X R M A

X-Ray Map Analyzer

GIS based tool for the statistical handling of X-ray Map

Requirement of the system and former user defined specification

This document provides an operative guide useful to perform X-Ray Map analysis using a Python script executable also from MS-DOS command prompt, largely based on ArcGIS® functions, without graphical user interface. The used ArcGIS® version (9.3 or later) should have ArcInfo license and support spatial analysis and data management extension.
This application born to statistically handle several types of multispectral images such as the X-ray maps which represent a multidimensional illustration of the elemental distribution within a selected domain (Fig.1). The procedure provides two different cycles of analysis, each of them able to be subdivided in four principal analytical steps (Figs.2-3). The first analytical cycle (First Cycle) have the aim to analyse the entire selected domain, finalizing the procedure by means of the classification of recognised mineral and non-mineral phases represented by means of arbitrary colours, useful to make for instance a mineral distribution map of selected domain. This first cycle is subdivided in two former steps of image processing consisting in a potential pre-filtering stage, followed by a RGB synthesis procedure accompanied by the principal component analysis (PCA). The two latter steps of the first analytical cycle are characterised by the image classification procedure by means of a maximum likelihood classification (MLC) algorithm, successively potentially finalized by a post-filtering stage, useful to minimise the weight of the isolated pixel (Fig.2). The second one (Second Cycle) (Fig.3) have the aim to deepen the analysis of a selected category recognised in the previous cycle with the aim to verify the possibility to recognise further potential subdivisions, such as those responsible for the presence of mineral zonation. This analytical cycle is subdivided in a first step of map algebra operation, finalized to the extraction of the selected classified class (e.g. a mineral phase) from the previously classified image (Fig.3). This is necessary to re-launch the multivariate statistical analytical step on a new X-ray maps array representative just of the elemental distribution within a single detected phase. A latter stage of a maximum likelihood image classification followed by a potential post-filtering stage is also present in the final steps of this second cycle (Fig.3).
The required software configuration useful to the functioning of the script provides the installation of ArcGis® 9.3 on your PC or Laptop and current tutorial refers to this version with Python 2.5.
Installation procedure:
Unzip □xrma.zip□in the root directory so as to obtain a path like □c:\xrma\□
Then it is necessary to set the initial configuration file xrma.cfg.
Open it with a textpad, thus edit the three parameters in function of your personal configuration of ArcGis, for instance, as follow:

```
#DOS Parameters
[DOS_info]
Dos_Shell_Cols=150
Dos_Shell_Lines=1000
```

#Path for ArcGis
[ARCGIS_info]
AGpath= C:/Program Files/ArcGIS/
Tbx_SA_path= C:/Programmi/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx
Tbx_DM_path= C:/Programmi/ArcGIS/ArcToolbox/Toolboxes/Data Management Tools.tbx
#Path for data workspace
[XRMA_info]
DataPath=c:/xrma/data/

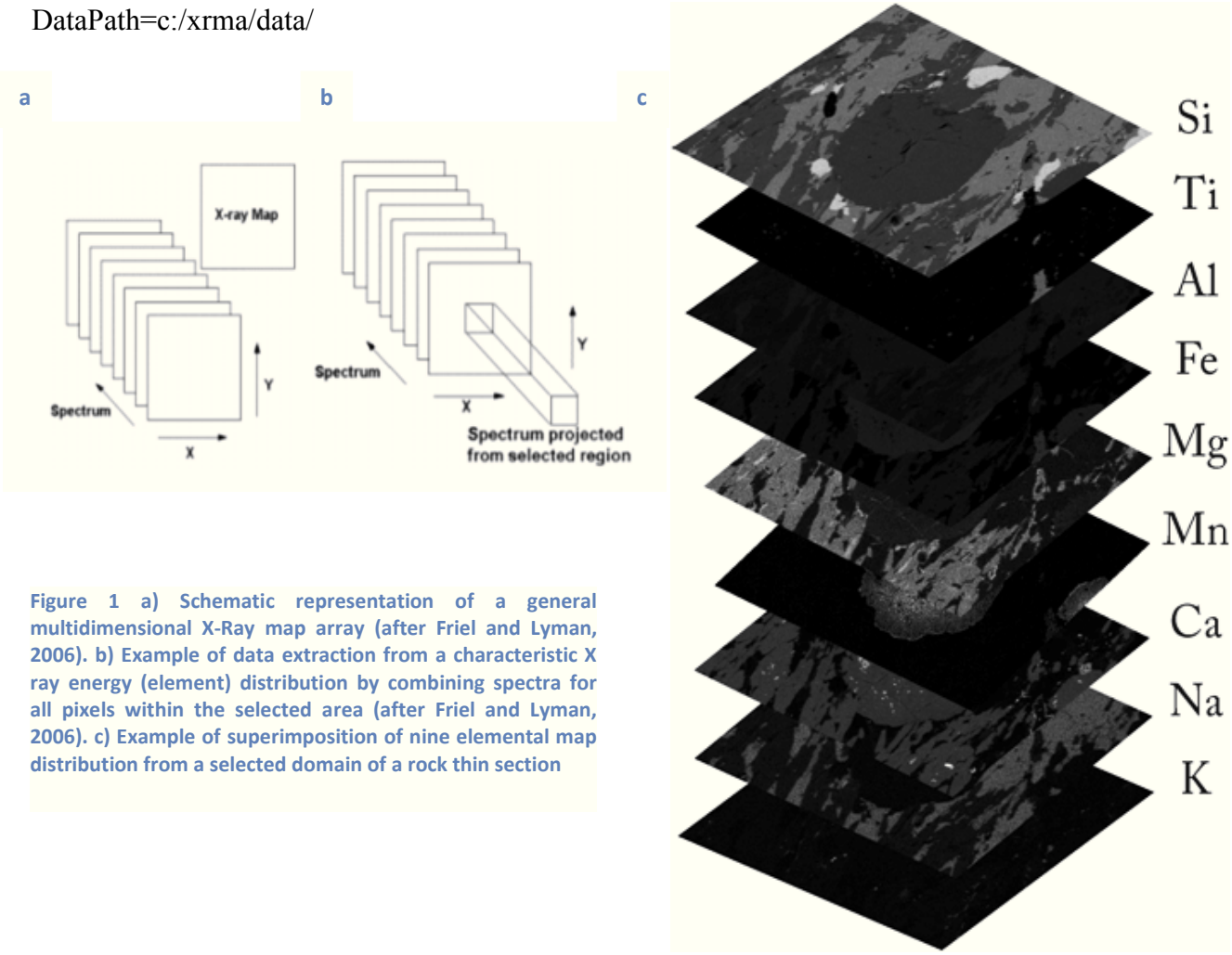


Figure 1 a) Schematic representation of a general multidimensional X-Ray map array (after Friel and Lyman, 2006). b) Example of data extraction from a characteristic X ray energy (element) distribution by combining spectra for all pixels within the selected area (after Friel and Lyman, 2006). c) Example of superimposition of nine elemental map distribution from a selected domain of a rock thin section

First Cycle

Before to begin, user can create a workspace data path for the input images (Fig.4) or can wait that the procedural instruction of the script ask to create the data path later.

Thus, it is necessary to put X-ray maps in □*.tif□ format within □c:\xrma\DataPath**MyWorkspace**\input_img□, possibly after have renamed each image with the name of the specific element (e.g. Al.tif), then open DOS cmd and digit:

1. □cd Python25□
2. □python c:\xrma\prog\xrma.pyc□

After this, the application starts recognising the operative system and the username. Then, begin to work and ask to set the workspace name to identify the user job or permit to create a new workspace (Fig.5).

