the naked eye, a preliminary exploration of the hyperspectrograms dataset by means of PCA allowed us to estimate the general sources of variability of the dataset independently of the specific bruise detection task, confirming the versatility of the proposed approach.

2. Materials and methods

2.1. Apple samples and bruising

'*Golden Delicious*' and '*Pink Lady*' apple varieties were selected for the present study in order to consider samples characterized by different peel colors. In particular, 'Golden Delicious' is a variety where bruises can be quite easily identified with the naked eye, while 'Pink Lady' is a variety where, due to the red colored peel, bruises are more difficult to be detected by common RGB imaging techniques. For each variety, 40 apples were collected and thoroughly inspected in order to verify that they were free from visible bruise defects; 20 samples were then randomly selected to be artificially bruised, while the 20 remaining apples were used as control samples. Several preliminary tests have been performed in order to define the proper bruising conditions, in a way to avoid the creation of defects immediately detectable at the naked eye. Based on these tests, the final experimental setup consisted in a pendulum with a 50 cm long wooden arm and a 250 g plastic cylinder (30 mm in diameter), that was lifted to a height of 6 cm with respect to the sample. Unfortunately, a quantitative comparison of the impact provided by our device with that provided

by the devices used in other literature articles is not possible, due to the lack of sufficient experimental details. However it must be underlined that, considering the combination of cylinder weight and impact height, the bruises induced in the present work should be considerably less intense than those considered in other similar research works. In order to evaluate in a more objective manner this aspect, RGB images were acquired on bruised 'Golden Delicious' and 'Pink Lady' samples in JPEG format using a Kodak EasyShare DX7630 zoom digital camera with a spatial resolution of 2856 × 2142 pixels. In order to have constant and homogeneous lighting conditions, the camera was mounted on a white painted wooden box, containing the apple to be photographed, equipped with 8 tungsten lamps [9]. As a matter of fact, the elaboration through Multivariate Image Analysis (MIA) of the RGB images did not allow to identify the presence of bruises in the majority of the cases, at least in the first days after bruising (data not reported).

2.2. Hyperspectral images acquisition

In order to evaluate the development over time of the bruised areas, hyperspectral images were acquired ten times at subsequent time intervals both on the bruised samples and on the control ones, starting from the day before bruising and 0 hours, 5 hours, 1, 2, 3, 6, 7, 8 and 9 days after bruising. On the whole, 800 non-congruent hyperspectral images were acquired.

Apples were imaged using a HSI system composed of a desktop NIR Spectral Scanner (DVOptic) embedding a Specim N17E reflectance imaging spectrometer, coupled to a Xenics XEVA 2608 camera (320 × 256 pixels) and covering the spectral range from 900 to 1700 nm (5 nm resolution). In order to enable a better evaluation of the stability of the system over time, a setup composed of a silicon carbide sandpaper as sample background, which is characterized by a very low and constant reflectance spectrum [24], a 99% reflectance standard and two ceramic tiles with

average reflectance values equal to 89% and 46%, respectively, were used for the acquisition of all the images.

The raw data were then converted into reflectance values by applying the instrument calibration procedure based on a high-reflectance standard reference and on the estimate of the dark current [25]. Furthermore, in order to minimize the residual variability among images over time, an additional internal calibration was performed [26]. Subsequently, the reflectance images were cropped to dimensions of 209 × 165 pixels with the two-fold aim of excluding the reference materials and of reducing the data size. Finally, in view of the low S/N ratio observed at the spectra extremes, only the 150 wavelengths between 955 and 1700 nm were considered for further analysis.

All the algorithms used for the elaboration of the hyperspectral image data and of the hyperspectrograms were written in MATLAB language (ver. 7.12, The Mathworks Inc., USA). Moreover, the MATLAB functions necessary to create and explore the hyperspectrograms dataset have been implemented into a Graphical User Interface (GUI), which is freely available upon request to the corresponding author. The classification and feature selection steps were performed using the PLS Toolbox (ver. 7.5, Eigenvector Research Inc., USA)..

2.3. Hyperspectral images segmentation

Before conversion into hyperspectrograms, all the acquired hypercubes were subjected to an image segmentation step mainly aimed at removing the pixels corresponding to the sandpaper sheet used as background by means of a fast thresholding procedure. In order to define the most discriminant wavelength, ROIs were manually selected on a few images to define two classes, i.e., 'Background' and 'Sample'. The most discriminant wavelength was then identified separately for each apple cultivar by maximizing the Fisher ratio (*Fr*), computed according to Eq. (1).

$$Fr = \left(m_{back} - m_{sample}\right)^2 / \left(V_{back} + V_{sample}\right) \quad (1)$$

where m_{back} is the average spectrum computed on the "Background" pixels, m_{sample} is the average spectrum of the "Sample" pixels, V_{back} is the variance of the "Background" pixels and V_{sample} is the "Sample" pixels. Therefore, the most discriminant wavelenegth, λ_t, was defined as the wavelength corresponding to the maximum value of Fr. In particular, λ_t = 1025 nm and λ_t = 1050 nm were used for 'Golden Delicious' and 'Pink Lady' datasets, respectively; in both cases, a threshold value equal to 0.1 reflectance units was considered for segmentation.

2.4. Creation of the hyperspectrograms

According to Ferrari et al. [6], an explorative analysis step by means of PCA was carried out at the pixel level on few images evaluating several spectral pretreatments, i.e., detrend, first and second derivatives, standard normal variate, and mean centering. These pretreatments were considered both separately and in different combinations, in order to search for the conditions allowing to highlight at best the apple surface defects, as well as to establish the optimal dimensionality of the PCA models. On the basis of this preliminary analysis, it was found that the best spectral pretreatment method to apply to each single hyperspectral image was linear detrend followed by mean centering. The optimal dimensionality of the PCA models calculated on the pretreated pixel spectra of each individual image was equal to 3 PCs, as evaluated on the basis of the respective scree plots.

Then, each hyperspectral image was individually converted into the corresponding hyperspectrogram (Scheme 1) using a Matlab function developed *ad hoc*, which goes through the following steps:

- the three-dimensional hypercube consisting of one spectral (λ) and two spatial (x and y) dimensions was unfolded to a two-dimensional matrix containing as many rows as the pixels retained after image segmentation and as many columns as the number of wavelengths;

- a PCA model was calculated on the pretreated spectra considering the previously defined number of PCs (3 PCs in the present case), and the corresponding score vectors, Q-residuals vectors, Hotelling T^2 vectors, and loading vectors were stored. . Therefore, each original image with dimensions equal to {209 × 165 × 150} (where 209 and 165 are the number of image pixels along the *x* and *y* directions, respectively, while 150 is the number of wavelengths) was reduced to a matrix of loading vectors with size {3 × 150} and a matrix with dimensions equal to {(209 × 165) × 5}, i.e., with the same number of pixels and including the PC1, PC2 and PC3 score vectors plus the Q residuals and Hotelling T^2 vectors. Thus, the data compression ratio performed in this first step is equal to (209 × 165 × 150) / [(3 × 150) + (209 × 165 × 5)] = 29.92;
- the frequency distribution curves of each score vector and of the Q residuals and Hotelling T^2 vectors were calculated, considering a number of bins equal to the number of spectral variables;
- each frequency distribution curve was normalized by the number of pixels;
- in order to avoid problems due to the sign indeterminacy of PCA decomposition, starting from the second analyzed image, for each PC the sign of each loading vector was defined in a way that the sum of the squared differences with respect to the corresponding loading vector calculated for the first image is minimum, and the sign of the corresponding score vector was defined accordingly;
- the hyperspectrogram of each image was then created by joining in sequence the frequency distribution curves of the scores vectors, of the Q residual vector and of the T2 vector; moreover, in order to maintain the most relevant spectral features of the hypercube data, the PC loading vectors were also joined in sequence at the end of the signal.

Therefore, a dataset of n hyperspectral images is converted in n mono-dimensional signals, the hyperspectrograms, each none composed of a number of variables m given by:

$$m = (2 \times A + 2) \times w \quad (2)$$

where A is the number of selected PCs, and w is the number of wavelengths of the spectra composing the hypercube. In the present case, therefore, given a number of PCs equal to 3 and a number of wavelengths equal to 150, each hyperspectrogram was a 1200 points-long signal.

For a more detailed description of the algorithm used to build hyperspectrograms, the reader is referred to Ferrari et al. [6].

As for the appropriate number of PCs to be retained in the PCA models used for the hyperspectrograms calculation, it has to be underlined that this does not really represent a crucial point. In fact, in case the dimensionality of the model is underestimated, the inclusion of the Q residuals ensures that all the potentially useful information is however kept. Conversely, in case of overestimation of the model dimensionality, it has to be considered that hyperspectrograms are usually intended to be subjected to a variable selection step where the PCs accounting for variability sources which are not useful for solving the problem at hand are discharged.

A further remark concerns the use of the frequency distribution curves of the PCs score vectors and of the Q residuals and Hotelling T^2 vectors: the use of such marginal distributions may have consequences on the performance of the hyperspectrogram approach in those cases where different clusters (e.g., a cluster of pixels of the defective area and a cluster of the remainder pixels) are oriented diagonally, for example in the PC1-PC2 score space. In a case like this, in fact, the distributions of the two clusters along PC1 and PC2 would be overlapped each other, thus making the identification of the defective area more difficult. On the other hand, the use of full A-dimensional distributions (with A equal to the number of selected PCs) could lead to a better separation of the clusters also in such critical situations, but this would lead to a size of the pixel-related part of the hyperspectrogram equal to w^A (or $w^{(A+2)}$, when considering also Q residuals and Hotelling T^2 vectors). This would make the calculation and further use of hyperspectrograms

computationally too demanding, not possible or not convenient, since even for a limited number of PCs the resulting signal would be too long to be efficiently used. However, based on some simulations made considering this kind of critical situation (results not reported for conciseness reasons), the further elaboration of the frequency distribution curves by means of PLS-DA led anyway to satisfactory results, likely thanks to the multivariate nature of the algorithm used for classification, which somehow takes into account the interactions between the different marginal distributions.

