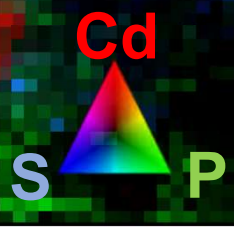
Image analysis of element distribution maps

Dr Paula Pongrac
paula.pongrac@bf.uni-lj.si

Instructions for MSc and PhD students and scientists looking to optimise visualisation of element distribution maps, to determine tissue-specific concentration of elements and to identify tissue-specific elemental profiles using multivariate statistics.



Univerza v Ljubljani
Biotehniška fakulteta

Institut "Jožef Stefan"
F2 / Odsek za fiziko nizkih in srednjih energij

Slovenian Research Agency

Title: Image analysis of element distribution maps
Author: Paula Pongrac
Publisher: Laboratory for Botany and Plant Physiology, Department of Biology, Biotechnical Faculty, University of Ljubljana
Designer: Paula Pongrac
Place of publication: Ljubljana, Slovenia
Year of publication: 2021
Number of editions: Electronic edition
Web site: https://repozitorij.uni-lj.si/info/index.php

Funding
The author was financially supported by the Slovenian Research Agency (Javna Agencija za raziskovalno dejavnost Republike Slovenije) through projects N7-0077 and N1-0105.

Acknowledgements
The author acknowledges continuous scientific support from Prof. Katarina Vogel-Mikuš and detailed revision of the text by Jure Mravlje.

Image on the title page: distribution of cadmium (Cd), phosphorus (P) and sulphur (S) in a stem cross section of *Gomphrena clausennii* grown hydroponically (Pongrac et al. 2018. Cadmium associates with oxalate in calcium oxalate crystals and competes with calcium for translocation to stems in the cadmium bioindicator *Gomphrena claussenii*. Metallomics 10, 1576–84. doi:10.1039/C8MT00149A). Author: Paula Pongrac

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 62029059
ISBN 978-961-6822-73-2 (PDF)

Content

You will need:

- Element-specific distribution maps in .csv formats.

  For the micro-particle induced X-ray emission (micro-PIXE) each file, when opened in Excel, is a matrix of 256×256 cells (=pixels). Each cell contains a value (element concentration in mg kg^{-1} dry weight). The matrix is the basis for the image generation and image analysis.

- ImageJ (Fiji): https://imagej.net/Fiji

- Microsoft Excel or similar (with settings for English punctuation)

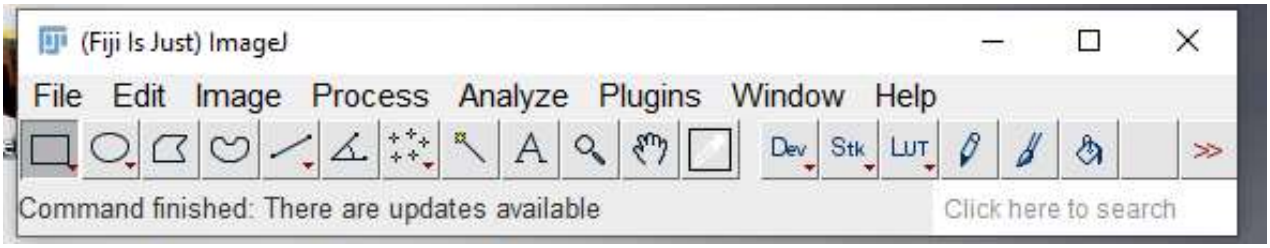
  In English punctuation decimal place is designated with a dot and thousands are separated with a comma. This differs from settings for e.g. Slovenian punctuation, where decimal place is designated with a comma and thousands are separated with a dot.

- Orange Data Mining: https://orangedatamining.com/

I. Optimising visualisation of element distribution maps

Place all distribution maps of a measurement in an easily accessible folder.

Install and open ImageJ (Fiji).

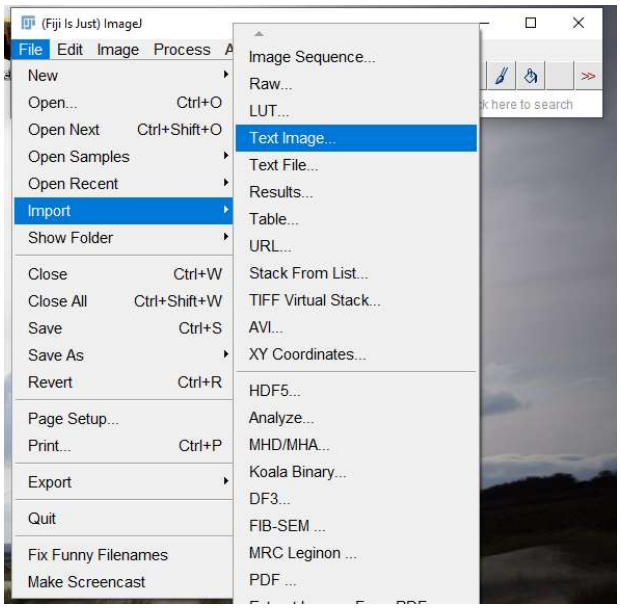


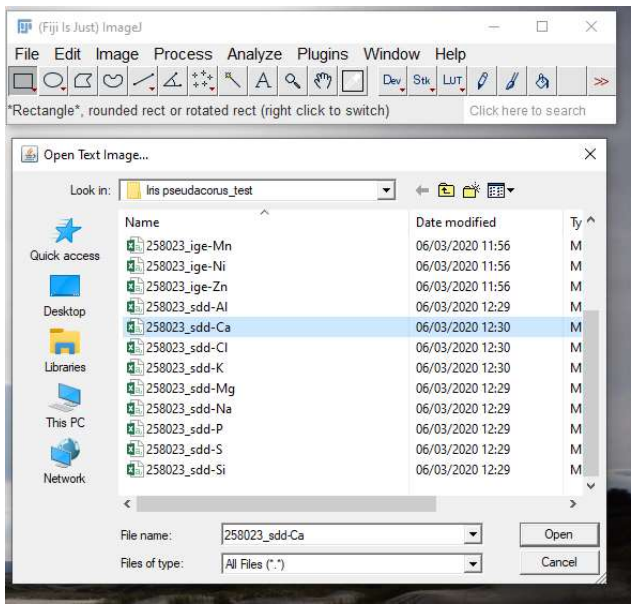
To load the distribution maps: **File > Import > Text Image** (select the .csv, i.e. „comma separated value“ file for each element separately) **> Open**.

For micro-PIXE distribution maps, the file name consists of

„run number_ige or _sdd-NameOfTheElement“.

e.g.: 258023_ige_Ca

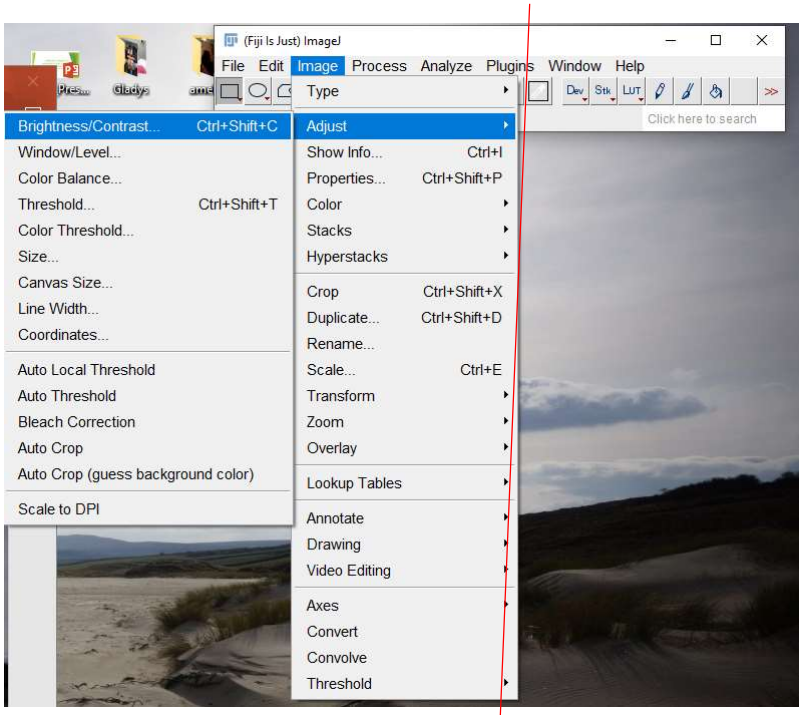


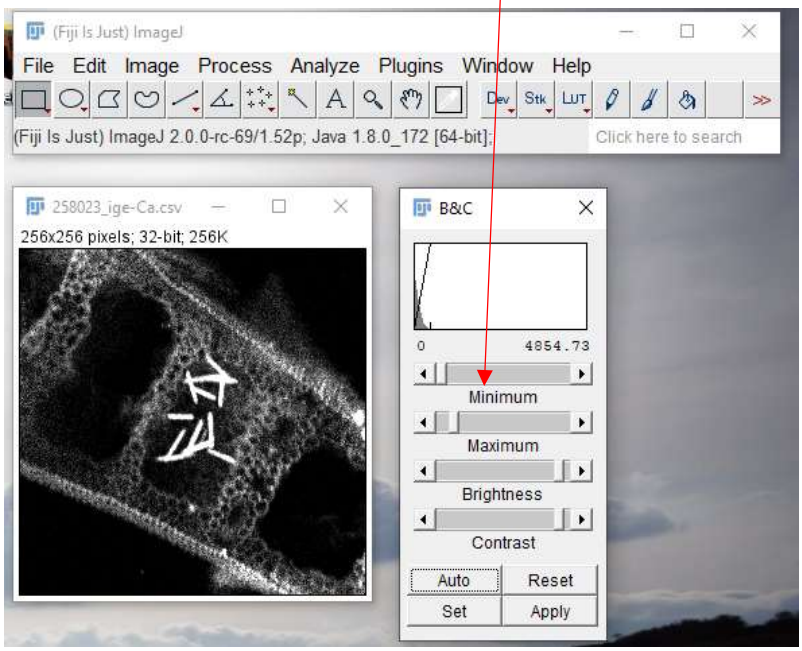


This sample is a leaf cross section of *Iris pseudacorus* L.

The Ca distribution maps is displayed, but the signal (distribution of concentration of Ca) in the image is poorly visible.

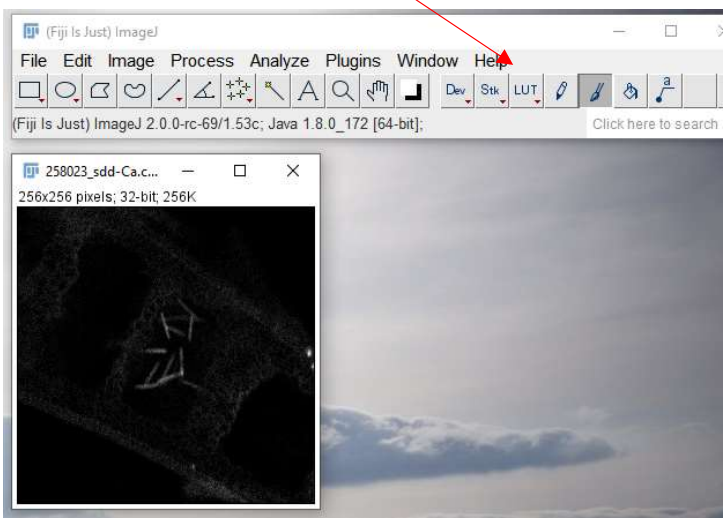
To increase contrast: **Image > Adjust > Brightness/Contrast** (adjust the values at Maximum by moving the slider to the left).

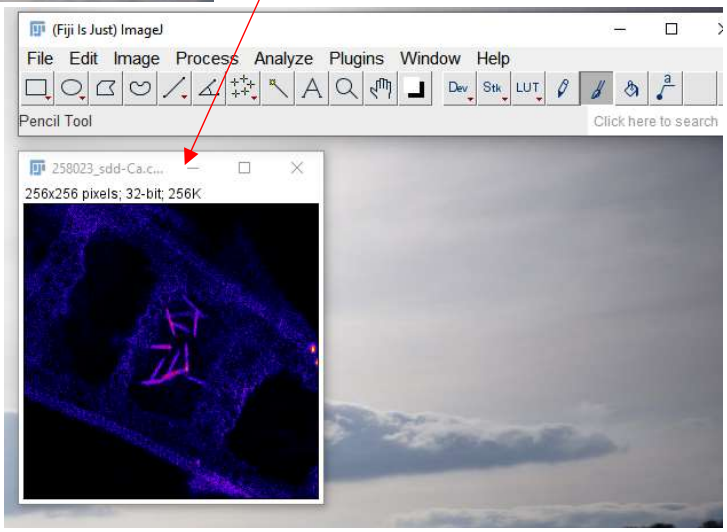




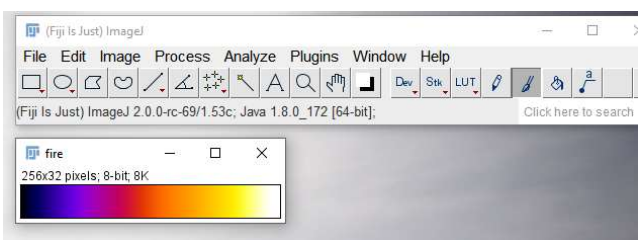
Note: now the signal in the image Ca distribution is more apparent.

To change the colour in your distribution maps, there are many options in ImageJ and they can be accessed through **Look-Up Table** (**LUT**). Select one of the LUT, e.g. Fire and your image changes the colour according to the colour scale unique to this LUT.

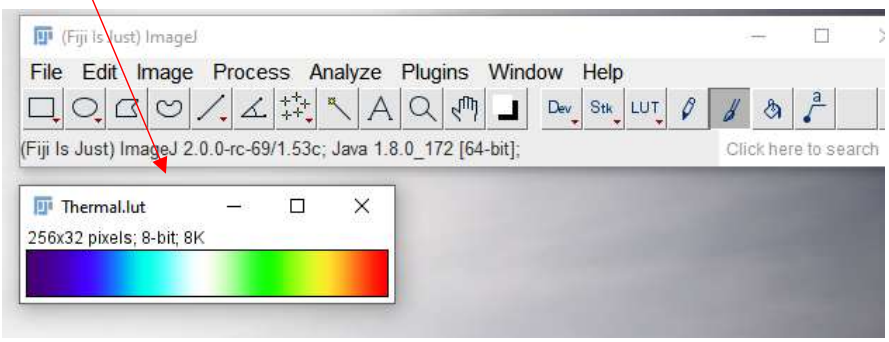




To save the colour scale of the LUT selected, close any distribution map(s) and click **LUT > Fire** again. The colour scale appears in a new window. You can save it: **File > Save As** (as e.g. .jpeg or .png). You will need this colour scale to be placed next to distribution maps.

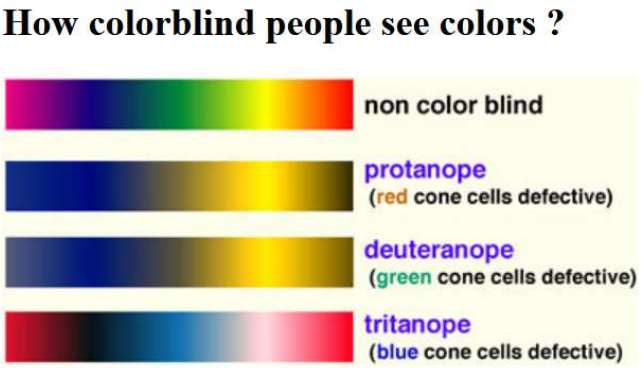


You can choose among several different LUTs. For example here is how Thermal colour scale looks like:



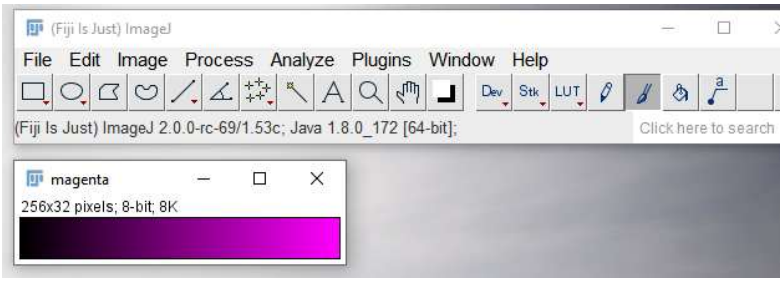
Note: the colours on the left hand side indicate smaller concentrations and those towards the right hand side indicate larger concentrations.