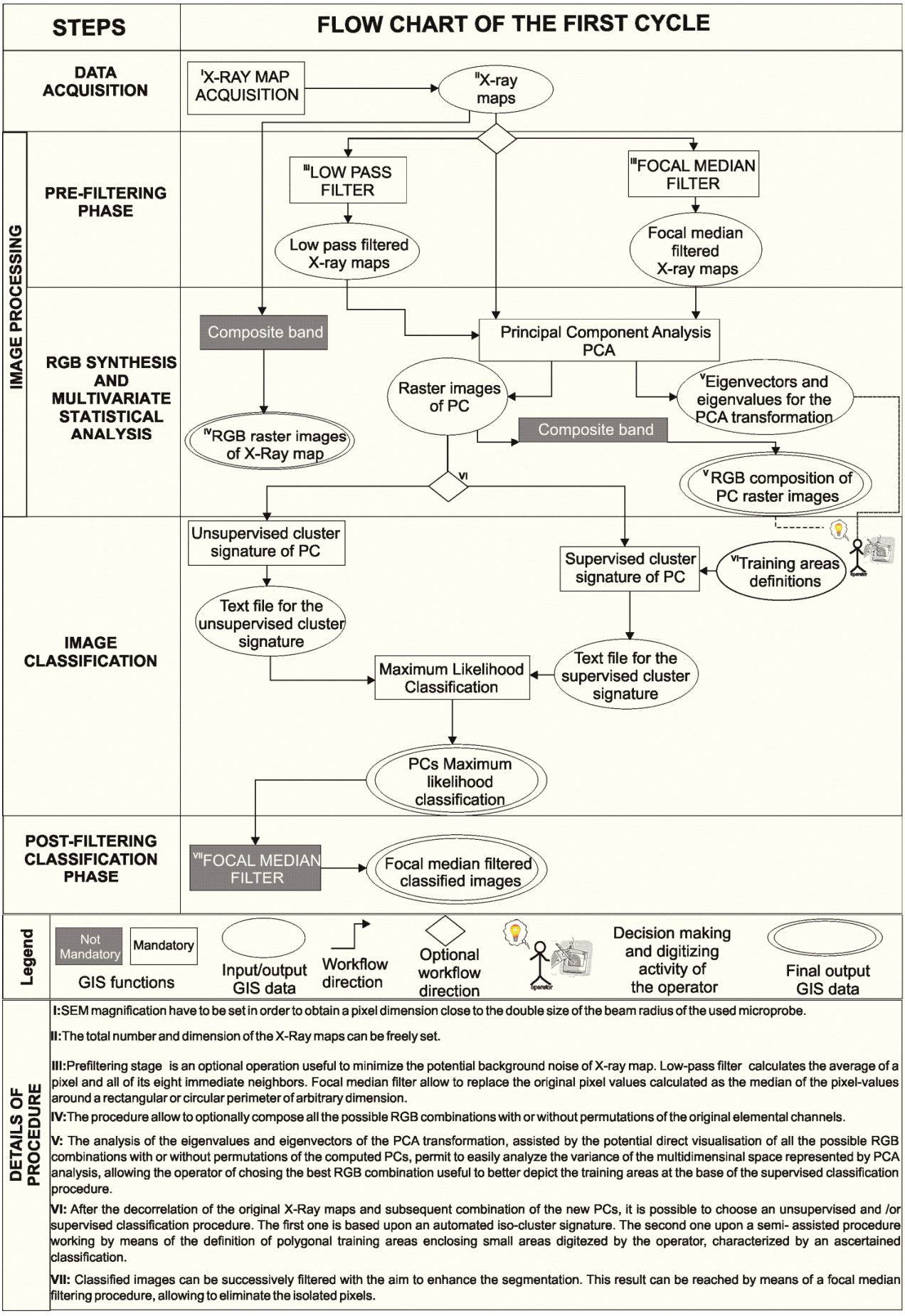


Figure 2 Flow chart of the first cycle of the geoprocessing procedure

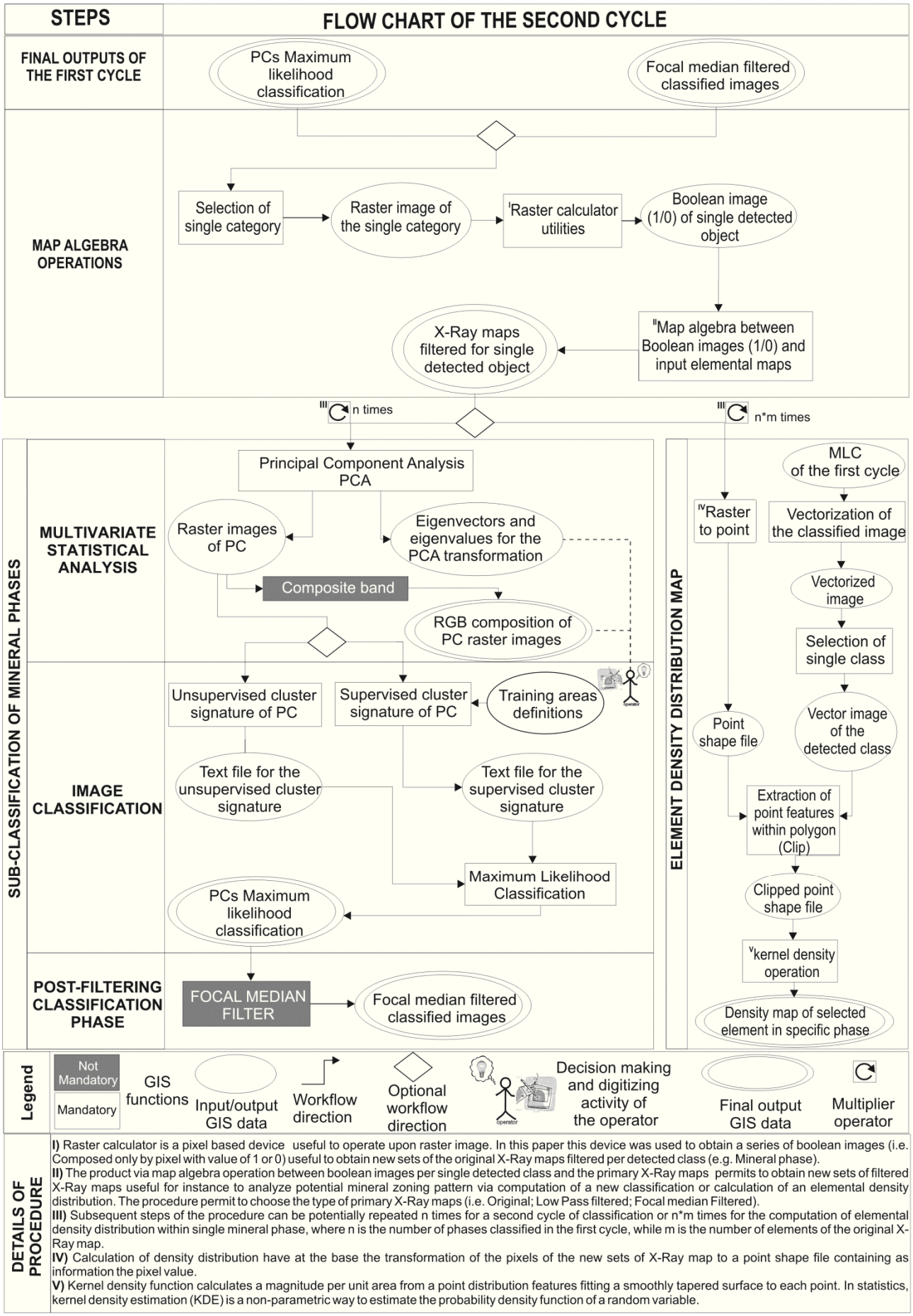


Figure 3 Flow chart of the second cycle of the geoprocessing procedure

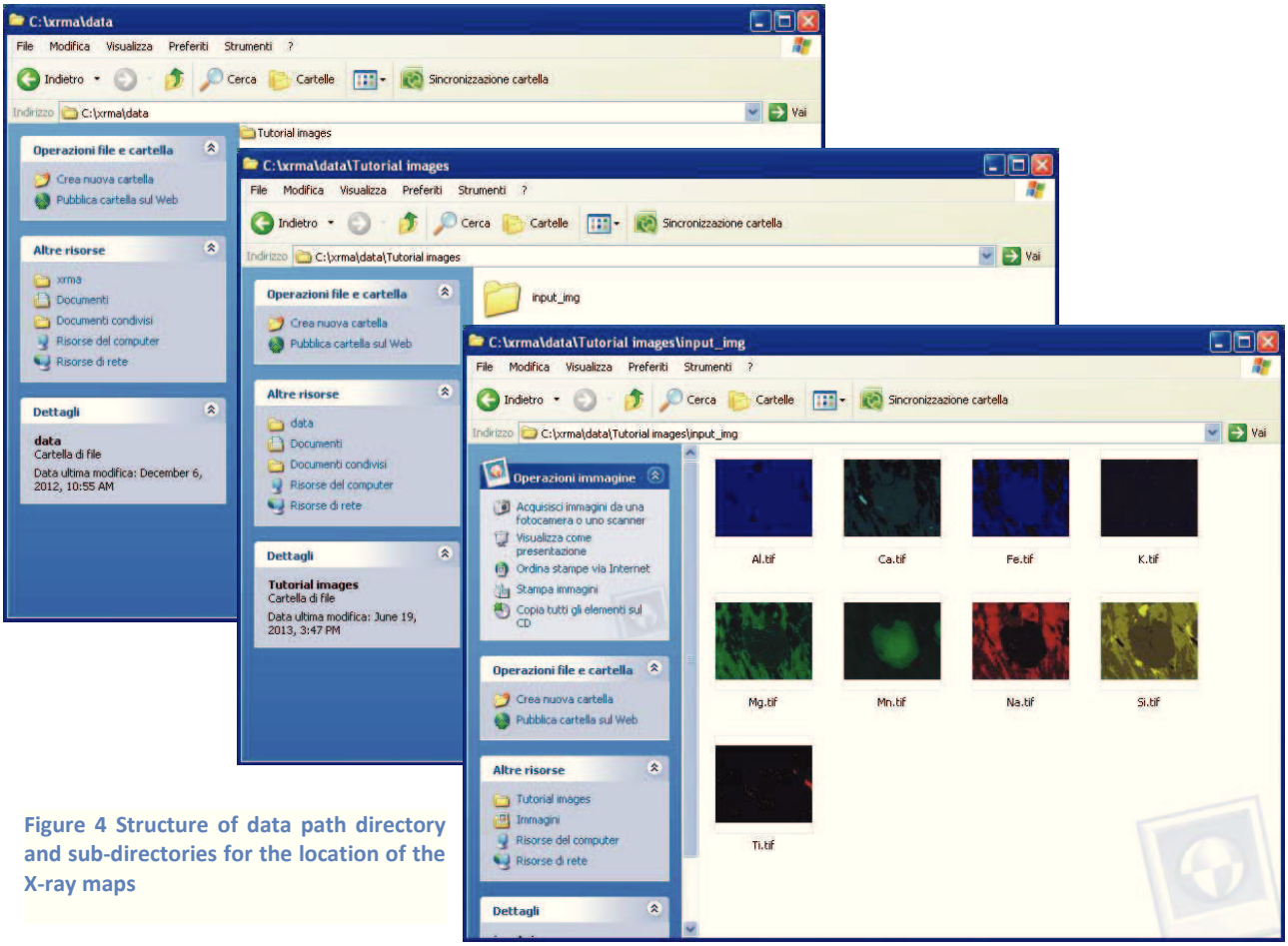


Figure 4 Structure of data path directory and sub-directories for the location of the X-ray maps

Pre-filtering phase

Following the procedure, user after have put the images to analyze into subdirectory c:\xrma\DataPath**MyWorkSpace**\input_img can choose which kind of filter, also no-filter (default choice), it needs to be applied in order to reduce the potential instrumental noise, such as those recognizable in the EDS X-ray map (Fig.6). The Low Pass Filter is applicable if the images appear very noisy. In alternative it is possible to apply a statistical filter such as Focal Median with the aim to reduce the weight of the isolated pixels.

In general, all the filtering procedures, calculates new pixel-values centring, for instance, a specified 3 x 3 matrix (i.e. filter) over each input raster cell (i.e. pixel). As the filter is passed over each cell, the centre is assigned the sum of the products of the cell value and the corresponding operand in the 3 x 3 filter. The output raster cell at the centre of the filter is assigned a new pixel-value based on the following formula:

$$N = Z1*F1 + Z2*F2 + Z3*F3 + ... + Z9*F9$$

Where N is the final matrix given by the new 3 x 3 matrix where each pixel is given by the product of the filter matrix Z with the original matrix F. When an input raster cell on the edge of the filter has a NoData value, the pixel-value of the cell is substituted for the missing values. On the edges of the raster, the filter lies partially outside the raster. When this occurs, the pixel-value of the cell at the centre of the filter is substituted for the missing values. In particular, a low pass filter smoothes the data by reducing local variation and removing noise calculating the average (mean) value for each 3 x 3 neighbourhood. The effect is that the high and low values within each neighbourhood will be averaged out, reducing the extreme values in the data.

