

PoliBrush – A user-friendly software to aid multivariate image analysis dissemination

Rodrigo Rocha de Oliveira^a, Cristina Malegori^a, Giorgia Sciutto^b, Paolo Oliveri^{a,*}

^a Department of Pharmacy, University of Genova, Viale Cembrano 4, 16148, Genova, Italy

^b Department of Chemistry, University of Bologna-Ravenna Campus, Via Guaccimanni, 42, 48121, Ravenna, Italy

ARTICLE INFO

Keywords:

PoliBrush
Hyperspectral image analysis
Image analysis
Brushing
Score plot
Principal component analysis

ABSTRACT

PoliBrush is a freely distributed, stand-alone software designed for teaching exploratory multivariate analysis in the frame of color RGB and spectral imaging. PoliBrush implements principal component analysis (PCA) as its core method. It features a single main window that provides users with essential tools for spectral image pre-processing and exploration. The software emphasizes an interactive brushing approach, enabling users to gain a comprehensive understanding of the relationships between PCA score space and image pixel space. In the present paper, PoliBrush usage and its operational procedure is demonstrated through two case studies, which are also available for download. The software has already been successfully used in various educational settings and workshops on multivariate approaches for spectral imaging, proving its effectiveness in teaching key concepts. With this original software publication, our aim is to make these tools also accessible to both experienced and novice image analysts.

1. Introduction

Software is a fundamental and essential tool for studying and operating in the chemometric field, since the very beginning of the discipline. Progress and expansion of chemometrics during the last decades determined an increased development of dedicated software solutions. Indeed, it is also possible to state the converse concept: namely, one of the reasons that supported the great diffusion of chemometrics can be found also in the significant advances of software tools and platforms.

The first software packages developed were functional to both didactic and research/application purposes. Gradually, a considerable development in both chemometric methods and computer performances led to very rich and complex software tools, whose many options may often disorient novice users and students that are approaching chemometrics for the first time. For this reason, availability of software that implements the basic methods and essential options, to be used for teaching chemometrics is crucial to make the learning process more efficient and more focused on the fundamentals of the discipline.

Nowadays, a wide assortment of software tools – both commercial and freeware – is available, especially for multivariate data analysis, but also for multivariate design of experiments.

Since the early introduction of chemical images [1–3], the need for

devoted software was recognized by the chemometric community [4–6], as demonstrated by the growing interest in the use of multivariate image analysis in chemistry (Fig. 1). Nonetheless, availability of software for multivariate analysis of hyperspectral images is still limited, if compared with general chemometric applications.

One of the first hyperspectral imaging software packages to be introduced was ENVI (environment for visualizing images), developed in 1994 by L3Harris Geospatial (Broomfield, CO, USA) in the remote-sensing framework, for visualizing, processing, and analyzing geospatial imagery [7]. ENVI is also a standard for hyperspectral image files, used nowadays in many different application fields.

Among the commercial software packages currently most diffuse within the chemical and chemistry-related communities, it is worth mentioning Evince and Breeze, by Prediktera AB (Umeå, Sweden) [8], the MIA_Toolbox – an add-on for PLS_Toolbox – by Eigenvector Research Inc. (Manson, WA, USA) [9], and KemoQuant by Middleton Spectral Vision (Middleton, WI, USA) [10].

During the last decade, a number of freeware solutions has been developed, the most diffuse of which are HYPER-Tools – a freely downloadable and free-of-use GUI operating under the Matlab environment [11] – and Scyven, a C++ API GUI developed by Scyllarus (Canberra, Australia) [12]. Function packages for Matlab, R and Python

* Corresponding author.

E-mail address: paolo.oliveri@unige.it (P. Oliveri).

<https://doi.org/10.1016/j.chemolab.2023.104918>

Received 19 May 2023; Received in revised form 14 July 2023; Accepted 18 July 2023

Available online 24 July 2023

0169-7439/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

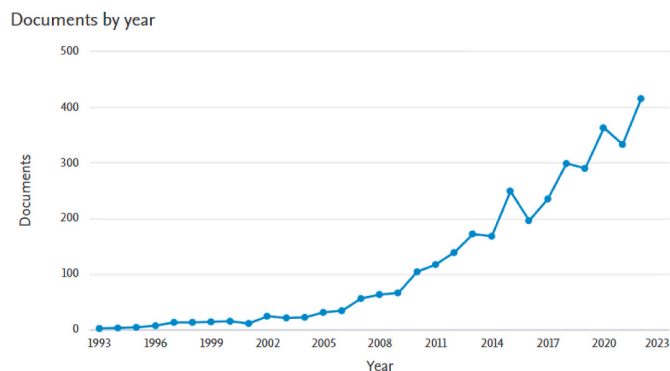


Fig. 1. Number of documents related to multivariate image analysis by year in the Scopus database over the past three decades in the subject area of chemistry.

users are also increasingly available from the repositories of the respective communities and on software sharing platforms.

The present paper is aimed at presenting and describing, with illustrative examples, functionalities implemented in PoliBrush, a standalone software purposely developed – with the Matlab App Designer platform (The MathWorks, Inc., Natick, MA, v. 2019b) – and freely distributed for teaching the basic concepts of exploratory multivariate image analysis (MIA) applied on RGB and spectral imaging data. The core method included in the software is principal component analysis (PCA) [13], and a special focus is dedicated to the so-called brushing approach [4], which allows the user to deeply understand the relationships between the PC score space and the image pixel space. Detailed descriptions of the exploration, preprocessing and analysis methods for multivariate images can be found elsewhere [4,13–16].

PoliBrush was already used within schools and workshops on multivariate approaches for spectral imaging, such as: National School of Chemometrics for Cultural Heritage (Ravenna, Italy, February 10–13th 2020), School of Chemometrics for the CHANGE (Cultural Heritage Analysis for New Generations) H2020 ITN-MSCA doctoral research programme (January 25–29th 2021), School of Chemometrics for ‘Raggruppamento Carabinieri Investigazioni Scientifiche’ RaCIS, Carabinieri (November 22–26th, December 13–17th 2021), Christmas Workshop ‘Spectral Imaging and Chemometrics for the Characterisation of Artistic Materials’ (Matera, Italy, December 15–16th 2022).

2. Software requirements

PoliBrush is a standalone user-friendly software developed using the MATLAB App Designer. The current version of PoliBrush can be installed on Windows operating systems and is freely available to download at the GitHub (<https://github.com/paololiveri/polibrush>) and at the Zenodo (<https://zenodo.org/record/8143341>) repositories. At the Zenodo repository, the user can also have access to the image datasets used in the present work. These images can be used not only for testing the software but also for teaching purposes.

PoliBrush hardware requirements are the same as the Matlab Runtime environment, R2019b (9.7), which can be found elsewhere [17]. There are no other specific needs.

3. Software structure and description

The PoliBrush’s graphical user interface (GUI) is characterized by a single main window, as shown in Fig. 2. The GUI is composed of two main elements: a) the *Toolbar menu*, with all the main operations and tools available, and b) the *Graphical output* interface, which enables the