Note: While appearance of the distribution maps is important, do not forget that not all people see colours in the same way. Therefore, have in mind the following „problematic" colours:



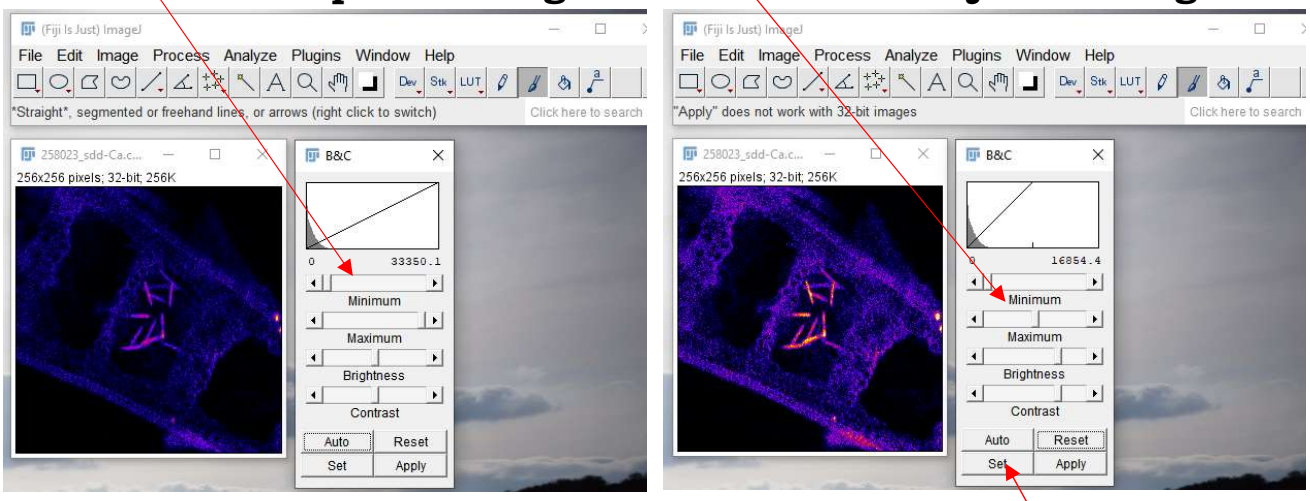
https://jfly.uni-koeln.de/color/

Thus, you can choose to use monochrome colours (gradients of a single colour), for example: Magenta:

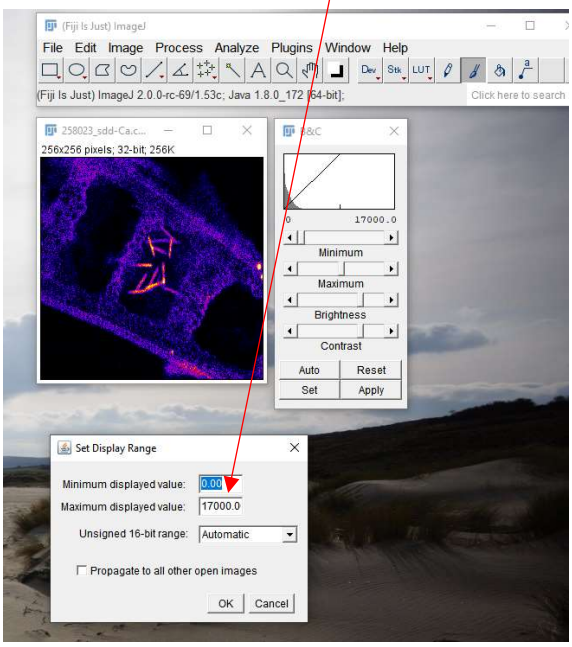


The displayed minimum and maximum value can be adjusted: **Image > Adjust > Brightness/Contrast:** by moving the slider at Maximum you can create a more contrast image (here LUT = Fire).

Imported image **Adjusted image**



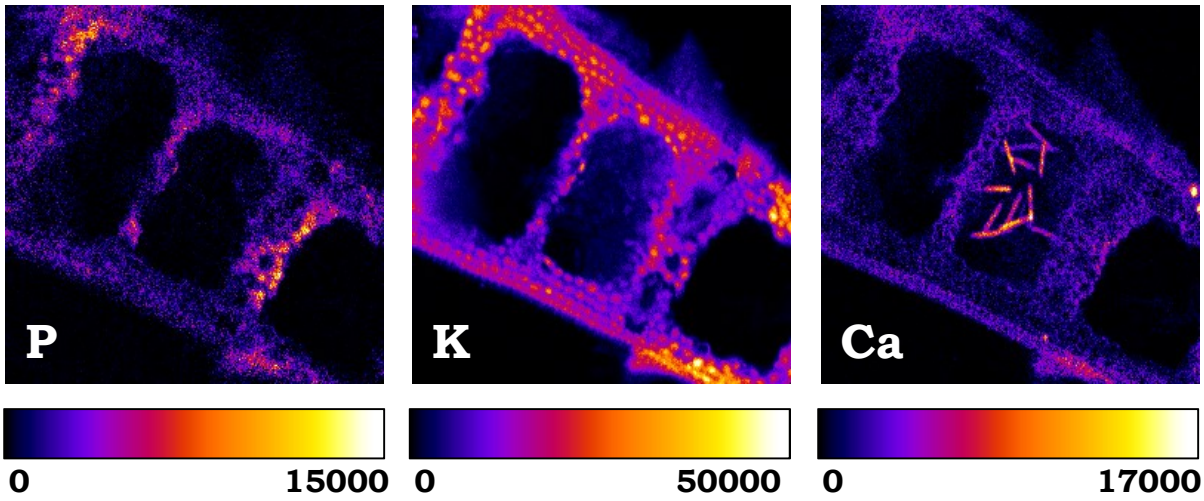
If you want an exact value to be your maximum (when comparing results of different treatments you will need to have distribution maps of individual element set to the same maximum) click **Set** and type in the exact value, e.g.17000 (this is concentration of Ca in mg kg^{-1} dry weight).



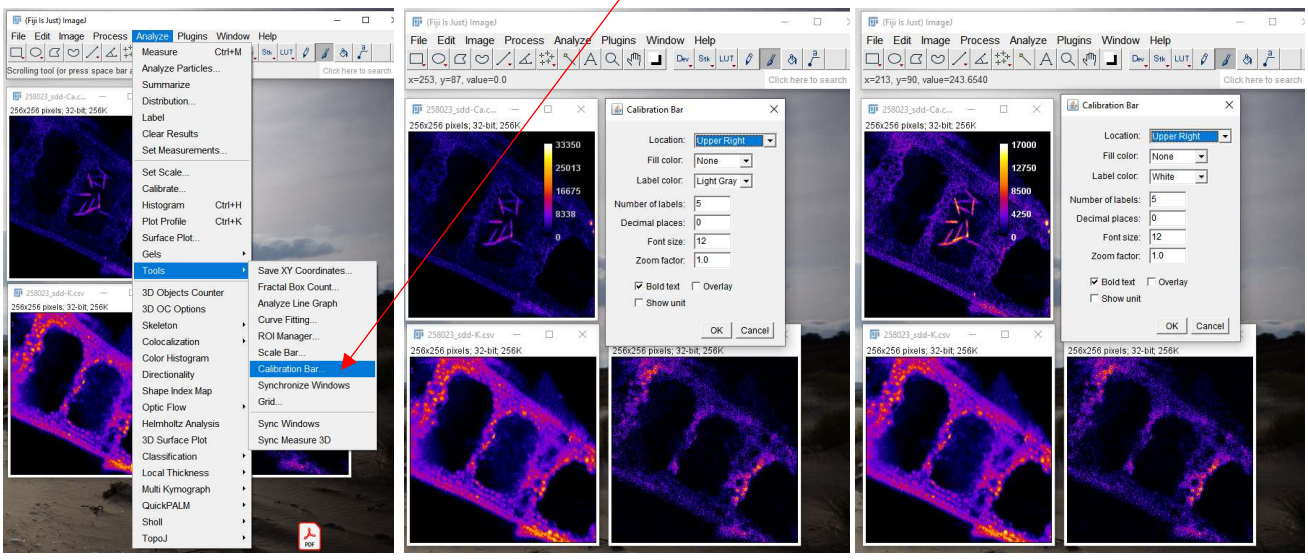
Note: By adjusting the colours displayed, you are not wrongfully changing the images, all the data is still there. All pixels whose values exceed the set maximum will appear white (in this LUT, of course).

Note: such adjustments are particularly useful, when there is small number of pixels with very large values (sometimes outside of the sample, e.g. a dust particle) and you need to adjust the colour range to have your sample visible.

When finished with adjustments, you need to save the image: **File > Save As** (as e.g. .jpeg or .png). When you have all the elements, you can make collage in a software of your choosing, e.g. Powerpoint, where you will need to add the name of the element whose distribution the image is displaying.

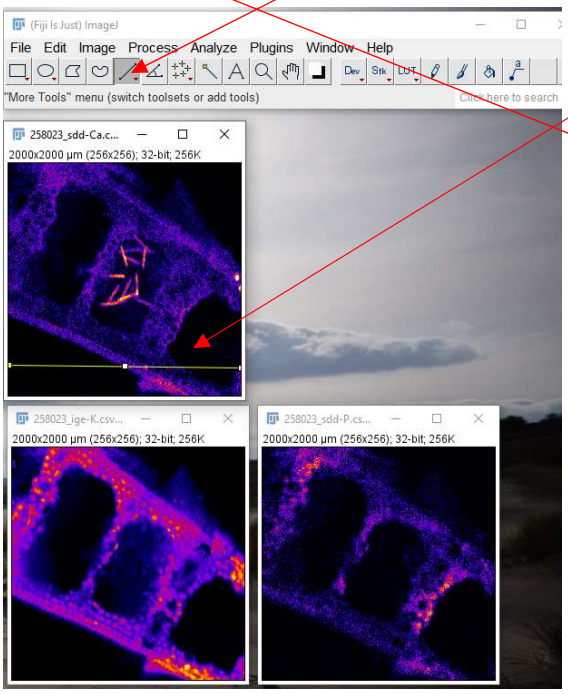


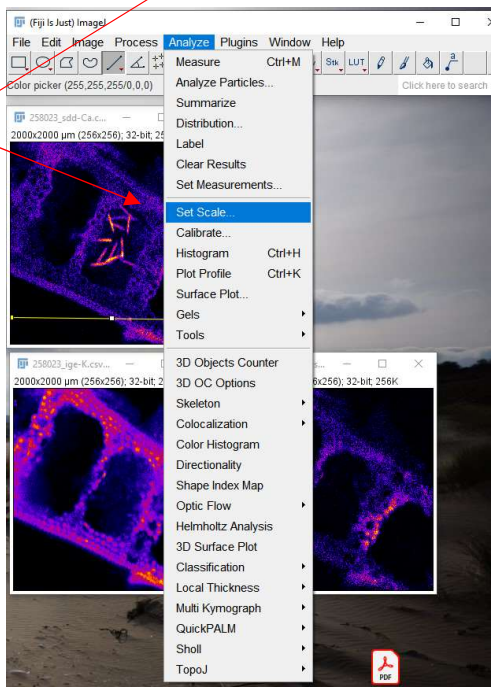
Alternatively, you can add colour scale directly on the distribution maps: **Analyze > Tools > Calibration Bar**. In the pop-up window, adjust the appearance of the image.



The size of the image is important, thus you need to add the scale bar on the image or write the size in the figure caption. Size of all micro-PIXE distribution maps is 256 × 256 pixels, but each pixel can in reality be of different size (anywhere between cca. 1 μm to 7.81 μm). The maximum window of scanning is 2000 × 2000 μm. Our example is: 1000 × 1000 μm.

Click the button **Straight** and draw the line across the image. **Analyze > Set Scale.**





In the pop-up window, fill in the required values:

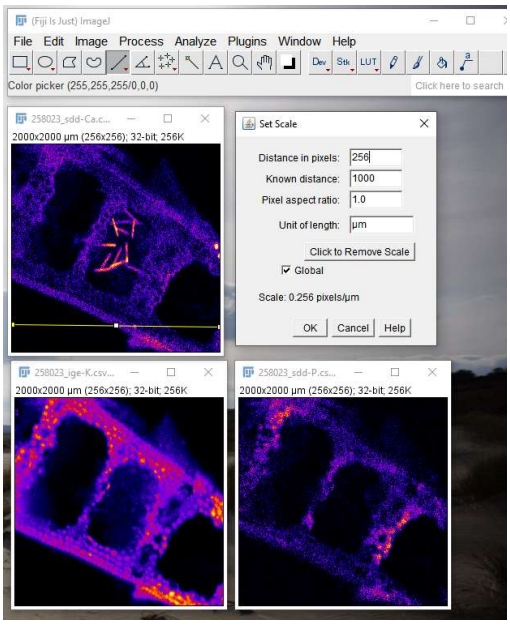
Distance in pixels: 256

Known distance: 1000

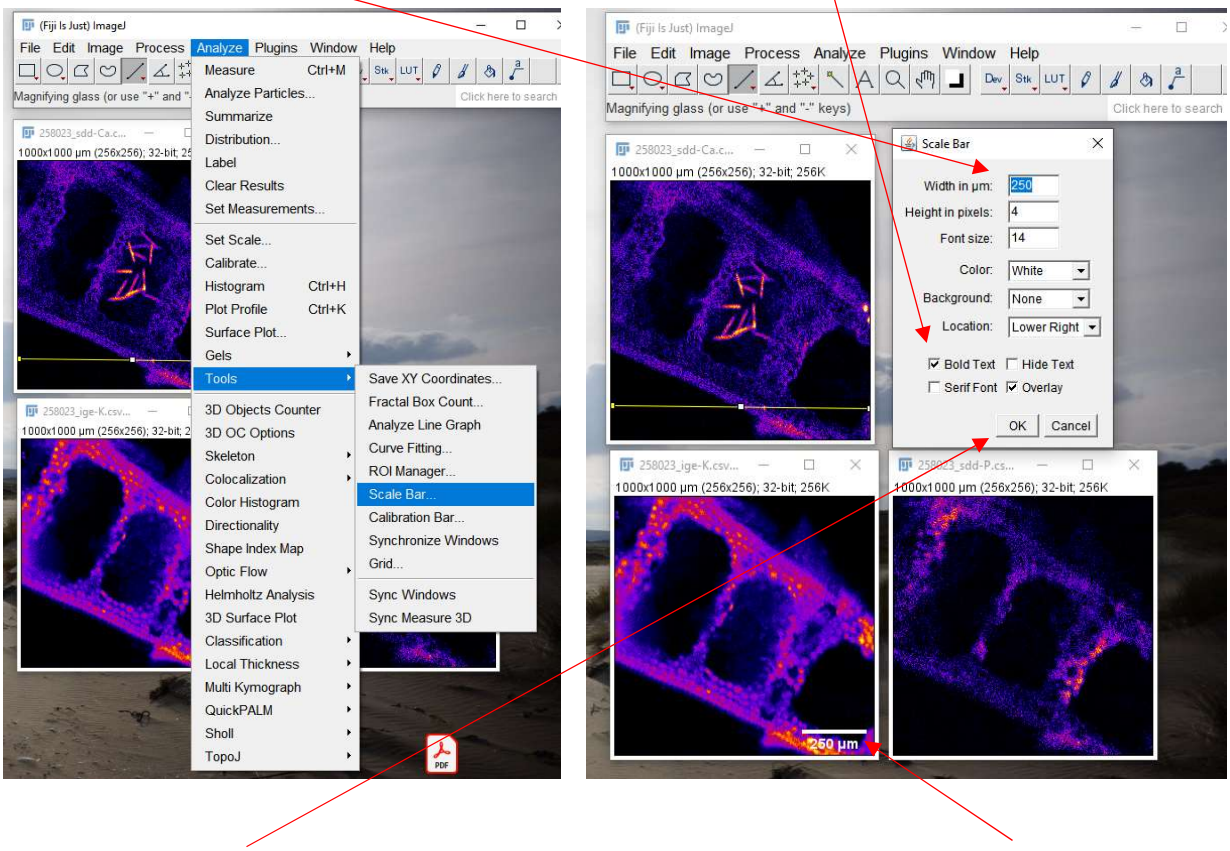
Pixel aspect ratio: 1

Unit length: μm

Tick the box in front of the **Global** if these values apply to all images open.