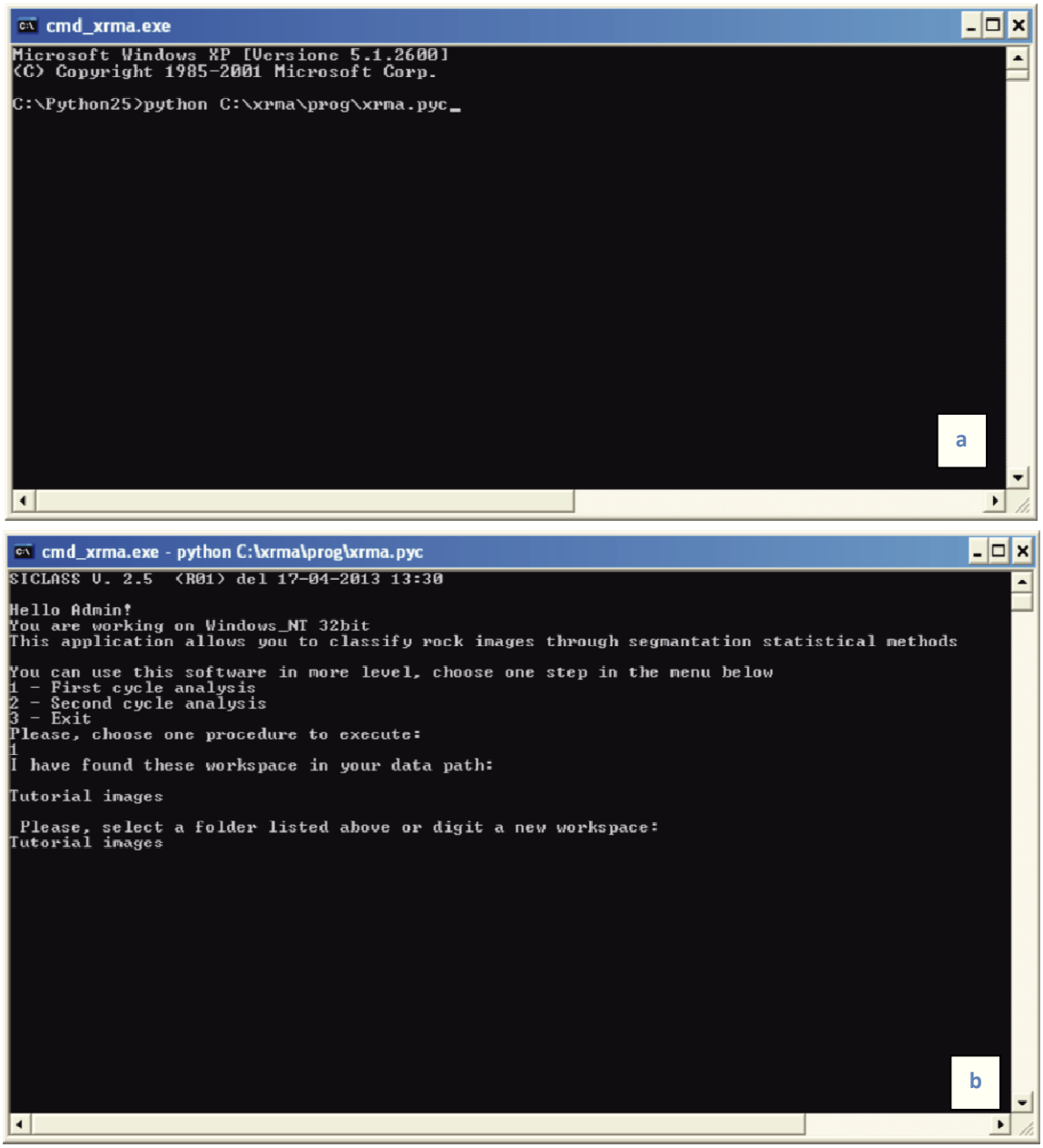
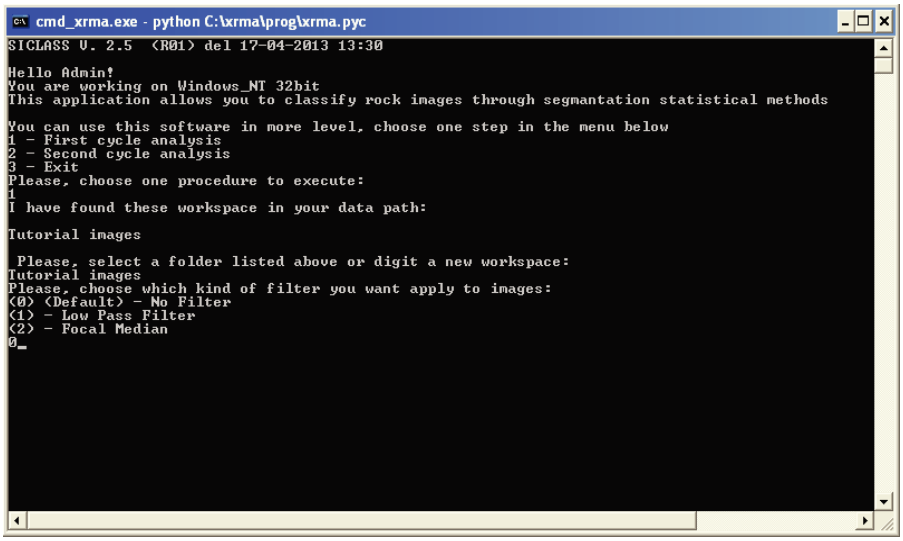


Figure 5 Example of the starting procedural steps of the script: a) starting instruction after opening of the DOS interface; b) choose of the data path directory

Figure 6 List of the applicable filtering method



The second applicable filtering process consists in a statistical procedure able to reduce the weight of the isolated pixel. In our procedure we choose as filtering method the focal median statistics, consisting in the transformation of the original pixel values matrix in a new matrix constituted by new pixel values derived by the calculation of the median value among all the neighbourhoods values around a specific geometry chosen by the user. In our case this geometric figure can be a square/rectangular shape of 1 to 9 pixel of length per side, or a circular geometry of 1 to 9 pixel unit of radius (Fig.7). A circular shaped filter of two pixels of diameter was chosen as default of our procedure.

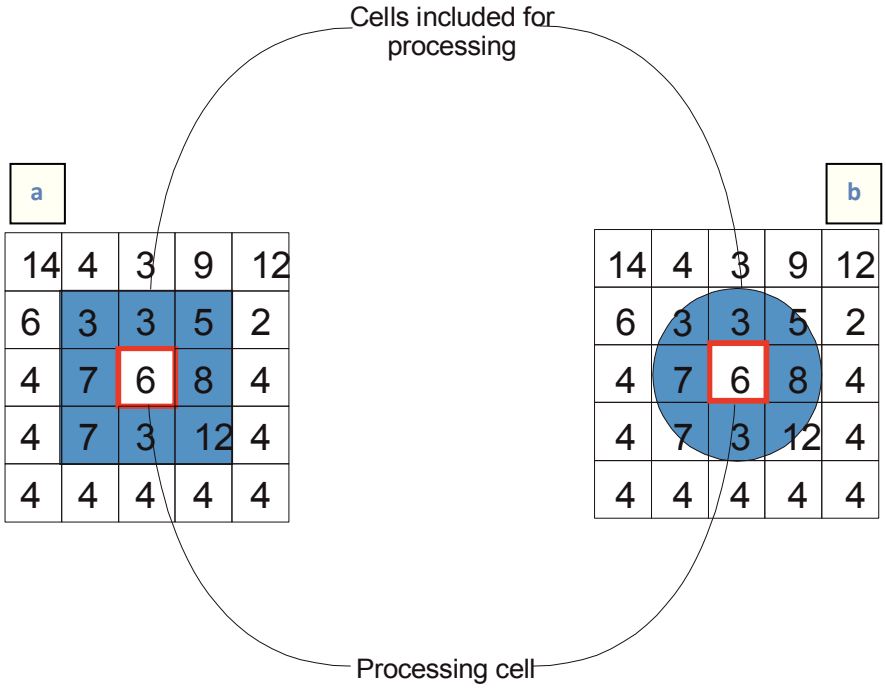


Figure 7 Schematic representation of the focal median principle used in the procedure: a) squared geometry with 3*3 cells per side; b) circled geometry with 3 cells of radius, any cells whose centre falls inside the radius of the circle will be included in processing neighbourhood

RGB synthesis and multivariate statistical analysis

The following operative function of the procedure (Fig.2) provides the possibility to make the RGB composition of the original X-ray map array making several colourmaps. These are used to display a single band raster per each available channel (i.e. one element per channel). In this way, each pixel value is associated with a colour, defined as a set of red, green, and blue (RGB) values (Fig.8). Since each value has a distinct colour associated with it, it will always display the same way each time you open it in a program that can read a raster with a colourmap. Colourmaps are capable of handling up to 8-bit data per channel (i.e. 256 colours).

In our procedure it is possible to chose no RGB composition or it is possible to do all the potential RGB composition following an automatic procedure consisting in a simple combination of the original channels without permutation or in a combination with all possible permutation (Fig.9). In particular, simple combination provides all the possible combination of the original images using each element just one time per channel, while combination with permutation provides all the possible combination with permutation of the primary channels taking care to avoid combination that use the same channel in two or three different primary colours. To this aim was implemented in python a recursive type algorithm which execute alternatively the two abovementioned functions, invoking in recursion the "Composite Band" function. Results will be automatically stored within the sub-directory: "c:\xrma\DataPath**MyWorkSpace**\out_img\FirstCycle\xr_rgb_comb" or "c:\xrma\DataPath**MyWorkSpace**\out_img\FirstCycle\xr_rgb_perm", respectively.