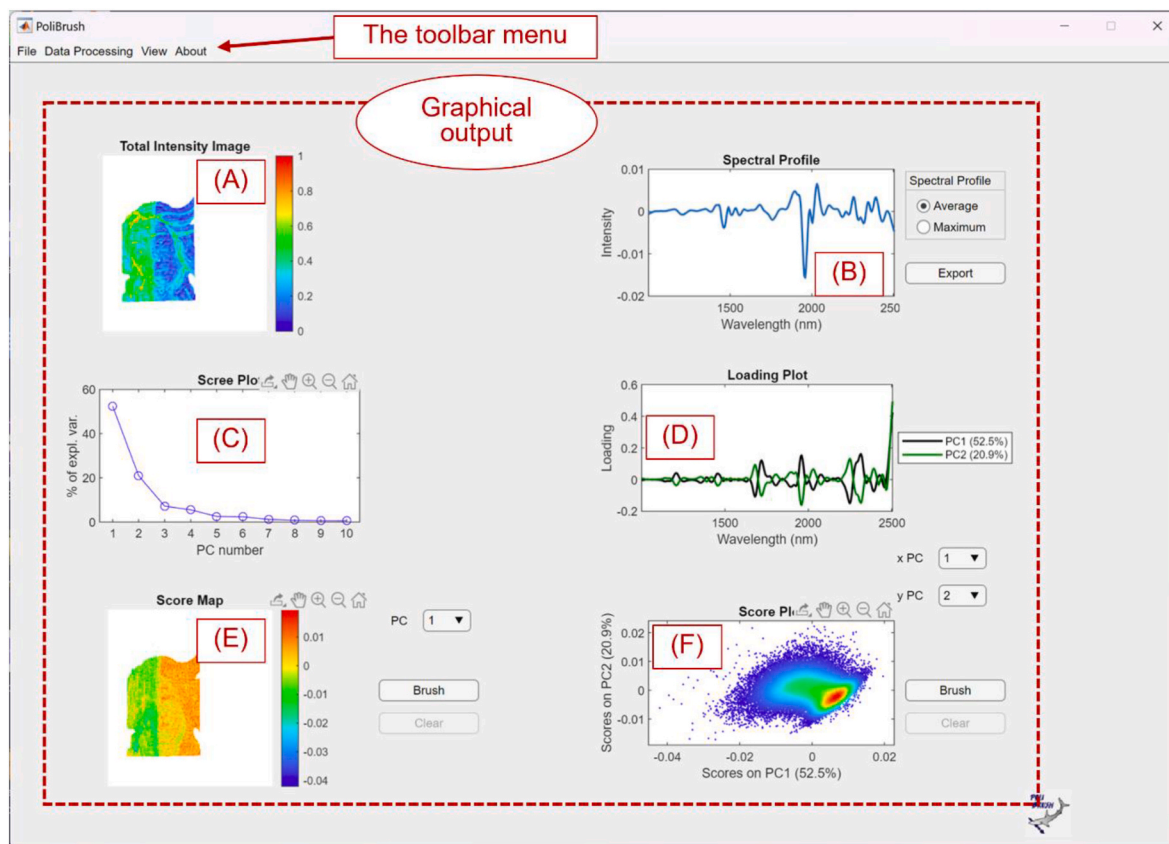


Fig. 2. Snapshot of the PoliBrush’s main window, showing the toolbar menu and the graphical output panels: total intensity image (A), average or maximum spectral profile (B), PCA scree plot (C), PCA loading line plot (D), PCA score map (E), and PCA score scatter plot (F).

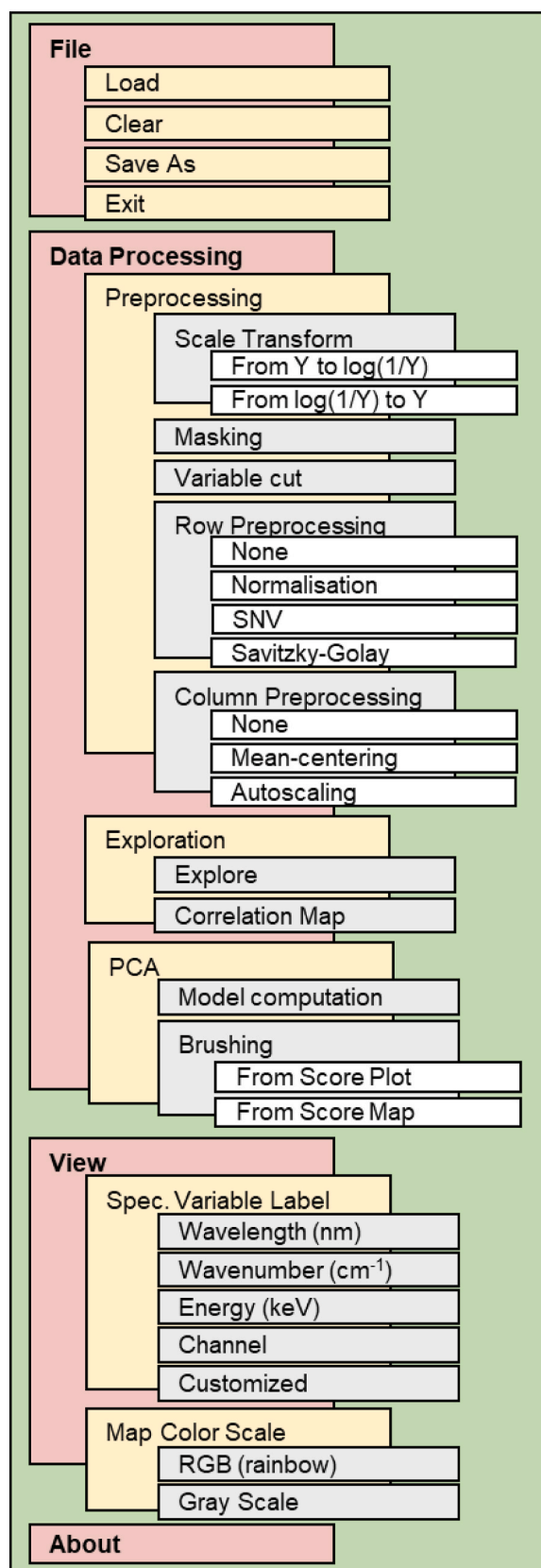


Fig. 3. Expanded PoliBrush's toolbar menu layout.

visualization of the HSI data and model outputs. Some of the menu features will display a pop-up window requiring user inputs or providing additional graphical outputs. A detailed description of the two main components of the GUI is covered in the following paragraphs.

3.1. The toolbar menu

The PoliBrush's toolbar menu is located at the upper left part of the main window, Fig. 2, and includes four tabs: *File*, *Data Processing*, *View* and *About*. Starting from these tabs, it is possible to access the sub-options menu, as represented in the expanded menu layout in Fig. 3. The detailed functionalities of each option available in PoliBrush are described below:

- 1) *File*. (Keyboard shortcut for the menu option is provided inside parenthesis).
 - a) *Load* (Ctrl + L). Clicking this button, a standard open file selection dialog box will pop up, allowing the user to browse the folder where the data file is located and to load it. The current version can open the most common chemical image file formats such as the Hierarchical Data Format 5 (.HDF5) [18] and ENVI-format images (.raw with related header file .hdr). It can also open RGB color images (.png, .jpg and .jpeg). Additionally, PoliBrush can load its own data file format (.pofi).
 - b) *Clear* (Ctrl + R). Clear the loaded data and model.
 - c) *Save As* (Ctrl + S). Save the preprocessed spectral image data in the PoliBrush's file format (.pofi) or Matlab format (.mat). Saving the HSI dataset after different preprocessing steps can be useful if the image analyst wants to compare different strategies before image exploration with PCA.
 - d) *Exit* (Ctrl + E). To close the software.
- 2) *Data Processing*
 - a) *Preprocessing*. Basic image and spectral preprocessing operations including the following:
 - Scale Transform*. Transforms the spectral signal, Y , applying the $\log_{10}(1/Y)$ or 10^{-Y} basic operations. Useful to transform spectra from reflectance to pseudo-absorbance, and vice-versa. Note that if non positive values are present in the dataset, this option will be disabled.
 - Masking*. Clicking this menu option, a pop-up window will open with the interactive masking tools interface, as shown in Fig. 4. Different criteria can be used to represent the spectral image, such as: *Total intensity* (sum of all the spectral intensities along the whole signal), *Single variable* (intensity of a specific spectral variable, manually selectable), *Ratio* (of the intensities of two manually selectable spectral variables), and *NDI* (Normalized Difference Index of two spectral variables, selectable manually [19]). Note that when zero values are present in the dataset, the *Ratio* and *NDI* criteria will be disabled. The corresponding image, obtained with the selected criterion, is represented in the top left panel of Fig. 4. Masking is carried out by defining the *Low* and *High* cut-off values for the selected criteria using the two sliding bars at the bottom right panel of Fig. 4. To aid this selection, a histogram, i.e., the representation of the frequency distribution of the pixel values, is also displayed. Additionally, the *Manual* criterion can be used to freehand draw a polygon region of interest (ROI).
 - Variable cut*. Used to remove a specific spectral range of the image dataset. A pop-up window allows the user to select the first and last variables of the range to be deleted. Fig. 5 shows the *Variable cut* interactive window with the spectral data in blue and the selected spectral range (to be deleted) in red.
 - Row Preprocessing*. Enables the application of basic spectral preprocessing in the direction of the rows, i.e., on each pixel spectrum of the image. The row preprocessing methods