To add the scale bar onto the images: **Analyze > Tools > Scale Bar**. In the pop-up window you can set the size of the scale bar on the image, e.g. 250 µm and select other options of appearance.

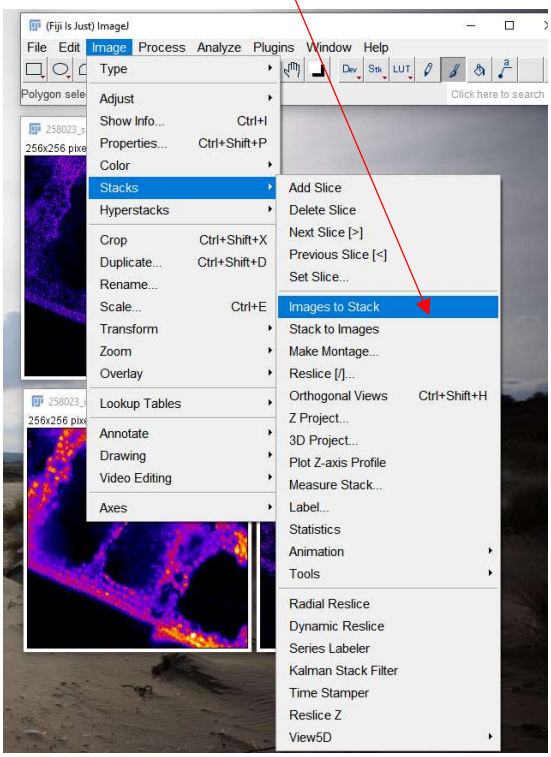


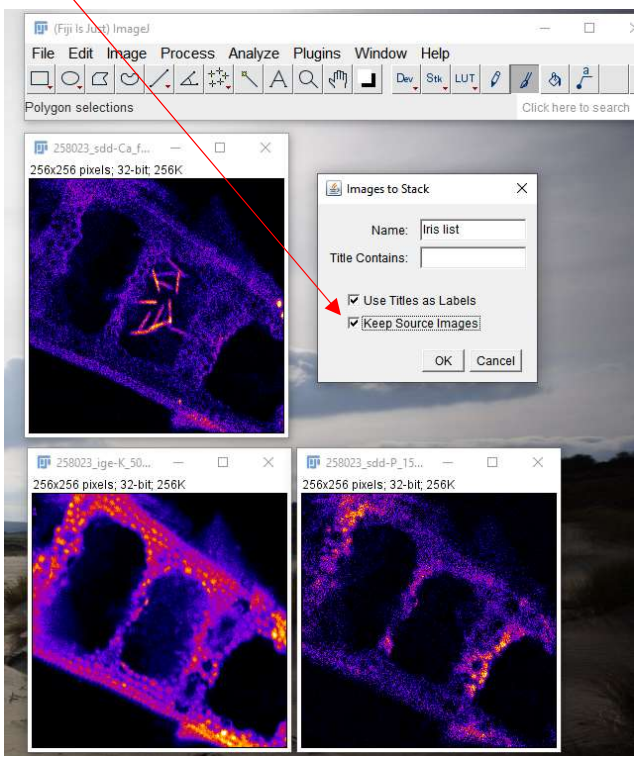
By clicking OK, the scale bar will appear on the image selected.

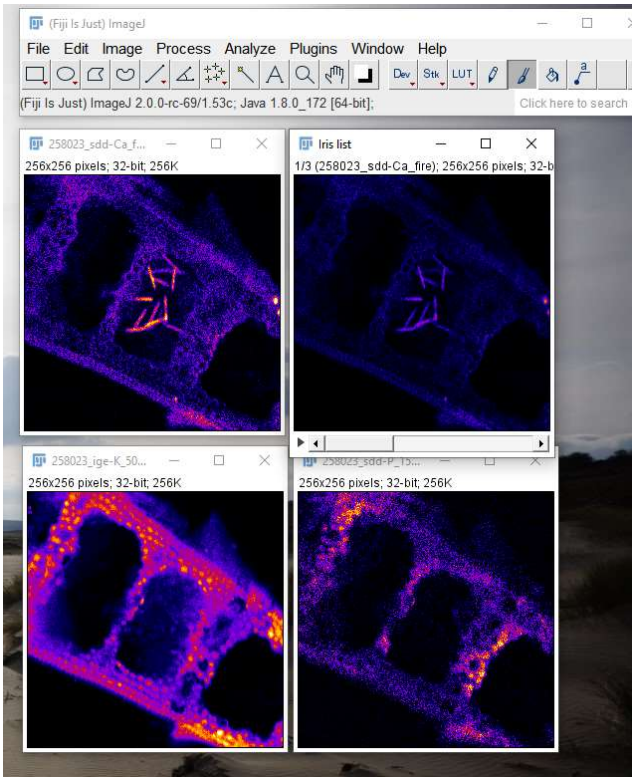
Note: you can add the scale bar to all images or select only one (e.g. the one placed on the top right in the Figure).

Note: there is no need to add values and units on the bar, especially if this extra text will cover important features of the image. If you keep only the bar, the value and the unit must be added to the figure caption.

There is another way to organise the images: **Image > Stacks > Images to Stack.** In the pop-up window add the name and tick the box in front of Keep the Source Images (so that the ImageJ does not close them).



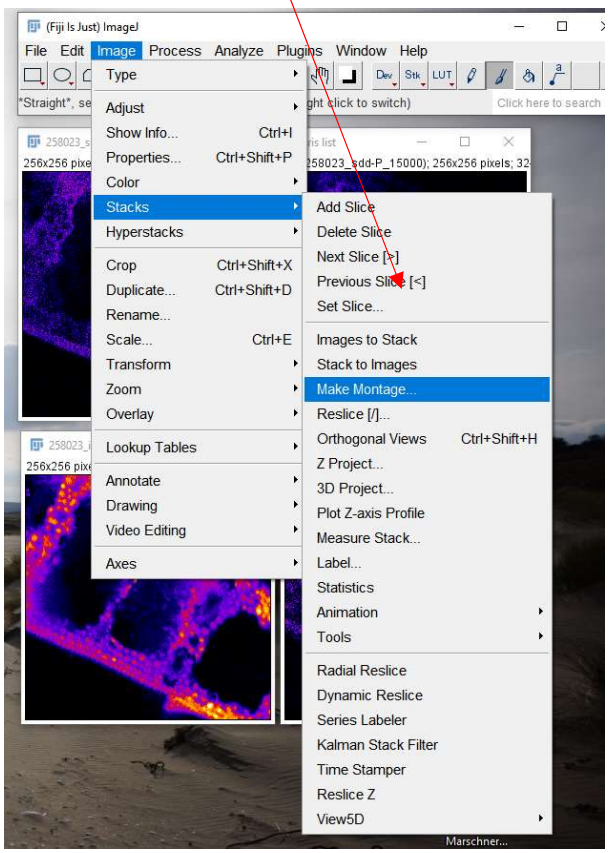


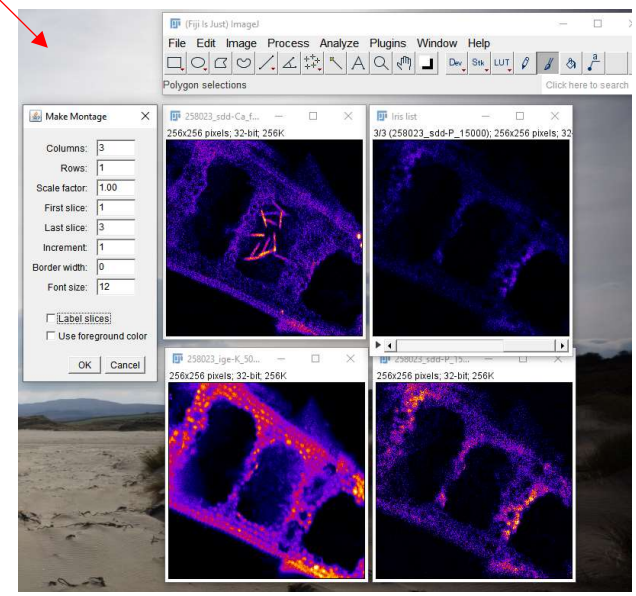


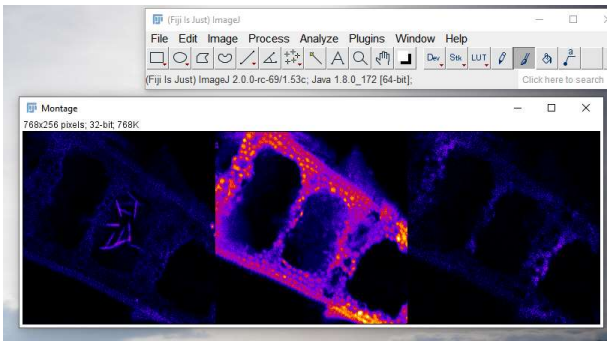
By clicking OK, a new window with images will appear. It has a slider on the bottom, by moving it you move between images (the order in which they appear depends on the order by which images were imported).

Note: the colour range in this window is the initial one and not the adjusted one.

To make collage out of these stacked images**: Image > Stacks > Make Montage**. In the pop-up window you can set the appearance of the collage. When finished, click OK and a Montage window appears. Save it: **File > Save As** (as e.g. .jpeg or .png).



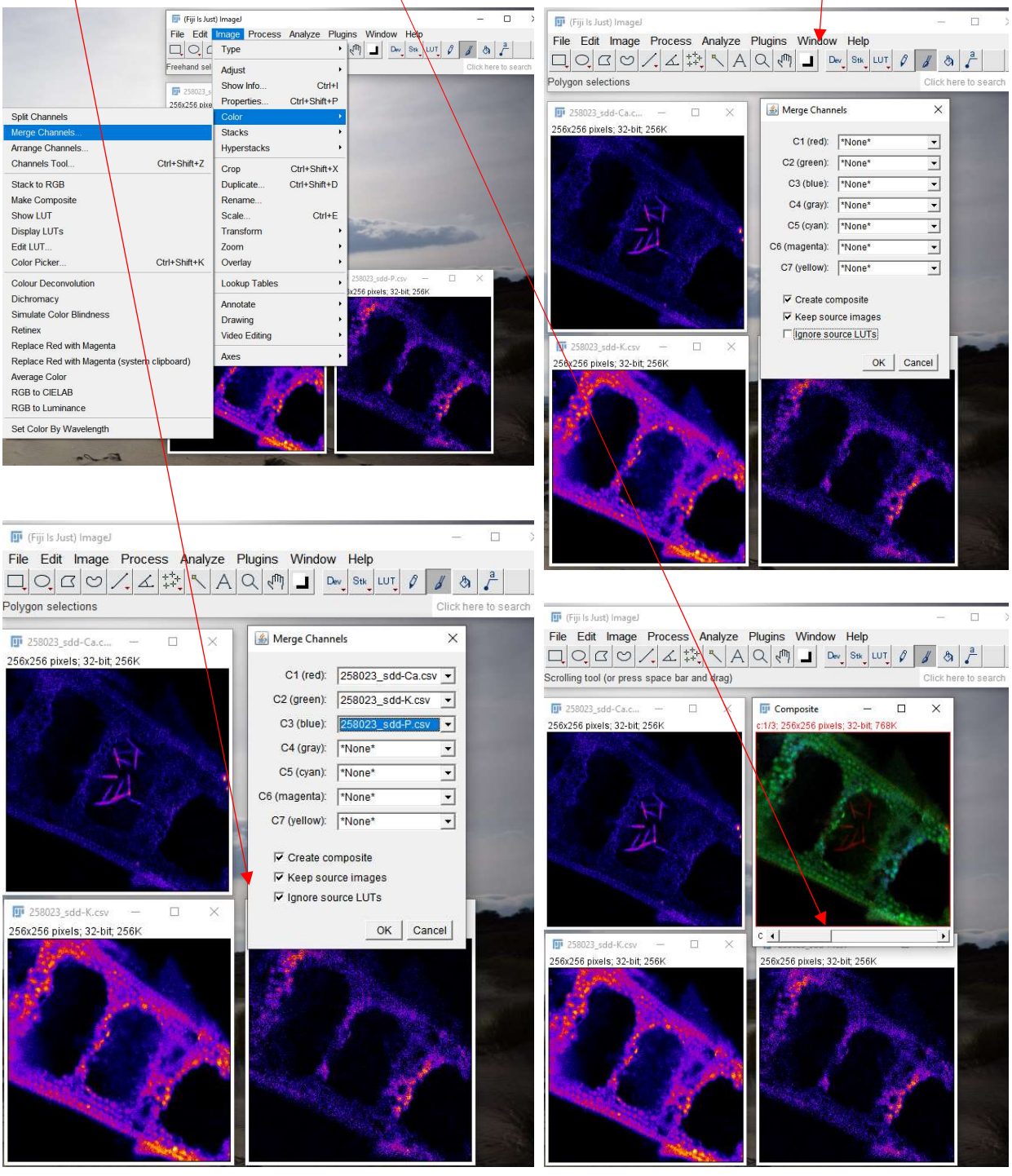




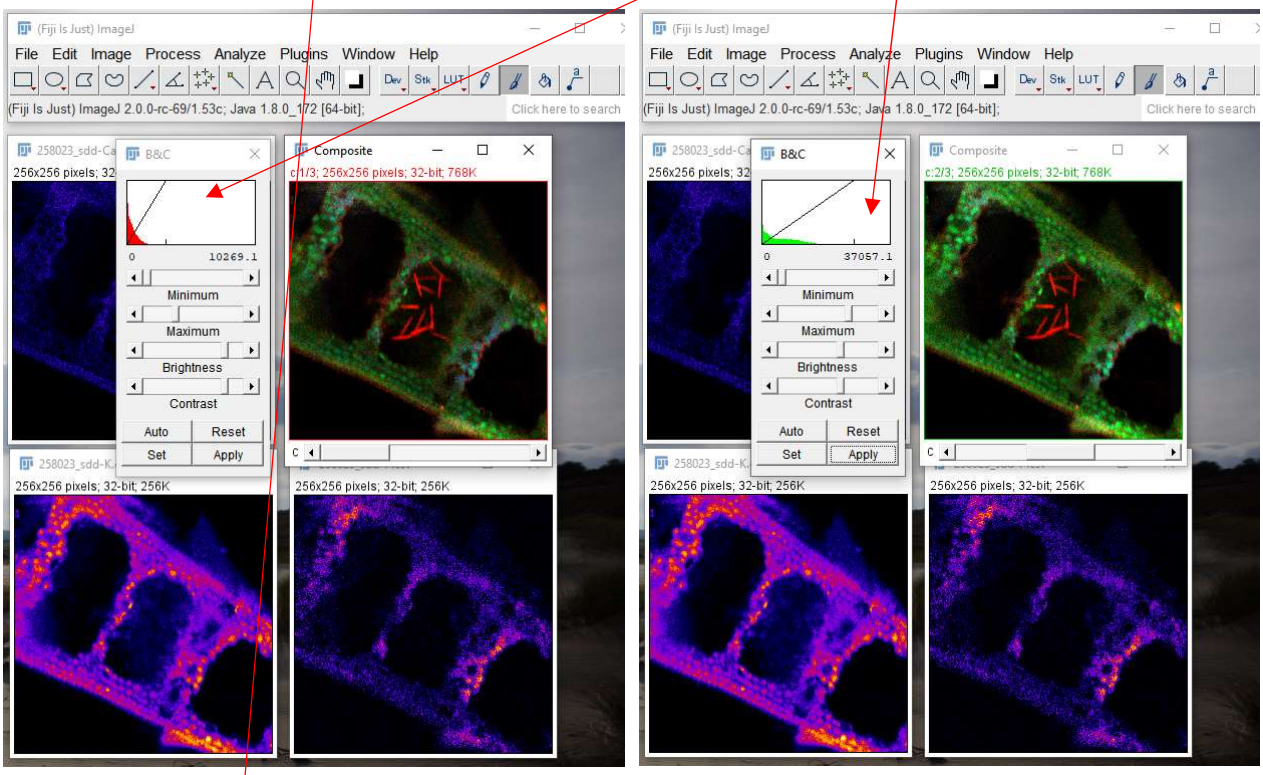
Note: now Brightness/Contrast changes all the maps simultaneously. In case you have large differences in concentrations (macroelements and microelements together) and want to make adjustments, you need to change the source images to 8-bit: **Image > Type > 8-bit** (this will convert your images to the same values: 255).

