Examination of the quality of spinach leaves using hyperspectral imaging technique

B. Diezma, L. Lleó, J.M. Roger, A. Herrero-Langreo, L. Lunadeid, M. Ruiz-Altisent

1   # Examination of the quality of spinach leaves using

2   # hyperspectral imaging technique

3

4   Belén Diezma[a]*, Lourdes Lleó[b], Jean Michel Roger[c], Ana Herrero-Langreo[c], Loredana

5   Lunadei[d], Margarita Ruiz-Altisent[a]

6

7   [a]Physical Properties and Advanced Techniques in Agrofood, LPF-TAG. Rural

8   Engineering Dept., Technical University of Madrid, c/Ciudad Universitaria s/n, 28040

9   Madrid, Spain

10   *tel: +34 913365623; fax: +34 913365845; e-mail address: belen.diezma@upm.ex

11   [b]Physical Properties and Advanced Techniques in Agrofood, LPF-TAG. Rural

12   Engineering Dept., Technical University of Madrid, c/Ciudad Universitaria s/n, 28040

13   Madrid, Spain

14   [c]CEMAGREF, 361 rue Jean-François Breton BP 5095, 34196 Montpellier Cedex 5,

15   France

16   [d]Rural Engineering Dept., Technical University of Madrid, c/Ciudad Universitaria s/n,

17   28040 Madrid, Spain

18

19

20

1

Abstract

3   The present research is focused on the application of hyperspectral images for the

4   supervision of quality deterioration in ready to use leafy spinach during storage

5   (Spinacia oleracea). Two sets of samples of packed leafy spinach were considered: a)

6   first set of samples was stored at 20°C (E-20) in order to accelerate the degradation

7   process; these samples were measured the day of reception in the laboratory and after

8   two days of storage; b) second set of samples was kept at 10°C (E-10); the

9   measurements were taken throughout the period of storage, beginning the day of

10   reception and repeating the acquisition of images three, six and nine days after.

11   Twenty leaves per testing date of test were analyzed. Hyperspectral images were

12   acquired with a push-broom CCD camera equipped with a spectrograph VNIR (400 to

13   1000 nm). Calibration set of spectra was extracted from E-20 samples, containing three

14   classes of degradation: class A (optimal quality), class B and class C (maximum

15   deterioration). Reference average spectra were defined for each class. Three models,

16   computed on the calibration set, with a decreasing degree of complexity were

17   compared, according to their ability for segregating spinaches at different quality

18   stages: spectral angle mapper distance (SAM), partial least squares discriminant

19   analysis models (PLS-DA), and a non linear index (Leafy Vegetable Evolution, LEVE)

20   combining five wavelengths included among the previously selected by CovSel

21   procedure. In sets E-10 and E-20, artificial images of the membership degree

22   according to the distance of each pixel to the reference classes, were computed

23   assigning each pixel to the closest reference class. The three methods were able to

24   show the evolution of the leaves along the time.

25

1

2   Keywords: Spinach leaves; Nondestructive assessment; Hyperspectral imaging;

3   Multivariate Analysis

4

1

1. Introduction

Fresh-cut fruit and vegetables, initially called minimally processed or lightly processed products, can be defined as any fresh fruit or vegetable that has been physically modified from its original state (by peeling, trimming, washing and cutting) to obtain 100% edible product that is subsequently bagged or prepackaged and kept in refrigerated storage (Martín-Belloso and Soliva-Fortuny, 2011).

The major preservation techniques applied to prevent or delay spoilage are chilling storage and modified atmosphere packaging, combined with chemical treatments and application of moderate heat treatments such as hot water or steam (Martín-Diana et al., 2007). Innovative sanitizers and preservation techniques are being introduced. However, change from use of conventional to innovative treatments requires knowledge of the benefits and restrictions as well as a practical outlook. These techniques must satisfy the consumers and maintain a balance between sensory and quality. Consequently, the sector is asking for innovative, fast, cheap and objective techniques to evaluate the overall quality and safety (or some of the specific quality parameters) of fresh-cut products in order to obtain decision tools for implementing new packaging procedures. Hyperspectral imaging technique could be a first approach.

In recent years, hyperspectral imaging technique has been regarded as an analytical tool for analyses conducted for quality evaluation of food products in research, control, and industries. The hyperspectral imaging system allows integrating spectroscopic and imaging techniques to enable direct identification of different components or quality characteristics and their spatial distribution in the tested sample (ElMasry et al., 2010). Measurement of the optical properties of food products has been one of the most successful nondestructive techniques for quality assessment to provide several quality details simultaneously. In these spectroscopic techniques, it is possible to obtain

4

1   information about the sample components based on the light absorption of the sample,

2   but it is not easy to extract the information on position/location. i. The combination of

3   the strong and weak points of visible/near-infrared spectroscopic techniques and vision

4   techniques is the hyperspectral imaging technique. Because hyperspectral imaging

5   techniques overcome the limits of spectroscopic techniques and vision techniques,

6   they have emerged as a powerful technique in agricultural and food systems. Based on

7   hyperspectral imaging techniques, multispectral imaging system can be built for real-

8   time implementations (Lee et al., 2005). It involves measuring the intensity of diffusely

9   reflected light from a surface at one or more wavelengths with relatively narrow band

10   passes. Since image data are considered two-dimensional, by adding a new dimension

11   of ''spectrum'' information, the hyperspectral image data can be perceived as a three-

12   dimensional data cube (Chao et al., 2001). Hyperspectral imaging, like other

13   spectroscopy techniques, can be carried out in reflectance, transmission or

14   fluorescence modes, being the reflectance the most usual mode.

15   One of the main challenges in the hyperspectral vision is the management and analysis

16   of large and complex databases to extract relevant information contained in them

17   (Fernandez Pierna et al., 2010). The starting point for this are the methods of spectral

18   pre-processing (normalization, smoothing, differentiation, etc.) and multivariate analysis

19   (correlation techniques, principal component analysis, discriminant analysis, etc.)

20   traditionally applied to spectroscopy (Gowen et al., 2007). In the case of hyperspectral

21   vision these procedures can be applied to the whole image or to sub-populations of

22   pixels representative of the variability of the samples. The projection of the whole

23   images onto new spaces generated by multivariate analysis or the computation of

24   indexes based on some wavelengths generate virtual images that must be analyzed

25   searching for similarity.

26   On hyperspectral imaging, spectral pre-processing techniques are applied to remove

27   non -chemical biases from the spectral information, e.g. scattering, temperature

1   influence (Hernandez-Sanchez et al., 2003) or device calibrations (Gowen et al., 2007);

2   (Fearn et al., 2009); (Rinnan et al., 2009). Geometric pre-processing methods are

3   applied to correct spectral data from drift baseline, non linearity, curvilinearity, as well

4   as additive and multiplicative effects. Some of these methods are the standard normal

5   variate transform (SNV), smoothing and differentiation (Zeaiter et al., 2005). However

6   pre-processing could also destroy valuable information and therefore these methods

7   should be applied carefully.

8   Researchers are often interested in finding the most relevant few wavelengths that

9   could influence the quality evaluation of the product (ElMasry et al., 2007). Several

10   strategies have been performed to select few wavelengths from hyperspectral data in

11   order to design the best adapted multispectral imaging system, such as general visual

12   inspection of the spectral curves and correlation coefficients (Keskin et al., 2004),

13   analysis of spectral differences from the average spectrum (Liu et al., 2003), stepwise

14   regression (Chong and Jun, 2005), orthogonal projection methods such as principal

15   component analysis (Mehl et al., 2004); (Xing and De Baerdemaeker, 2005) or CovSel

16   method (Roger et al., 2011), a priori knowledge of pigment spectral signatures and/or

17   comparison of different spectral indexes, (Lu and Peng, 2006; Lleo et al., 2011; Mehl et

18   al., 2004; Merzlyak et al., 2003; Zude, 2003)

19   Image analysis can be implemented in several food products characterization (Du and

20   Sun, 2006); (Sun, 2010); (Cubero et al., 2011); Gowen et al. (2007). Regarding

21   hyperspectral in fruits and vegetables safety, there are many published results

22   (ElMasry et al., 2012); some of them are related to the detection of sour skin

23   (Burkholderia cepacia) in infected onions (Wang et al., 2012), fecal contamination in

24   apples (Liu et al., 2007) and contamination by fungi in maize (Del Fiore et al., 2010).