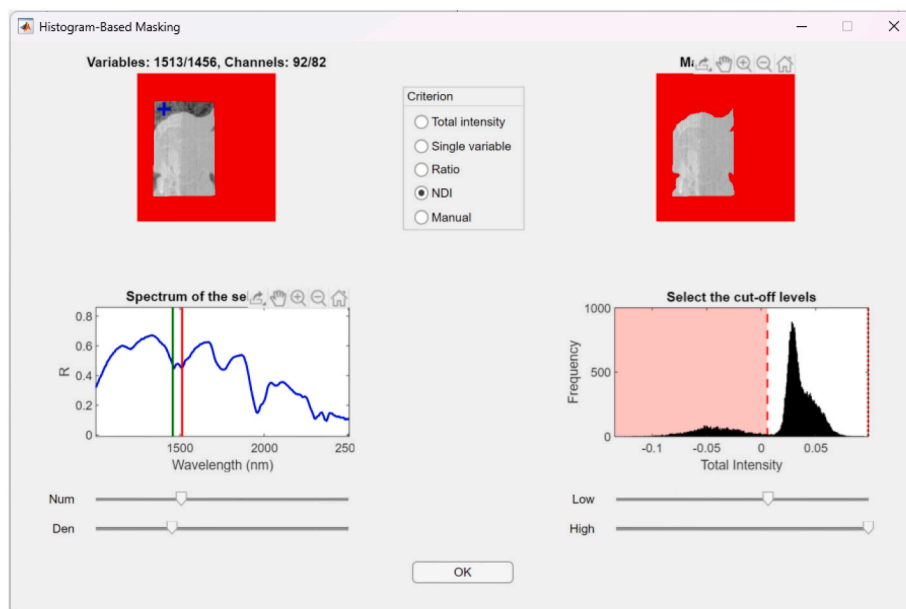


Fig. 4. Snapshot of the “Image Masking” pop-up window using the NDI criterion. Top left: representation of the spectral image with the selected criterion. Top right: masked image. Bottom left: in blue, the spectrum of the selected pixel (white cross in the top left panel) and the selected variables (red and green vertical lines related to the numerator and denominator, respectively). Bottom right: image histogram and the adjustable Low and High cut-off levels (red vertical lines).

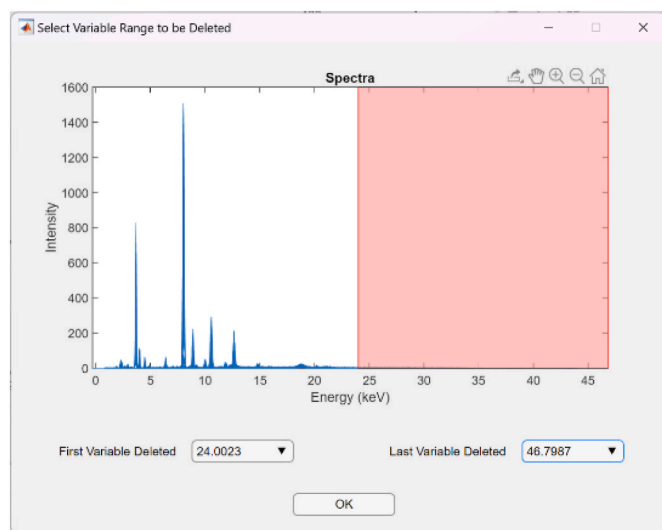


Fig. 5. Snapshot of the “Variable cut” pop-up window. The blue curves show the spectra from randomly selected pixels of the image, while the semi-transparent red area indicates the selected variable range to be deleted.

available are the normalization of spectral intensities (to sum = 1), the standard normal variate (SNV) transform and the Savitzky-Golay (SavGol) smoothing and derivative [20]. A pop-up window shall open when preprocessing parameters must be set. No row preprocessing is selected by default (none).

Column Preprocessing. Spectral preprocessing in the direction of the columns, i.e., variables. Column mean-centering (default) and autoscaling are available. No column preprocessing is applied when “none” is selected.

b) Exploration

Explore. This menu option allows the user to interactively explore the spectral image using the pop-up window shown in Fig. 6. The spectrum of a manually selected pixel is depicted at

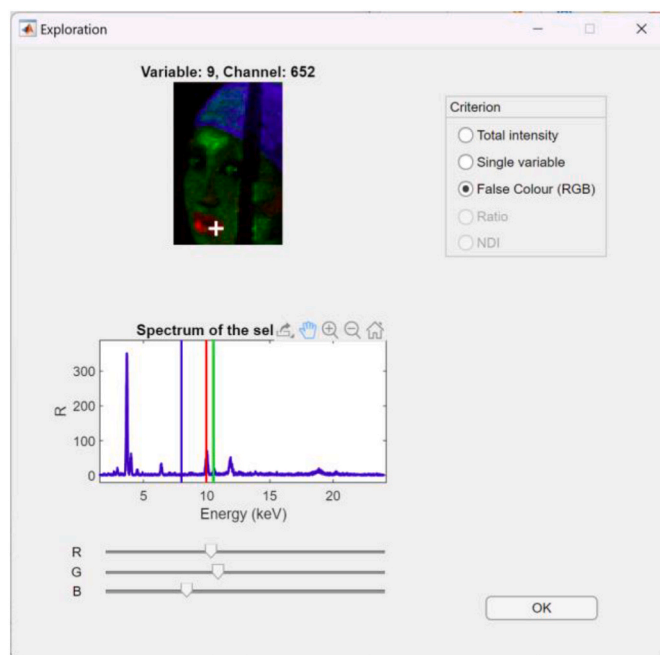


Fig. 6. Snapshots of the “Exploration” pop-up window with the “False Colour (RGB)” criterion selected.

the bottom panel. At the upper left panel, the spectral image can be represented using the same user-selectable criteria previously mentioned for the *Masking* tool. Additionally, the false-color (RGB) criterion represents the image using three spectral variables simultaneously, as depicted in Fig. 6. More details of the *Explore* tool operational procedure are presented in the next section, Case 1.

Correlation Map. A correlation analysis map – which encodes by a gray-level or color scale the correlation coefficient between pairs of spectral variables – is displayed in a pop-up

window. If a row pre-processing transform is selected, correlation is computed on pre-processed data.

c) PCA

Model computation. Runs principal component analysis (PCA) on the preprocessed image data. PCA results are displayed on the graphical outputs C to F in Fig. 2. See the “Graphical output” section for a detailed description.

Brushing. The brushing procedure allows studying the relationships between the scores in the PCA score scatter plot and the position of the corresponding pixels in the image [4, 21, 22]. Brushing can be performed by drawing a customizable ROI polygon either on the PCA scatter score plot (From Score Plot) or the inverse, i.e., from the score image representation. Quick access brushing button located on the right side of the PCA score image and scatter plot can also be used (bottom part of Fig. 2).

3) View

a) **Spectral Variable Label.** It allows changing the spectral axis label among different default options – e.g., Wavelength (nm), Energy (keV) – or to type it manually.

b) **Map Color Scale.** This View menu option allows changing the image colormap schemes to represent the pixel intensity values (from low to high) in two colormap types: color scale (from blue to red) or gray scale (from black to white).

4) **About.** This menu option opens a pop-up window with information about PoliBrush developers and version.

3.2. Graphical output

Fig. 1 shows a snapshot of the PoliBrush main window with the graphical output panels marked with letters from A to F. The graphical output panels are grouped in two sections: the first related to the spectral image (panels A and B); and the second related to the PCA model output (panels C to F). A description of the data outcome reported in the different panels is detailed below.

Panel A gives an image representation of the HSI dataset loaded. It displays the total intensity image, i.e., the sum of the spectral variables of the hyperspectral image. Two color scales can be used to display the distinct levels of the total intensity image which can be adjusted in the View menu. When color RGB images are analyzed, the image is represented as such.

Panel B depicts the spectral representation of the HSI dataset as the averaged (or maximum) spectral profile of the hyperspectral image. This plot is updated every time a row spectral preprocessing is applied and to show the average/maximum spectral profile of a ROI selected with brushing. The “Export” button next to this panel (Fig. 2 upper right) allows the user to export the spectral profile and the related spectral variables to an ASCII file.

Panel C shows the PCA scree plot depicting the percentage of explained variance as a function of the number of principal components (PCs).

Panel D displays the loading profiles of a selected pair of PCs.

Panel E displays the score map of a selected PC, represented in the original image spatial coordinates.