2.5. Explorative analysis and classification of the hyperspectrogram datasets

As a first step, for each apple variety, the corresponding dataset of 400 hyperspectrograms was subjected to explorative analysis by means of PCA, in order to obtain an overview of the whole dataset structure and to identify possible outlier samples/images. In this context, the effects of two column pretreatments, i.e., mean centering and autoscaling, were evaluated.

Partial Least Squares – Discriminant Analysis (PLS-DA) [26, 27] was then used to discriminate the two classes i.e., bruised and control samples, both for the '*Golden Delicious*' and for the '*Pink Lady*' datasets, evaluating also in this case both mean centering and autoscaling as signal pretreatments. For each model, the optimal number of Latent Variables (LVs) was chosen on the basis of the minimum value of the classification error in cross-validation, calculated using a customized cross-validation procedure in order to include in the same deletion group all the hyperspectrograms calculated from replicate images of couples of randomly selected apple samples, in a way to avoid over-optimistic results. The samples of each dataset were randomly divided into a training set including about 2/3 of samples (corresponding to 280 hyperspectral images) and a test set composed of the remainder 1/3 of samples (corresponding to 120 images). The classification results are reported in terms of Classification Efficiency (EFF), i.e., the geometric mean of

Sensitivity (percentage of objects of the modeled class correctly accepted by the class model) and Specificity (percentage of objects of other classes correctly rejected by the class model) [26, 28].

Subsequently, a variable selection step was performed on both hyperspectrogram datasets with the two-fold aim of further reducing data size and, most important, of identifying the signal regions most relevant to discriminate between the 'bruised' and 'control' classes. In particular, the interval Partial Least Squares-Discriminant Analysis (iPLS-DA) algorithm [6, 26, 29, 30] was used to this purpose; iPLS-DA essentially consists in subdividing the whole signal range in a user-defined number of intervals of equal length and in calculating classification models by iteratively adding or removing intervals, depending on whether the *forward* or the *reverse* search strategy is used. The intervals most useful for classification are identified by minimizing the value of the classification error in cross-validation. The selection of the number of intervals retained in the final model is therefore performed automatically by the algorithm in a blind way, with no a priori assumptions made on the basis of the nature of the problem at hand.

In the present work, forward iPLS-DA was applied considering two different interval sizes, consisting of 10 and 25 variables. The same subdivision between training and test sets and the same cross-validation procedure already used for the whole PLS-DA models was also used for the calculation and subsequent validation of the iPLS-DA models.

Finally, the classification model results obtained using hyperspectrograms were compared to the results obtained using a conceptually simpler, but still automated (i.e., not requiring any manual selection of ROIs) approach. To this aim, the average spectrum of each segmented hyperspectral image was computed, and for each dataset the corresponding matrix of size {400, 150} was then used to calculate PLS-DA and iPLS-DA models considering the same training and test sets as well as the same cross-validation procedure used for the hyperspectrograms matrix. In this case, in addition to mean centering and autoscaling, also linear detrend followed by mean centering was considered as signal pretreatment, since it was the same used for the hyperspectrograms creation. It has to be underlined that the method based on average spectra was selected as comparison method

in light of the main aim of the hyperspectrogram approach, which is to investigate the dataset structure at image-level and with no in-depth knowledge about the system. However, we are aware that this method, although very effective in the case of homogeneous samples, does not perform properly when there is the need to identify spatially localized features within the image scene since, conversely to hyperspectrograms, average spectra do not account for spatial variability.

2.6. Image reconstruction using the selected pixel-related features

In order to gain a better understanding of the choices made automatically by iPLS-DA, the hyperspectrogram regions selected on the frequency distribution curves (i.e., the pixel-related features that resulted to be useful for classification) were projected back into the original image domain. The procedure adopted to reconstruct the selected hyperspectrogram features can be summarized as follows:

1. for each frequency distribution curve included in the hyperspectrogram, the values of PCA scores / Q residuals / Hotelling T^2 corresponding to the hyperspectrogram portions selected by iPLS-DA were stored;

2. for each sample image considered for reconstruction, among the corresponding PCA score / Q / T^2 images, only those were kept, for which at least one portion of the hyperspectrogram was selected;

3. for each one of the PCA score / Q / T^2 images that were kept, only the pixels with a value falling within the range selected at point 1 were maintained, and represented as separate greyscale images, or altogether as false-color images.

For example, let us suppose that iPLS-DA has led to the selection of portions of frequency distribution curves corresponding to PC1 score values from 10 to 30 and to Q residuals values from 0 to 60, while no regions have been selected for PC2 scores and for T^2. Only PC1 scores and Q

residuals are considered in the following step, where all the PC1 score values lower than 10 and greater than 30, and all the Q values greater than 60 are discarded. Finally, two greyscale images, i.e., the PC1 score image and the Q residuals image are generated, where only the retained pixels are represented. Alternatively, a false-color image of the retained pixels can be created, with e.g., the PC1 score values in the red channel and the Q residual values in the green channel.

The images obtained with this procedure allowed therefore to localize in the original image domain the features responsible for the identification of the bruise defect, which represents the main advantage of the imaging techniques.

3. Results and discussion

3.1. Creation of the hyperspectrograms and explorative analysis

By applying the hyperspectrogram approach, the original dataset composed of 800 hyperspectral images and with a size of about 12 GB was converted into a new matrix of 1200-points long signals with a total size of 4.7 MB. As an example, the hyperspectrograms obtained from the 'Golden Delicious' and 'Pink Lady' datasets are reported in Figure 1a, and Figure 1b, respectively. After the calculation of a first PCA model on the whole dataset, that showed a clear separation of clusters corresponding to the two apple varieties (data not reported for conciseness reasons), PCA models were calculated on the 'Golden Delicious' and 'Pink Lady' datasets separately.

3.1.1 'Golden Delicious' dataset

The hyperspectrogram dataset computed on the 'Golden Delicious' images was initially investigated by means of PCA considering mean center as signal pretreatment. This PCA model was found to have an optimal dimensionality equal to 4 PCs, accounting for about 70% of the total

variance. Considering the confidence limit equal to 99.7%, the Q residuals vs. Hotelling T^2 plot showed 13 outlier signals, 10 of which belonging to only two apple samples. The investigation of the score plots (as an example, the PC1-PC2 score plot is reported in Figure 2a) allowed to verify that no effects related to instrumental drifts over the different days of acquisition was present, but did not show the existence of a clear distinction between bruised and control samples.

The same hyperspectrogram dataset was also investigated by means of PCA considering autoscale as signal pretreatment, and selecting 2 PCs. In this case, a partial separation between the two classes was visible in the PC1-PC2 score plot, as shown in Figure 2b. Furthermore, the outlier signals identified in this figure correspond to the same samples identified also on the Q residuals vs. Hotelling T^2 plot of the mean centered hyperspectrograms; in the case reported in Figure 2b, a progressive change of these samples over time was also clearly visible. The two outlier samples were a control apple and a bruised apple, which were the only two samples showing clearly visible signs of rot in the last days of acquisition, while in the former days they were not distinguishable with the naked eye from other apples with bruises or other variations in the pigmentation of the surface. As an example, the RGB images of the bruised sample G20 taken at different acquisition times are reported in Figure 3a, where it can be noticed that although one day after bruising the bruised area was clearly visible, it was however similar to bruised areas observed on other non-outlier images, like those taken from sample G10 (Figure 3b). The fact that the only two rot samples were clearly identified as outliers much time before the effective appearance of the signs of rot, suggests that it could be also possible to use HSI as a technique for early detection of this problem. However, since this was beyond the aims of the present work, and due to the fact that only two rotten samples were present into the whole dataset, all the measurements related to the two outlier samples were removed before further processing the data, in view of detecting solely the defects due to bruising.

3.1.2 'Pink Lady' dataset

The PCA model obtained on the mean centered hyperspectrograms of the 'Pink Lady' dataset was found to have an optimal dimensionality equal to 4 PCs, accounting for 69.28 % of the total variance, while for the model calculated on the autoscaled hyperspectrograms only 2 PCs were found to be significant. Similarly to what observed for the 'Golden Delicious' dataset, the PC1-PC2 score plot of the autoscaled signals showed a better separation between the two classes with respect to the mean centered ones. Considering the Q residuals vs. Hotelling T^2 plot, no outliers for T^2 values were identified in both cases, while samples with Q residual values beyond the 99.7 % confidence limits were observed when considering the autoscaled data. However these samples were not removed from the dataset, since they were corresponding to bruised samples, mainly imaged in the last days of acquisition; therefore, the Q residual values could have been related to informative variability.

3.2. Hyperspectrograms classification by PLS-DA and iPLS-DA

3.2.1 'Golden Delicious' dataset

The results of the PLS-DA models calculated on the 'Golden Delicious' dataset considering both the whole signal range and the variables selected by iPLS-DA are reported in Table 1. In accordance with the results of PCA, the comparison between the Classification Efficiency (EFF) values obtained with the mean centered and with the autoscaled hyperspectrograms highlights that autoscaling leads in general to better classification performances. Concerning the results obtained with variable selection, it can be noticed that the same number of variables was selected for both the considered interval size values independently of the processing method; however, a lower model dimensionality as well as higher classification efficiency values were obtained when using the autoscaled hyperspectrograms. This can be explained considering that, since in the hyperspectrograms the pixel-related quantities are reported as frequency distribution curves, pretreating the signals by autoscaling allows to increase the contribution of the parts where low

frequency values are observed, i.e., those parts corresponding to a restricted number of pixels, such as for example the pixels corresponding to a surface defect. Autoscaling the hyperspectrograms results therefore particularly useful when dealing with a defect detection issue, like in the present case.