25   However, few applications of hyperspectral vision system have been focused on ready

26   to use leafy vegetables until now. As an example, there is a research concerning rapid

1   detection of Escherichia coli contamination in packaged fresh spinach using

2   hyperspectral imaging (Siripatrawan et al., 2011).

3   The present work aims at proposing a) the development and optimization of a

4   hyperspectral vision system for monitoring the evolution of minimally processed

5   spinaches during storage, b) the establishment of multivariate analysis procedures to

6   identify and classify the damages that occur over the lifetime of the product, like the

7   changes of pigments and structure and incipient rot. Models with decreasing degree of

8   complexity are compared according to their ability for segregating spinaches at

9   different quality stages. In a first step all the wavelengths of the spectra were taken into

10   account, computing spectral angle mapper distance to the reference spectra (SAM), in

11   a second step partial least squares discriminant analysis models (PLS-DA) were

12   performed, and finally the CovSel procedure for variables selection was implemented

13   and a non linear index based on a subset of the selected wavelengths is proposed.

14 2. Materials and Methods

15   *2.1. Sample collection*

16   Two sets of samples of spinach (Spinacia oleracea) minimally processed, packed in

17   sealed plastic bags (200 g) were considered for further tests and analysis, with the goal

18   of generating sufficient variability in the rate of deterioration. A single leaf of spinach

19   was considered as one sample unit. Two sets of experiments were conducted on two

20   sets of samples. In the first experiment, packed leafy spinach purchased from a local

21   wholesale produce distributor was stored at 20°C (E-20), in order to accelerate the

22   degradation process, and measured the day of reception of samples in the laboratory

23   (T0) and after 2 days of storage (T2). In the second experiment, other set of samples

24   with the same origin (E-10), was left at 10°C; the measurements were taken throughout

25   the period of storage, beginning the day of reception (T0) and repeating the acquisition

26   of images 3, 6 and 9 days after (T3, T6 and T9 respectively). 20 leaves per day of test

1   were analyzed; in every date, leaves coming from two different packages were

2   considered. Unopened bags of spinach were used for each date. Figure 1 shows RGB

3   images of representative leaves of the four days of measurements of the E-10 test.

4   *2.2. Hyperspectral imaging*

5   The hyperspectral vision system consisted of a push-broom EMCCD Luca-R camera

6   (AndorTM Technology, Northern Ireland) equipped with a spectrograph Hyperspec®

7   VNIR (spectral range: 400 to 1000 nm;). The spectral binning was configured to obtain

8   189 wavelengths (spectral resolution 3.17 nm). The acquisition and the storage of the

9   images were made through specific software (Headwall Hyperespec®, Headwall

10   Photonics Inc, USA). The illumination was provided by two halogens lamps with

11   regulated and variable intensity. Each individual leaf was placed on a platform that

12   moved under the camera (MoCo DC motor controller, Micos, USA). The sample was

13   scanned line by line according to the movement (push-broom system). The spatial

14   resolution was 260 µm.

15   The leaves were placed on a black platform to acquire the greater dimension (direction

16   of the main nerve) and a width of 20 mm. The images were acquired from the beam of

17   the leaves. Once the raw images were acquired, the corresponding relative reflectance

18   hypercube was computed, containing the relative reflectance spectrum of each pixel of

19   the image with respect to a reference (mean spectrum of a barium sulfate white

20   reference).

21   *2.3. Calibration set*

22   Calibration set was extracted from E-20 samples, which contained the deterioration

23   extreme states. Three classes of degradation were identified on those images: Class A,

24   optimal quality, fresh tissue from first day (T0); Class B, from non-fresh tissues

25   belonging to samples of the second date of measurements (T2), but without visible

26   deterioration; Class C, regions with visible deterioration on T2 samples. On the

1   hyperspectral images, areas belonging to different leaves were manually selected and

2   their pixels assigned to one of the three defined classes. All the spectra of those

3   regions composed the calibration set that was constituted by 3600 spectra (1200

4   spectra of each class). The average spectrum of each class was computed.

5   This calibration set was considered for the computation of PLS-DA model and the

6   application of CovSel procedure.

7   *2.4. SAM distance*

8   One of the most applied strategies for material mapping is the use of similarity

9   measures. Frequently, studies make use of a deterministic similarity measure to

10   compare an unknown pixel spectrum with a library of reference spectra (Keshava,

11   2004). Spectral Angle Mapper (SAM), is a common distance metric, which compares

12   an unknown pixel spectrum t to each spectrum r of the K considered spectra of

13   reference, and assigns t to the class reference having the smallest distance. The

14   reflectance spectra of individual pixels can be described as vectors in an n-dimensional

15   space, where n is the number of spectral bands. Each vector has a certain length and

16   direction. The length of the vector represents brightness of the pixel, while the direction

17   represents the spectral feature of the pixel. Variation in illumination mainly affects the

18   length of the vector, while spectral variability between different spectra affects the

19   angle between their corresponding vectors (Kruse et al., 1993). The more similar the

20   two spectra are, the smaller the spectral angle between them. The spectral angle can

21   have values between 0 and $\pi/2$ and is calculated by the formula derived from the inner

22   product of two vectors, $\Theta = \cos^{-1}(\sum t_i r_i /(\sum t_i^2 \sum r_i^2)^{1/2})$, where the sum is extended to all

23   the spectral bands, t the reflectance of the actual spectrum and r the reflectance of the

24   reference spectrum. The standard Spectral Angle Mapper (SAM), available in most

25   image processing software packages, uses the average spectrum of each region of

26   interest as spectrum reference (Luc et al., 2005).

1   SAM distances were computed between the average spectrum of each quality class

2   and each anonymous spectrum of the images belonging to E-20 and E-10 sets. From

3   the distance values the membership degree to each class was computed in order to

4   obtain virtual images with pixels having values between 0 and 1. Each pixel was

5   assigned to the reference class to which it computed the maximum membership

6   degree. Artificial images of the assignation of the pixels to the classes were computed.

7   *2.5. Partial Least Square-Discriminant Analysis (PLS-DA)*

8   In spectroscopy, in order to create calibration models, a (generally) linear relationship

9   is sought between spectra and reference measurements (responses). Until few years

10   ago discrimination from spectra was less common, in spite of a great number of

11   potential applications: defect detection, object or product recognition, outlier detection,

12   etc. In discrimination, the variable to predict is qualitative, i.e. it takes its values in an

13   unordered discrete set. Except in the simple case of 2 classes, which is analogous to a

14   quantitative response case, the factorial regression methods are unsuited. The

15   discriminant methods, which solve the issues of dimensioning and conditioning,

16   proceed similarly to factorial regression: a classical discriminant analysis (DA) is

17   performed on latent variables, provided either by a principal component analysis (PCA-

18   DA), or by a PLS between the spectra and the class membership (PLS-DA). As far as

19   regression is concerned, PLSR is generally more powerful than PCR, since the latent

20   variable design takes into account the relationship between the spectra variables and

21   the responses. Due to the same reason, in the discrimination case, PLS-DA is

22   generally more efficient than PCA-DA (Barker and Rayens, 2003).

23   Partial least square and discriminant analysis, PLS-DA, was applied on the calibration

24   set (n=3600) between two matrices X and Y. The spectra constituted the X matrix, n x

25   p, being n the total number of spectra of the calibration set, and p the number of

26   wavelengths. Y matrix had n rows and three columns corresponding to the three

1   classes A, B, C of the calibration set. In the Y matrix, each pixel was codified by three

2   numbers corresponding to "membership values", one for each class, with value of 0 or

3   1; e.g., a response encoded (0 1 0) means that the sample belongs to class B. Firstly,

4   a PLS-2 computed k latent variables, which maximize the covariance between X and Y,

5   transforming X in a n x k matrix. On this new reduced matrix, a classical linear

6   discriminant analysis was performed to determine the most discriminant subspace.

7   Each sample was then projected onto this subspace, yielding scores.

8   The prediction results are displayed in a confusion matrix, presenting the number of

9   samples assigned to each class. These results allow the computation of the error of the

10 model (Roussel et al., 2003). A cross validation process was applied, splitting the

11 population in ten parts. The resulting errors allowed us to determine the optimal value

12 of latent variables. Further the procedure generates discrimination vectors, which allow

13 the projection of any anonymous individual on the space generated by these vectors

14 obtaining their new coordinates (scores) in this space.