Panel F shows the density-based score scatter plot [23] of the same pair of PCs selected for Panel D. Clicking in the dropdown list next to the panels D to F, the user can select the PC or pair of PCs to be visualized. The color scale in Panel F represents the density of points (pixels) in the score plot going from blue (low density) to red (high density).

The user is encouraged to use the figure manipulation icons, which include zoom in and out, pan, restore the original view, and export graphical output. The export graphical output allows copying to clipboard as image or vector image and saving to different file formats (.png, .jpg, .tif and .pdf).

4. Datasets available

Two HSI datasets collected using two different hyperspectral imaging techniques from artwork painting reproductions will be used in the next section to illustrate the operational procedure. The datasets can be downloaded from the PoliBrush download link and are divided in two case studies according to the imaging technique used as follows:

4.1. Case 1: X-ray fluorescence (XRF) mapping

This dataset contains the XRF mapping of painting reconstruction, inspired by Johannes Vermeer’s “Girl with a Pearl Earring” produced in the Microchemistry and Microscopy Art Diagnostic Laboratory (M2ADL) of the University of Bologna (2018). The painting was obtained by applying pigments mixed with linseed oil on a commercial canvas. Pigments were selected considering their elemental signals detectable by XRF spectroscopy. In a more detail, azurite ($\text{Cu}_3(\text{CO}_3)_2(\text{OH})_2$) was used for the creation of the veil, lead white ($(\text{PbCO}_3)_2 \text{Pb}(\text{OH})_2$) for the flesh tone and vermilion (HgS) for the lips. (Fig. 7A). The XRF map was collected with the micro-XRF portable spectrometer (ELIO) produced by BRUKER. The XRF system is equipped with a rhodium-target X-ray tube. The voltage range is between 10 and 50 kV, and the anode current range is between 5 and 200 μA . The X-ray fluorescence is revealed by a Peltier-cooled silicon drift detector (SDD) with CUBE technology, with an active area of 50 mm^2 . The typical energy resolution for Mn-K α radiation is < 140 eV. A motorized XY stage is mounted on a tripod for mapping analysis with a total travel of 100 mm \times 100 mm. The collimator used for the analysis was of 0.5 mm.

4.2. Case 2: push-broom NIR-HSI

The dataset related to this case study contains the push-broom scanning near infrared HSI (NIR-HSI) of Sandro Botticelli’s Venus painting detail, produced in the Microchemistry and Microscopy Art Diagnostic Laboratory (M2ADL) of the University of Bologna (2018). The painting was executed with egg tempera on a wood panel prepared with a layer of gypsum and glue. Ancient and modern pigments were selected and employed according to their vibrational signals, detectable by NIR spectroscopy. In a more detail, earth-based pigments were used for the hair and for the flesh tone, while zinc white (ZnO) was applied in the white areas of the eyes and in the hair ribbon. Finally, dammar varnish was applied to half of the painted surface, according to ancient practices which involved the use of a terpenic varnish to improve color rendering and protect the underlying painted layer (Fig. 8A). The painting was scanned with a SWIR3 hyperspectral push-broom camera working in the 1000–2500 nm spectral range, at 5.6 nm spectral resolution (Specim Ltd, Finland). The scanning system is also characterized by three halogen lamps (35 W, 430 lm, 2900 K, each) used as illumination sources and a horizontal moving stage (40 \times 20 cm) on which the painting was scanned. The scanning parameters were set as follows: scan rate equal to 0.7 mm/s, push-broom camera frame rate equal to 50.00 Hz and exposure time equal to 9 ms. The hyperspectral system was controlled with the Lumo Scanner software (Specim Ltd, Finland). Before the scan, dark (with the camera shutter closed) and white (using a Spectralon reference tile) reference images were acquired, to compute reflectance values (see Fig. 9).

5. Operational procedure

In this section, the operational procedure and usage of PoliBrush will be demonstrated using the two datasets described in the previous section.

5.1. Case 1: analysis of an X-ray fluorescence (XRF) image

Analysis of the XRF image dataset using PoliBrush aims at the

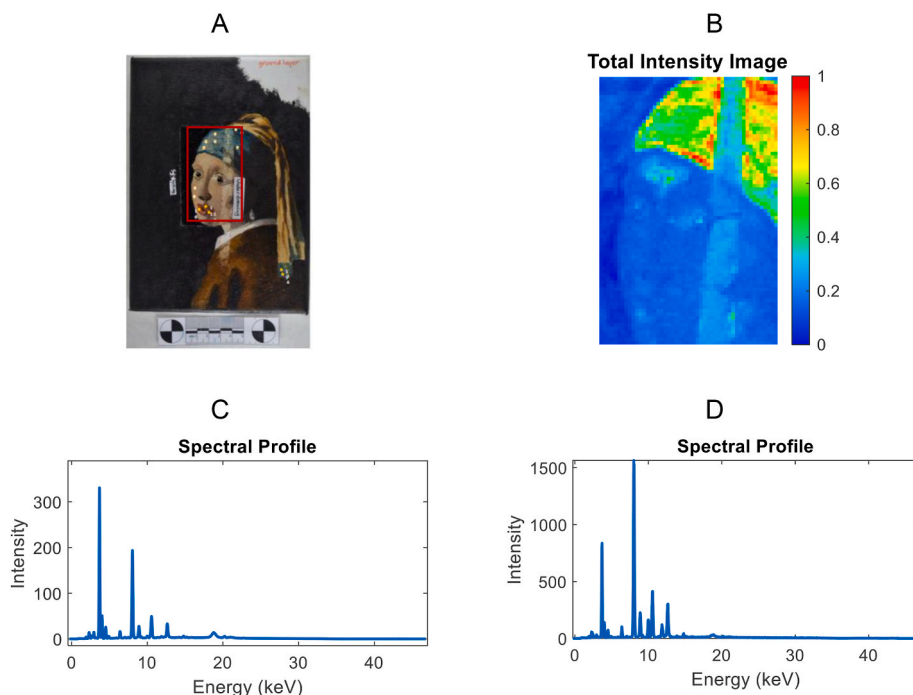


Fig. 7. Case 1 – Reproduction of Johannes Vermeer’s “Girl with a Pearl Earring” painting. RGB image (A) showing the area mapped with XRF (add rectangle). XRF total intensity image (B), average (C) and maximum spectral profile (D).

identification and localization of the elements constituting the painting materials and their mixtures [24]. This dataset can be analyzed in PoliBrush after loading the “vermeer.HDF5” downloadable file.

After loading the dataset, the PoliBrush’s main window initially shows the total intensity image and the average spectral profile, as reported in Fig. 7B–C, respectively. Looking at the average spectral profile, it can be observed that low (-0.27 to 1.5 keV) and high (24 to 47 keV) energy ranges contain not relevant information in the dataset. The user can also confirm this by switching to the spectral profile of the maximum intensity for each variable, Fig. 7D. These uninformative spectral channels can be removed using the *Variable Cut* menu option. The popup window in Fig. 5 enables the selection of the first and last variable of the range, which can be deleted clicking on the “OK” button at the bottom of the window. Other variable ranges can also be deleted repeating this procedure as many times as required.

5.1.1. Image exploration using the “explore” tool

After removing the non-informative variables from the dataset, the user can explore the dataset visualizing intensity maps of the different peaks present in the XRF image. In PoliBrush, this can be carried out using the *Explore* option tool (*Data Processing/Exploration ... /Explore*), which pops-up the related window, as shown in Fig. 6. When the single variable criterion is selected, clicking on the image (white cross in the top plot of Fig. 6A), the user can select a pixel and, to choose a spectral channel, slide the control bar below its spectral profile representation (bottom part of Fig. 6A). After an initial visual inspection of the data and identification of the XRF emission lines (according with their energy values), the user can visualize the spatial distribution of each element detected and understand the elemental composition of the paint components used within the mapped area. Fig. 8 shows the distribution maps related to some peaks selected from different areas of the painting.

To explore the spectral image dataset, the user can start visualizing the most intense peak of the average spectral profile (see Fig. 7C) located at 3.69 keV. The peak intensity distribution map is depicted in Fig. 8A and the spectral profile for a pixel select in a high-intensity area (top left of Fig. 8A) is shown in Fig. 8B. This peak is related to Ca from the preparation layer.