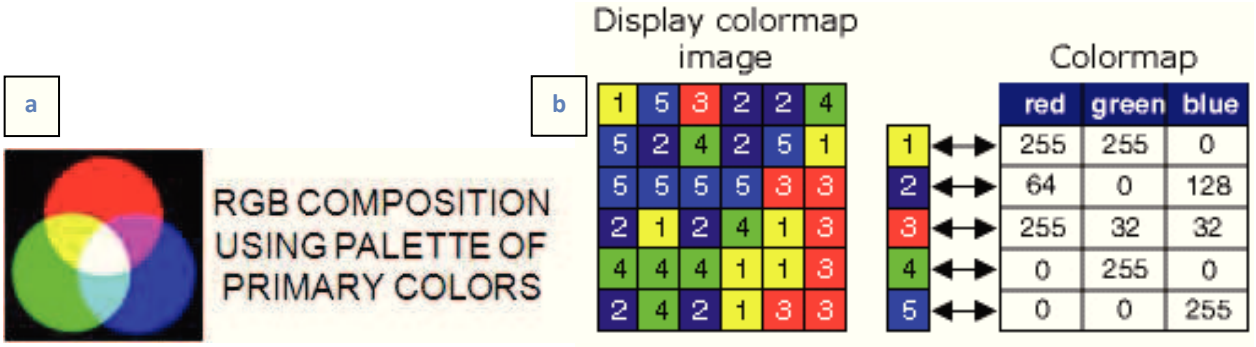
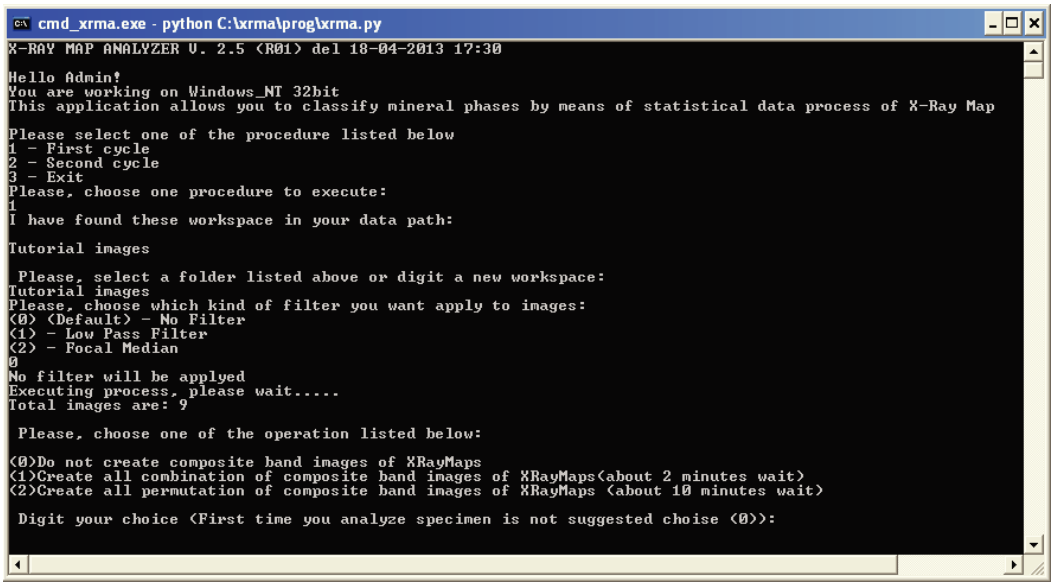


Figure 8 a) Schematic representation of the colour composition using primary colors and b) relative color composition on a map (After ArcGis tutorial)

Figure 9 List of the applicable RGB composition of the X-ray map array



After this last optional operative function, the procedure provides a mandatory function consisting in the principal component analysis of the original set of X-Ray map. The PCA is used to transform the data attributes of an original multiband raster, such as our X-ray map array, from the input multivariate attribute space to a new multivariate attribute space whose axes are rotated with respect to the original space (see Fig.10 as schematic example). The final result of a PCA is a new multiband raster with the same number of bands as the original raster (one band per axis in the new multivariate space), where the first principal component will have the greatest variance, the second will show the second most variance not described by the first, and so forth (Fig.10). A further output of the PCA consists to the construction of the matrix of the eigenvalues and eigenvectors reported in the second raw and one per column, respectively in the reported Table 1. This output can be visualized directly in the shell of the promt (Fig.11) or in □*.txt□form at the subdirectory: □c:\xrma\data**MyWorkSpace**\pca\pca_report□ The columns represent the new axes of the multidimensional coordinate system expressed as vectors oriented respect to the previous primary elemental map. The eigenvalues expressed as percentage of the total (see raw 3 in Table 1), highlights as most of the information is concentrated, in the first three to five principal components. In our case, for instance, calculated results concentrate about the 96% of information within the first five principal components. The main reason to transform original data distribution displayed in the n-elemental space to a new multidimensional space represented by PCA is to compress data by eliminating redundancy, emphasize the variance within the bands of a raster, and make the data more interpretable (Fig.12). This can aid the operator in recognizing the potential inter-dependence

(e.g. the specific covariance) between elemental variability linked, for instance, with the recognition of the mineral phases (Table 1).

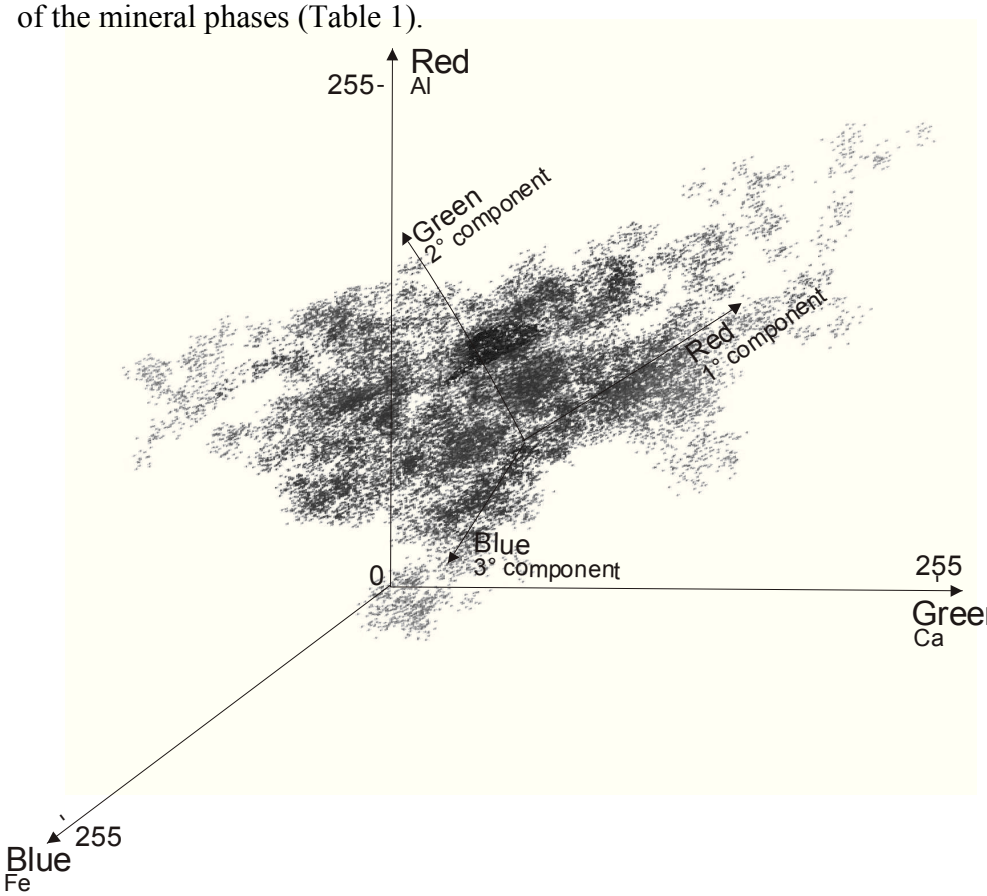


Figure 10 Example of schematic 3D representation of the decorrelation function known as principal component analysis, useful to re-project pixel values variability into a new coordinate system given by the axes of the principal components. This new coordinate system is able to minimize the redundancy (i.e. the superposition) of the original distribution, emphasizing class separability

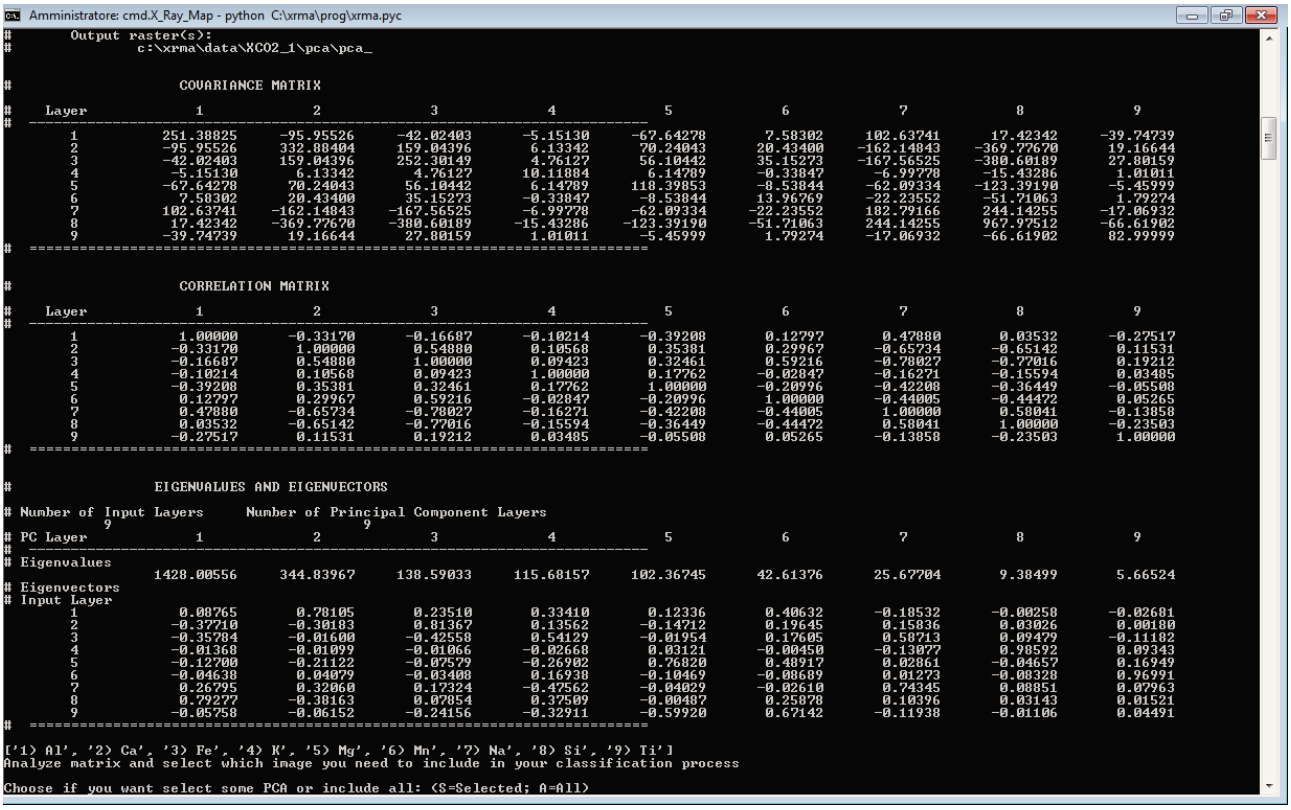


Figure 11 Visualization of the calculated covariance and correlation matrix within the prompt of the procedure. These are at the base for the calculation of the eigenvalues and eigenvectors matrix, which report, one per column, the coordinates of the new rotated axes which representing the principal components.