15 The obtained model from calibration set was applied to the totality of the pixels of the

16 hyperspectral images E-20 and E-10, and therefore the artificial images of the scores

17 were computed. The distance of Mahalanobis was computed between the new

18 coordinates of the pixel (scores) and the centroid of each quality class (A, B, C) of the

19 calibration set. Each pixel of the images was assigned to the class with the maximum

20 membership degree computed according to the distance of Mahalanobis.. Artificial

21 images of the assignation of the pixels to the classes were computed.

22 *2.6. CovSel computation and LEVE index*

23 CovSel is an algorithm which performs variable selection step by step on the basis of

24 their global covariance with all the responses. Each variable selection is followed by

25 the projection of the data orthogonally to the selected variable. Therefore the selected

26 variables are independent each other as much as possible. At the exit of this procedure

11

1   a set of a predetermined number k of variables is generated. CovSel can be used for

2   multi-response regression or for discrimination as well.

3   CovSel algorithm was applied to the spectra of calibration set aiming at obtaining the

4   best combination of wavelengths regarding the process of deterioration. For that, the

5   same matrices X, Y than in the mentioned above case of PLS-DA were employed for

6   such selection. As a result, a group of k wavelengths of X were selected producing a

7   smaller matrix X* of dimension n rows and k columns.

8   Further, a non linear index was proposed, resulting of the combination of some of the

9   most relevant wavelengths selected by CovSel. Artificial images containing the values

10   of this index were obtained for E-10 and E-20 sets. Analogously to the procedure

11   explained in previous paragraphs, the distances and the membership degrees of the

12   pixels to each quality class was computed, considering the centroid values of the index

13   for each class. The corresponding artificial images of the assignation of pixels were

14   also obtained for this procedure.

15   *2.7. Comparison between procedures*

16   Several procedures were carried out in order to compare the performance of each

17   proposed procedure and its concordance in the assignation of pixels and leaves.

18   The artificial images of membership degrees to class C obtained by the procedures

19   proposed were studied by means of Analysis of Variance (ANOVA) in order to compare

20   them for the detection of the evolution of the leaves. All pixels of leaves were pulled

21   together for the first day (first group n=197,660 corresponding to T0 of E-10 set) and

22   for the last day (second group n=182,531 corresponding to T9 of E-10 set).

23   Additionally, concordance in the assignation of pixels between methods was tested

24   taking into account the percentage of pixels assigned to the same class by the different

25   procedures in E-20 and E-10 sets.

1 Finally, each leave was assigned to the class for which it presented the maximum

2 relative frequency (e.g.: number of pixels of Class X/total number of pixels of the leave)

3 according to SAM, PLS-DA and CovSel+LEVE index; so, three assignations (to

4 classes A, B or C) were computed and compared for each sample.

5 All the analyses were made by means of routines generated in Matlab 7.0

6 (MathWorks).

3. Results and discussion

8 Preliminary analysis (data not shown) showed that the best performance of the models

9 was obtained considering raw spectra; further on, no pre-processing techniques were

10 applied to the spectra before the multivariate data analysis.

11 Figure 2a shows the average spectra of the 3 classes A, B, C calculated on the

12 calibration set (n=1200 for each class). The reflectance spectrum corresponding to

13 class C presents a general decrease in the whole range from 450 till 1000 nm, being

14 particularly high in the region of NIR from 700 till 1000 nm. In addition, the slope of the

15 spectrum between 710 and 900 nm is increasing from class A to class C, being this

16 change especially relevant for the most deteriorated samples (class C). Regarding the

17 VIS range of the spectra, a general decrease in reflectance values from the soundest

18 to the most deteriorated pixels was observed. The spoilage in these leaves induced

19 spectral changes different from the ones caused by the aging and senescence typically

20 described in leaves by remote sensing researches (Asner, 1998); (Liy et al., 2010).

21 These authors stated that aging of vegetation induces a decrease in water content and

22 in foliar pigments; as a result, global reflectance increases in all the visible and near-

23 infrared range. The high decrease in global reflectance in the spectra of class C could

24 indicate that the main effect of the evolution is not the water loss, suggesting that the

25 spoilage suffered by packed leaves is different from the typical deterioration observed

26 in vegetation by remote sensing. In fact, the deteriorated areas observed in packed

1   leaves seemed to be smashed and moist (Figure 1, treatment T9) whereas in remote

2   sensing the areas affected by deterioration show yellowness, loss of water, chlorophyll

3   and some other pigments, but not a moist or smashed aspect.

4   Jacquemoud and Baret (1990) stated that the interaction of the light with plant leaves

5   depends on the chemical and physical characteristics of the tissues. Absorption is

6   essentially a function of changes on molecules (chlorophyll a and b, carotenoids,

7   water... etc) and is responsible for the holes in the reflectance spectrum. The refractive

8   index discontinuities within tissue induce scattering which acts on the global trend of

9   the reflectance spectrum. The internal structure of the leaves affects the scattering of

10   the light, therefore it controls the reflectance of the whole spectrum, being clearer this

11   effect where the absorption is low, especially in NIR.

12   The main deterioration effect in the present study is probably the destruction of the

13   internal structure of the tissue (wall cells) inducing an increase in free water. It could be

14   supposed that refractive index discontinuities decrease within the leaves, leading to a

15   reduction of the light scattering and consequently a decrease in the reflectance of the

16   spectra. This decrease in reflectance was observed on class C spectra, being

17   especially pronounced in the NIR region (above 700 nm) according to Jacquemoud

18   and Baret (1990). Additionally the presence of free water facilitates deeper penetration

19   of the light in the tissue, increasing the path length and making more probable the

20   absorption of the light during its travel by the pigments, which decreases the

21   reflectance, mostly in the visible range. Around red edge (670-720 nm) both

22   phenomena occurred: a) the general decrease of global reflectance in NIR region and

23   b) the specific decrease in reflectance at red edge induced by the absorption of the

24   chlorophyll. The high slope at NIR region (700–900 nm) in the average spectrum of

25   class C (Figure 2a) could be caused by the differences in the intensities of these two

26   mentioned phenomena in the red edge range and in the rest of the NIR range where

27   the effect of the absorption could be considered negligible. The average spectrum of

14

1   class B showed an intermediate stage that could denote an incipient phase of the

2   degradation of the structure of the leaves.

3   *3.1. PLS-DA / CovSel*

4   A total of 6 latent variables were chosen for the PLS-DA by means of a cross-validation

5   procedure (Geisser, 1993). In addition, it has been checked that the score images

6   showed less noise with 7 than with 8 latent variables; in images obtained considering 8

7   latent variables structure details of the leaves, such as veins, were not clearly identified

8   (image not shown).

9   Figure 2b shows the score plot of the calibration set. Each point of this graph

10   corresponds to the projection of one spectrum of this set onto the 2-dimensional

11   discriminant space spanned by the discriminant vectors of the PLS-DA (vp_1, vp_2). Thus,

12   the abscissa and the ordinates of the scores can be interpreted with regard to the

13   combination of spectra features (peaks, slopes, etc.) and the shape of the discriminant

14   vectors. However, PLS-DA discriminant vectors are normalized but not orthogonal; in

15   our case they are almost collinear, making their interpretation difficult. Trying to

16   improve the interpretation of the discriminant vectors, a new vector (u) was computed

17   by orthogonalizing vp_2 against vp_1: $u = (I - vp_1^T vp_1) vp_2$. Figure 2c and Figure 2d

18   present the two vectors vp_1 and u, respectively.