For comparison purposes, the classification efficiency values obtained from the PLS-DA and iPLS-DA models calculated considering the average spectra of the segmented hyperspectral images instead of the hyperspectrograms are reported in Table 2, where it can be noticed that this simpler approach did not lead to satisfactory classification performances. Indeed, the best model obtained using linear detrend followed by mean center as spectra pretreatment and iPLS-DA with an interval size equal to 25 variables resulted in EFF values equal to 57.54% and 40.39% in cross-validation and in prediction on the external test set, respectively.

Concerning hyperspectrograms, the overall best performing model was based on 30 variables only (Table 1, 4^{th} column), that were all selected on the signal region accounting for the spatial variability (i.e., in the part of the hyperspectrogram that contains the frequency distribution curves, *see* Figure 1a), and which led to an EFF value in prediction on the external test set equal to 92.42%. In particular, the three selected intervals fall within the frequency distribution curves of PC2 scores, PC3 scores and Q residuals, respectively.

A visual interpretation of the selected features was obtained by reconstructing for each selected interval the corresponding grey scale image, as described in Section 2.6. As an example, the PC2 scores, PC3 scores and Q residuals pseudo-color images of the same sample measured at two different times after bruising are reported in Figure 4, together with the corresponding reconstructed grayscale images, where only the pixels corresponding the selected hyperspectrogram regions are reported. Considering the sample imaged 5 hours after bruising, it can be noticed that the defective area results evident on the PC2 score image, while no information relevant for classification is present neither on the PC3 score image nor on the Q residual image. Consistently with these results, the reconstructed PC2 score image shows that mainly the pixels related to the

defective area are selected, while no features are reported on the PC3 and Q residual reconstructed images. In contrast to the image taken after 5 hours, where high PC2 score values (in red color) correspond to the newly damaged area, in the case of the image taken 6 days after bruising the high PC2 score values account for the region surrounding this area. The information related to the originally impacted area is instead clearly distinguishable on the PC3 score image. Also in this case, it can be noticed that the features retained in the images reconstructed using only the selected variables are fully consistent with these findings. As far as the Q residuals are concerned, their interpretation is less clear, since very few pixels are selected only on the image taken 6 days after bruising.

A more compact representation of the features used for classification was obtained by superimposing the three grey scale reconstructed images to form a unique false-color image showing in the red, green and blue channels the information related to PC2, PC3 and Q residuals, respectively. Figure 5 reports the results obtained on another bruised sample at four different times of acquisition. In particular, the PC2/PC3/Q residual false-color images (Figure 5b) show that the bruise, although visible also in the RGB images (Figure 5a), results much more evident in the hyperspectral ones, and that additional information about its development with time can be deduced. Moreover, the comparison of the false color images of Figure 5b with the corresponding reconstructed images (Figure 5c) confirms that the classification model is actually based on these bruise-related features. The few sparse pixels that are visible in the left image of Figure 5c (sample before bruising) are likely due to the fact that, within the interval selected by iPLS-DA for PC2 (corresponding to the red channel in Fig 5c), the corresponding hyperspectrogram values for a non-bruised sample are not necessarily always equal to zero, i.e., there could be some pixels whose PC2 values lie within the selected interval, but their number is however much lower than for the image of a bruised sample.

Similarly, a check made considering the reconstructed images of the control samples has shown that the selected features are not present at all, or only few sparse pixels not accounting for

any specific sample area are selected. As an example, the corresponding images taken from a control sample are shown in Figure 6, where no significant variation can be appreciated over time.

Furthermore, a detailed inspection of each single misclassified image was done by MIA at the pixel level, which revealed that one image taken on an apple sample before bruising was not assigned to the control class due to the presence of preexisting, not visible, bruises. Moreover, MIA also revealed that two samples that were not predicted as bruised did not actually show the presence of a bruised area in the first hours after bruising, while this appeared in the following days.

3.2.2 'Pink Lady' dataset

As far as the 'Pink Lady' dataset is concerned, the classification results obtained on the whole hyperspectrogram range as well as using only the regions retained by the feature selection algorithm are summarized in Table 3. Although also in this case in general better results were obtained when using the autoscaled hyperspectrograms, the effect of the signal pretreatment on the classification efficiency values was less evident compared to the 'Golden Delicious' dataset. Moreover, the variable selection step led to PLS-DA models with similar classification performances and comparable dimensionality with both the considered interval sizes. For conciseness reasons, only the results of the model obtained considering an interval size of 25 variables are discussed here. This classification model was based on 275 variables (highlighted in Figure1b) and led to an EFF value in prediction on the external test set equal to 94.04% (5 LVs). In contrast to the 'Golden Delicious' dataset, in this case two intervals were selected also on the hyperspectrogram region accounting for the spectral features (i.e., the loadings part of the hyperspectrogram). The fact that the automated iPLS-DA algorithm led to the selection of a different number of intervals and of intervals located in different signal regions when applied to 'Pink Lady' and Golden Delicious' is likely to be related to the differences existing between the two varieties, in terms of juiciness and phenolic composition, which greatly affect the browning process [31, 32].

In particular, the PC1 loadings between 1580 and 1700 nm and the PC2 loadings between 1080 and 1200 nm were selected, which are ascribable to the C-H stretch first overtone and C-H second overtone, respectively. Considering that both these spectral regions are related to aromatic structures [33], the selection is likely to account for the changes which occur in polyphenolic molecules as a consequence of bruising. As far as the spatially-related features are concerned, different signal regions were selected on the frequency distribution curves of PC1, PC2 and PC3 scores, as well as on the frequency distribution curve of the Hotelling T^2 values. The evaluation of the reconstructed score images, obtained for each frequency distribution curve by projecting back in the original image domain only the pixels with a value falling within the selected interval, showed that PC1 accounted for the round shape of the samples. In fact, in all the samples and independently of the acquisition time, a ring of pixels at equal sample height was selected. Similarly to the 'Golden Delicious' dataset, bruise-related features evolving over time were highlighted in the reconstructed score images of PC2, PC3 and, in this case, also of the T^2 values. As an example, in Figure 7b, the false-color images obtained by superimposing the reconstructed score images of PC2, PC3 and T^2 (corresponding to the red, green and blue channels, respectively) are shown for the images acquired on the same bruised sample at different times of acquisition. By comparing them with the corresponding RGB images (Figure 7a), it can be observed that in this case bruises are very difficult to distinguish with the naked eye, while they are clearly identified when the reconstructed score images are considered. Furthermore, the information about the development of the defective area over time, which is not available at all in the RGB images, is instead well evident in the reconstructed score images. For comparison, the RGB images and the corresponding reconstructed score images obtained on a control sample are shown in Figure 7c and Figure 7d, respectively, where no changes can be observed over time.

An alternative representation of samples evolution over time, which takes advantage of the hyperspectrogram approach, can be obtained by projecting the bruised sample signals on the principal components space calculated considering only the hyperspectrograms of the sound apples

which were used as control. In this way, in fact, it is possible to evaluate how the defective apples diverge with respect to the remainder ones by looking at Q residuals and at Hotelling T^2 plots. In order to apply this procedure to the 'Pink Lady' dataset, a PCA model was computed on the hyperspectrograms corresponding to the control apples already included in the previous training set (140 signals) and using autoscaling as signal pretreatment. The PCA model, which was found to have an optimal dimensionality equal to 4 PCs, was then used to estimate all the remaining hyperspectrograms (380 signals). As expected, the results showed that the control samples of the test set actually did not deviate from those included in the model, while a significant divergence was observed when the bruised apple signals were projected on the PC subspace. As an example, the Q residuals chart and the Hotelling T^2 chart are reported in Figure 8a and Figure 8b, respectively, for the signals acquired at ten time intervals on two test set samples, i.e., a control sample (on the left with respect to the dash-dotted line) and a bruised one (on the right of the dash-dotted line). Conversely to the control samples, which in both cases lie within the 95% confidence limits, it can be observed how the evolution of the bruised samples over time is effectively summarized by the complementary information conveyed by the Q residuals and the Hotelling T^2 charts. Indeed, starting from the signal representing the apple sample before bruising and which is therefore correctly lying within the confidence limits, as soon as the bruise is induced the corresponding signal is immediately detected as outlier by the Q residuals chart. Although the deviation of the bruised apple from the model actually decreases over time, the deviation of the same sample within the model increases over time and therefore starting from 6 days after bruising it becomes easily detectable on the Hotelling T^2 chart. The evaluation of the Q residuals and Hotelling T^2 charts also allowed to easily highlight the nonlinear development of the bruised area over time and, in particular, the different behavior of the newly bruised samples with respect to those with the less recent bruises. These results suggested therefore the potentiality of the system to effectively discriminate not only the sound samples with respect to the bruised ones, but also the recently bruised samples (0-3 days after bruising) from those with older bruises (6-9 days after

bruising). As reported in Baranowski et al. [23], the information about the time after damage could be valuable from a practical point of view, since it could allow to sort and grade fruits more effectively as well as to obtain a better prediction of fruits susceptibility to infections. Therefore, in order to verify the ability of the system to discriminate recent from old bruises, additional PLS-DA models were computed on the 'Pink Lady' dataset by applying the same classification and variable selection procedure described in Section 2.5, but considering in this case three classes, i.e., 'control', 'recent bruises' (0-3 days after bruising) and 'old bruises' (6-9 days after bruising). The classification results confirmed the effectiveness of the approach (Table 4) and in particular the best performing model resulted in EFF values for the external test set equal to 91.33%, 91.27% and 98.44% for the 'control', 'recent bruises' and 'old bruises' classes, respectively. Nicely, also the hyperspectrogram regions selected by iPLS-DA are very similar to those previously selected for the two-class problem (results not reported for conciseness reasons).