Table 1: Example of calculated eigenvectors and eigenvalues through the principal component analysis with relative elemental and mineral recognition per eigenvector

Principal Components	1th	2nd	3rd	4th	5th	6th	7th	8th	9th
Eigenvalues	1428.0	344.839	138.59033	115.68157	102.37	42.61376	25.6770	9.385	5.665
Eigenvalues %	64.53	15.58	6.263	5.22	4.62	1.92	1.16	0.42	0.25
Element	Eigenvectors								
Al	0.0876	0.78105	0.2351	0.3341	0.1233	0.40632	-0.1853	-0.003	-0.027
Ca	-0.377	-0.3018	0.81367	0.13562	-0.147	0.1964	0.1583	0.0302	0.002
Fe	-0.358	-0.016	-0.42558	0.54129	-0.02	0.17605	0.5871	0.0947	-0.112
K	-0.014	-0.0109	-0.0106	-0.0266	0.0312	-0.0045	-0.1307	0.9859	0.093
Mg	-0.127	-0.2112	-0.07579	-0.26902	0.7682	0.48917	0.0286	-0.046	0.169
Mn	-0.046	0.04079	-0.03408	0.16938	-0.104	-0.08689	0.0127	-0.083	0.97
Na	0.2679	0.3206	0.17324	-0.47562	-0.040	-0.0261	0.7434	0.0885	0.08
Si	0.7927	-0.3816	0.07854	0.37509	-0.004	0.25878	0.1039	0.0314	0.015
Ti	-0.058	-0.0615	-0.24156	-0.32911	-0.599	0.67142	-0.1193	-0.011	0.045
Elemental brightness	Si-Na-Al	Al-Na-Mn	Ca-Al-Na-Si	Fe-Si-Al-Mn-Ca	Mg-Al	Ti-Mg-Al-Si-Ca-Fe	Na-Fe-Ca	K-Fe-Na	Mn-Mg
Mineral recognition	**Qtz+Pl**	**Pl+Grt**	**Cc+Ap+Ttn**	**Grt-Qtz**	**Amph+ Chl**	**Ilm+Chl+Ttn**	**Pl+FeOx**	**Wmca**	**Chl**

Mineral phases can be recognised, indeed, by the fact that the pixel brightness in the new multidimensional space is due to the eigenvector orientation values ranging from -1(pure black) to 1(pure white) (Table 1) (Fig.12). In the first principal component for instance, the brightest value is given by Si which account for 0.792, followed by Na (0.269) and Al (0.087) (Table 1) (Fig.12b). In this case, is then reliable that the major contribution to the final brightness is given by those pixels representative for quartz and albite. The following second principal component is due instead by the fundamental elemental contribution of Al □Na □Mn, emphasizing the visualization of albite and garnet (Table 1) (Fig.12b). This new graphical representation permit then to emphasize the maximum separability between recognisable classes, because most of the information can be directly visualized in the first three to five principal components. In this way, for instance, the RGB synthesis of three selected principal components is majorly informative than the variability expressed from a maximum of three element per RGB synthesis, usable for the composition of the elemental channel (Fig.12a).
In the following operative function, the prompt ask the user to produce or not a series of RGB combination of the new principal components.

Please, choose one of the operation listed below:

(0)Do not create composite band images
(1)Create automatically just one composite band image with PCA123
(2)Create all combination of composite band images with all PCA (about 2 minutes wait)
(3)Create all permutation of composite band images with all PCA (about 10 minutes wait)

The output results of this last operation can be visualized in *.tif format in the subdirectory

C:\xrma\data**MyWorkSpace**\out_img\FirstCycle\rgb_comb
or
C:\xrma\data**MyWorkSpace**\out_img\FirstCycle\rgb_perm

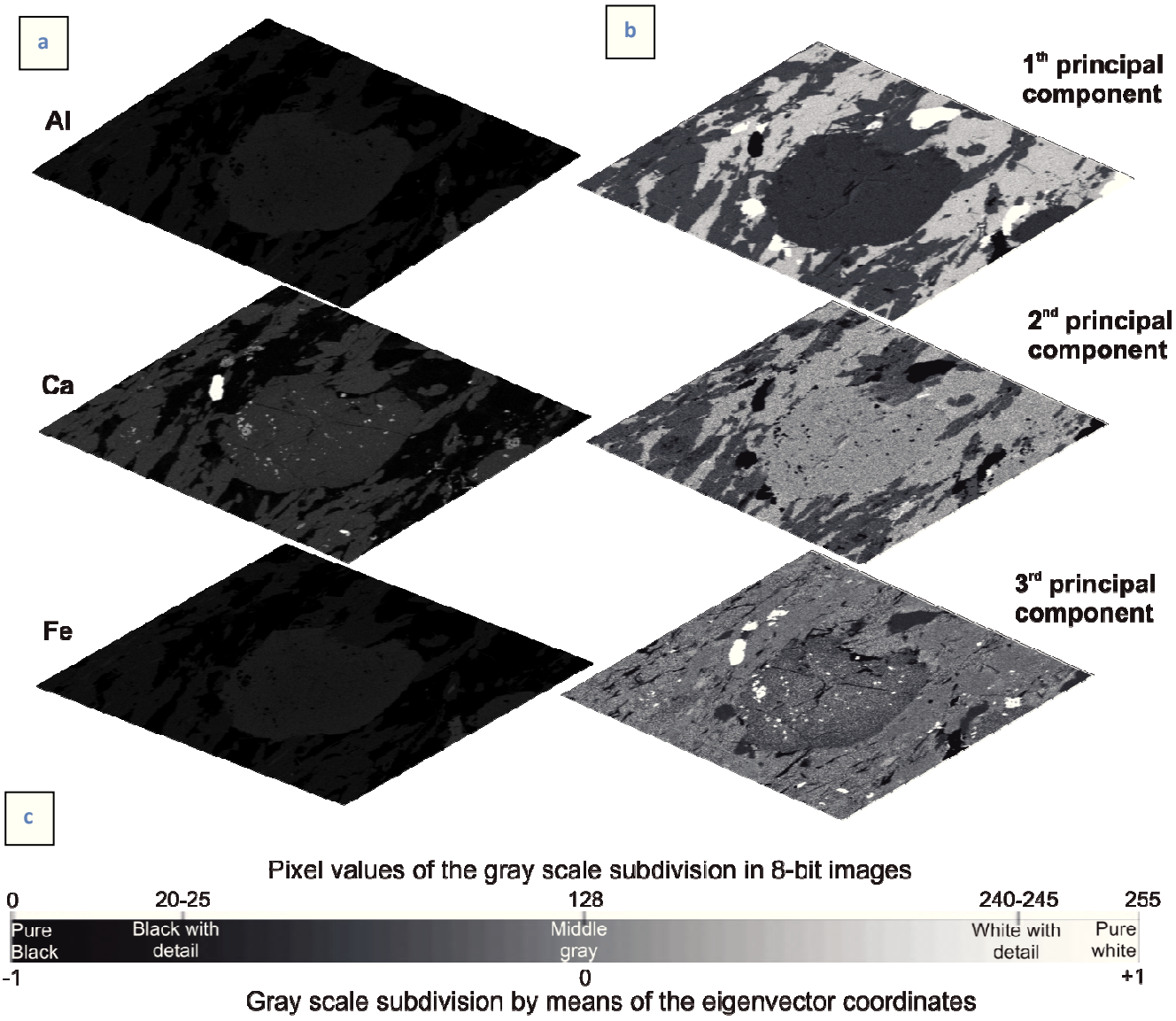


Figure 12 a) Superposition of three elemental channels expressed as gray-scale images; b) Superposition of the first three principal components calculated from the original Al, Ca, Fe, K, Mg, Mn, Na, Si, Ti X-ray maps array; c) Generic gray scale bar of 8-bit images, in the upper side the subdivision is expressed in the 0 (pure black) - 255 (pure white) classical range, in the lower side the subdivision is expressed in the range -1 (pure black) - +1 (pure white) as a function of the eigenvector coordinates of the specific principal component vector; d) RGB composition from Al-CA-Fe primary element map; e) RGB composition of the first three principal components calculated on the basis of the original nine elemental distribution map (i.e. Al, Ca, Fe, K, Mg, Mn, Na, Si, Ti)

Image classification

First cycle procedure continues with the possibility to make the classification of all recognizable classes present in selected micro-domain. During this processing stage the prompt asks to operator which kind of classification he wants to execute:

- □U□ - Unsupervised Classification, without interaction by user. In this case the algorithm creates a signature file stored in: □C:\xrma\data\MyWorkSpace\sign\isoclus1.gsg□based on an iso-cluster function useful to determine the characteristics of the natural groupings of cells in multidimensional attribute space.
- □S□ - Supervised Classification, where the user has to prepare some training areas assisted by the visualization of the original X-ray map array and/or by the new principal components channels, finalized to a partial mineral pre-recognition. This training areas are then used by the signature function to prepare the □signature file□ stored in the sub-directory: □C:\xrma\data**MyWorkSpace**\sign\signpca_cycle1.gsg□ This file is then used by the Maximum Likelihood Classification (MLC) algorithm, which assumes that each spectral class can be described by a multivariate normal distribution. Therefore, MLC takes advantage of both the mean vectors and the multivariate spreads of each class.