19   The score plot of Figure 2b shows that vp_1 clearly orders the 3 classes from C

20   (negative scores) to A (positive scores), through B (null and little positive scores). On

21   this axis, classes A and B are overlapping. The discriminant vector vp_1 (Figure 2c)

22   shows three main features that explain that ordering:

23   -   A large positive peak around 550 nm contributes positively to the score for

24   spectra presenting a high reflectance in this zone, as those of classes A and B,

25   meaning that the leaves are less and less green while aging

15

1 -   A sharp positive peak at 700 nm computes a part of the scores which

2 corresponds to the position of the red edge, which collapse when the leaves age

3 -   The combination of a negative peak at 720 nm and a positive zone around 760

4 nm computes the slope of the red edge, which decreases when the leaves age

5 The second axis of the score plot (Figure 2b) discriminates the low scores of classes A

6 and C from the high scores of the class B. The vector u (Figure 2d), which represents

7 the part of vp2 not included in vp1, shows also three main features:

8 -   A negative zone between 400 and 530 nm combined with a positive one

9 between 530 and 660 nm realizes a colorimetric balance between blue/green and

10 yellow/red colour of the leaves.

11 -   A series of peaks and holes in the NIR region, between 800 and 1000 nm. This

12 shape cannot be finely analyzed, but may be due to subtle changes in chemicals.

13 Indeed, this spectral zone is affected by sugars, proteins and water (Osborne and

14 Fearn, 1986).

15 -   A negative peak at 707 nm, which cannot be directly interpreted relatively to

16 score plot.

17 The six wavelengths chosen by the CovSel algorithm were: 752, 902, 538, 717, 430

18 and 647 nm, as reported on Figure 2a. With the perspective of a multispectral

19 approach, a new non linear index was tested based on the most relevant wavelengths

20 selected by CovSel and on the loadings of the first discriminant vector of PLS-DA.

21 Leafy Vegetable Evolution index (LEVE index) focuses on two spectral regions: NIR

22 region (750 -900 nm) and VIS region (506 – 614 nm). LEVE index is defined by the

23 expression $((R_{900}-R_{750})/(R_{900}+R_{750}))/((R_{519}+R_{646})/R_{538})$, where R_x corresponds to the

24 relative reflectance at x nm.

1   The numerator of LEVE index located at NIR region computed a kind of normalized

2   slope between 750 and 900 nm, including the two first wavelengths selected by

3   CovSel. The denominator located at VIS region (including two of the wavelengths

4   selected by CovSel) was an approximation to the minus normalized second derivative

5   at 538 nm according to the proposal by Lleó et al. (2011). LEVE index increased from

6   class A to class C.

7   *3.2. Virtual images of membership degree*

8   In Figure 3 and Figure 4, the virtual images of membership degree to class C are

9   showed for the three described classification procedures (i.e. based on SAM, PLS-DA

10   and CovSel plus LEVE index) applied on samples of set E-20 and E-10 respectively.

11   All the procedures identified the regions belonging to class C very accordingly. More

12   accentuated evolution could be observed on samples from set E-20, which is

13   concordant with the storage conditions. Inside each leaf it can be observed pixels with

14   a very different rate of deterioration. In addition, leaves belonging to the same storage

15   conditions presented different rates of degradation. This fact is especially evident on

16   the second date (second row of images of Figure 4) of E-20, where the third, fourth and

17   sixth leaves (from left of the row) contained little regions close to class C, whereas the

18   last five leaves showed patches of very deteriorated pixels. Regarding E-10, on the

19   third line at Figure 5 (third measurement date) four leaves presented an appearance

20   similar to the corresponding to the first date of measurements. That suggests that the

21   evolution rate of the leaves is very heterogeneous inside commercial batches, which

22   could be explained by the original condition of the vegetal material on harvest and

23   postharvest chain (Medina et al., 2012).

24   For ANOVA computation the membership degrees to class C obtained by the three

25   methods was considered, comparing the first date of measurements of E-10 spectra

26   with the last date (n=197,660 pixels for first date; n= 182,531 pixels for last date). All

1   methods sense the evolution with a better performance for PLS-DA (F value 582,705 at

2   p level 0), which could be expected taking into account that this model includes all the

3   wavelengths and it is based on the covariance matrix between the spectra (X) and the

4   visual assignation to the deterioration classes A, B and C (Y). Similar F values were

5   obtained for PLS-DA and LEVE index, 333,099 and 320,215 respectively.

6   *3.3. Virtual images of assignation to quality class*

7   As mentioned before all pixels of the leaves were assigned to their closest class

8   according to membership degree; the assignation was to the class with highest

9   membership degree. Figure 5 shows, such as example, the virtual images for

10   assignation to the classes according to LEVE index. Similar pattern of distribution of

11   regions in the three artificial images of pixels classification was observed in calibration

12   (E-20) and validation (E-10) sets of leaves (data not shown). The analysis of this

13   observation was performed quantifying the percentage of pixels with the same

14   assignation for the different procedures (Table 1).

15   The assignation of the pixels to one of the three classes showed high concordance

16   when comparing the three methods of analysis. The highest levels of concordance

17   occurred between the assignations based on SAM distance and on LEVE index (93%

18   and 87% for E-20 and E-10 sets respectively), whereas the most discrepant

19   assignations corresponded to the comparison between PLS-DA and LEVE index (76%

20   and 75% for E-20 and E-10 sets). Consequently, LEVE index could be employed

21   instead of PLS-DA or SAM distance, presenting the advantage of using only 5

22   wavelengths and not all the spectra, which could be integrated in a simpler

23   multispectral vision system.

24   In addition, classification of leaves into the class in which the relative frequency of

25   pixels is the highest reinforced the fact of concordance between methods (Table 2): the

26   82.5% and the 92.5% of leaves of E-20 and E-10 respectively, were classified in the

1 same category by the three methods. The highest level of coincidence was achieved

2 between SAM distance and LEVE index with the 95% (E-20) and 97.5% (E-10) of the

3 leaves assigned to the same quality class. Regarding the discrepancies, LEVE index

4 assigned higher membership degree to class C which allowed segregating better

5 between class C and class B. Figure 6 presents the images of membership degree to

6 class C of one leave of E-10 set assigned to class B according to SAM distance (left)

7 and to class C according to LEVE index (right). It suggests that if the identification of all

8 the possible problematic leaves is crucial, it could be better to use LEVE index instead

9 of SAM distance.

10 ## 4. Conclusions

11 The main differences observed in the reflectance spectra over storage time was an

12 overall decline in the intensity of VIS and NIR ranges and an increase in the slope of

13 between 710 and 900 nm, which was particularly pronounced for the most deteriorated

14 quality category. A hypothesis has been formulated for the explanation of these effects:

15 the destruction of the internal structure of tissue cause an increase in free water

16 content and consequently the decrease of the refractive index discontinuities within the

17 leaves and the reflectance of the spectra. Furthermore, the same fact also facilitates

18 the deeper penetration of light increasing the probability of being absorbed by pigments

19 in the VIS range, which also decrease the reflectance. The differences in the intensities

20 of these two phenomena (increase in free water content and increase in absorption of

21 the light by pigments) in the red edge could explain the high slope at the NIR region in

22 the average spectrum of the most deteriorated class.

23 The performed tests showed the ability in discriminating between different storage

24 periods of virtual images resulting from the application of three analytical techniques

25 (SAM, PLS-DA, CovSel and non linear index) to the hyperspectral images.

19

1   Due to the high intra-date and intra-leaf variability, the study of a first calibration set

2   composed according to dates was unsuccessful. Therefore, a calibration population

3   (3600 spectra) was defined by a supervised selection visually of pixels according to

4   their evolution stage.

5   The implementation of different multivariate techniques on the calibration sub-

6   population generated classification models in three categories that responded to the

7   main states of the evolution, followed by the product in the tests. The projection of the

8   hyperspectral images onto the generated subspaces and the assignation of each pixel

9   to one of the defined categories, allowed the identification of regions with different

10   states of evolution in the leaves. The system made possible to determine the

11   percentages of the areas of these regions and to establish decision rules about the

12   overall quality of the leaf, which is a clear advantage compared to the

13   spectrophotometric methods that analyze a small area of the sample.

14   The method CovSel, dedicated to the problem of variable selection for highly

15   multivariate data, allowed the identification of the most relevant (and independent) 10

16   wavelengths, which were the starting point for the definition of a non linear index, LEVE

17   index, combining 5 of the 10 wavelengths selected by CovSel, achieving similar results

18   that SAM distance analysis and PLS-DA, these last two procedures using the complete

19   spectra. More than the 95% of the leaves were classified into the same quality class by

20   SAM distance and by LEVE index.

Acknowledgements

1

2   Table 1.Percentage of concordance between methods in the assignation of the pixels

3   of leaves to the classes A, B and C for E-20 and E-10 samples

4   Table 2. Percentage of concordance between methods in the assignation of the leaves

5   to the three classes A, B and C for E-20 and E-10 samples.