Similarly, the user can continue exploring selecting the most intense pixel area of the total intensity image (see Fig. 7B) which is located at the top right side of the painting. From the spectral representation of this pixel, Fig. 8D, the variable related to the most intense peak, located at 8.05 keV, is selected to display the intensity map shown in Fig. 8C. This peak is related to emission line of Cu, which constitutes the blue pigment azurite used to represent the veil.

Another of PoliBrush’s exploration features is the false-color image representation of three variables mapped to the RGB (red, green, blue) color scale. This requires choosing the “False Colour (RGB)” criterion and selecting the three variables to be represented sliding the three horizontal bars of the *Exploration* pop-up window, as shown in Fig. 6B, Fig. 8E reports this type of false-color map when the intensity of the peaks at 9.99 keV (Hg), 10.55 keV (Pb) and 8.05 keV (Cu) are mapped to RGB, respectively. The colored vertical lines in Fig. 8F indicate the location of the selected variables. The variable assigned to the red color is related to the emission of Hg present in the vermilion pigment used to represent the lips of the painting subject, which is shown as red pixels in Fig. 8E. The green color can be assigned to the emission line of Pb in the lead white pigment abundant on the face and eyes of the subject. Finally, the blue color can be related to emission line of Cu present in the azurite pigment, as it is highlighted also in Fig. 8C–D.

5.2. Case 2: analysis of a push-broom NIR-HSI

The analysis of this case study aims at the demonstration of the operational procedure required for the application of essential pre-processing steps and the PCA-based MIA with the brushing techniques available in the presented software.

From the available dataset the user can find the “venere_NIRHSI” folder which contains the ENVI hyperspectral files (.raw and .hdr) related to the scanned sample, white reference, and dark reference. To load the data into PoliBrush environment, the user should use the *File/Load ...* menu option and, in the “Select File to Open” window, change the file format to (*.raw). Then, browser to the folder containing all the files and open the ‘venere.raw’ file. To properly load the NIR-HSI data, all the ENVI files generated by the imaging acquisition software,

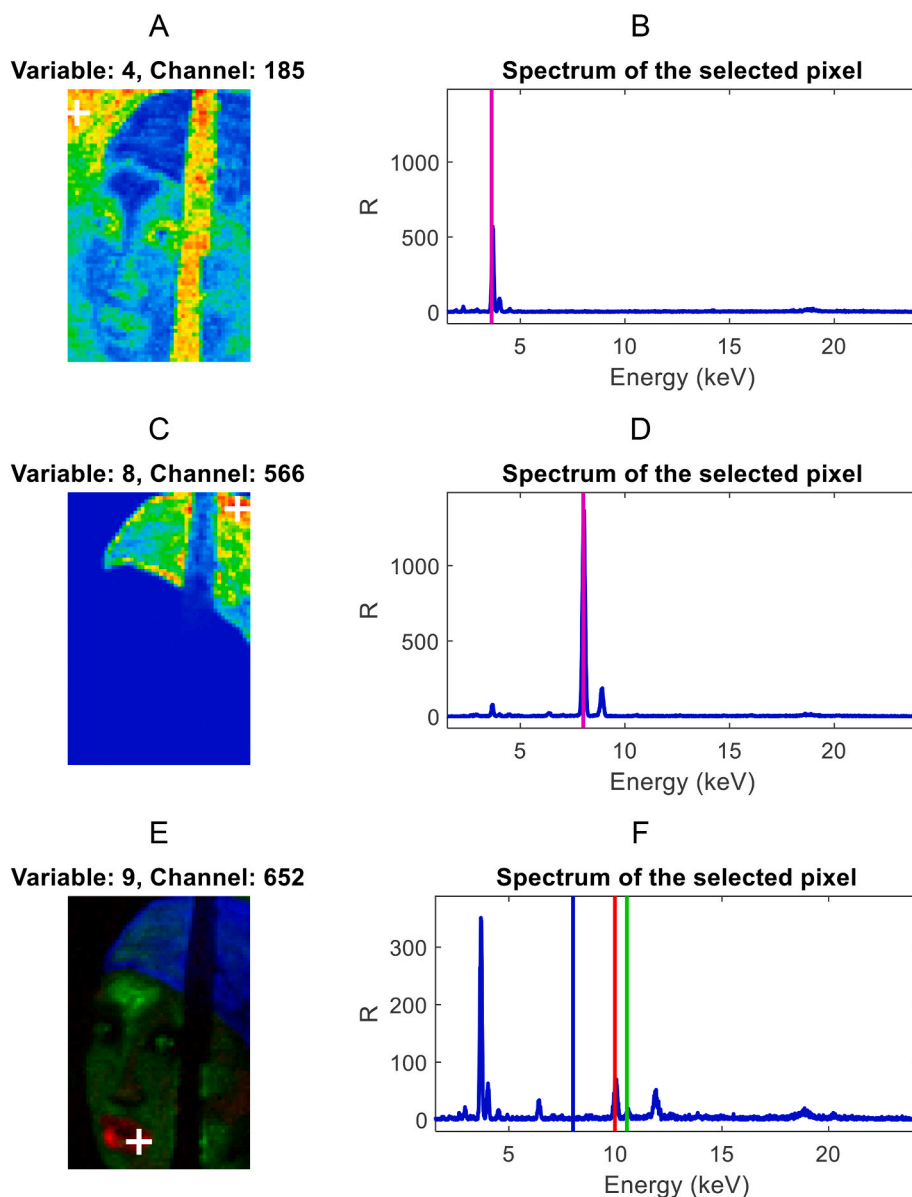


Fig. 8. Spectral image exploration graphical outcomes. XRF intensity map of the Ca emission peak (A) and spectrum of a pixel placed at the dark background of the painting (B). XRF intensity map of the Cu emission peak (C) and spectrum of a pixel placed at the blue veil of the painting (D). False colour RGB composition with the overlapped intensity maps of Hg (red), Pb (green) and Cu (blue) (E) and the spectrum of the pixel at the lower lip of the painting subject showing the three variables used to represent the RGB composition.

including the dark and white references, must be in the same folder of the sample. When loading the data, the software will automatically compute the spectral reflectance value for each pixel and wavelength.

5.2.1. Selection of a ROI and image masking using the “masking” tool

Fig. 8B shows the representation of the HSI dataset as total reflectance intensity image using the gray scale colormap. The colormap can be changed to gray scale using the menu option: *View/Map Color Scale/Gray Scale*. For this demonstration, an initial region of interest (ROI) focusing on the area with the highest variation has been selected. The ROI is extracted using the masking tool with the manual criterion (see Fig. 4). The user must draw the masking polygon around the painting subject. Fig. 9C shows the manually selected ROI perimeter in gray overlayed onto the total intensity image, and Fig. 8D reports the outcomes of the masking procedure, with the removed area highlighted in red.

When the contour of the ROI to be selected is complex or a more advanced masking is required, the user can explore other criteria and use the histogram-based masking tools. The masking procedure using the Normalized Difference Image (NDI) [19] of two spectral channels selected manually will be demonstrated to remove the blue background

of the painting detail (see original color image in Fig. 9A). Starting from the manually selected ROI, the user should open the masking tools and select the NDI criterion. At this point, the user must perform iterative trials to find a pair of spectral variables that gives the highest contrast between the object of interest and background to be masked out. This process includes the following steps:

- clicking on the image (Fig. 4 upper left), look for the spectral bands whose intensity changes significantly when comparing the spectrum of the object of interest in respect to the area of the background;
- using the horizontal sliding bars (Fig. 4 bottom left), select a pair of spectral variables related to the bands identified;
- looking at the NDI image, adjust the spectral variables until the highest contrast is observed between the two areas;
- create the mask using the two horizontal sliding bars below the image histogram (Fig. 4 bottom right), to adjust the low and high cut-off values while looking at the resulting masked image until the desired masking is achieved.