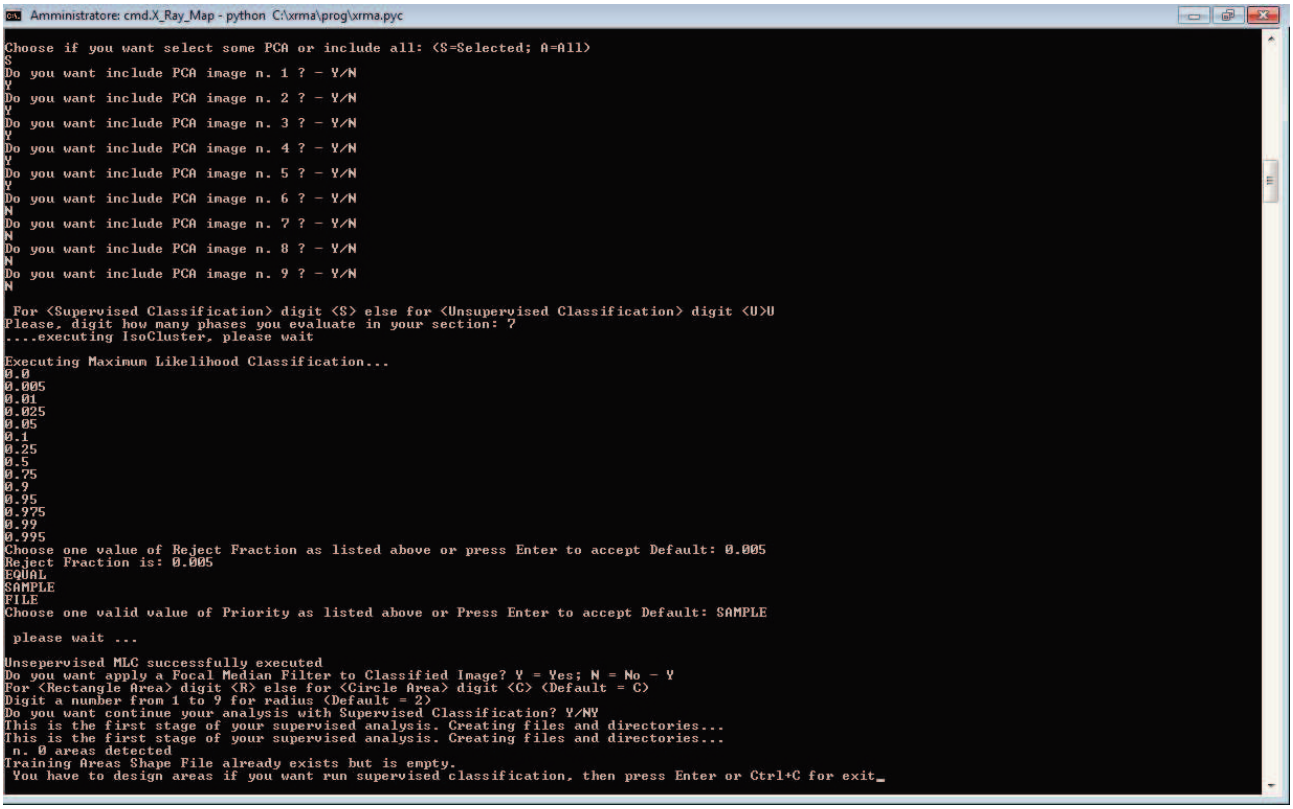


Figure 13 Visualization of the promt interface reporting the selected choice of the principal components considered for the MLC, followed by the unsupervised classification procedural step characterized by the classification default options relatively to the reject factor (i.e. 0.005) and priority classification mode (i.e. SAMPLE). The following supervised classification mode allow to be create the training areas

If user digit □U□

system performs an automatic classification of the selected X-ray mapset, asking simple informations to complete the cycle, such as the number of PC images to include in the calculation thread, the approximate number of expected classes of subdivisions (e.g. mineral

phases), the rejected fraction of unclassified pixels, how a priori probabilities will be determined[1] (Fig.13). The result of the unsupervised classification procedure will be then stored in the sub-directory "c:\xrma**MyWorkspace**\mlc\mlc_uns_pca".

The output image of unsupervised MLC function can be successively filtered with a post classification focal median filter, with the potential aim to obtain a balanced image without presence of noisily pixels. This last output can be found at the sub-directory "c:\xrma**MyWorkspace**\mlc\mlc_uns_focal"(Fig.14).

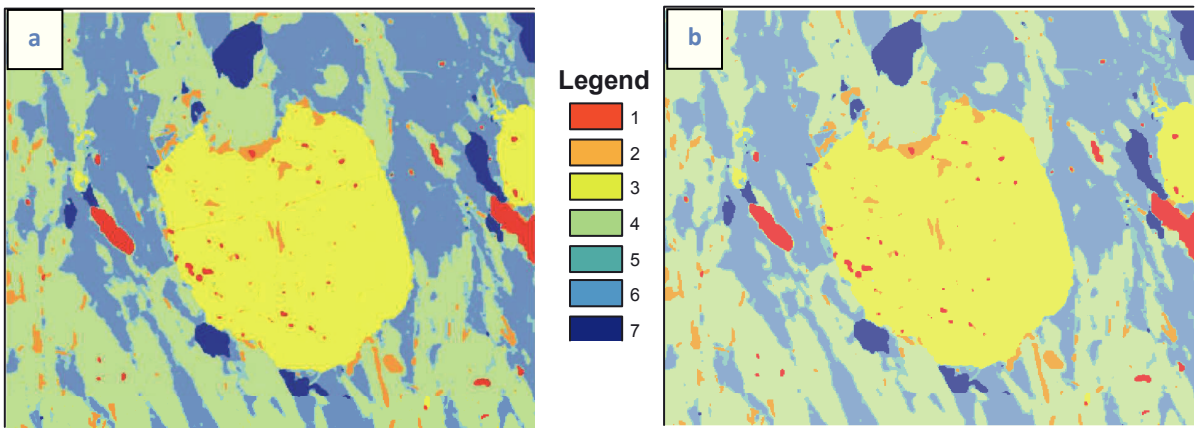


Figure 14 Unsupervised classification results: a) without focal median post-classification filtering; b) with focal median post-classification filtering

If user digit "S":

system verifies if the directory "c:\xrma\input_vect"is empty or contains valid training areas files, so in this case continues his calculation, differently the system ask to the user to create training areas, remaining in standby operative mode (Fig.13).

In order to create training areas user have to open ArcMap and edits "TrArea1.shp"placed in "c:\xrma**MyWorkspace**\input_vect" (see informative video). After this step, it is just necessary press enter to continue the procedure. At the end of this part of the procedure, the command prompt ultimate the supervised classification and obtain a classified image called mlc_1_sup. Also in this case the output image of MLC function can be filtered with a focal median post-processing filter, so it's possible to obtain a balanced image without presence of noisily pixel stored into sub-directory: "c:\xrma**MyWorkspace**\mlc\mlc_sup_focal" (Fig.15).

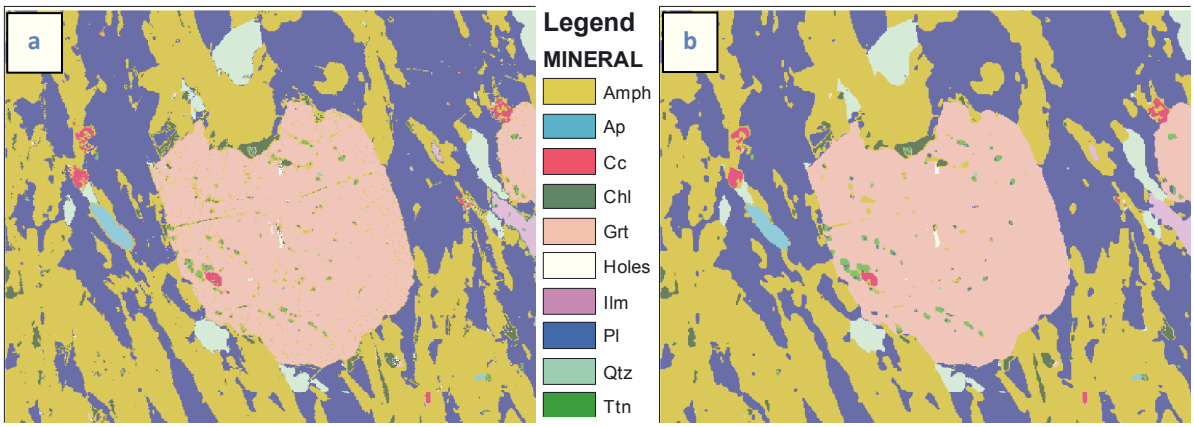


Figure 15 Supervised classification results: a) without focal median post-classification filtering; b) with focal median post-classification filtering

[1] When the EQUAL a priori option is specified, each cell is classified to the class to which it has the highest probability of being a member. By choosing the SAMPLE a priori option, the a priori probabilities assigned to all classes sampled in the input signature file will be proportional to the number of cells captured in each signature.

Second Cycle

In order to launch the second cycle, it is necessary open the MS-DOS cmd within the C:\Python25 directory then select "Second Cycle" on the input menu.
The procedure allows in this case the user to investigate, with more detail, within an arbitrary selected phase recognized in the first cycle. At the end of the first cycle it is indeed possible use the results of the MLC, in order to select one or more phases to highlight for instance the potential existence of chemical gradient within a mineral phase as in the case of zonation. At the base of the second analytical cycle there is the creation via map algebra operations of new sets of X-Ray maps characteristics per single detected phase. After this precursor operative step, the cycle proceeds with two distinct analytical possibilities. The first one (sub-classification of mineral phases) allows to make the sub-classification of a selected mineral phase, while the second one (element density distribution map) calculates the density distribution of an element within a selected phase (Fig.3). In this view, the procedure can be potentially repeated n times for the sub-classification procedure or n*m times for the computation of elemental density distribution within single mineral phase, where n is the number of phases classified in the first cycle, while m is the number of the elements detected from original X-Ray map.