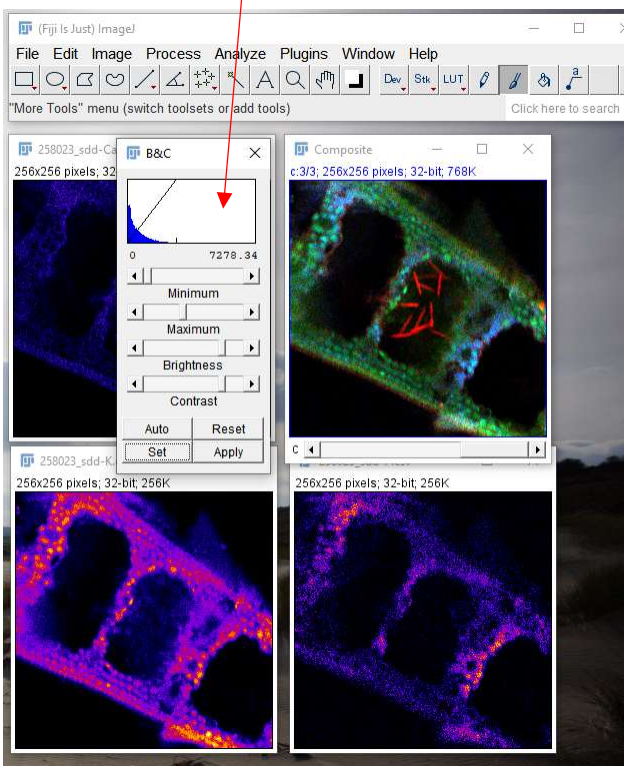
Note: changing LUTs at this stage (or other parameters) is easy, as all changes are performed simultaneously for all images in the stack.

Overlaying images is a good way to demonstrate tissue-specific distribution of elements. ImageJ enables combining up to seven images: **Image > Color > Merge Channels**. In the pop-up window select which element should be displayed in which channel (=colour). Remember the order. Do not forget to tick boxes Keep the source images and Ignore source LUTs. By clicking OK a new window appears in which slider takes you to a different channel.



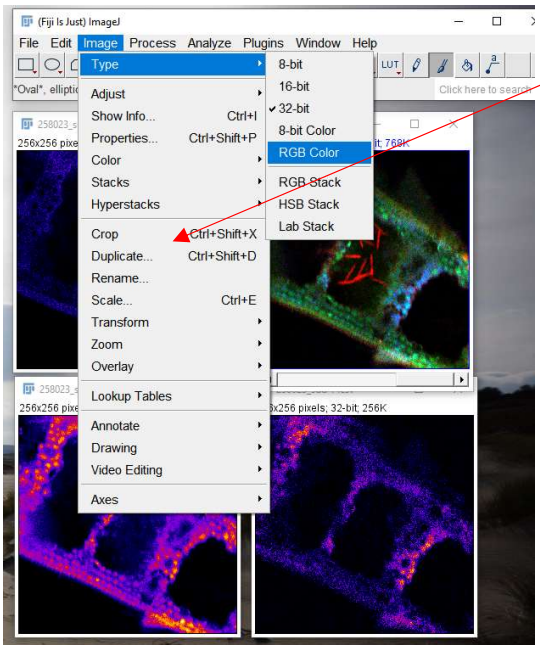
Using Brightness/Contrast you can again adjust the maximum for each of the image. In this example the red channel shows Ca distribution, the green channel shows K distribution and the blue channel shows P distribution.

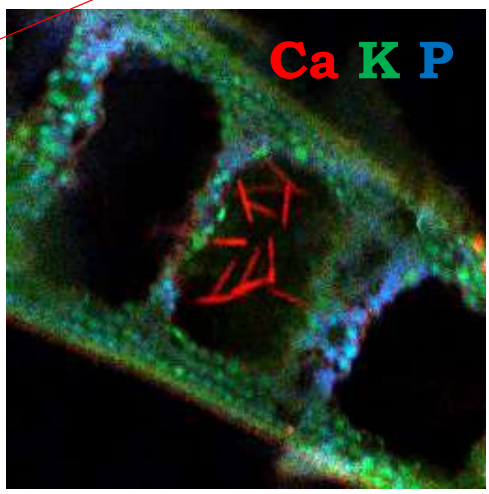




Note: it is a good idea to write down what was the maximum set value for each of the distribution map.

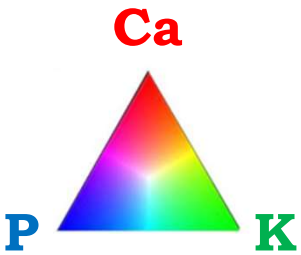
When the combined image is of a good contrast, you can merge it into a single image: **Image > Type > RGB Color or Image > Color > Stack to RGB.**

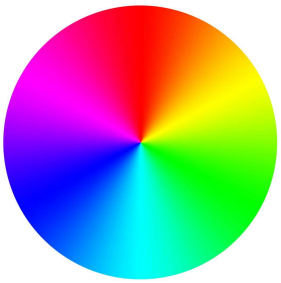




When saving the final image (**File > Save As**) the name should indicate which element is displayed in which colour. e.g. CaR_KG_PB, where R stands for red, G for green and B for blue.

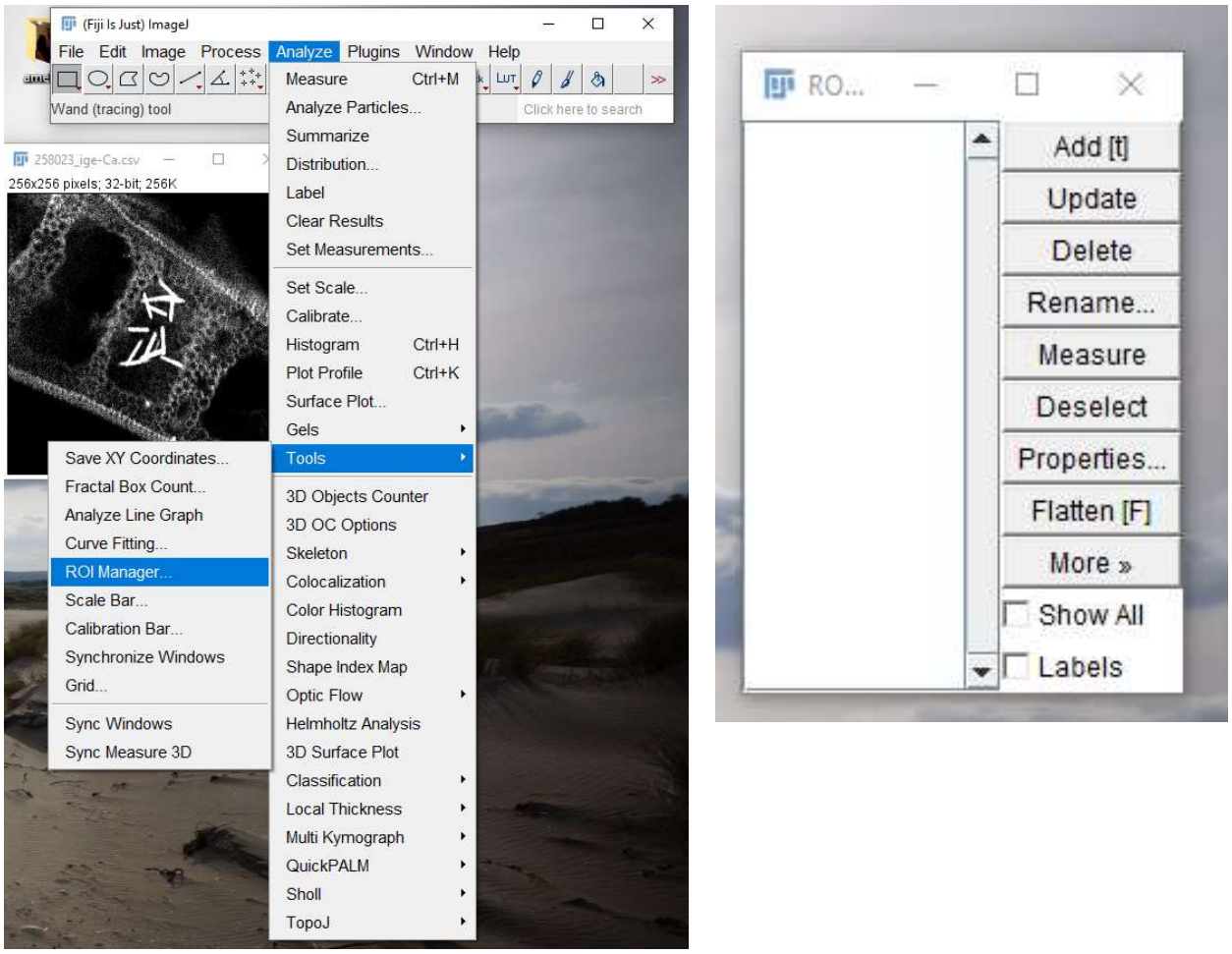
Note: in these images, co-localisation of a particular element in a pixel results in colour mixing. This is why it is good to include the triangle or the circle of co-localisation of colours on or next to the image.





II. Determining tissue-specific concentration of elements

To begin with tissue selection in the distribution maps: **Analyze > Tools > ROI Manager** (ROI = Region Of Interest).

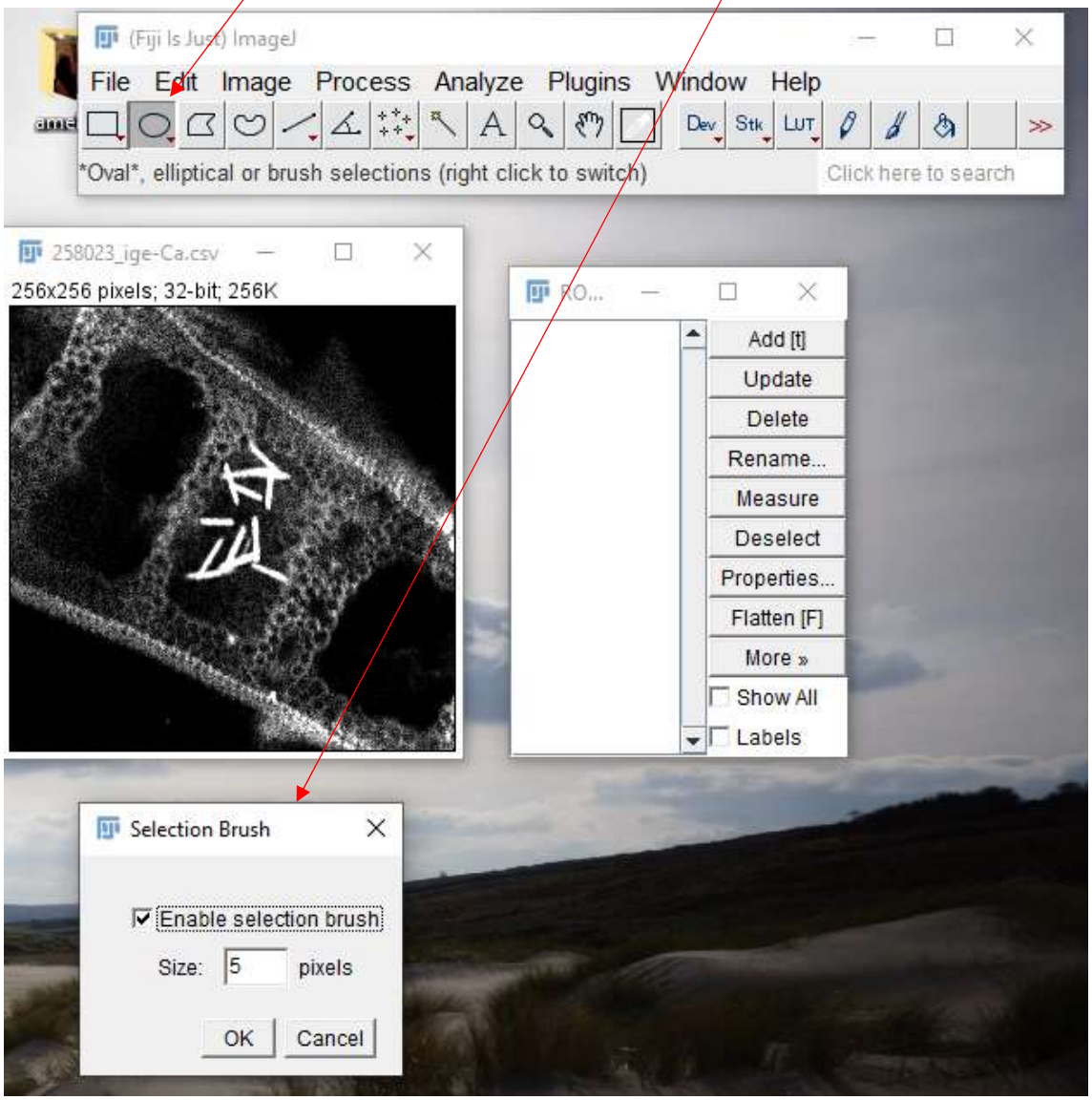


ROI Manager enables saving the selected tissue in the image and calculating the average Gray values (in micro-PIXE this is the average concentration in the selected tissue).

The tool for tissue selection is „**Brush**“ and it is found among options of the second button of the main ImageJ window (the so called „Oval“, elliptical or brush selections).

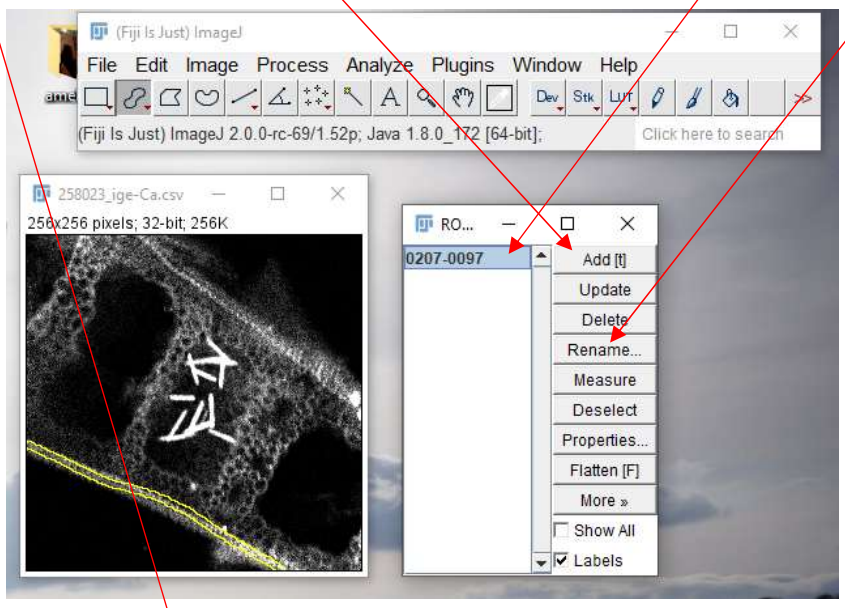
Right-click on this button to open pop-up window (Selection Brush) where the size of the brush can be selected.