Finally, it has to be underlined that, similarly to what was observed for the 'Golden Delicious' dataset, the use of average spectra did not lead to acceptable results (Table 2) also for the 'Pink Lady' dataset. The overall best performing model, in fact, which was obtained using linear detrend combined with mean centering as spectra pretreatment and iPLS-DA with an interval size equal to 25 variables, resulted in classification efficiency values equal to 73.44% and to 49.19% in cross-validation and in prediction of the external test set, respectively. This further confirmed that the approach based on hyperspectrograms is definitely more efficient in the identification of surface defects, although the calculation of hyperspectrograms is nearly as fast as the calculation of the average spectra: as an example, using a personal computer running with Microsoft Windows 7–64 bit ® and equipped with an Intel Core ® i7-2600 CPU @ 3.40 GHz processor and 4.00 GB RAM, the average time for the calculation of a single hyperspectrogram resulted equal to 1.55 s, while the corresponding time required to calculate an average spectrum was equal to 1.16 s.

4. Conclusions

The potentiality of the hyperspectrograms-based approach for the fast exploration and classification of a large dataset of NIR hyperspectral images has been investigated. Hyperspectrograms calculated on a dataset of 800 hyperspectral images of 'Golden Delicious' and 'Pink Lady' apples were used to classify bruised and sound fruits imaged over a period of 10 days.

Before classification, PCA was applied as explorative analysis. This permitted to verify the absence of effects related to instrumental drifts over the different days of acquisition. Moreover, outlier images were clearly identified, which were ascribable to phenomena different from the presence of bruises: in particular, apples showing clear signs of rot in the last days of acquisition were detected as outliers already from the first days of measurement, when, with the naked eye, they were apparently indistinguishable from the other samples.

The application of a variable selection algorithm to the hyperspectrograms allowed to achieve classification efficiency values in prediction of the external test sets equal to 92.42% and 94.04% for the 'Golden Delicious' and on the 'Pink Lady' datasets, respectively. Moreover, a critical evaluation of the choices made by the feature selection algorithm was made possible by reconstructing as images the spatial features selected on the frequency distribution curves of the hyperspectrograms. The evaluation of the reconstructed images confirmed that in both datasets the automated selection performed on hyperspectrograms has led to the exact localization of the spatial features related to the onset and to the subsequent evolution with time of the bruise defect.

The results confirmed the robustness of the approach when dealing with datasets composed of a very large number of hyperspectral images, and also indicate that hyperspectrograms could represent an effective tool for the investigation of dynamic processes.

The proposed method, which is intended as a fast exploration tool for large datasets of hyperspectral images, could be further improved in order to enhance its flexibility and the interpretation of the results in chemical terms, e.g. by considering the pure components spectra and

the concentration profiles calculated with MCR-ALS instead of the PCA loading and score vectors for the calculation of the hyperspectrograms.

A further possibility, which is being implemented within the hyperspectrograms GUI, consists in the calculation of the hyperspectrograms considering a global PCA model (i.e., calculated on all the images together). This could constitute an effective trade-off between the ability of considering the average properties of each image (as with the average spectra) and the possibility of evaluating the spatial variability-related aspects, which is maximized when considering the separate PCA models calculated on each image like in the present work.

References

1. H. Grahn, P. Geladi, Techniques and applications of hyperspectral image analysis, John Wiley & Sons Ltd., Chichester, 2007.
2. A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends Food Sci. Tech. 18 (2007) 590-598.
3. D.-W. Sun (Ed.), Hyperspectral imaging for food quality analysis and control, Academic Press, London, 2010.
4. J. Burger, A.A. Gowen, Data handling in hyperspectral image analysis. Chemom. Intell. Lab. Syst. 108 (2011) 13-22.
5. A.A. Gowen, F. Marini, C. Esquerre, C. O'Donnell, G. Downey, J. Burger, Time series hyperspectral chemical imaging data: challenges, solutions and applications. Anal. Chim. Acta 705 (2011) 272–282.
6. C. Ferrari, G. Foca, A. Ulrici, Handling large datasets of hyperspectral images: reducing data size without loss of useful information. Anal. Chim. Acta 802 (2013) 29-39.
7. A. Antonelli, M. Cocchi, P. Fava, G. Foca, G.C. Franchini, D. Manzini, A. Ulrici, Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm, Anal. Chim. Acta. 515 (2004) 3–13.
8. G. Foca, F. Masino, A. Antonelli, A. Ulrici, Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques, Anal. Chim. Acta. 706 (2011) 238–245.
9. A. Ulrici, G. Foca, M.C. Ielo, L.A. Volpelli, D.P. Lo Fiego, Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams, Innov. Food Sci. Emerg. Technol. 16 (2012) 417–426.
10. J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: A review with applications. Chemom. Intell. Lab. Syst. 107 (1) (2011) 1-23.
11. A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis, TrAC Trends Anal. Chem. 23 (2004) 70–79.
12. J.M. Amigo, Practical issues of hyperspectral imaging analysis of solid dosage forms. Anal. Bioanal. Chem. 398 (2010) 93-109.
13. J.X. Wu, S. Rehder, F.V.D. Berg, J.M. Amigo, J.M. Carstensen, T. Rades, C.S. Leopold, J. Rantanen, Chemical imaging and solid state analysis at compact surfaces using UV imaging. Int. J. Pharm. 477 (1-2), (2014) 527-535.
14. J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications. Chemom. Intell. Lab. Syst. 140 (2015) 1-12.
15. M. Van Zeebroeck, V. Van linden, H. Ramon, J. De Baerdemaeker, B.M. Nicolaï, E. Tijskens, Impact damage of apples during transport and handling. Postharvest Bio. Tec. 45 (2007) 157-167.
16. Farh. Geoola, Fara. Geoola, U.M. Peiper, A spectrophotometric method for detecting surface bruising on "Golden Delicious" apples. J. Agr. Eng. Res. 58 (1994) 47-51.

17. J. Xing, S. Landahl, J. Lammertyn, E. Vrindts, J. De Baerdemaeker, Effects of bruise type on discrimination of bruised and non-bruised ‘Golden Delicious’ apples by VIS/NIR spectroscopy. Postharvest Bio. Tec. 30 (2003) 249-258.

18. J. Xing, C. Bravo, D. Moshou, H. Ramon, J. De Baerdemaeker, Bruise detection on ‘Golden Delicious’ apples by VIS/NIR spectroscopy. Comput. Electron. Agr. 52 (2006) 11–20.

19. R. Lu, Detection of bruise on apples using near-infrared hyperspectral imaging. T. ASAE 46 (2003) 523-530.

20. J. Xing, C. Bravo, P.T. Jancsók, H. Ramon, J. De Baerdemaeker, Detecting bruises on ‘Golden Delicious’ apples using hyperspectral imaging with multiple wavebands. Biosyst. Eng. 90 (1) (2005) 27-36.

21. J. Xing, W. Sayes, J. De Baerdemaeker, Combination of chemometric tools and image processing for bruise detection on apples. Comput. Electron. Agr. 56 (2007) 1-13.

22. G. ElMasry, N. Wang, C. Vigneault, J. Qiao, A. ElSayed, Early detection of apple bruises on different background colors using hyperspectral imaging. LWT-Food Sci. Technol. 41 (2) (2008) 337-345.

23. P. Baranowski, W. Mazurek, J. Pastuszka-Wo´zniak, Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. Postharvest Bio. Tec. 86 (2013) 249–258.

24. J. Burger, P. Geladi, Hyperspectral NIR image regression part II: Dataset preprocessing diagnostics. J. Chemom. 20 (2006) 106-119.

25. J. Burger, P. Geladi, Hyperspectral NIR image regression part I: Calibration and correction. J. Chemom. 19 (2005) 355-363.

26. A. Ulrici, S. Serranti, C. Ferrari, D. Cesare, G. Foca, G. Bonifazi, Efficient chemometric strategies for PET–PLA discrimination in recycling plants using hyperspectral imaging. Chemom. Intell. Lab. Syst. 122 (2013) 31-39.

27. S. Chevallier, D. Bertrand, A. Kohler, P. Courcoux, Application of PLS-DA in multivariate image analysis. J. Chemom. 20 (2006) 221-229.

28. M. Forina, P. Oliveri, H. Jäger, U. Römisch, J. Smeyers-Verbeke, Class modeling techniques in the control of the geographical origin of wines. Chemom. Intell. Lab. Syst. 99 (2009) 127-137.

29. L. Nørgaard, A. Saudaland, J. Wagner, J. P. Nielsen, L. Munck, S.B. Engelsen, Interval Partial Least-Squares Regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Appl. Spectrosc. 54 (2000) 413-419.

30. R. Leardi, L. Norgaard, Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions. J. Chemom. 18(11) (2004) 486–497.

31. J.L. García, M. Ruiz-Altisent, P. Barreiro, Factors influencing mechanical properties and bruise susceptibility of apples and pears, J. Agr. Eng. Res., 61(1) (1995) 11-17.

32. M.J. Amiot, M. Tacchini, S. Aubert, J. Nicolas, Phenolic composition and browning susceptibility of various apple cultivars at maturity. J. Food Sci., 57 (1992) 958–962.

33. J.S. Shenk, J.J. Workman, M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, Third ed., Marcel Dekker, New York, 2008, p. 356.

Captions of schemes and figures

Scheme 1: Key steps of the hyperspectrograms-based approach.

Figure 1: Hyperspectrograms calculated on the 'Golden Delicious' dataset (a) and on the 'Pink Lady' dataset (b); the variables selected by iPLS-DA are highlighted in grey.

Figure 2: PC1-PC2 score plots from the PCA models calculated on the mean centered (a) and autoscaled (b) 'Golden Delicious' hyperspectrograms.

Figure 3: RGB images of the outlier 'Golden Delicious' sample G20 (a) and of another severely bruised sample (G10) (b).

Figure 4: Images of PC2 scores, PC3 scores and Q residuals (red = high values, blue=low values) together with the corresponding grey scale images of the reconstruction of the selected features, obtained from the analysis of two hyperspectral images acquired on a bruised sample at different times. For comparison purposes, the RGB images of the two samples are also reported on the top of the figure.