Map algebra operation
This function provides a pixel based raster calculation useful to obtain a series of new X-ray maps array per single detected object aiming to focus in the subsequent operative steps the pixels relatively to one single phase. After selection of the workspace, the system creates a set of image files representing each chemical element in each mineral phase, applying a map algebra operation and then shows a list of the distinct mineral phases indentified in the first cycle (Fig.16). This procedure starts to calculate a series of boolean images (i.e. composed only by pixel with value of 1 or 0). These are useful to obtain the new sets of filtered X-Ray maps via the product between boolean images per single detected class and the primary X-Ray maps (i.e. original, low pass filtered or focal median filtered) (Fig.16).
To do this, the prompt asks to the user if prefer to make the new sets of X-ray map per single detected mineral phase by means of the classes recognized during the maximum likelihood classification procedure or, those classes recognized by the focal median filtered classified image. After this operation able to produce several sets of new X-ray maps as many as were the original ones (see e-p - element in phase sub-directory), it is possible to choose one phase at a time, in order to investigate deeply about it, obtaining more details about potential inner chemical distribution.

Sub-classification of mineral phases
After the choice of Maximum Likelihood Classification of a mineral phase (Fig.17), the promt ask to choice a specific phase where to execute again the principal component analysis on the specific X-ray map array representative for the single selected class (Fig.17). Applying then the PCA function for those elements which potentially show the most variable elemental distribution recognizable in the selected class (e.g. Ca; Fe; Mg; Mn for Grt) (Fig.18), will be possible to perform new images useful for a new maximum likelihood classification in the same manner produced for the first cycle, but working just on the pixels of the single detected mineral phase (Fig.19). It is worth noting that the percentage of the eigenvalue concentrated in the first principal component drastically rise up respect to the previous first cycle of analysis.

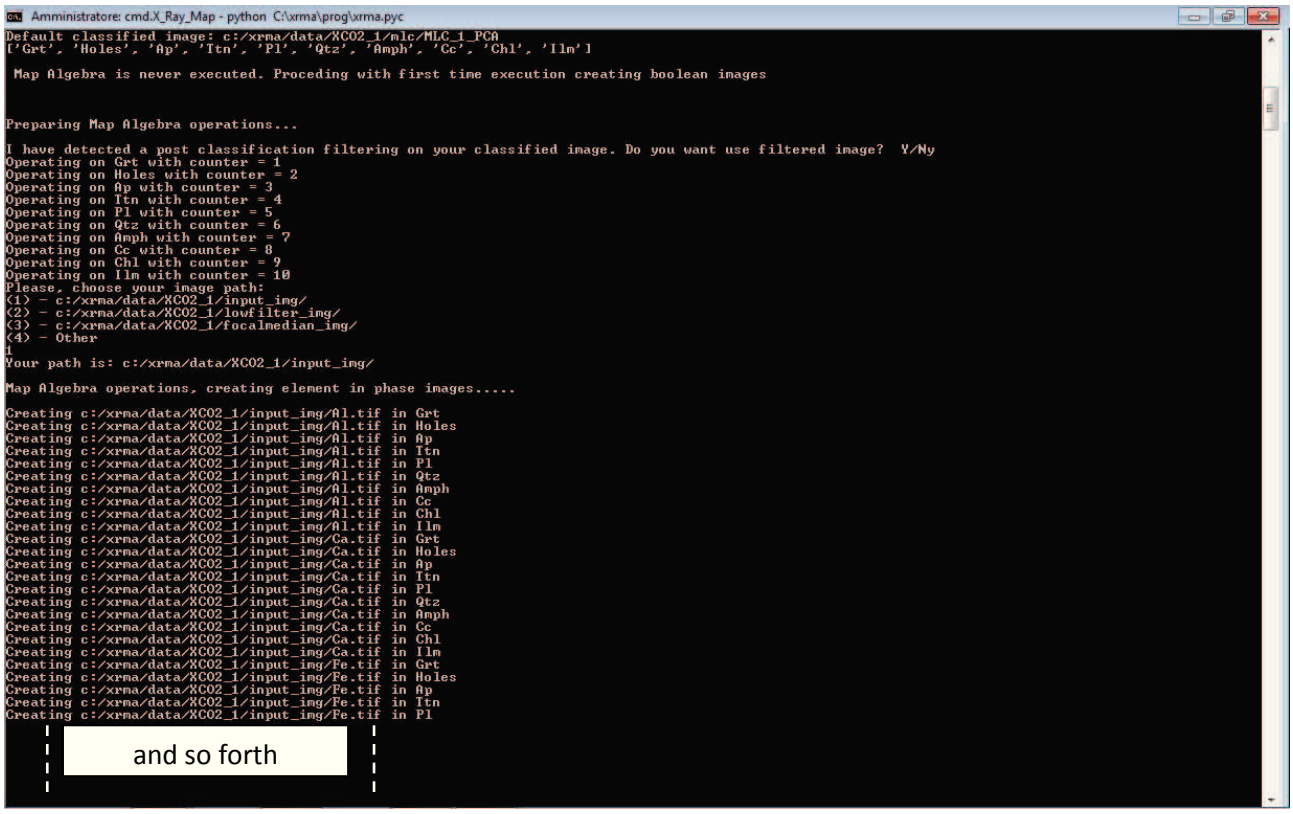


Figure 16 Visualization of the map algebra operation execute at the beginning of the second analytical stage

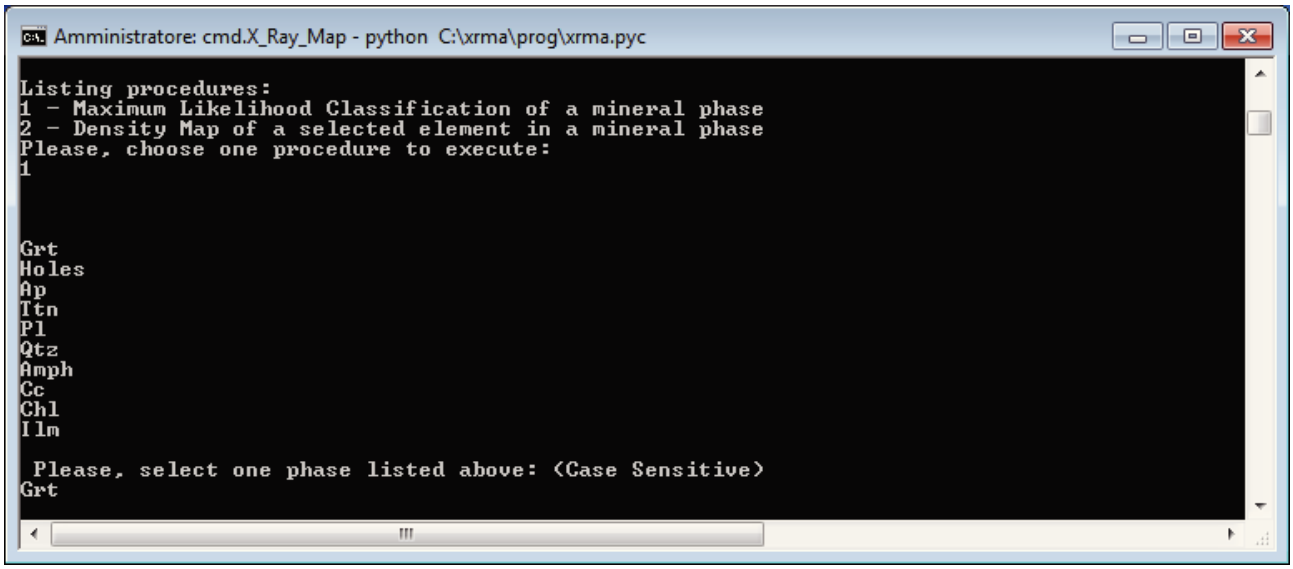


Figure 17 Chose of the sub-classification analytical procedure to be execute in the second cycle for garnet maximum likelihood classification

Like in the first cycle, user can choose if execute unsupervised or supervised classification. Nevertheless, commonly is better execute directly the supervised one, due to the easy possibility to depict the training areas editable within a new TrainingAreas file (TrArea2.shp) automatically stored inside the sub-folder: "c:\xrma**MyWorkspace**\SecCycle\"+Selected Phase\input_vect". After editing and saved the new "*.shp" file, it is only necessary press just enter to continue the procedure or restart xrma.pyc, choosing "Second Cycle", if previously shut down the software. Yielded results, stored in the path "c:\xrma**MyWorkspace**\SecCycle\"+Selected Phase\mlc", can be then post-classified with a focal median filtering stage to minimize the weight of isolated pixels. Obtained results can be opened in ArcGis and exported in one of the most used raster formats (Fig.20) as well as used to extrapolate quantitative results by means of ArcGis functions (e.g. histograms of the volume percentages) (see informative video).

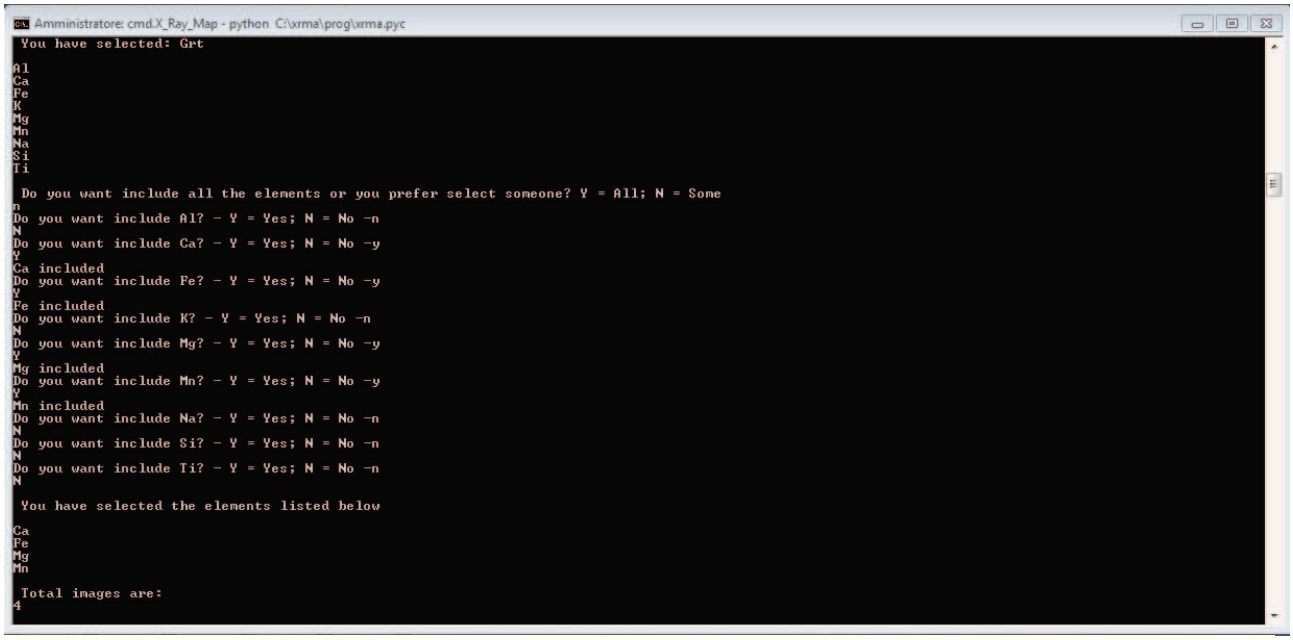


Figure 18 Chose of the elemental images to be included in the second cycle of principal component analysis for garnet