6   Figure 1. Representative leaves of the E-10 set along the dates of measurements.

7   Figure 2. a: Average spectra of the 3 classes A, B and C (n=3600 spectra in total);

8   vertical lines indicate the 6 wavelengths selected by CovSel; b: scores of the 3 classes

9   calculated by the PLS-DA on the learning set; c: first discriminant vector (vp1); d:

10 residual of the second discriminant vector to the first one (u).

11 Figure 3. Visualization of the membership degree to class C, in E-20 samples, for the

12 different proposed procedures (each row of the table). Redder pixels are closest to

13 class C (more deteriorated areas) whereas deep blue pixels are far from class C

14 (healthy tissues). Each image contains the leaves corresponding to first date, first row,

15 ant last date of measurements, second row

16 Figure 4. Visualization of the membership degree to class C, in E-10 samples, for the

17 different proposed procedures. Redder pixels are closest to class C (more deteriorated

18 areas) whereas deep blue pixels are far from class C (healthy tissues). Each image

19 contains the leaves corresponding to the four dates of measurements: each row

20 corresponds to one of the dates of measurements.

21 Figure 5. Visualization of the assignation of each pixel to one of the three quality

22 classes considering the maximum membership degree. Blue color corresponds to class

23 A, orange color to class B and color brown to class C.

1   Figure 6. Images of membership degree to class C of one leave of E-10 set assigned

2   to class B according to SAM distance (left) and to class C according to LEVE index

3   (right).

4   ## References

5   Asner G.P. (1998) Biophysical and biochemical sources of variability in canopy

6   reflectance. Remote Sens. of Environ. 64, 234-253.

7   Barker M., Rayens W. (2003) Partial least squares for discrimination. Journal of

8   Chemometrics 17:166-173. DOI: 10.1002/cem.785.

9   Cubero S., Aleixos N., Molto E., Gomez-Sanchis J., Blasco J. (2011) Advances in

10   Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits

11   and Vegetables (vol 4, pg 487, 2011). Food and Bioprocess Technology 4:829-830.

12   DOI: 10.1007/s11947-011-0585-8.

13   Chao K., Chen Y.R., Hruschka W.R., Park B. (2001) Chicken heart disease

14   characterization by multi-spectral imaging. Applied Engineering in Agriculture 17:99-

15   106.

16   Chong I.G., Jun C.H. (2005) Performance of some variable selection methods when

17   multicollinearity is present. Chemometrics and Intelligent Laboratory Systems 78:103-

18   112.

19   Del Fiore A., Reverberi M., Ricelli A., Pinzari F., Serranti S., Fabbri A.A., Bonifazi G.,

20   Fanelli C. (2010) Early detection of toxigenic fungi on maize by hyperspectral imaging

21   analysis. International Journal of Food Microbiology 144:64-71. DOI:

22   10.1016/j.ijfoodmicro.2010.08.001.

1   Du C.J., Sun D.W. (2006) Learning techniques used in computer vision for food quality

2   evaluation: a review. Journal of Food Engineering 72:39-55. DOI:

3   10.1016/j.jfoodeng.2004.11.017.

4   ElMasry G., Sun D.-W., Professor Da-Wen S. (2010) Principles of Hyperspectral

5   Imaging Technology, Hyperspectral Imaging for Food Quality Analysis and Control,

6   Academic Press, San Diego. pp. 3-43.

7   ElMasry G., Wang N., ElSayed A., Ngadi M. (2007) Hyperspectral imaging for

8   nondestructive determination of some quality attributes for strawberry. Journal of Food

9   Engineering 81:98-107.

10   ElMasry G., Kamruzzaman M., Sun D.-W., Allen P. (2012) Principles ans Applications

11   of Hyperspectral Imaging in Quallity Evaluation of Agro-Food Products: A Review.

12   Critical Reviews in Food Science and Nutrition 52:25. DOI:

13   http://dx.doi.org/10.1080/10408398.2010.543495.

14   Fearn T., Riccioli C., Garrido-Varo A., Guerrero-Ginel J.E. (2009) On the geometry of

15   SNV and MSC. Chemometrics and Intelligent Laboratory Systems 96:22 - 26.

16   Fernandez Pierna J.A., Vermeulen P., Dardenne P., Baeten V. (2010) Integration of

17   chemometric tools in hyperspectral imaging data: contaminant detection, International

18   Association for Spectral Imaging. IASIM-10, Dublin, Ireland.

19   Geisser S. (1993) Predictive Inference Chapman and Hall, New York, NY.

20   Gowen A.A., O'Donnell C.P., Cullen P.J., Downey G., Frias J.M. (2007) Hyperspectral

21   imaging - an emerging process analytical tool for food quality and safety control.

22   Trends in Food Science & Technology 18:590-598.

23   Hernandez-Sanchez N., Lurol S., Roger J.M., Bellon-Maurel V. (2003) Robustness of

24   models bases on NIR spectra for sugar content prediction in apples. Journal of Near

25   Infrared Spectroscopy 11:97-107.

1   Jacquemoud S., Baret F. (1990) Prospect- a Model of Leaf Optical Properties Spectra.

2   Remote Sensing of Environment 34:75-91.

3   Keshava N. (2004) Distance metrics and band selection in hyperspectral processing

4   with application to material identification ans spectral libraries. IEEE Transactions on

5   Geoscience and Remote Sensing 42:1552-1565.

6   Keskin M., Dodd R.B., Han Y.J., Khalilian A. (2004) Assessing nitrogen content of golf

7   course turfgrass clipping using spectral reflectance. Applied Engineering in Agriculture

8   20:851-860.

9   Kruse F., Lekoff A., Boardman J., Heidebrecht K., Shapiro A., Barloon P., Goetz A.

10   (1993) The spectral image processing system (SIPS) - interactive visualization analysis

11   of imaging spectrometer data. Remote Sensing of Environment 44:145-163.

12   Lee K.J., Kang S., M. K., Noh S.H. (2005) Hyperspectral imaging for detecting defect

13   on apples., in: ASAE (Ed.), ASAE Annual International Meeting, ASAE, Tampa, Florida,

14   USA.

15   Liu Y., Windham W.R., Lawrence K.C., Park B. (2003) Simple algorithms for the

16   classification of visible/near-infrared and hyperspectral imaging spectra of chicken

17   skins, feces, and fecal contaminated skins. Applied Spectroscopy 57:1609-1612.

18   Liu Y.L., Chen Y.R., Kim M.S., Chan D.E., Lefcourt A.M. (2007) Development of simple

19   algorithms for the detection of fecal contaminants on apples from visible/near infrared

20   hyperspectral reflectance imaging. Journal of Food Engineering 81:412-418. DOI:

21   10.1016/j.jfoodeng.2006.11.018.

22   Liy Z.Y., Shi J.J., Wang D.C., Huang J.F. (2010) Discrimination and spectral response

23   characteristic of stress leaves infected by rice Aphelenchoides besseyi Christie. Guang

24   Pu Yu Guang Pu Fen Xi (Spectroscopy and Spectral Analysis) 30:710-714.

1   Lu R., Peng Y. (2006) Hyperspectral Scattering for assessing Peach Fruit Firmness.

2   Biosystems Engineering 93:161-171.

3   Luc B., Deronde B., Kempeneers P., Debruyn W., Provoost S. (2005) Optimized

4   Spectral Angel Mapper classification of spatially heterogeneous dynamic dune

5   vegetation, a case stude along the Belgian coastline, in: ISPMSRS (Ed.), The 9th

6   International Symposium on Physical Measurements and Signatures in Remote

7   Sensing.

8   Lleo L., Roger J.M., Herrero-Langreo A., Diezma-Iglesias B., Barreiro P. (2011)

9   Comparison of multispectral indexes extracted from hyperspectral images for the

10   assessment of fruit ripening. Journal of Food Engineering 104:612-620.

11   Martín-Belloso O., Soliva-Fortuny R. (2011) (Ed.)^(Eds.) Advances in Fresh-cut Fruits

12   and Vegetables Processing, CRC Press Taylor & Francis Group. pp. Pages.