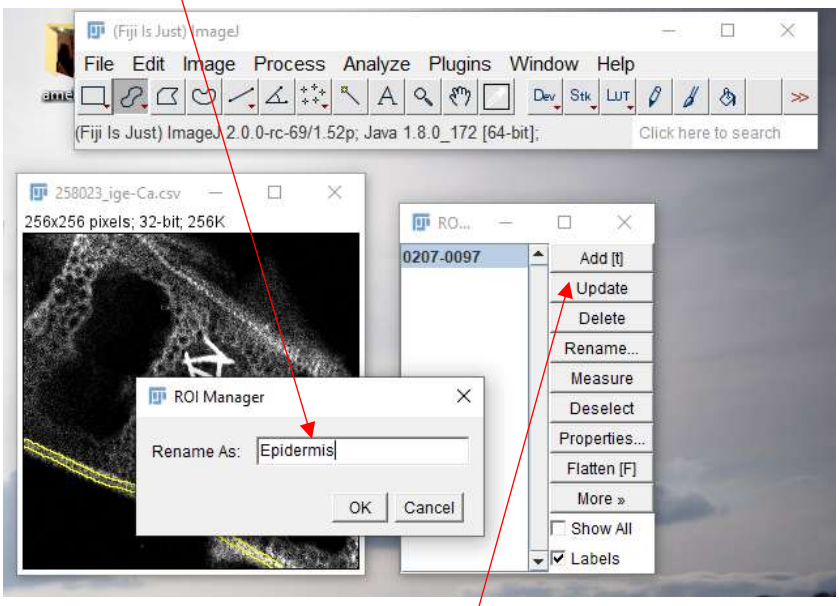
Tick the box at Enable selection brush and set the size of the brush (units: pixels) and adjust it according to the size of the tissue/ cells in the distribution map. When the size is selected **> OK**.



To use the selection tool, slide over the area of the image with keeping the left mouse button pressed. When you are happy with the selection, click **Add** in the ROI Manager window: a number will be added to the left hand side of the ROI Manager window.

To change the name, click on the number and then click **Rename...** Write the name of the tissue.

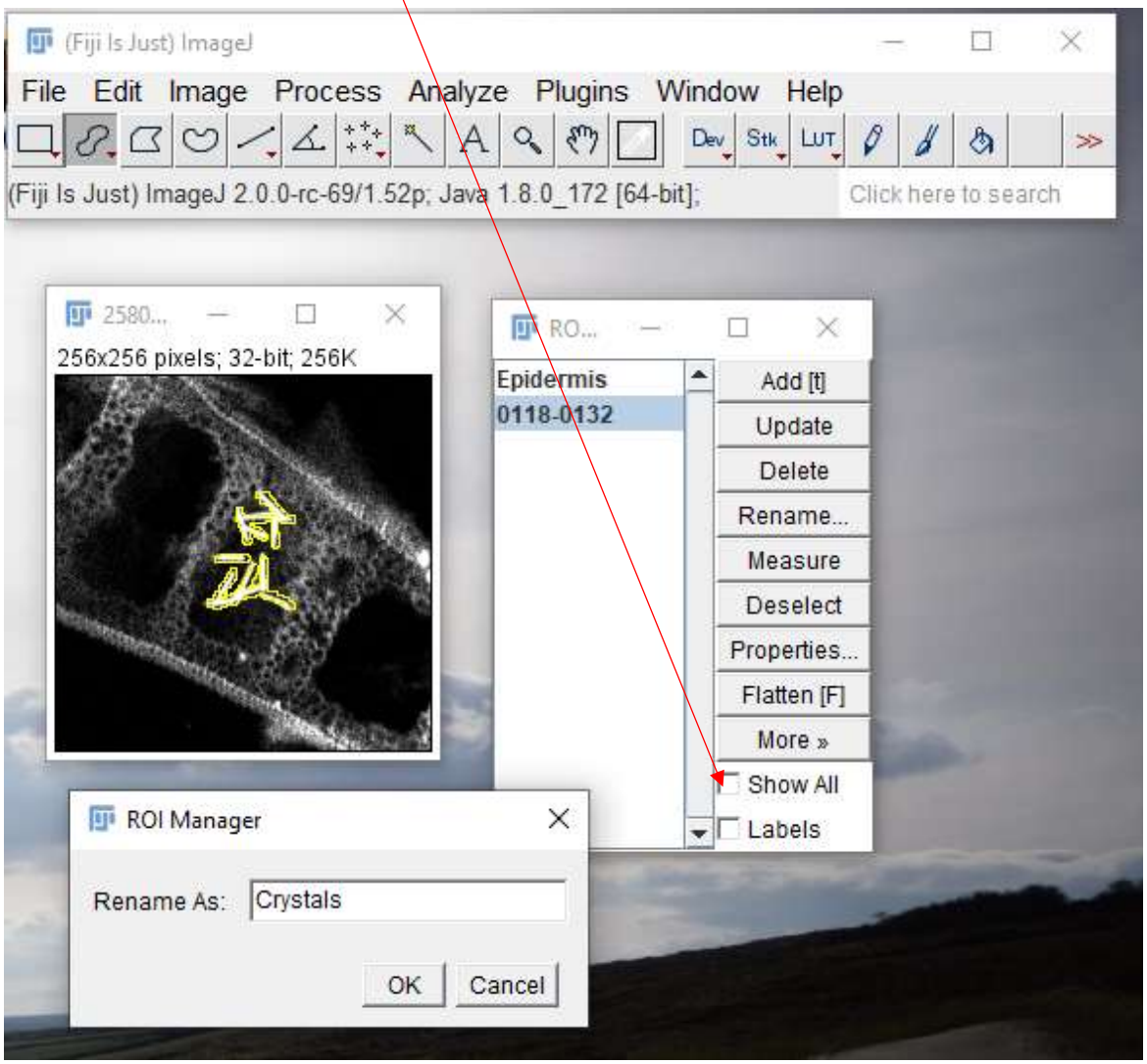




Note: to change the selection after it has been added, always click **Update** to save changes.

To continue with adding new selection, you need to double click the box next to **Show all**.

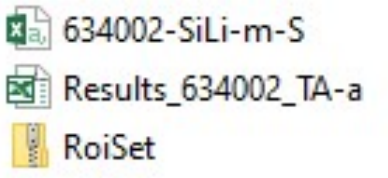
Now, select the second tissue, add it to the ROI Manager and rename it (e.g. Crystals).



In case you have discontinuous tissue, press and hold „**Shift**“ when you begin sliding at different position. Repeat as needed.

Continue with tissue selection and adding & renaming them in the ROI Manager.

When the selection is completed, save the ROI Manager for later use. Make sure you save all tissues (if a name is highlighted, which you do by clicking on it, you will save only the selection). You do this by selecting all or none of the names. The file is saved as RoiSet (zip). Later you open it as **File > Open**.



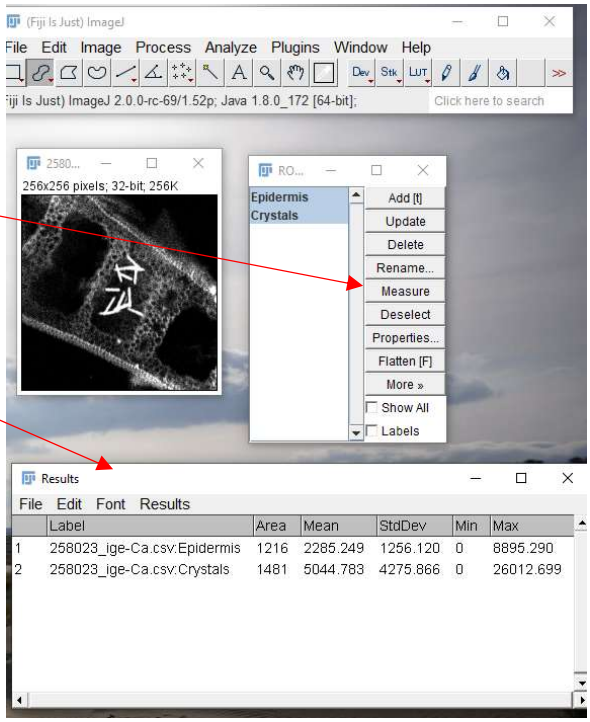
Note: during tissue selection you might want to check other distribution maps. Some tissues are apparent only in some distribution maps.

Note: the tissue added to ROI Manager will be displayed as the area on any distribution maps you open (click on the name in the ROI Manager window).

To get results, select the tissue of interest (or all at once) and click **Measure**.

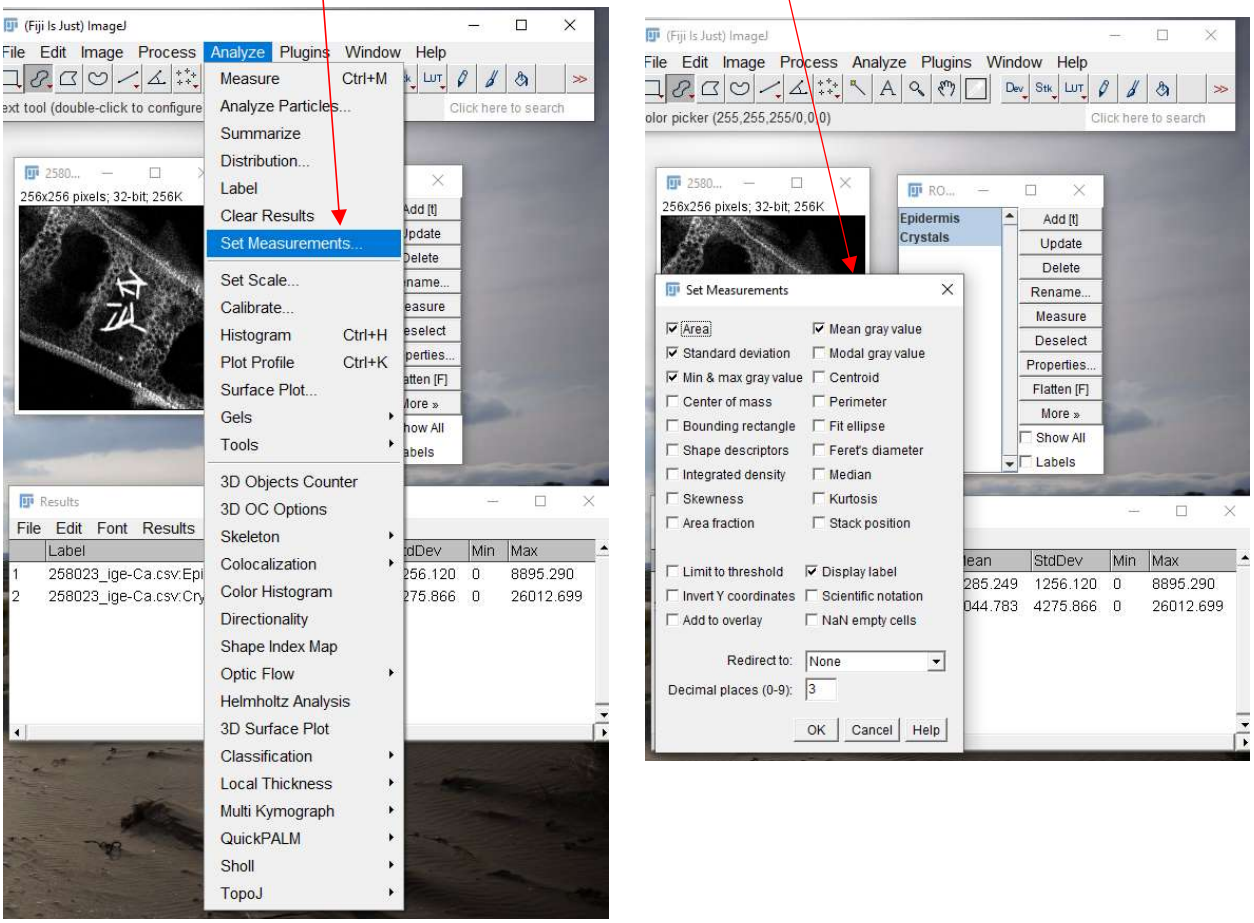
„Result“ window appears, which contains:

- **Label** = sample name.
- **Area** = number of the pixels selected.
- **Mean** = average value, i.e. concentration in the tissue selected in mg kg^{-1} dry weight.
- **StdDev** = standard deviation of the mean (SD).
- **Min** and **Max** = minimum and maximum values.



At first, your „Results“ window might not offer these parameters. To remedy this: **Analyze > Set Measurements** and by ticking boxes in front of names you select which parameters you want calculated.

Note: once you select these parameters, ImageJ will remember them. Only manual unticking at Set Measurement changes this.

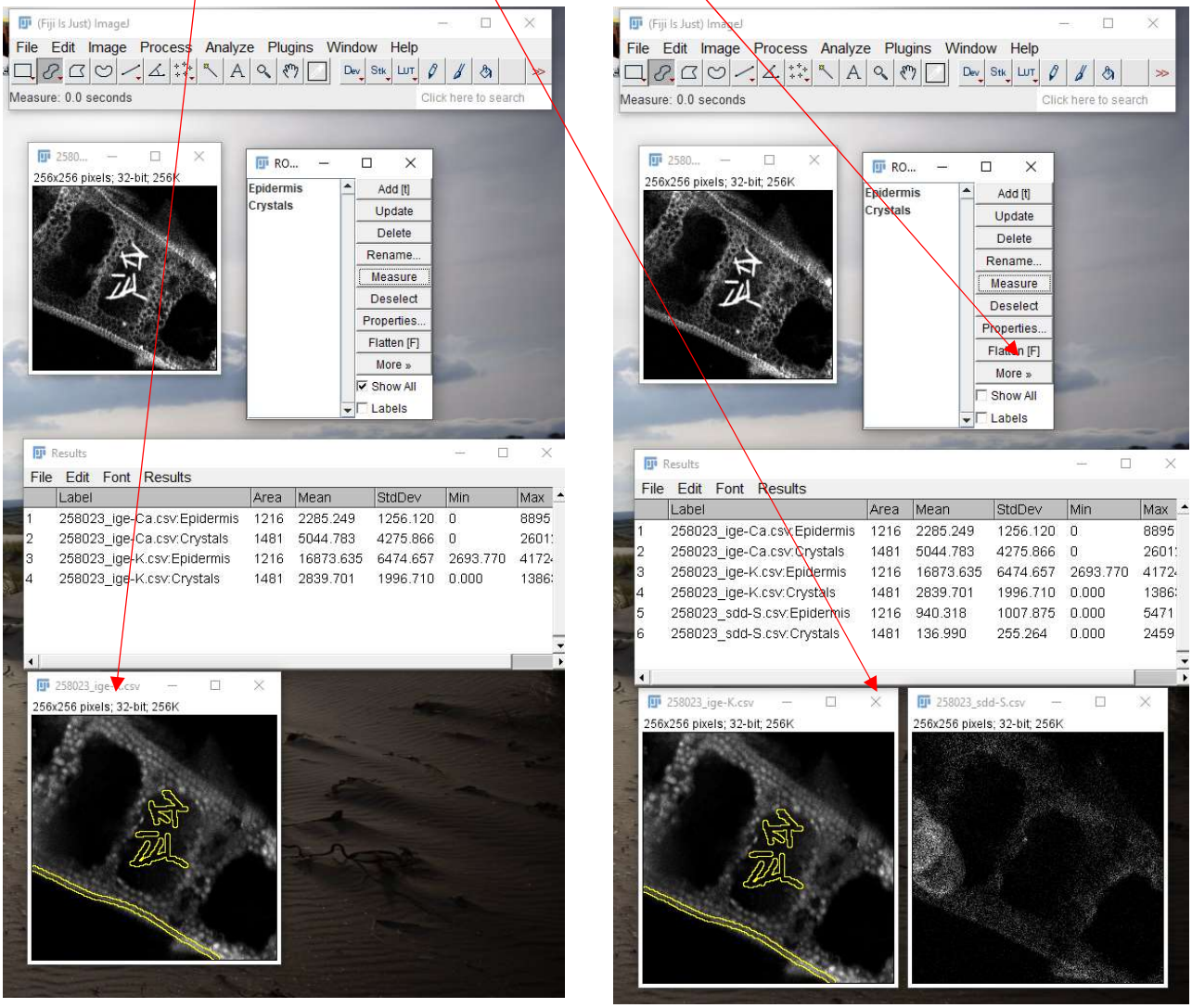


Note: If you want create graphs showing e.g. different concentration of elements in different tissues (from this one measurement), you will need to calculate standard errors (SE) of means. ImageJ provides you with SD, which can be a very big number, because the number of pixels (=n) used to calculate it is large (larger n leads to larger SD).