Figure 5: RGB images acquired on a bruised sample at different times (a); corresponding false-color images of the whole (b) and of the reconstructed (c) PC2 score image (red channel), PC3 score image (green channel) and Q residual image (blue channel).

Figure 6: RGB images acquired on a control sample at different times (a); corresponding false-color images of the whole (b) and of the reconstructed (c) PC2 score image (red channel), PC3 score image (green channel) and Q residual image (blue channel).

Figure 7: RGB images and corresponding false-color images of the reconstructed PC2 score image (red), PC3 score image (green) and HotellingT^2 image (blue) obtained on a bruised sample (a, b) and on a control sample (c, d).

Figure 8: Q residuals (a) and Hotelling T^2 values (b) charts for the images taken at subsequent time intervals on two test set samples: the left part of both figures reports the values of the control sample, while the right part reports the values obtained for the bruised sample.

Table 1: Classification results obtained for the 'Golden Delicious' dataset. The best model selected in cross-validation is indicated in italic characters.

Pretreatment		Whole signals		Interval size 10		Interval size 25	
		Mean center	**Autoscale**	**Mean center**	**Autoscale**	**Mean center**	**Autoscale**
No. of LVs		1	1	5	*1*	8	1
No. of variables		1200	1200	30	*30*	125	125
Efficiency (%)	**CV**	69.73	72.10	92.70	*93.15*	89.37	88.85
	Pred.	64.80	64.23	82.79	*92.42*	56.14	80.36

Table 2: PLS-DA and iPLS-DA results obtained on‘Golden Delicious’ and ‘Pink Lady’ datasets considering the average spectra calculated on the HSI images. The best models selected in cross-validation are indicated in italic characters.

		Golden Delicious			Pink Lady		
		Whole signals	**Interval size 10**	**Interval size 25**	**Whole signals**	**Interval size 10**	**Interval size 25**
Mean Center	**# Variables**	150	10	25	150	40	50
	# LVs	2	1	4	4	4	5
	Cal.	49.11	56.61	60.44	71.79	75.41	75.05
	CV	35.09	54.41	37.90	62.56	69.64	67.01
	Pred.	17.61	17.94	40.39	64.09	58.52	46.04
Autoscale	**# Variables**	150	20	50	150	40	25
	# LVs	6	4	6	4	4	4
	Cal.	69.83	57.81	67.14	72.62	74.71	75.05
	CV	42.30	38.15	42.64	63.65	68.93	67.98
	Pred.	41.47	37.37	40.59	64.09	61.83	46.72
Linear Detrend + Mean Center	**# Variables**	150	10	*25*	150	30	*100*
	# LVs	5	2	*1*	3	5	*6*
	Cal.	64.11	59.20	*52.18*	75.75	74.95	*78.25*
	CV	41.43	50.43	*57.54*	65.59	71.35	*73.44*
	Pred.	45.75	44.03	*40.39*	56.81	46.04	*49.19*

Table 3: Classification results obtained for the 'Pink Lady' dataset. The model showing the overall best performances is indicated in italic characters.

Pretreatment		Whole signals		Interval size 10		Interval size 25	
		Mean center	**Autoscale**	**Mean center**	**Autoscale**	**Mean center**	**Autoscale**
No. of LVs		3	2	10	4	4	*5*
No. of variables		1200	1200	130	200	50	*275*
Efficiency (%)	**CV**	93.29	94.83	99.05	100.00	94.96	*100.00*
	Pred.	93.27	92.30	96.23	89.22	86.46	*94.04*

Table 4: PLS-DA results obtained on the 'Pink Lady' dataset for the classification of 'control', 'recent bruises' and 'old bruises' classes. The model showing the overall best performances is indicated in italic characters.

			Whole signals		Interval size 10		Interval size 25	
Pretreatment			**Mean center**	**Autoscale**	**Mean center**	**Autoscale**	**Mean center**	**Autoscale**
No. of LVs			4	3	9	4	7	4
No. of variables			1200	1200	80	120	175	125
EFF (%)	**CV**	Control	93.69	95.53	96.87	*98.55*	90.65	92.55
		Recent bruises	95.94	96.65	96.89	*97.83*	93.60	96.38
		Old bruises	89.72	95.51	84.82	*98.89*	93.08	92.83
	Pred.	Control	91.65	93.10	94.87	*91.33*	83.87	87.36
		Recent bruises	93.11	90.72	92.55	*91.27*	90.21	90.90
		Old bruises	93.01	94.07	96.98	*98.44*	94.07	94.66

Scheme 1

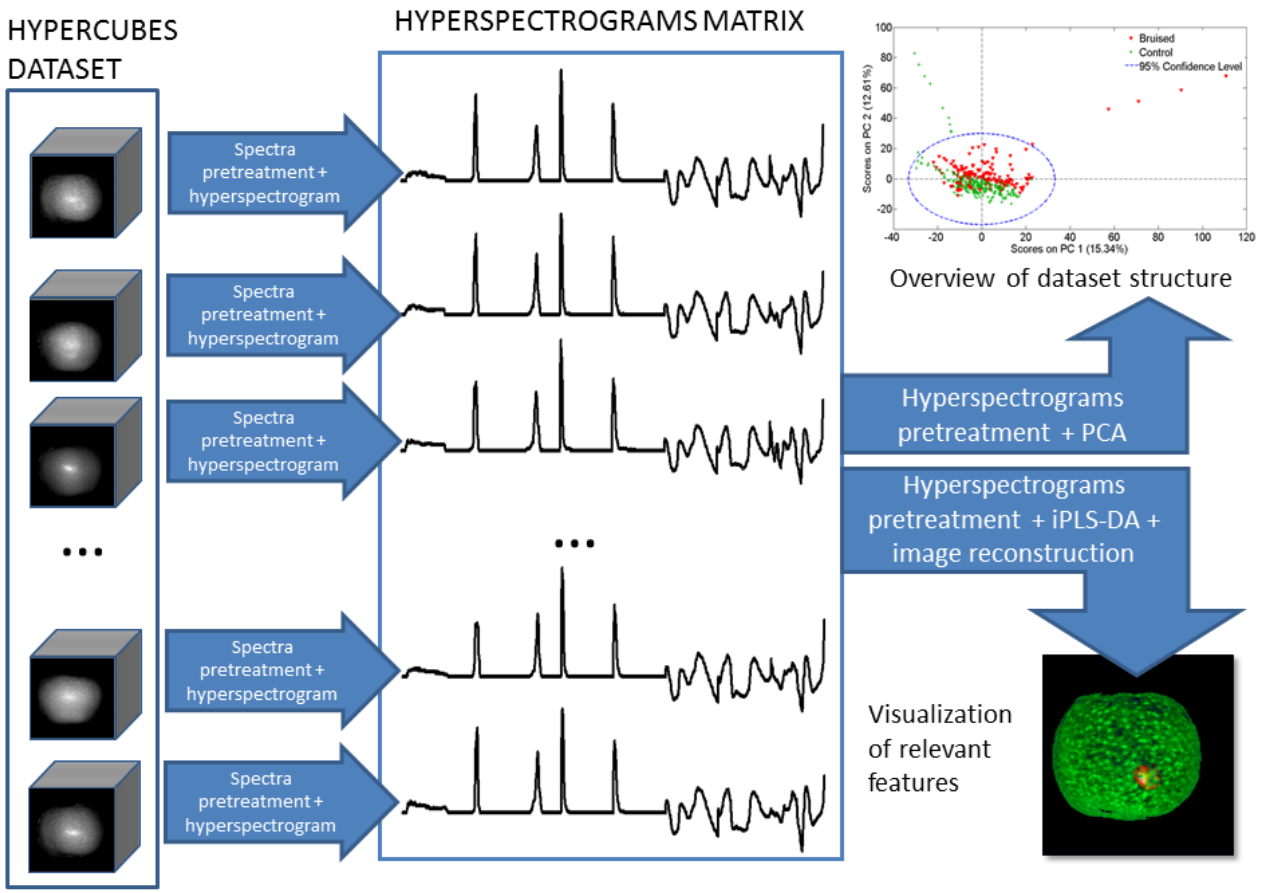

HYPERCUBES DATASET
HYPERSPECTROGRAMS MATRIX
Spectra pretreatment + hyperspectrogram
Spectra pretreatment + hyperspectrogram
Spectra pretreatment + hyperspectrogram
Spectra pretreatment + hyperspectrogram
Spectra pretreatment + hyperspectrogram
Bruised
Control
95% Confidence Level
Scores on PC 2 (12.61%)
Scores on PC 1 (15.34%)
Overview of dataset structure
Hyperspectrograms pretreatment + PCA
Hyperspectrograms pretreatment + iPLS-DA + image reconstruction
Visualization of relevant features