For the present case study illustration, the pair of spectral variables found to provide a sufficient contrast between the background and the

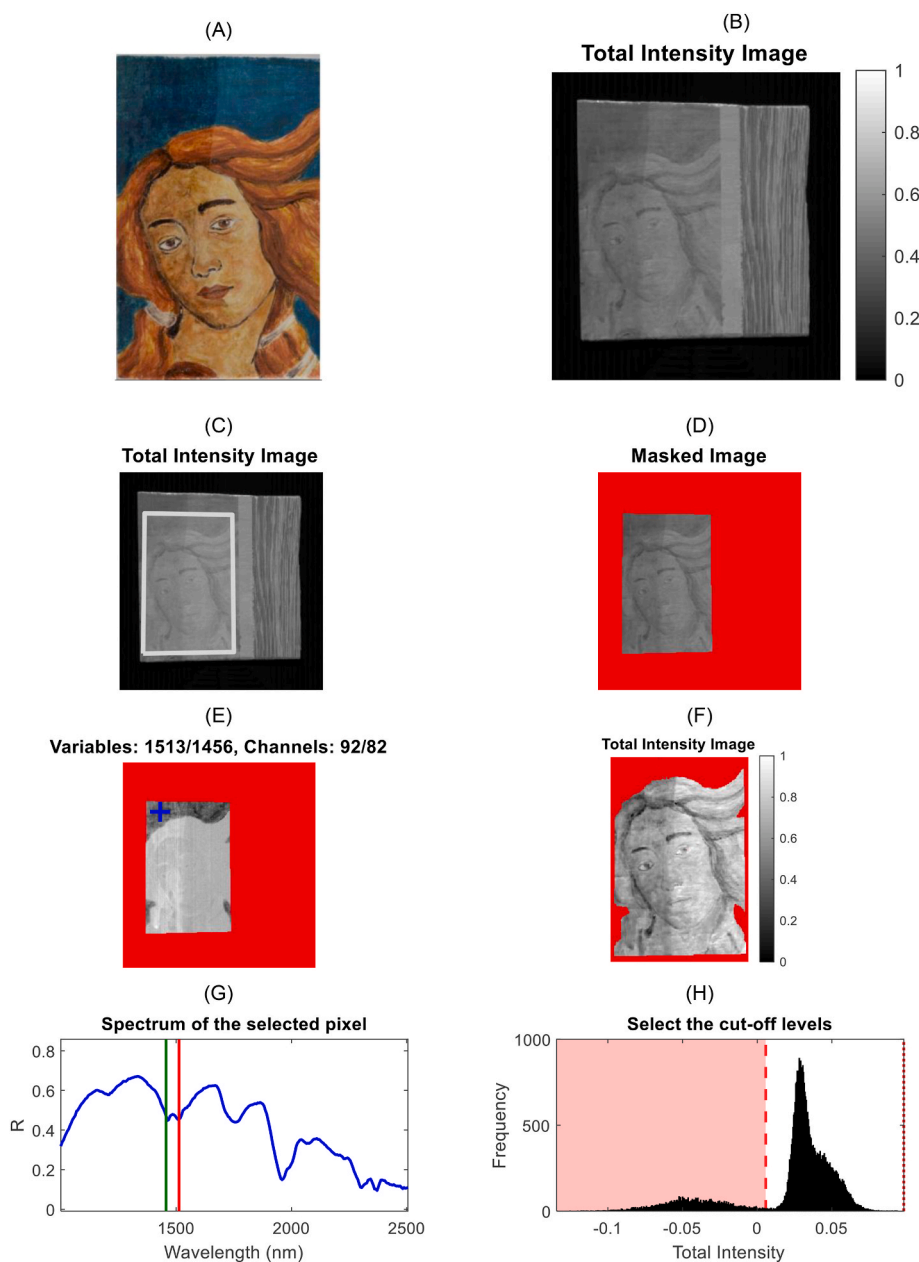


Fig. 9. Case 2, detail of Sandro Botticelli's Venus detail as color RGB image (A). Total intensity image of the raw NIR-HSI dataset (B) with the manual ROI masking polygon (C) and after manual masking (D). Image representation using the NDI masking criterion (E). Total intensity image after NDI masking (F). Reflectance spectrum of the selected pixel (G). Histogram of the NDI image (H). Gray scale goes from black (minimum value) to white (maximum value).

object of interest (in this case, Venus' head) using the NDI criterion is related to the two bands at 1456 nm (OH first overtone gypsum, green line) and 1513 nm (OH first overtone azurite, red line), depicted in Fig. 9G, Fig. 9E shows the NDI image representation using the selected variables, where dark (low NDI) and bright (high NDI) pixels are related to the background to be masked out and to the desired ROI, respectively. The NDI histogram in Fig. 9H shows that the NDI values range between -0.15 and 0.1 , where the mode (visually identifiable as a relative maximum in the histogram profile) related to background pixels (lower NDI) is sufficiently separated from the mode of the painting subject (higher NDI). The *Low* NDI cut-off value, placed between the two modes, is indicated with the red-dashed vertical line in Fig. 9H. The *High* cut-off value was set at the maximum for the present case study. The semi-transparent red area represents the NDI values below or above the *Low* or *High* cut-off values, respectively. Therefore, all the pixels with the NDI value outside the cut-off range are masked out. Fig. 9F shows the

final masked image after removing the background.

The user is encouraged to explore other pairs of variables to improve the contrast, to change the masking criteria. Similar procedures can be carried out using the *Ratio* and *single variable* criteria. When the *total intensity* criterion is selected, only the histogram cut-off values should be adjusted. The selection of the criteria will depend on the difficulty level in achieving the best contrast between the two areas. It is advisable to start with a simpler criterion, such as manual, total intensity or single variable; then, if needed, to use the ratio or the NDI criteria.

5.2.2. The spectral "preprocessing" tool

Before modelling the NIR-HSI data with PCA, the user may apply different spectral preprocessing steps using the menu option: *Data Processing/Preprocessing* Fig. 10 shows the average spectral profile of the masked image before (Fig. 10A) and after two typical preprocessing steps (Fig. 10B–C). The first step (Fig. 10B) consisted of the logarithmic

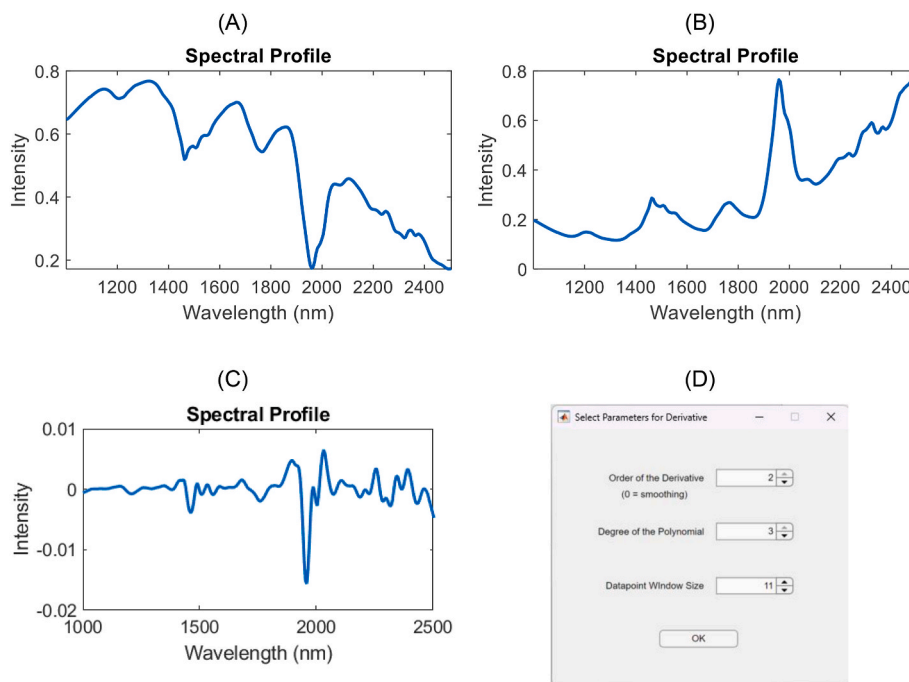


Fig. 10. Case 2 – Row preprocessing steps. Average raw reflectance spectrum (A), after transforming to pseudoabsorbance (B), and after the application of the second order Savitzky-Golay derivative (C). Savitzky-Golay derivative parameters settings pop-up window (D).

scale transformation from reflectance to pseudoabsorbance (... /Scale Transform/From Y to $\log(1/Y)$), see. The second preprocessing step (Fig. 10C) consisted of the application of the Savitzky-Golay second derivative [20] (... /Row Preprocessing/Savitzky-Golay) to remove scattering artifacts from the NIR signal. The derivative parameters settings were adjusted as shown in the pop-up window in Fig. 10D. Column mean-centering of the spectral variables was also applied before PCA.