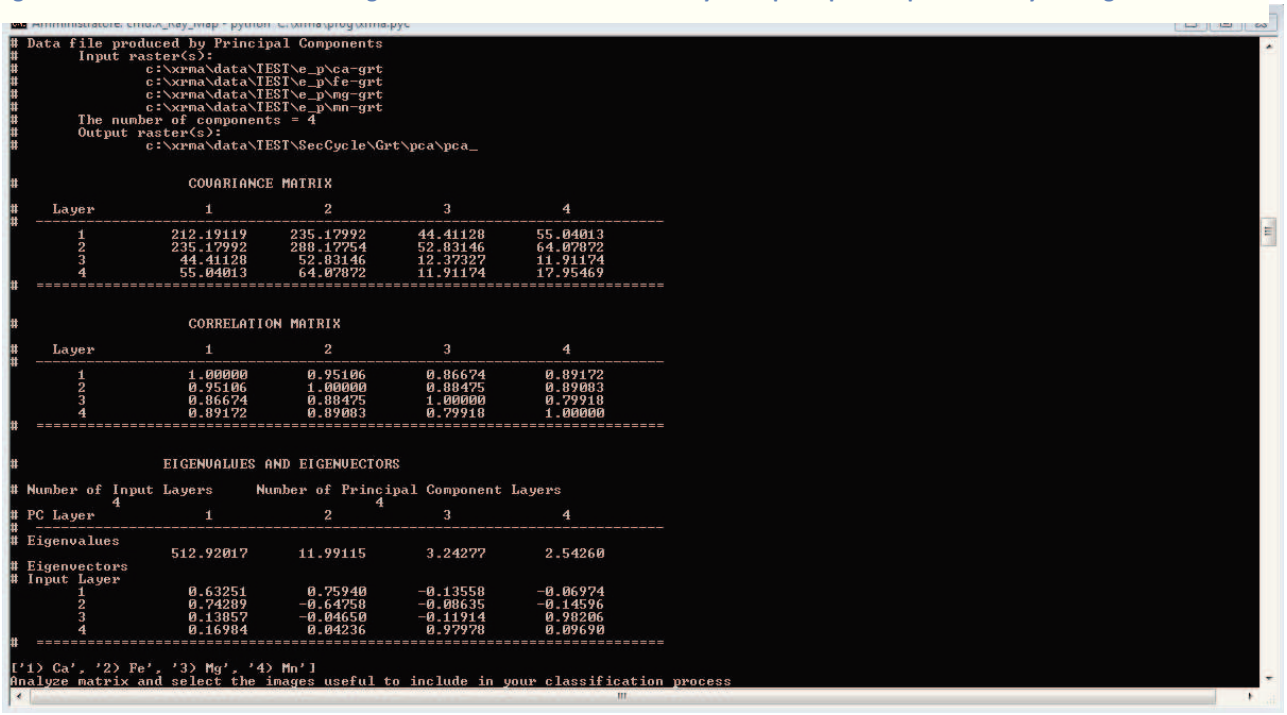


Figure 19 Principal component analysis of the selected elemental images within garnet

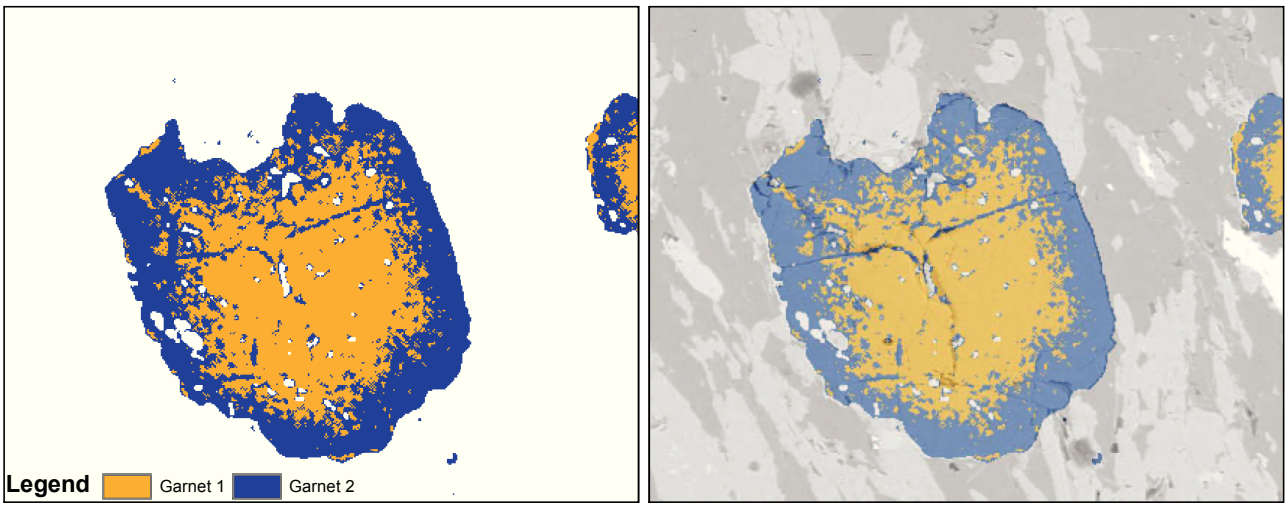


Figure 20 Supervised classification of the second analytical cycle for garnet, a) without and b) with superposition on BSE image

Density map classification procedure

This stage of the second cycle is very simple to execute, after have been chosen at the beginning of the second cycle the procedure n° 2: "Density Map of a selected element in a mineral phase" fig.17), it is just necessary to choose one phase and one chemical element to obtain an elaboration of density distribution of a selected element within selected phase by means of a kernel density operational mode. The results files will be written automatically to subfolder: "C:\xrma\data\XCO2_1\refine\element name-phase name" Kernel density function calculates a magnitude per unit area from a point distribution features fitting a smoothly tapered surface to each point. In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable (Fig.21).

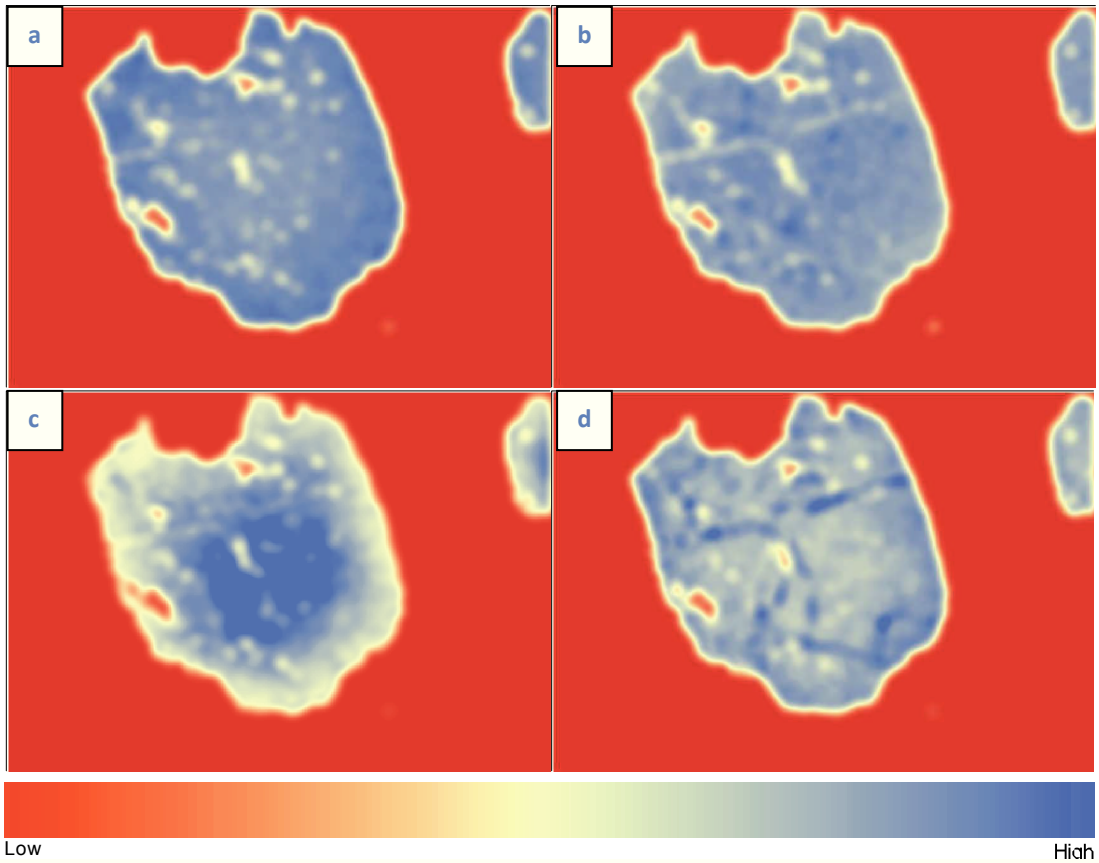


Figure 21 Distribution of elemental concentration density within garnet calculated via kernel density algorithm: a) iron; b) calcium; c) manganese and d) magnesium distribution. It is worth noting that the colour palette can be set arbitrarily using ArcGis symbology function