Figure 1

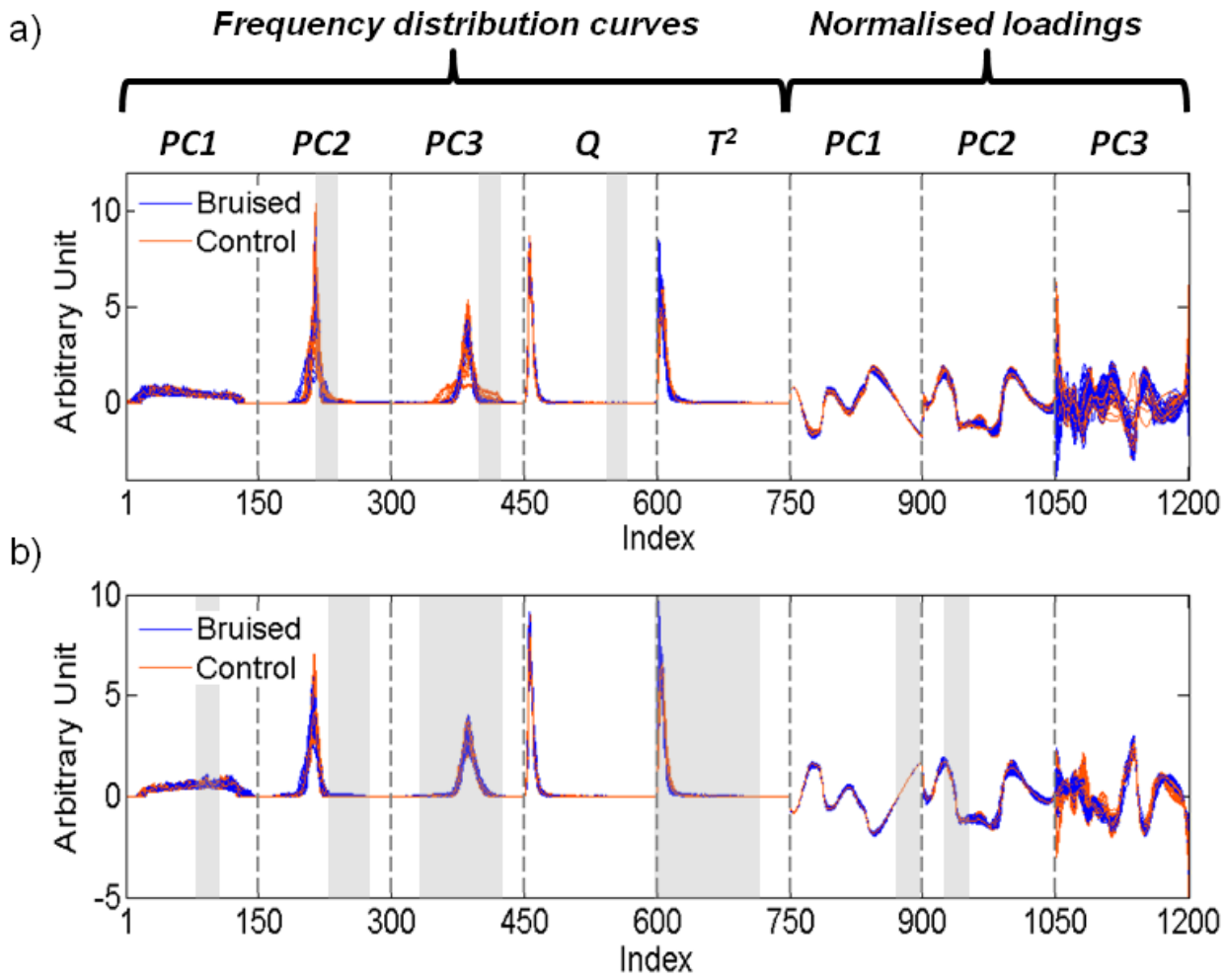

a)
Frequency distribution curves
Normalised loadings
PC1
PC2
PC3
Q
T²
PC1
PC2
PC3
Bruised
Control
Arbitrary Unit
10
5
0
1
150
300
450
600
750
900
1050
1200
Index
b)
Bruised
Control
Arbitrary Unit
10
5
0
-5
1
150
300
450
600
750
900
1050
1200
Index

Figure 2

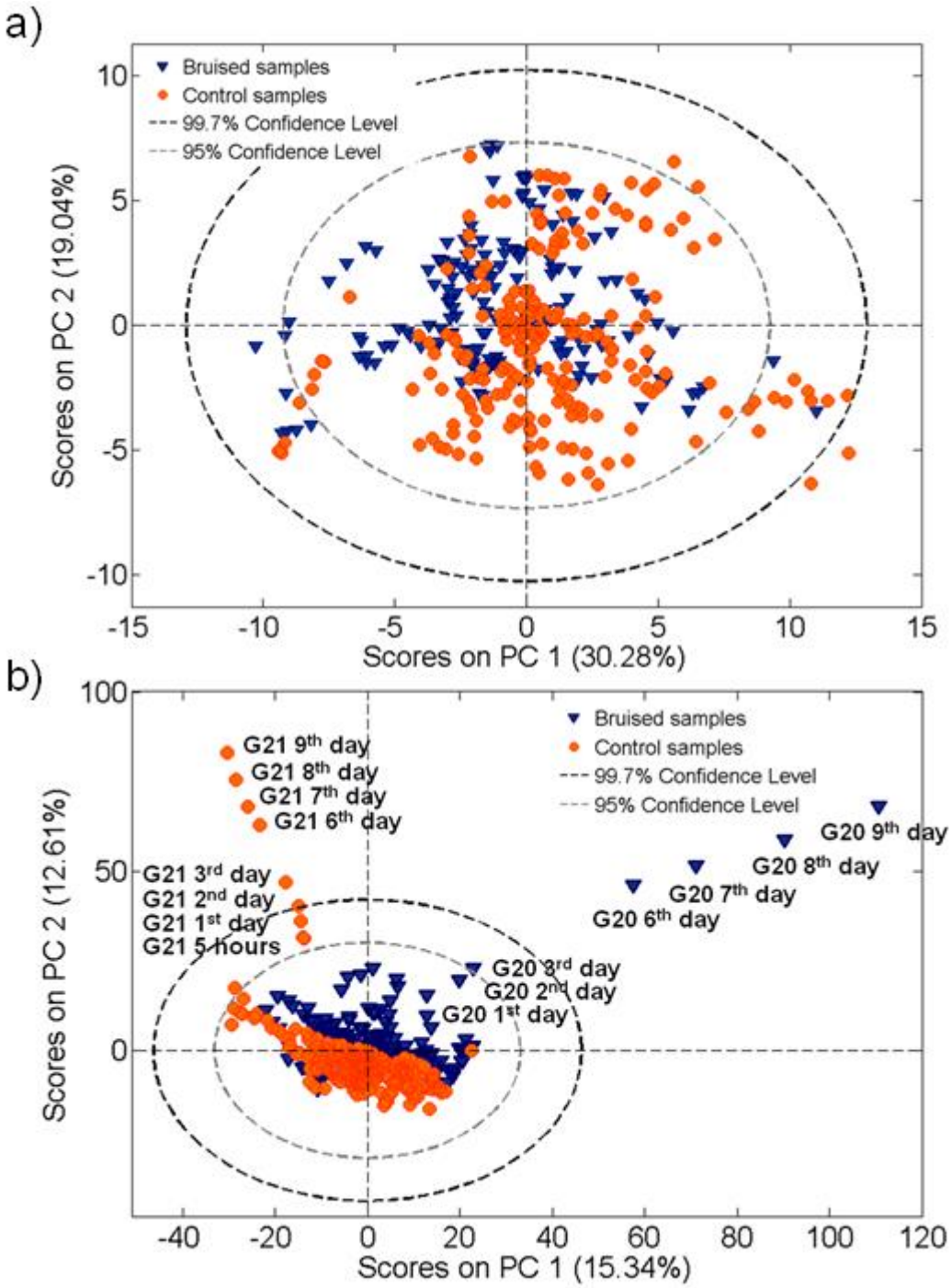

a)
Bruised samples
Control samples
99.7% Confidence Level
95% Confidence Level
Scores on PC 2 (19.04%)
Scores on PC 1 (30.28%)
b)
Bruised samples
Control samples
99.7% Confidence Level
95% Confidence Level
G21 9th day
G21 8th day
G21 7th day
G21 6th day
G21 3rd day
G21 2nd day
G21 1st day
G21 5 hours
G20 9th day
G20 8th day
G20 7th day
G20 6th day
G20 3rd day
G20 2nd day
G20 1st day
Scores on PC 2 (12.61%)
Scores on PC 1 (15.34%)