5.2.3. Principal component analysis

To run the PCA model, the user must select the menu option *Data Processing/PCA/Model computation*. The PCA model is calculated using a maximum of ten principal components (PCs). The graphical output for the MIA of the preprocessed NIR-HSI data is shown in Fig. 11. The PCA scree plot in Fig. 11A shows the variation of the explained variance in percentage, for each PC. The first two principal components (PC1 and PC2) loading profiles are plotted as function of the spectral variables in Fig. 11B. Correspondingly, the density-based score scatter plot of the first two PCs is shown in Fig. 11D. The score image of PC1 is shown in Fig. 11C. To display a different PC score map or pair of PCs represented in the score and loading plots, the user must select the PC number using the button on the right side of the respective PCA graphical output to expand the dropdown list with the PCs (see Fig. 2).

The PC1 score image, Fig. 11C, clearly shows the enhanced difference between the two sides of the painted surface due to the presence of the dammar varnish on the left side. Thus, when focusing on the left side of the PC1 score image, patterns with different score values are detected, which are related to the uneven varnish layer. A qualitative assessment of the score image suggests that the yellow-orange and blue vertical strokes are related to areas with thin and thick varnish layers, respectively. Evenly spread green areas are assumed to have intermediate layer thickness. The right side of the score image in Fig. 11C shows patterns related to the painting details: in this case, lower and higher score values are related to bright and dark paint areas, respectively. Analyzing the PC2 vs PC1 score scatter plot (Fig. 11D), two main clusters of pixels are noticed. The trend along the PC1 axis suggests a relationship between the clusters and the PC1 score image.

5.2.4. The “brushing” tool

To aid the understanding of the relationship between score image and score plot, the user might use the brushing tool. Brushing can be activated by using either the menu option *Data Processing/PCA/Brushing ...* or the Brush buttons next to the score map and score plot graphical output (see the bottom of the PoliBrush’s main window in Fig. 2). A video with the brushing procedure is available in the supporting information (SI_Brushing.mp4). Brushing applied to the score map consists of manually drawing a polygon to select a ROI on the score map. Subsequently, the points in the score plot related to the pixels inside the ROI will be automatically highlighted. The inverse brushing selection, i.e., when a ROI is drawn in the score plot, works similarly but, in this case, pixels are selected by the user in the score plot, and automatically highlighted in the score map. The clear button (Fig. 2 bottom part) can be clicked to clean the applied brushing. For demonstration, Fig. 11E shows the PC1 score image after application of the brushing tool to select a ROI covering the right side of the painting detail. The corresponding brushed pixels in the PC1 score image are highlighted as gray dots in the scatter score plot in Fig. 11F. The representation of the brushed pixels in the score plot indicates that they cover the cluster located at the right side of the score plot. The inverse brushing is demonstrated by brushing the main cluster of the PC4 vs. PC3 score plot, as reported in Fig. 11H. The corresponding brushed points are highlighted on the PC4 score map in Fig. 11G, while pixels outside of the brushed area in the score plot are displayed in black. The highlighted pixels on the PC4 score map are mostly related to the dark orange paint areas all over the painting detail, despite the presence/absence of the varnish layer.

6. Conclusions

PoliBrush is a software freely distributed for the purpose of teaching to beginners the principles of multivariate analysis applied to RGB and spectral imaging data. For this reason, it implements basic exploratory methods, such as PCA, and interactive tools such as brushing, which allow the user to efficiently recognize relationships between the image pixel space and the PC score space, fully understanding multivariate visualization and analysis of multidimensional images. Moreover,

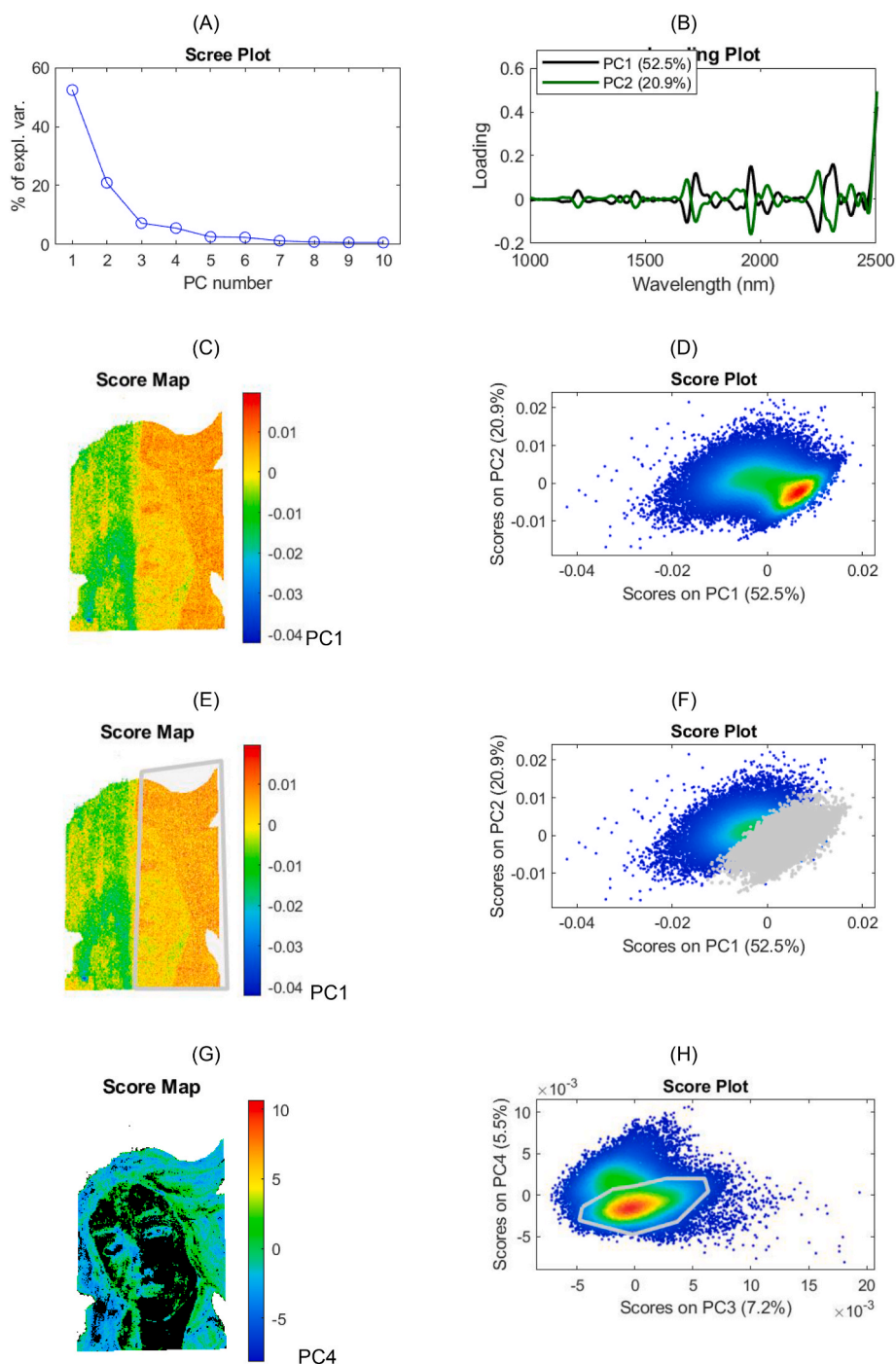


Fig. 11. Case 2 – PCA and brushing graphical output. PCA scree plot (A), PC1 and PC2 loadings plot (B), PC1 score image (C), density-based PC2 vs PC1 score scatter plot (D), brushing applied to the PC1 score image with brushed area in gray transparent area (E), PC2 vs PC1 score plot with brushed pixels in gray (F), PC4 score image after applying brushing to the score plot, pixels not brushed are displayed in black (G), PC4 vs PC3 score plot with brushing applied (H).

PoliBrush also facilitates a thorough interpretation of the multivariate outcomes obtained. This makes its use particularly suitable not only for didactic purposes but also in multidisciplinary research contexts, where users may lack expertise in multivariate analysis.

The operational procedure of PoliBrush has been demonstrated through the application on data obtained from two distinct HSI techniques on separate case studies. Each of the two application examples described effectively highlights the relevant information within the dataset, while also assisting in the appropriate utilization and exploitation of multivariate image analysis.