13   Martín-Diana A.B., Rico D., Barry-Ryan C., Frías J.M., Henehan G.T.M., Barat J.M.

14   (2007) Efficacy of steamer jet-injection as alternative to chlorine in fresh-cut lettuce.

15   Postharvest Biology and Technology 45:97-107.

16   Medina M.S., Tudela J.A., Marín A., Allende A., Gil M.I. (2012) Short postharvest

17   storage under low relative humidity improves quality and shelf life of minimally

18   processed baby spinach (Spinacia oleracea L.). Postharvest Biology and Technology

19   67:1-9.

20   Mehl P.M., Chen Y.-R., Kim M.S., Chan D.E. (2004) Development of hyperspectral

21   imaging technique for the detection of apple surface defects and contaminations.

22   Journal of Food Engineering 61:67-81.

23   Merzlyak M.N., Solovchenko A.E., Gitelson A.A. (2003) Reflectance spectral features

24   and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in

25   apple fruit. Postharvest Biology and Technology 27:197-211.

1   Osborne B.G., Fearn T. (1986) Near Infrared Spectroscopy in Food Analysis Ed.

2   Longman Scientific and Technical, Harlow, UK.

3   Rinnan A., Van Den Berf F., S.B. E. (2009) Review of the most common pre-

4   processing techniques for near-infrared spectra. Trends in Analytical Chemistry

5   28:1201 - 1222.

6   Roger J.M., Palagos B., Bertrand D., Fernandez-Ahumada E. (2011) CovSel: variable

7   selection for highly multivariate and multi-response calibration. Application to IR

8   spectroscopy. Chemometrics and Intelligent Laboratory Systems 106:216-223.

9   Roussel S., Bellon-Maurel V., Roger J.-M., Grenier P. (2003) Authenticating white

10 grape must variety with classification models based on aroma sensors, FT-IR and UV

11 spectrometry. Journal of Food Engineering 60:407-419.

12 Siripatrawan U., Makino Y., Kawagoe Y., Oshita S. (2011) Rapid detection of

13 Escherichia coli contamination in packaged fresh spinach using hyperspectral imaging.

14 Talanta 85:276-281.

15 Sun D.-W. (2010) Hyperspectral Imaging for Food Quality Analysis and Control,

16 Hyperspectral Imaging for Food Quality Analysis and Control, Academic Press, San

17 Diego. pp. xiii.

18 Wang W., Li C., Tollner E.W., Gitaitis R.D., Rains G.C. (2012) Shortwave infrared

19 hyperspectral imaging for detecting sour skin (Burkholderia cepacia)-infected onions.

20 Journal of Food Engineering 109:38-48.

21 Xing J., De Baerdemaeker J. (2005) Bruise detection on 'Jonagold' apples using

22 hyperspectral imaging. Postharvest Biology and Technology 37:152-162.

23 Zeaiter M., Roger J.M., Bellon-Maurel V. (2005) Robustness of models developed by

24 multivariate calibration. Part II: The influence of pre-processing methods. Trends in

25 Analytical Chemistry 24:437-445.

1   Zude M. (2003) Comparison of indices and multivariate models to non-destructively

2   predict the fruit chlorophyll by means of visible spectrometry in apple fruit. Analytica

3   Chimica Acta 481:119-126.

4

Suggested Reviewers:

Douglas Rutledge
Laboratoire de Chimie Analytique, AgroParisTech
rutledge@agroparistech.fr
Expert in spectroscopy and chemometrics.

Fernando Riquelme
CEBAS-CSIC
fernando.riquelme@csic.ess
Expert in harvest and postharvest systems

Moon S. Kim
USDA, ARS, BARC, EMFSL
moon.kim@ars.usda.gov
Research Physicist. Expert in hyperspectral and multispectral imaging technologies

Table

		SAM	PLS-DA	*LEVE* index
	SAM	100%	77%	93%
E-20	**PLS-DA**	77%	100%	76%
	LEVE **index**	93%	76%	100%
	SAM	100%	80%	87%
E-10	**PLS-DA**	80%	100%	75%
	LEVE **index**	87%	75%	100%

Table 1.Percentage of concordance between methods in the assignation of the pixels

of leaves to the classes A, B and C for E-20 and E-10 samples

Table

		SAM	PLS-DA	*LEVE* index
E-20	**SAM**	100%	85%	95%
	PLS-DA	85%	100%	85%
	***LEVE* index**	95%	85%	100%
E-10	**SAM**	100%	93.75%	97.5%
	PLS-DA	93.75%	100%	93.75%
	***LEVE* index**	97.5%	93.75%	100%

Table 2. Percentage of concordance between methods in the assignation of the leaves

to the three classes A, B and C for E-20 and E-10 samples.

Figure

Figure 1. Representative leaves of the E-10 set along the dates of measurements.
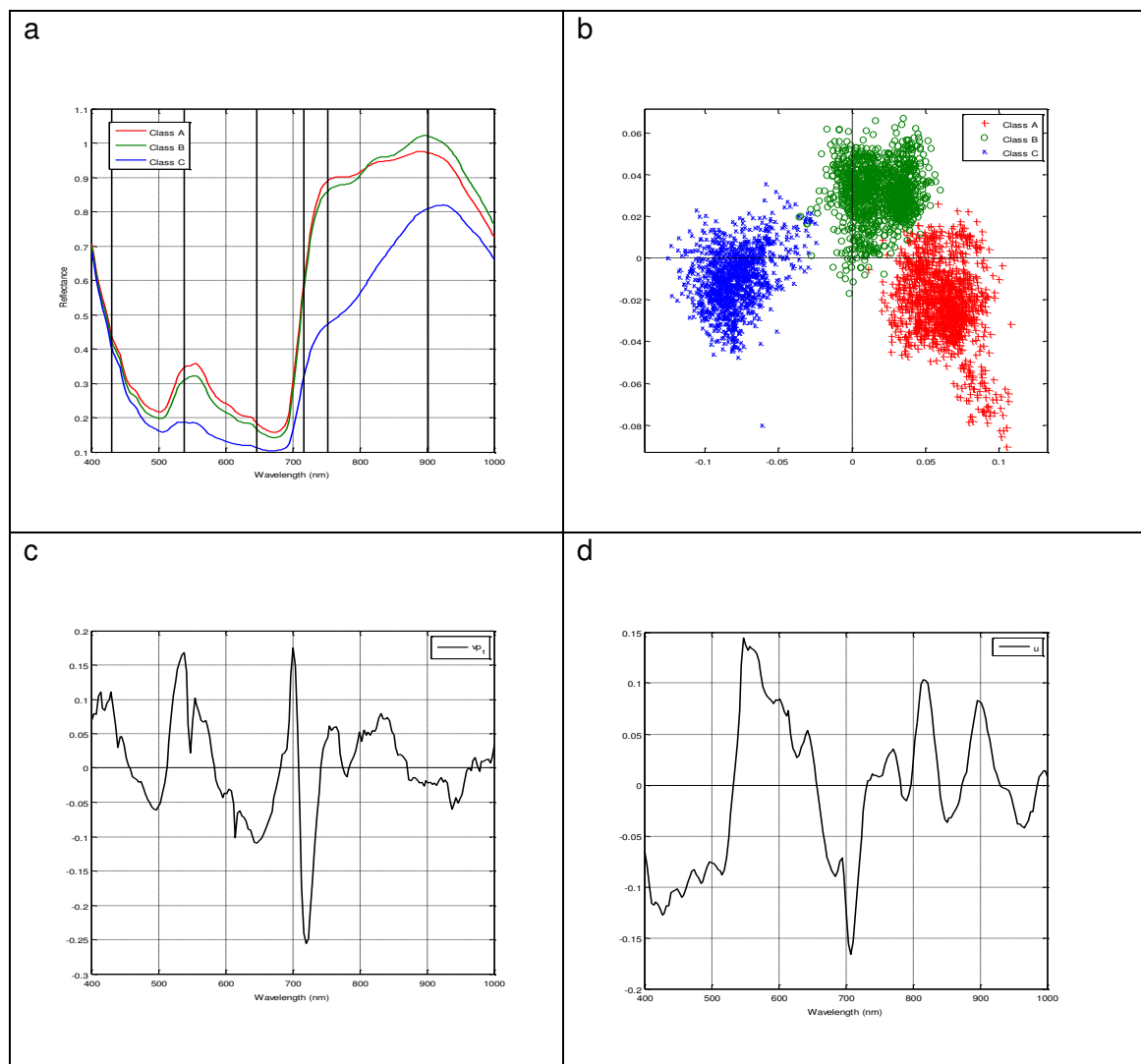
Figure 2. a: Average spectra of the 3 classes A, B and C (n=3600 spectra in total);

vertical lines indicate the 6 wavelengths selected by CovSel; b: scores of the 3 classes

calculated by the PLS-DA on the learning set; c: first discriminant vector (vp_1); d:

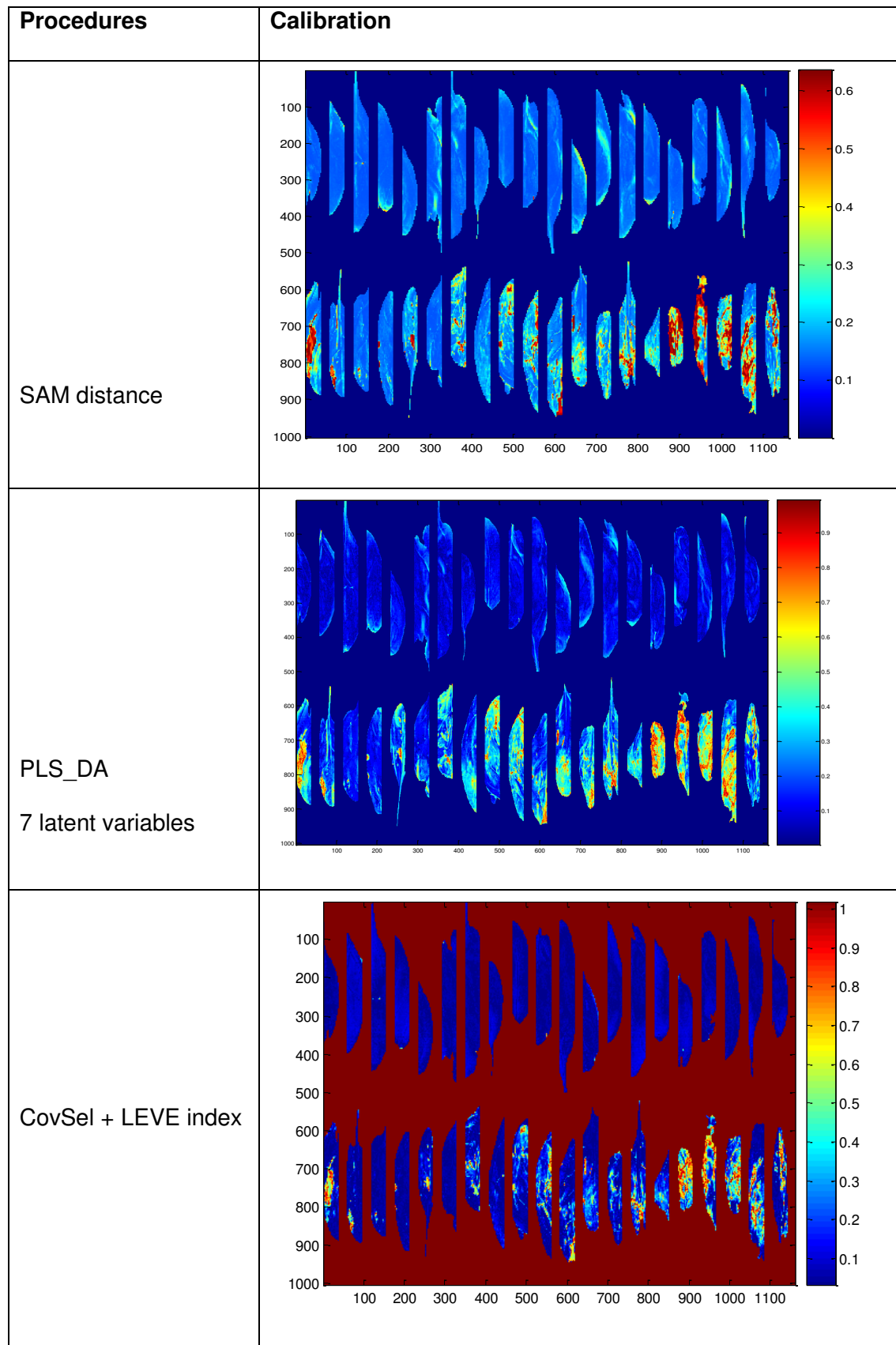residual of the second discriminant vector to the first one (u).

Figure

Procedures	Calibration
SAM distance	
PLS_DA 7 latent variables	
CovSel + LEVE index	

Figure 3. Visualization of the membership degree to class C, in E-20 samples, for the

different proposed procedures (each row of the table). Redder pixels are closest to

class C (more deteriorated areas) whereas deep blue pixels are far from class C

(healthy tissues). Each image contains the leaves corresponding to first date, first row,

ant last date of measurements, second row.