Figure 3

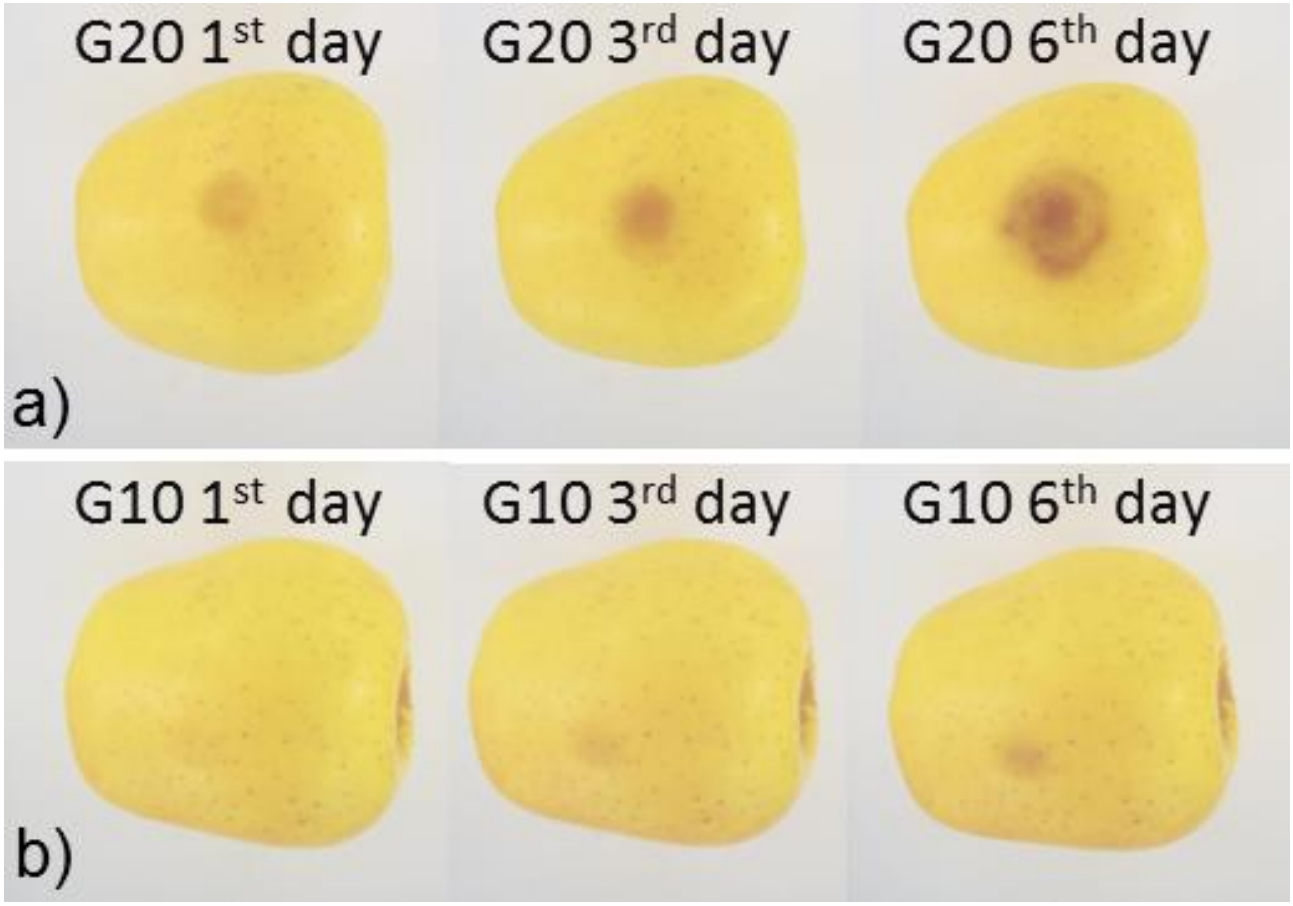

G20 1st day
G20 3rd day
G20 6th day
a)
G10 1st day
G10 3rd day
G10 6th day
b)

Figure 4

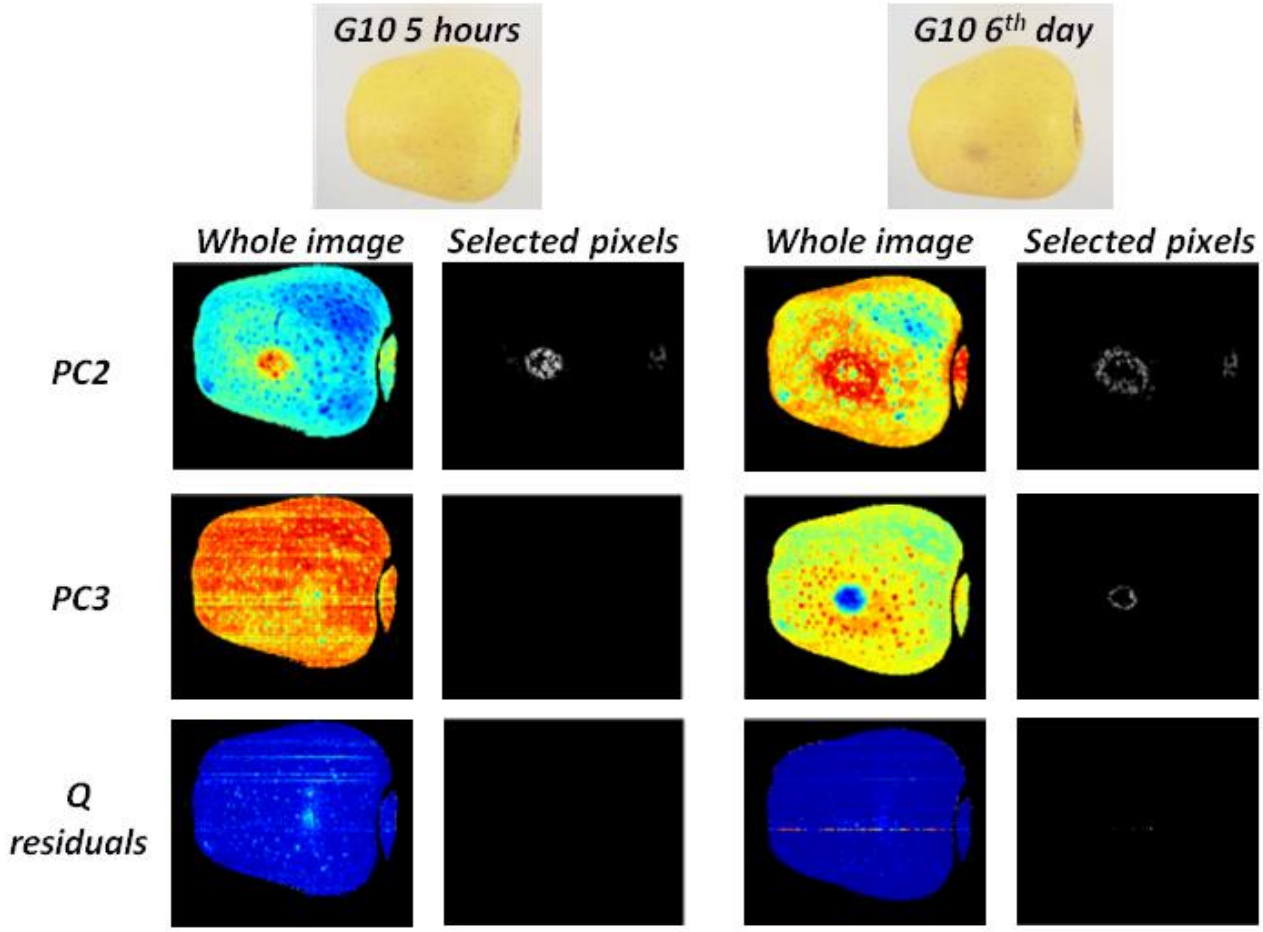

G10 5 hours
G10 6th day
Whole image
Selected pixels
Whole image
Selected pixels
PC2
PC3
Q residuals