To calculate SE: **SE=SD/sqrt(n)**

To perform calculations for other element(s), simply click on the image opened and then click **Measure**. Calculations for this element will be added to the Results window. The Results can be saved either by **File > Save As** or by highlighting the results, copying them and pasting them to Excel or other suitable data management software.

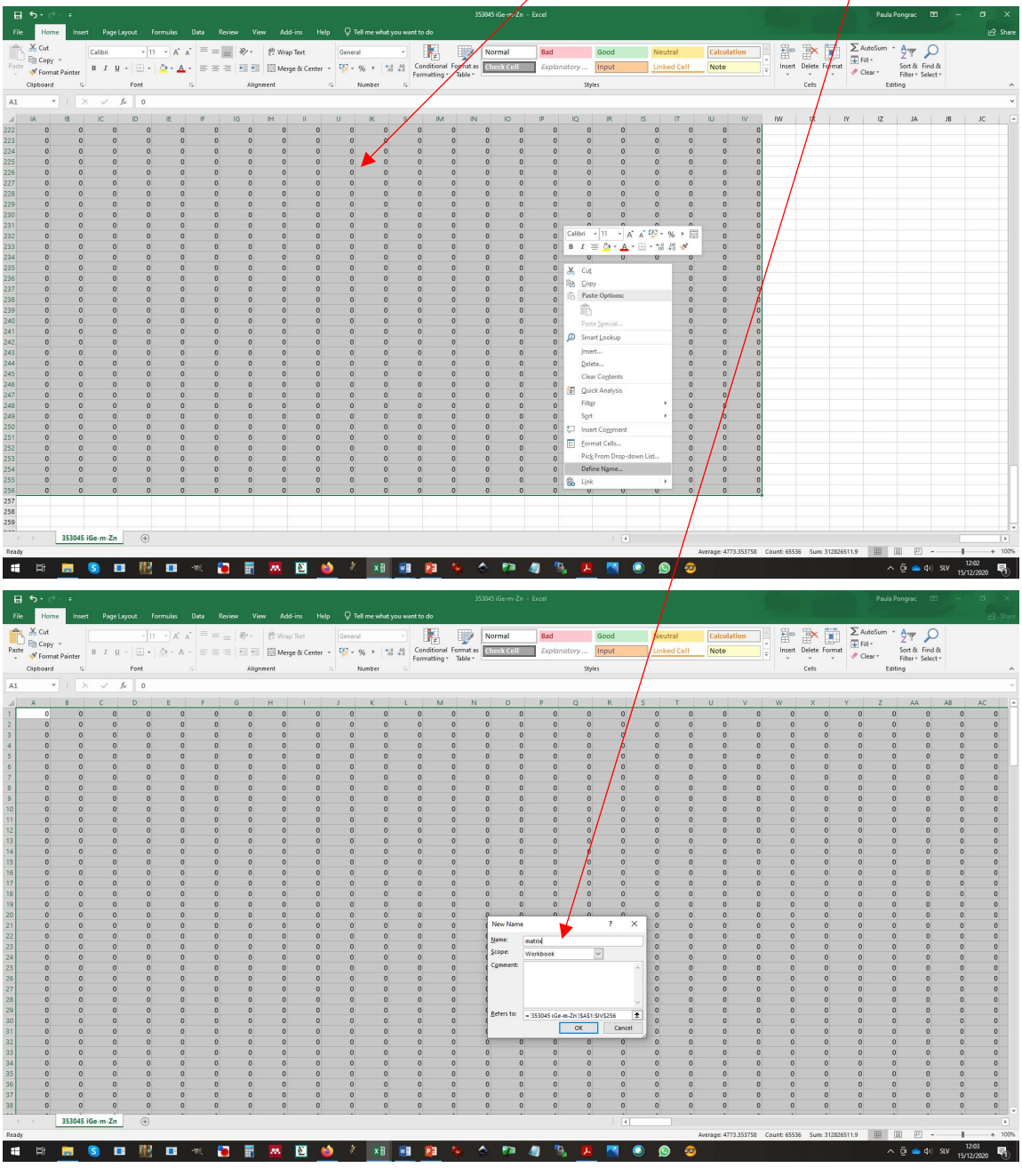
Note: To save an image with the tissues selected displayed, click Show all in ROI Manager > Flatten and then save the image: **File > Save As** (as e.g. .jpeg or .png).



III. Identifying tissue-specific element profiles using multivariate statistics

Before starting the multivariate analysis in Orange Data Mining, your data needs to be formatted for this specific purpose. The initial format is 256 × 256 pixels, which needs to be transformed into a single column. The sample used to demonstrate this is a leaf cross section of *Thlaspi praecox* Wulf.

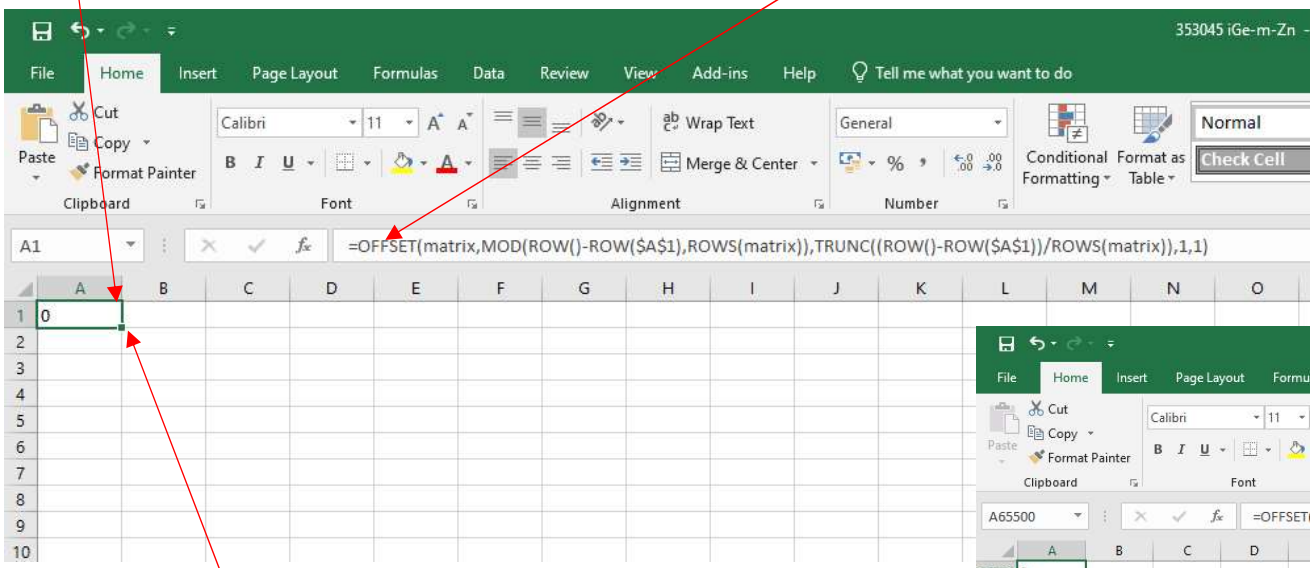
Open each matrix (here Zn is shown) in Excel and use Ctrl+a to highlight it whole. Right click on the mouse and in the pop-up window select „Define name“. Name the selection „matrix“ **> OK**.

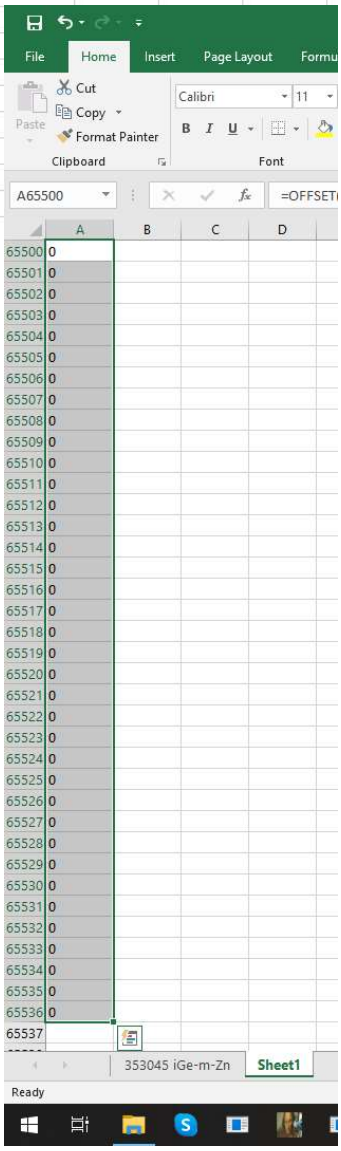


Open new sheet in the same file and paste (Ctrl+v) the following formula to the cell A1:

=OFFSET(matrix,MOD(ROW()-ROW(A1),ROWS(matrix)),TRUNC((ROW()-ROW(A1))/ROWS(matrix)),1,1)

You will see this formula also in the formula box > **Enter.**

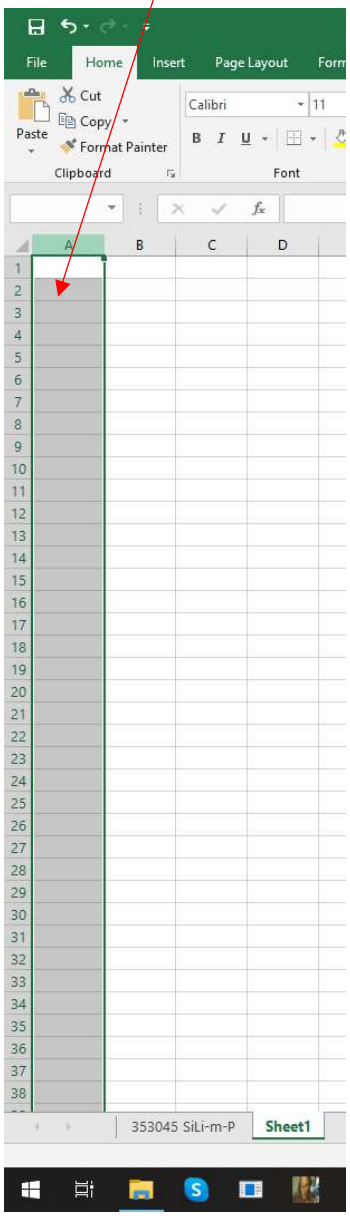


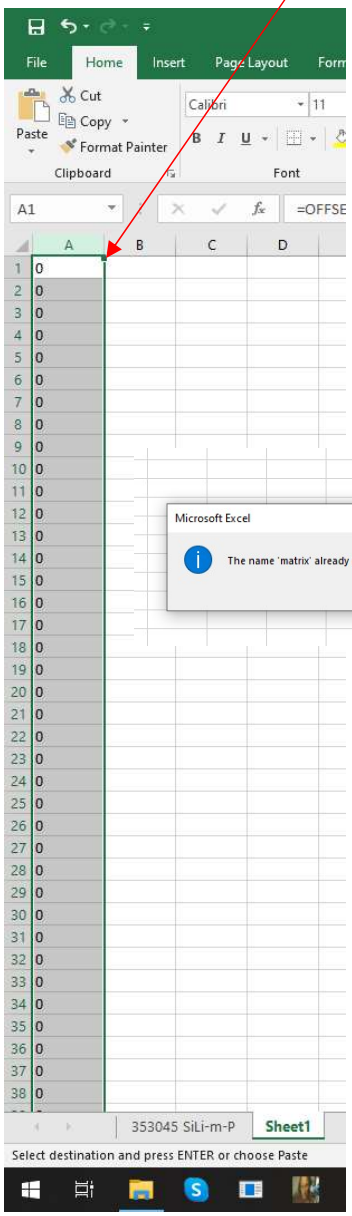


You need to copy this formula to the row 65536 (256×256=65536) by dragging the green box positioned in the bottom right corner of the A1 cell.

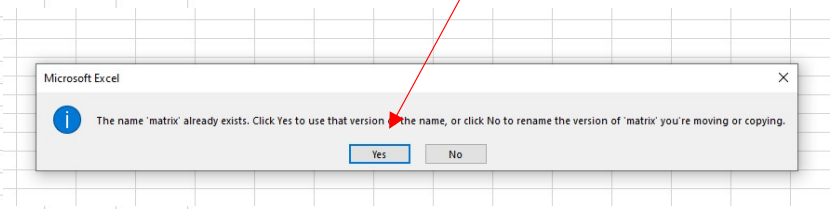
Once you are finished, whole column will be populated with values originating from the matrix (the formula used appends column under column until all 256 columns from he matrix are in a single column in this new sheet).

The created column contains formula in each cell. This makes it useful for copying and pasting it in files of all other elements: open another matrix (here for P distribution), highlight the matrix and define its name (matrix). In the new sheet in the P file, click on the column A (to have it highlighted) and paste the column containing formulas (from the first attempt).

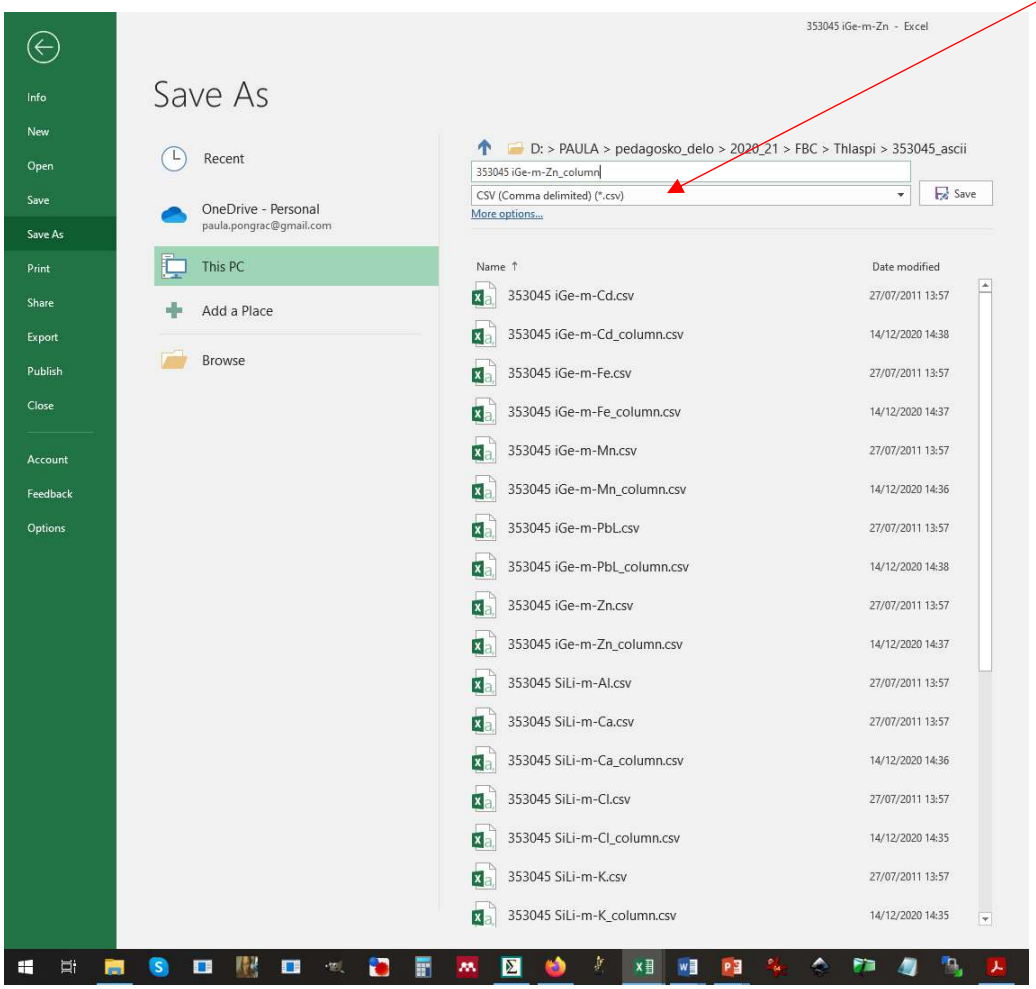




Upon pasting, you will be notified that the name matrix already exists and if you want to keep the name click YES and if you want to change it to click NO. Select **Yes**.