The software will be maintained and improved following the

suggestions that hopefully will arise from users, especially from students, but always trying to keep it at an entry level, with simple options and straightforward procedures implemented.

Validation

The software was independently tested by:

- **Dr. Oxana Rodionova**, Federal Research Center for Chemical Physics, RAS.

PoliBrush GUI v. 3.2 was installed and tested on a Windows 10 PC.

This is a free standalone software, which allows a wide range of users

to apply it on their PCs. At the same time, as PoliBrush is designed using Matlab compiler, the presence of Runtime library (R2019b) is necessary. If it is not installed or an earlier version of the Runtime Library is present, then R2019b is installed during the installation procedure for PoliBrush.

The interface is easy to use and intuitive. Single main window provides a quick start for even inexperienced users. Both hyper-spectral files and color images can be analyzed. All functionality consists of three main stages, Data preprocessing, Exploration and Principal Component Analysis (PCA).

Basic image preprocessing is available including manual selection of ROI. Main variable row-wise and column-wise preprocessing methods are presented in the software. In the Exploration step, among other options, it is possible to select manually a pixel and depict the corresponding spectrum. The PCA computations are presented by four main windows, Scree plot, Loading plot, and two presentations of Scores: Score map and Score plot. The relationship between the pixels in the score plot and their position in the image can be studied using the brushing procedure.

The paper describes all features of the software in detail and demonstrates software possibilities with the help of two real examples available for the reader. The brushing procedure is explained by means of a short video-tutorial. No bugs or malfunctioning were encountered.

It can be concluded that PoliBrush is a useful software, which can be highly recommended for the quick start when exploration of hyper-spectral or image data is required.

- **Dr. Anna de Juan**, Associate Professor in the Department of Chemical Engineering and Analytical Chemistry at the Universitat de Barcelona, Spain.

PoliBrush is a very user-friendly interface oriented to the exploration of hyperspectral images (HSI) by Principal Component Analysis (PCA). It is a standalone product and, hence, the parent commercial package MATLAB is not required for their use. It can interpret common commercial formats of HSI and provides a good span of options in terms of preprocessing, cropping, image visualization and PCA output visualization. The menu structure is clear and is easy to walk through. There are interesting options that connect plots, such as the possibility to interconnect scatter score plots with score maps or the option of selecting specific wavelengths to develop false RGB images. Some of the graphical plots obtained can be additionally stored in several image or pdf formats. Following the data sets provided, showing very different examples of XRF and NIR images, it is easy to reproduce the results as described in the related paper. An improvement would be the possibility to connect more than one spectral preprocessing action. Summarizing, a recommendable good freeware to make a first quick exploration of hyperspectral images.

CRedit authorship contribution statement

Rodrigo Rocha de Oliveira: Methodology, Formal analysis, Validation, Visualization, Writing – original draft. **Cristina Malegori:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Giorgia Sciutto:** Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – review & editing. **Paolo Oliveri:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests

Paolo Oliveri reports financial support was provided by University of Genoa.

Data availability

A link for the download of data described in the present paper is provided.

Acknowledgements

Financial support provided by Università degli Studi di Genova (Research Project Curiosity Driven 2020: “3Depth – From 2D to 3D hyperspectral imaging exploiting the penetration depth of near-infrared radiation”, CUP: D34G20000100005) is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemolab.2023.104918>.

References

- [1] A.F.H. Goetz, B. Curtiss, Hyperspectral imaging of the earth: remote analytical chemistry in an uncontrolled environment, *Field Anal. Chem. Technol.* 1 (1996) 67–76, [https://doi.org/10.1002/\(SICI\)1520-6521\(1996\)1:2<67::AID-FACT2>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1520-6521(1996)1:2<67::AID-FACT2>3.0.CO;2-B).
- [2] D. Briggs, M.J. Hearn, Analysis and chemical imaging of polymer surfaces by secondary ion mass spectroscopy, *Spectrochim. Acta Part B At. Spectrosc.* 40 (1985) 707–715, [https://doi.org/10.1016/0584-8547\(85\)80120-8](https://doi.org/10.1016/0584-8547(85)80120-8).
- [3] P. Geladi, E. Bengtsson, K. Esbensen, H. Grahn, Image analysis in chemistry I. Properties of images, greylevel operations, the multivariate image, *TrAC, Trends Anal. Chem.* 11 (1992) 41–53, [https://doi.org/10.1016/0165-9936\(92\)80118-](https://doi.org/10.1016/0165-9936(92)80118-).
- [4] K. Esbensen, P. Geladi, Strategy of multivariate image analysis (MIA), *Chemometr. Intell. Lab. Syst.* 7 (1989) 67–86, [https://doi.org/10.1016/0169-7439\(89\)80112-1](https://doi.org/10.1016/0169-7439(89)80112-1).
- [5] P. Geladi, H. Grahn, K. Esbensen, E. Bengtsson, Image analysis in chemistry II. Multivariate image analysis, *TrAC, Trends Anal. Chem.* 11 (1992) 121–130, [https://doi.org/10.1016/0165-9936\(92\)85010-3](https://doi.org/10.1016/0165-9936(92)85010-3).
- [6] P. Geladi, K. Esbensen, Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools, *J. Chemom.* 5 (1991) 97–111, <https://doi.org/10.1002/cem.1180050206>.
- [7] ENVI, <https://www.nv5geospatialsoftware.com/Products/ENVI>. (Accessed 25 July 2023).
- [8] Prediktera, <https://prediktera.com/>. (Accessed 25 July 2023).
- [9] MIA Toolbox, https://eigenvector.com/software/mia_toolbox/. (Accessed 25 July 2023).
- [10] KemoQuant, https://www.middletonspectral.com/product/kemoquant_analysis_software/. (Accessed 25 July 2023).
- [11] HyperTools, <https://www.hypertools.org/>. (Accessed 25 July 2023).
- [12] Scyven, <https://scyllarus.data61.csiro.au/software/scyven/>. (Accessed 25 July 2023).
- [13] C. Malegori, P. Oliveri, Principal component analysis, in: *Hyperspectral Imaging Anal. Appl. Food Qual.*, CRC Press, 2018, pp. 85–107, <https://doi.org/10.1201/9781315209203-6>.
- [14] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. Essential steps before image analysis, *Chemometr. Intell. Lab. Syst.* 117 (2012) 138–148, <https://doi.org/10.1016/j.chemolab.2012.05.009>.
- [15] J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: a review with applications, *Chemometr. Intell. Lab. Syst.* 107 (2011) 1–23, <https://doi.org/10.1016/j.chemolab.2011.03.002>.
- [16] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, *Anal. Chim. Acta* 896 (2015) 34–51, <https://doi.org/10.1016/j.aca.2015.09.030>.
- [17] Mathworks, Matlab system requirements - release 2019b - windows, <https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/support/sysreq/files/system-requirements-release-2019b-windows.pdf>. (Accessed 25 July 2023).
- [18] S. Koranne, Hierarchical data format 5 : HDF5, in: *Handb. Open Source Tools*, Springer US, Boston, MA, 2011, pp. 191–200, https://doi.org/10.1007/978-1-4419-7719-9_10.
- [19] C. Malegori, E. Alladio, P. Oliveri, C. Manis, M. Vincenti, P. Garofano, F. Barni, A. Berti, Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics, *Talanta* 215 (2020), 120911, <https://doi.org/10.1016/j.talanta.2020.120911>.
- [20] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627–1639, <https://doi.org/10.1021/ac60214a047>.
- [21] P. Oliveri, C. Malegori, M. Casale, E. Tartacca, G. Salvatori, An innovative multivariate strategy for HSI-NIR images to automatically detect defects in green coffee, *Talanta* 199 (2019) 270–276, <https://doi.org/10.1016/j.talanta.2019.02.049>.

- [22] M. Manley, C.M. McGoverin, P. Engelbrecht, P. Geladi, Influence of grain topography on near infrared hyperspectral images, *Talanta* 89 (2012) 223–230, <https://doi.org/10.1016/j.talanta.2011.11.086>.
- [23] P.H.C. Eilers, J.J. Goeman, Enhancing scatterplots with smoothed densities 20 (2004) 623–628, <https://doi.org/10.1093/bioinformatics/btg454>.
- [24] XRF spectra, <http://www.xrfresearch.com/xrf-spectra/>. (Accessed 6 January 2023).