Figure 5

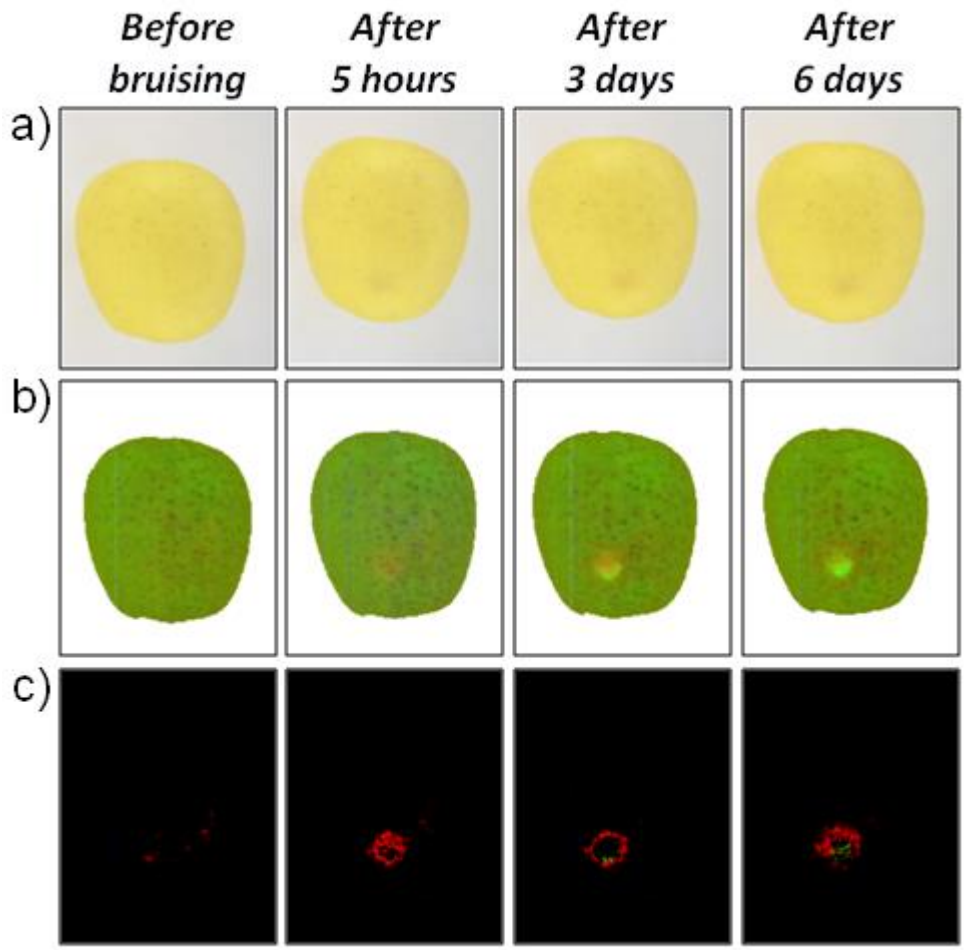


Figure 6

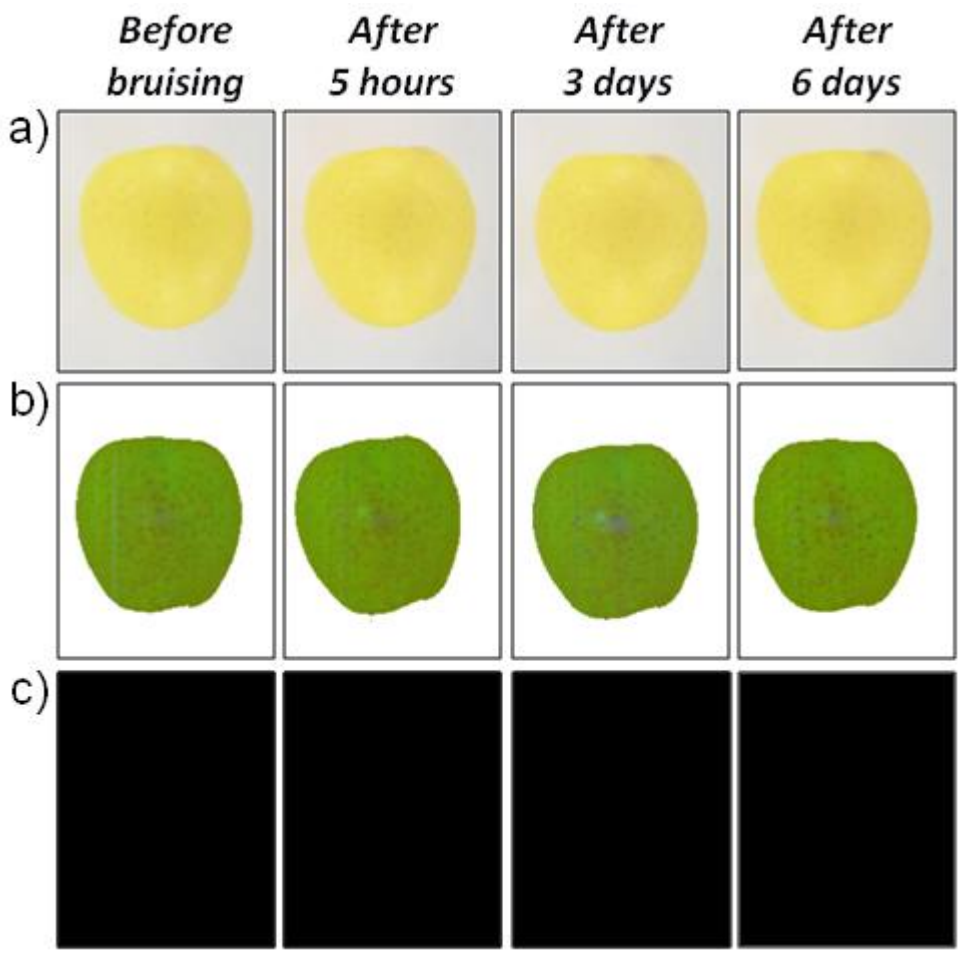


Figure 7

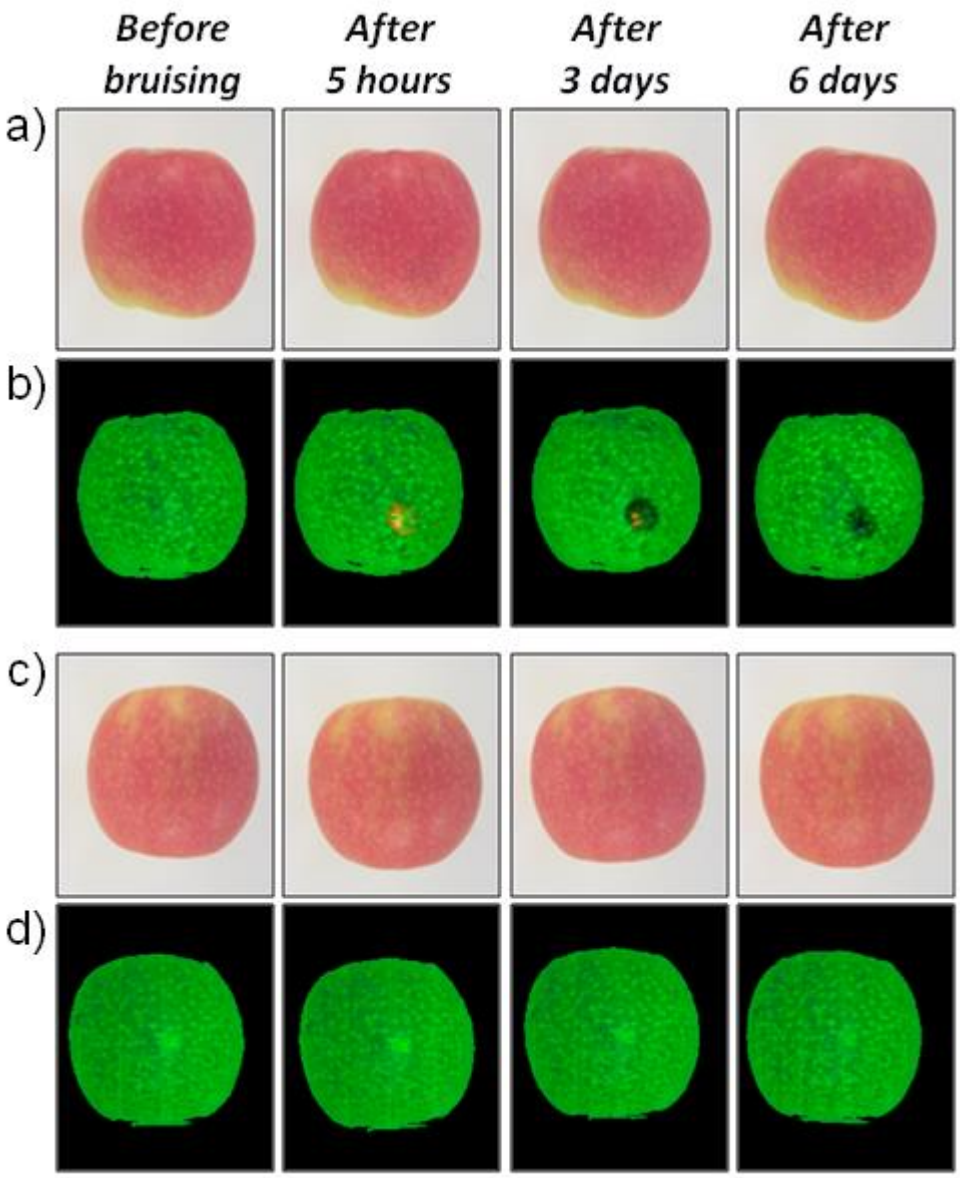
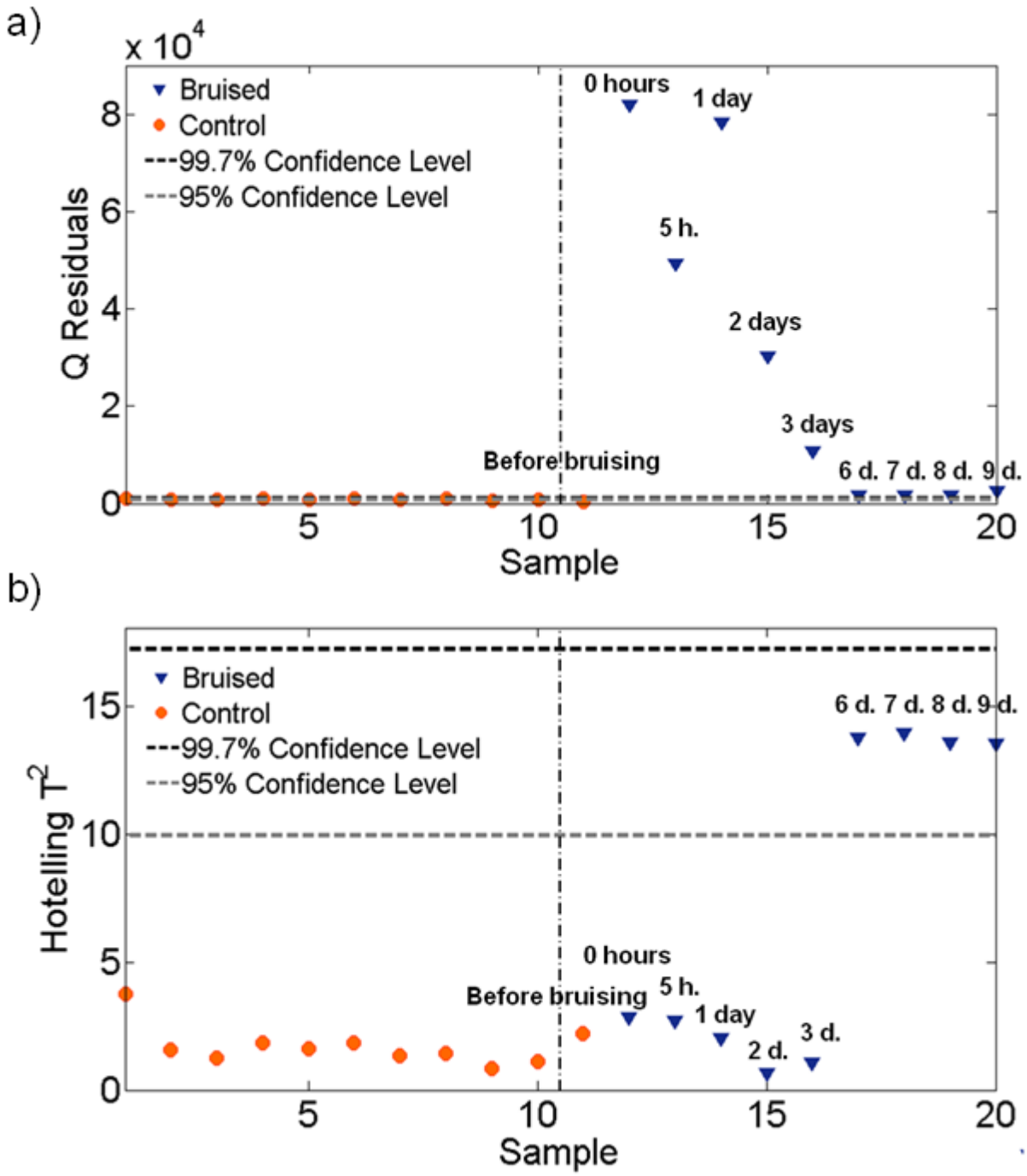


Figure 8


a)
x 10^4
Bruised
Control
---99.7% Confidence Level
---95% Confidence Level
Q Residuals
0 hours
1 day
5 h.
2 days
3 days
Before bruising
6 d. 7 d. 8 d. 9 d.
Sample
b)
Bruised
Control
---99.7% Confidence Level
---95% Confidence Level
Hotelling T^2
6 d. 7 d. 8 d. 9 d.
0 hours
Before bruising
5 h.
1 day
2 d.
3 d.
Sample