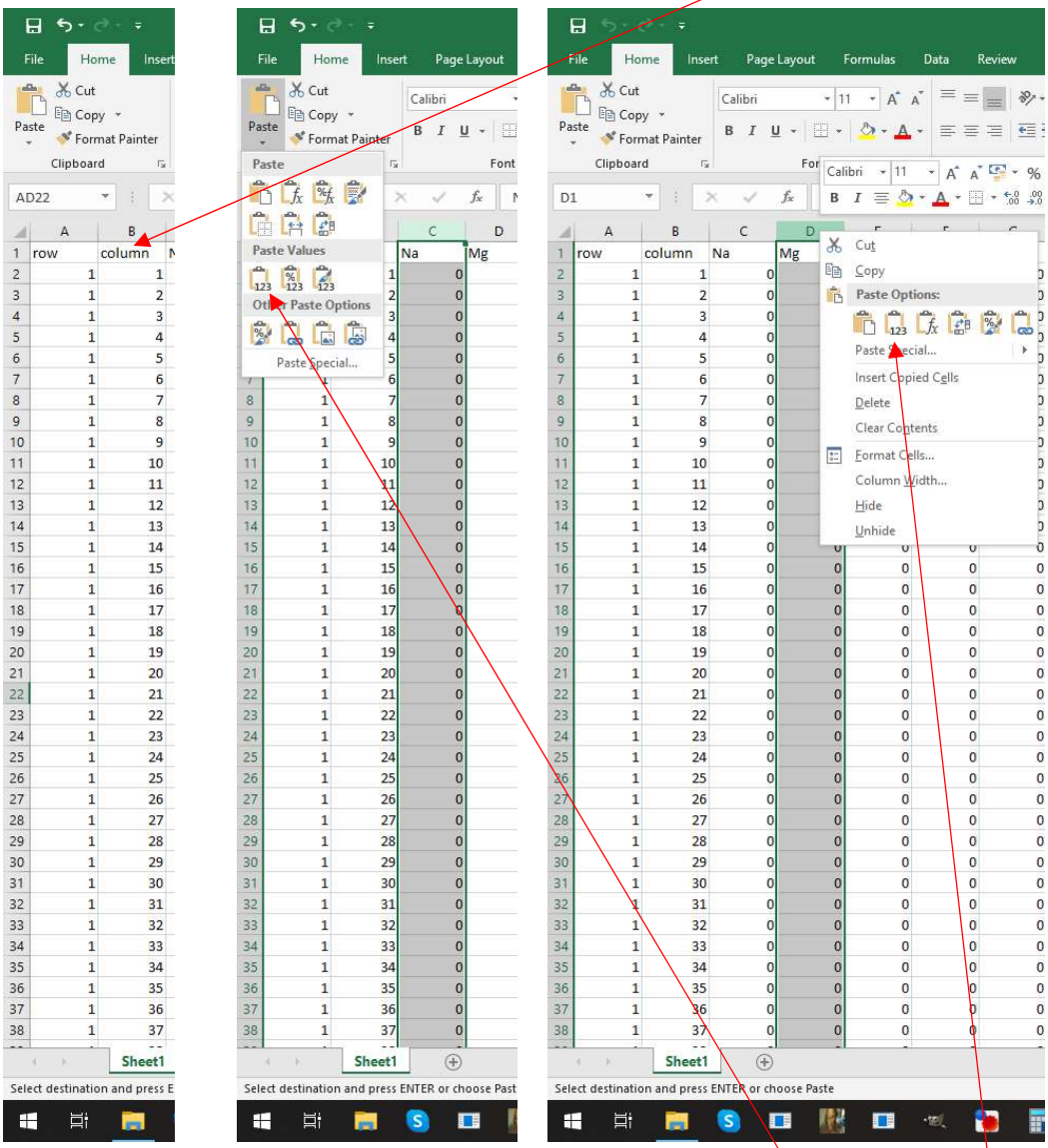


All the columns should be saved: **File > Save as**. Use the .csv format.



Note: .csv format can contain only one sheet. Therefore, when saving you must save it under a different-from-matrix name (e.g. add an extension _column) or save all files with columns into a different folder. In case you fail to do this, you will overwrite the matrix and thus loose it.

Now combine all the columns into a single Excel file next to the first two columns being filled in with coordinates (column × rows values).



If you have xxx_column files closed, you need to open them, click on the column A (to highlight it) > **Ctrl+c** and **Ctrl+v** in the combined file to move the column.

If you have files with columns open, make sure you paste only the values (and not the formulas) > **Paste Special** > **Paste values** (= the button with 123).

Name the columns (element name) after inserting one row at the top of the columns.

When columns for all elements are copied, you need to calculate z-normalised values (this value indicates how far from the average is a value in a cell:

$$=(x-\mu)/\sigma$$

x, value in each cell/ pixel

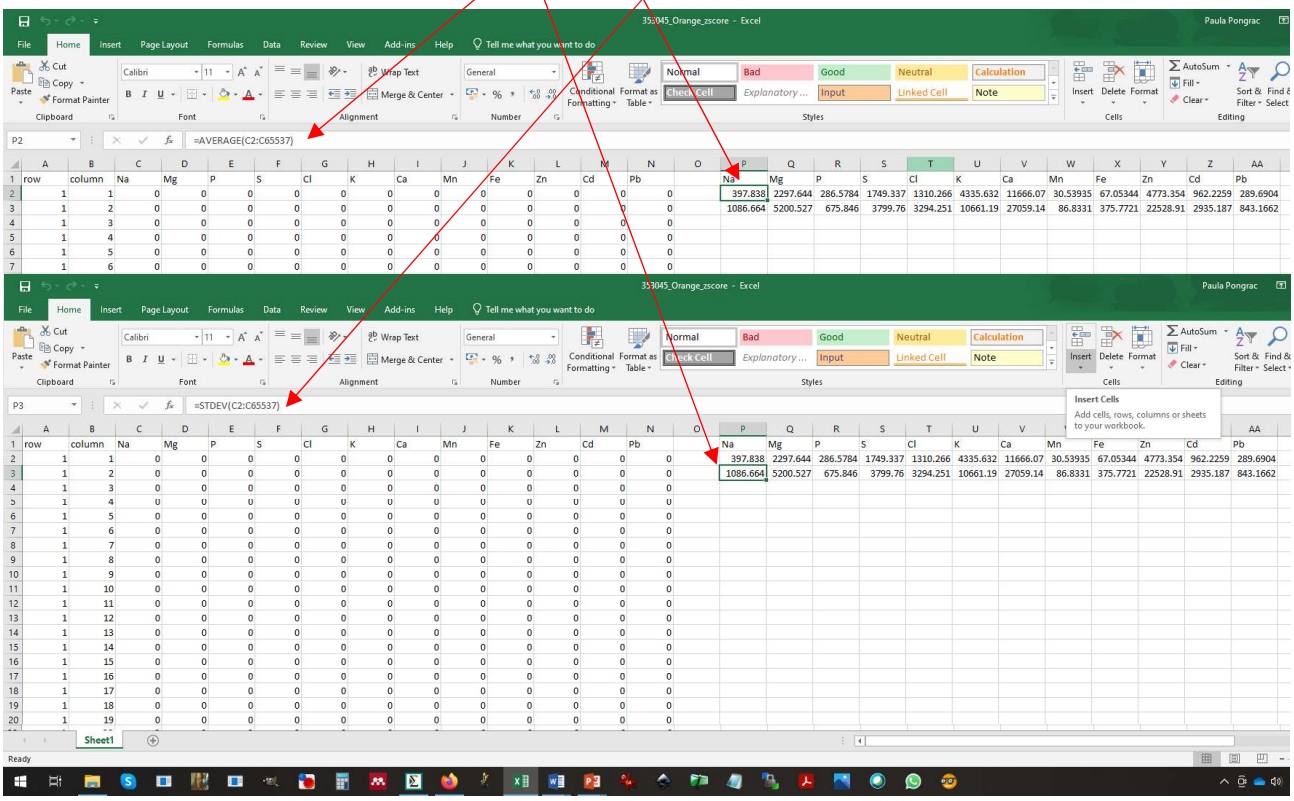
μ, average value in each column

our example, for Na in column C: =average(C1:C65536)

σ, standard deviation for all values in a column

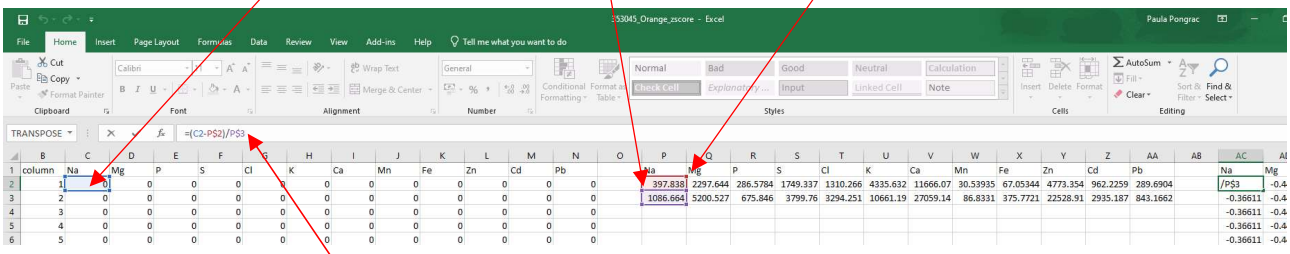
our example, for Na in column C: =stdev(C1:C65536)

Therefore, you need to calculate the average and the standard deviation for each element.



Now calculate z-value for each cell in each column using the formula =(x-μ)/σ by starting in a new column. The first data point (in our example) is C2 (this is x), average is in P2 (this is μ) and standard deviation is in P3 (this is σ).

If you click on the cell with the formula, the cells used in the calculations are shown in colour: blue, red and purple.

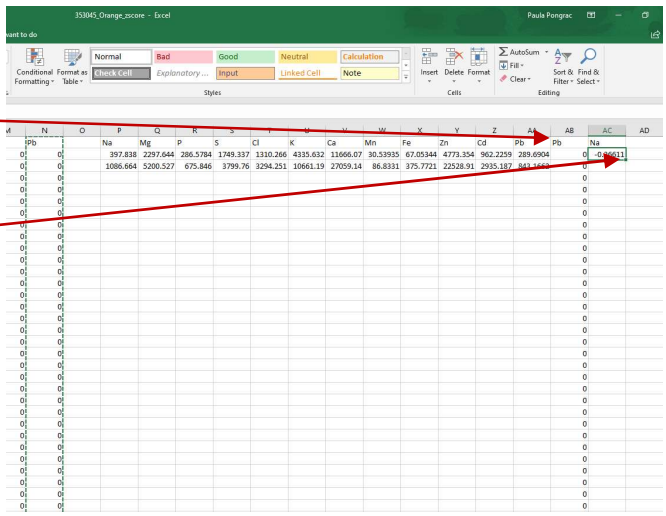


You will need to drag and copy this formula to the row 65536.

Note: values for μ and σ are the same for a single element. To have this memorised in the formula, add dollar sign $ in front of the numbers 2 and 3: P$2 and P$3. Now values in these two cells will always be used when calculating z-values.

Note: to avoid dragging the formula all the way to 65536 you can use this trick:

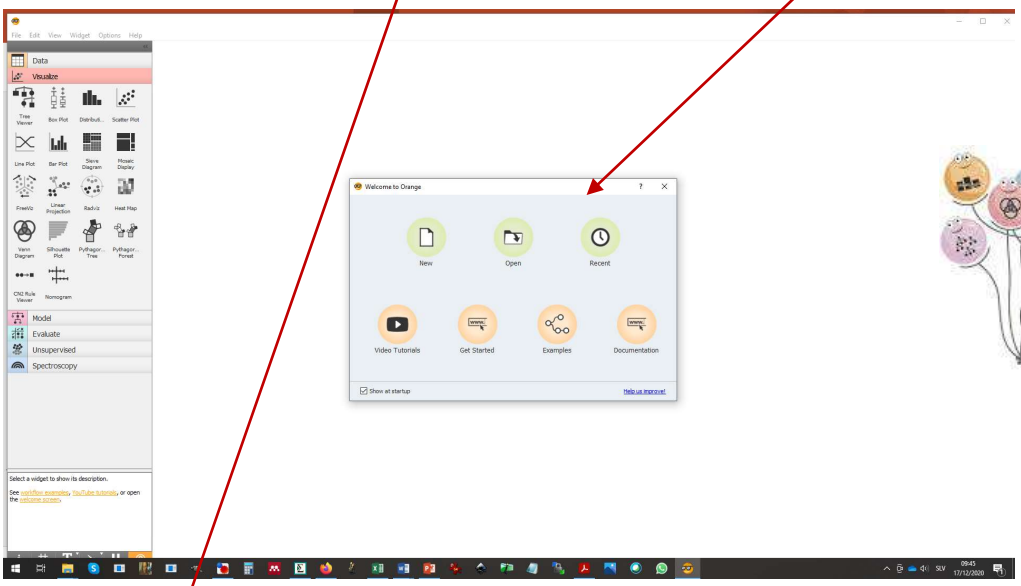
paste a column which already has the values all the way to row 65536 next to the new column (the Pb column for example). By double clicking on the green square on the bottom right corner of the cell, the formula will be copied as far along the column as there are values in the neighbouring column.

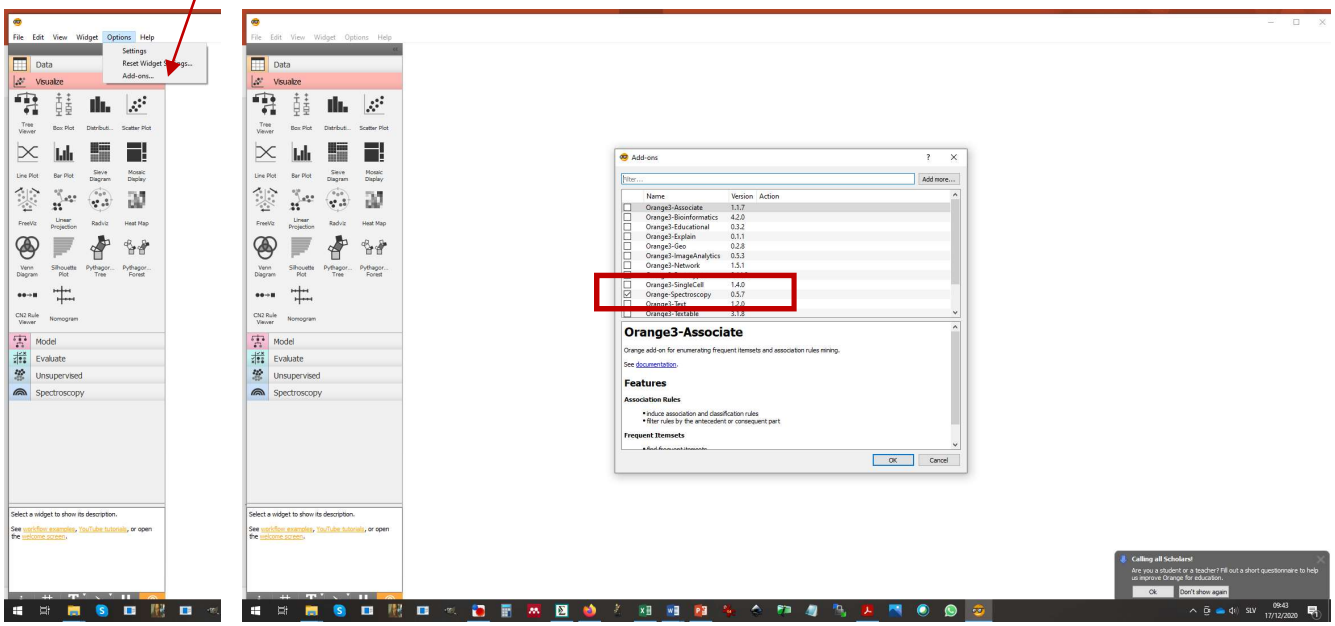


Now calculate all z-values for all elements and save the file in the .xlxs or .csv.