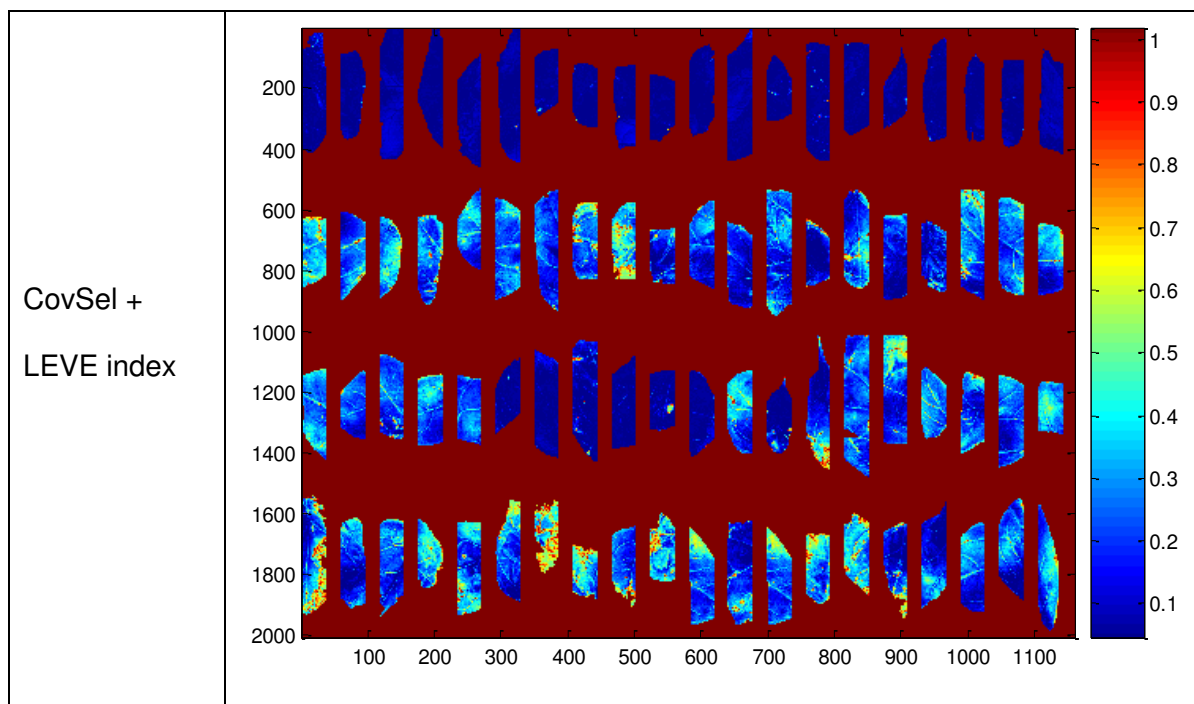| Procedures	Validation
SAM distance	
PLS_DA 7 latent variables	
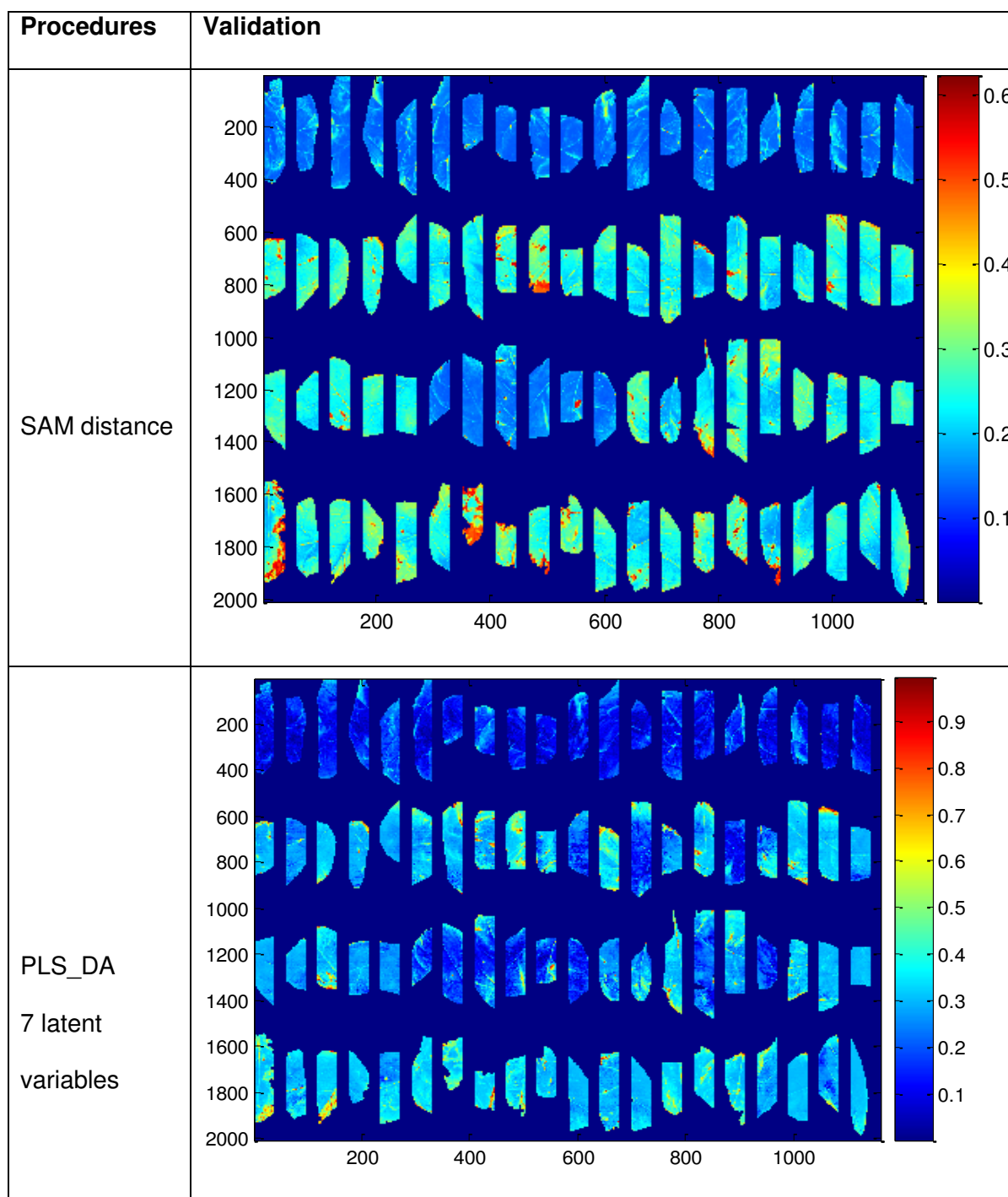
Figure 4. Visualization of the membership degree to class C, in E-10 samples, for the different proposed procedures. Redder pixels are closest to class C (more deteriorated areas) whereas deep blue pixels are far from class C (healthy tissues). Each image contains the leaves corresponding to the four dates of measurements: each row corresponds to one of the dates of measurements.
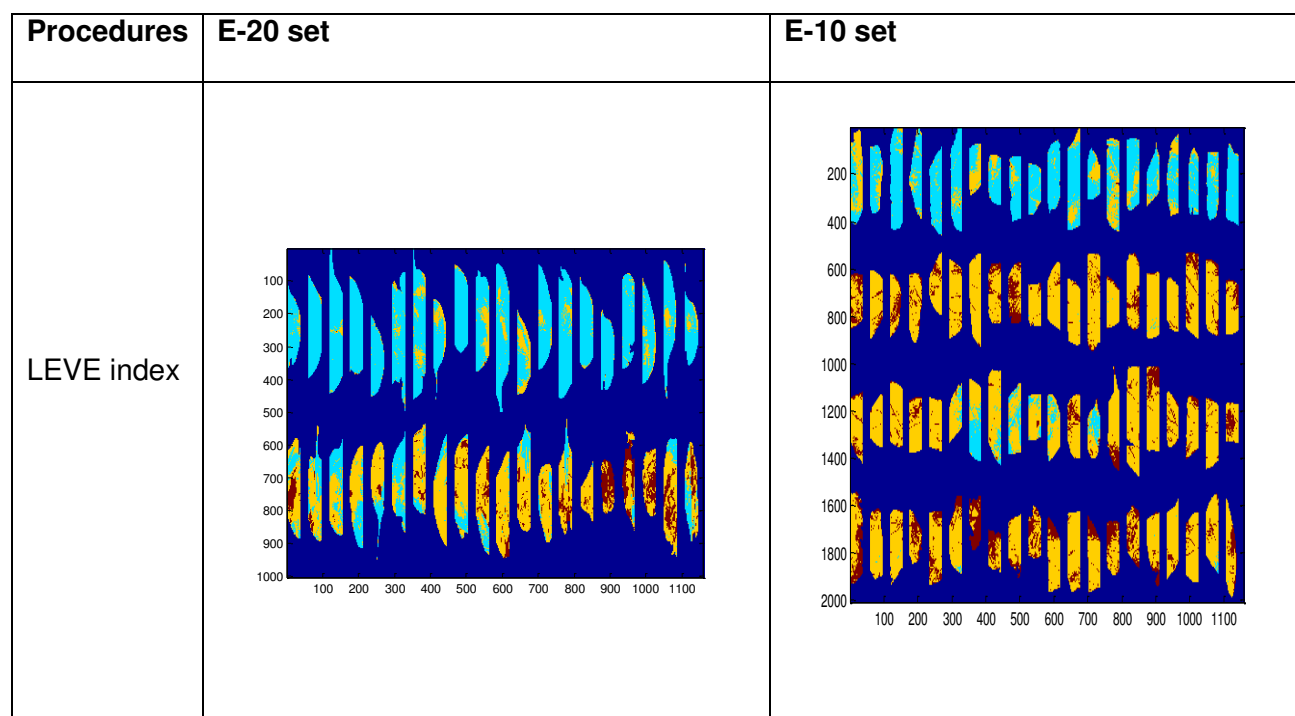
Figure

Figure 5. Visualization of the assignation of each pixel to one of the three quality classes considering the maximum membership degree. Blue color corresponds to class A, orange color to class B and color brown to class C.
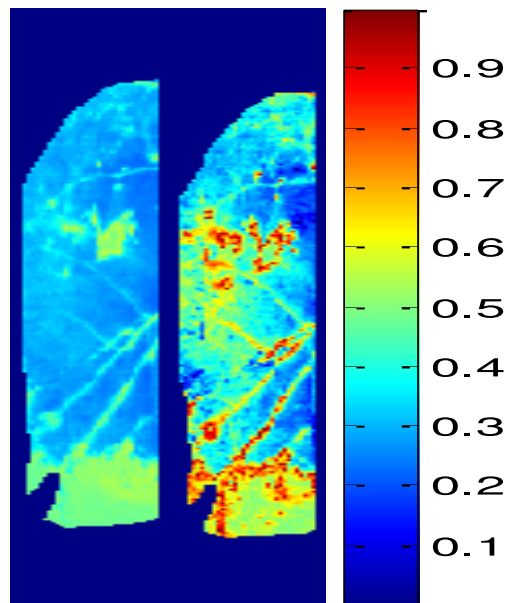
Figure

Figure 6. Images of membership degree to class C of one leave of E-10 set assigned

to class B according to SAM distance (left) and to class C according to LEVE index

(right).