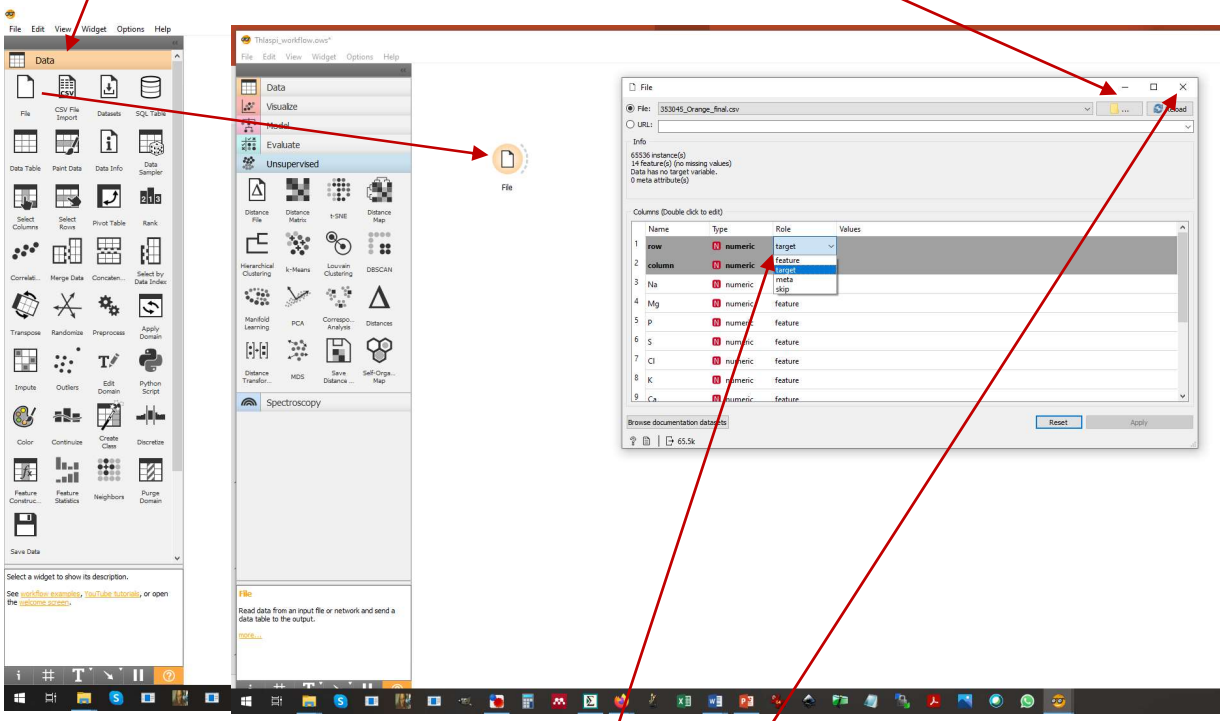
Henceforth we will only need the z-values, thus copy the z-values (next to the column and row columns) to a new Excel (make sure you paste only values and not also formulas) and save it. Both .csv and .xlxs can be imported into Orange Data Mining.

Open Orange Data Mining and close the first window appearing and go to **Options > Add-ons > Orange Spectroscopy > OK**. Restart Orange.





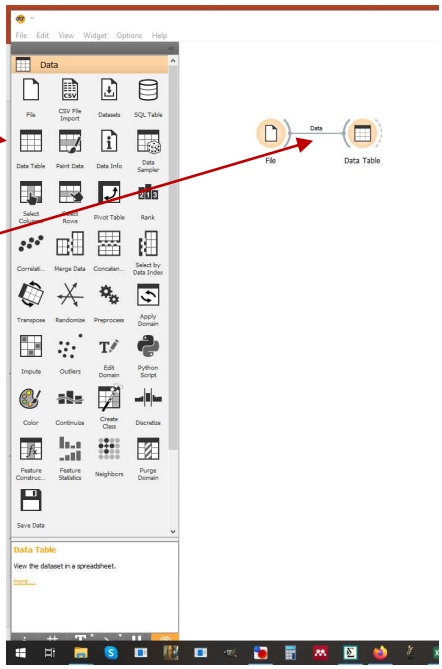
Go to **Data** and click on **File**. This creates a widget File on canvas. This will be where you import your data (the z-values). Double click on the File widget and in a pop-up window browse to find your file.



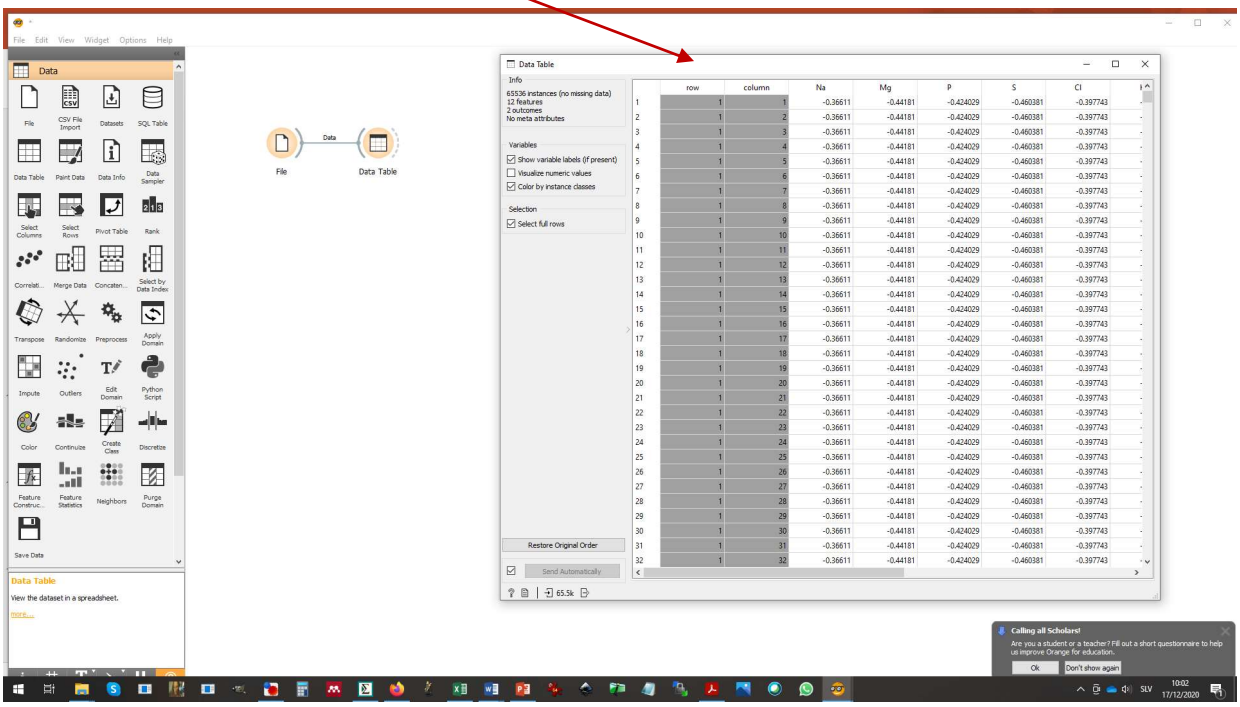
Once the file is imported, the table describing your data will populate with factors form your data file. In the Role column, select Row and Column to be „target“, all other factors should be „feature“. Close the File window (data is not lost).

To inspect the content of the data file imported click on the Data Table and a new Widget will appear.

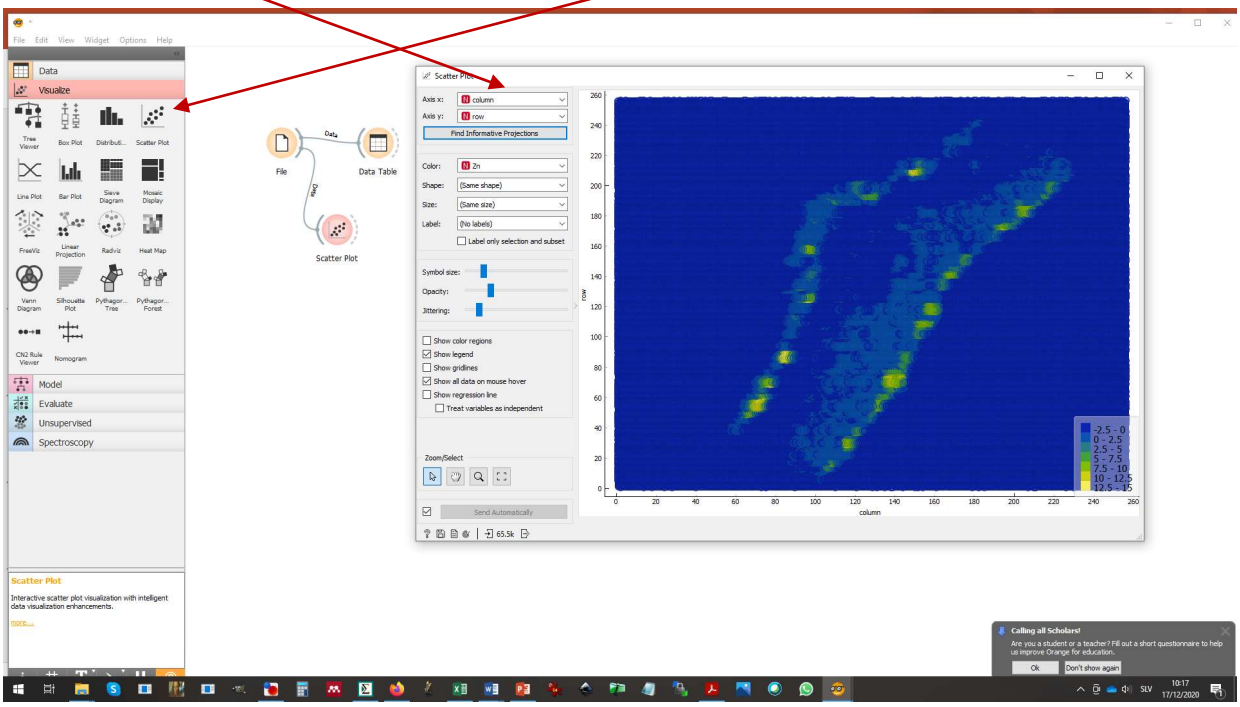
By connecting (with the left button on the mouse pressed drag a line) the File and the Data Table widgets they begin communicating.



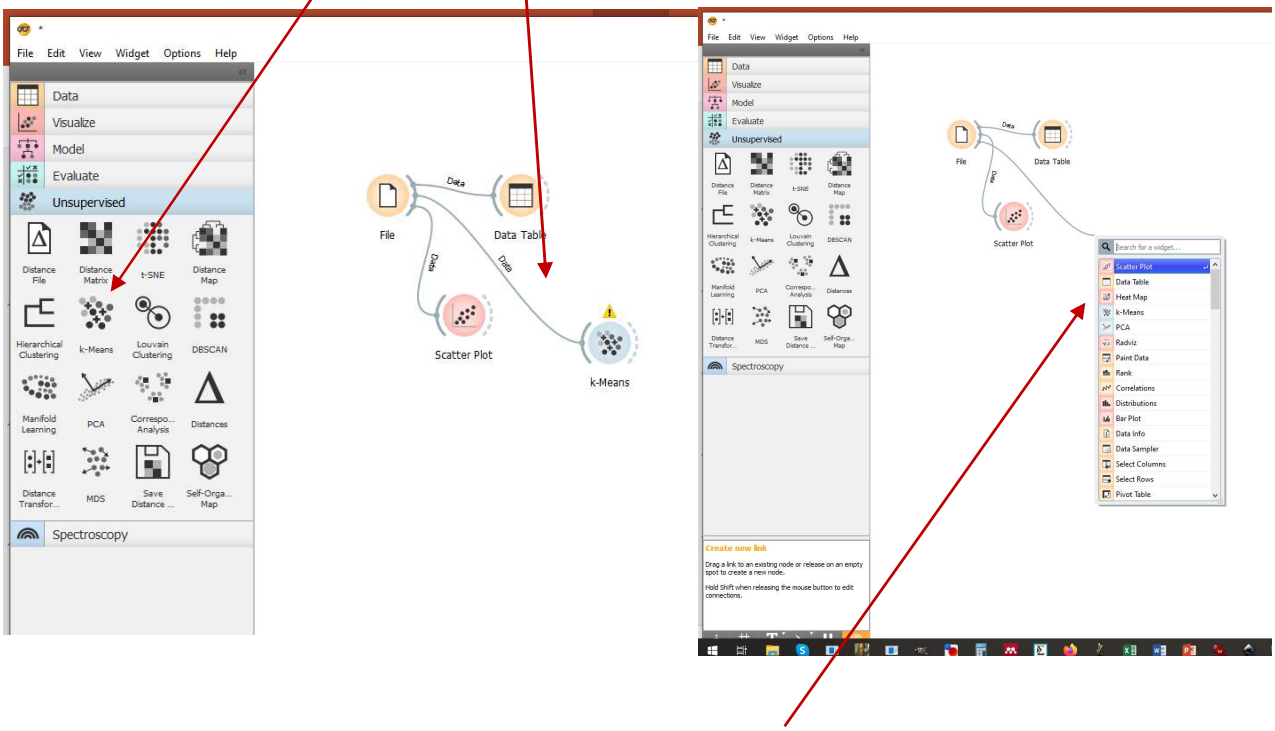
To inspect the content of the Data Table just double click the widget and a pop-up window will appear displaying data. The first two columns are highlighted (this indicates their role as „targets“).



You can perform initial visualisation of the data: **Visualise** > **Scatter Plot**. Connect the File and the Scatter Plot widgets and in a pop-up window select an element to be displayed. These are not concentrations but z-scores for Zn.



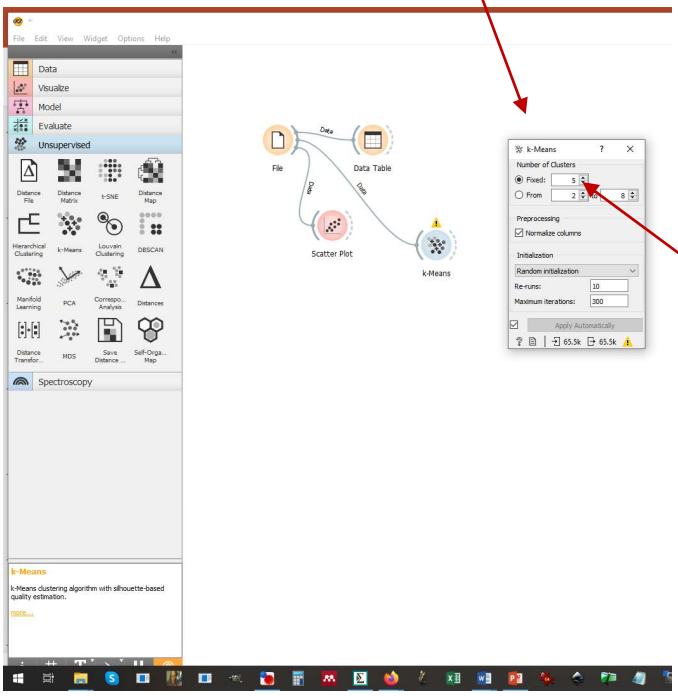
In our data, we want to find tissues with specific elemental profiles (taking all elements into account). You will be looking for clusters using k-Means statistics. You can find the widget under **Unsupervised > k-Means**. Link the File widget with the k-Means widget.



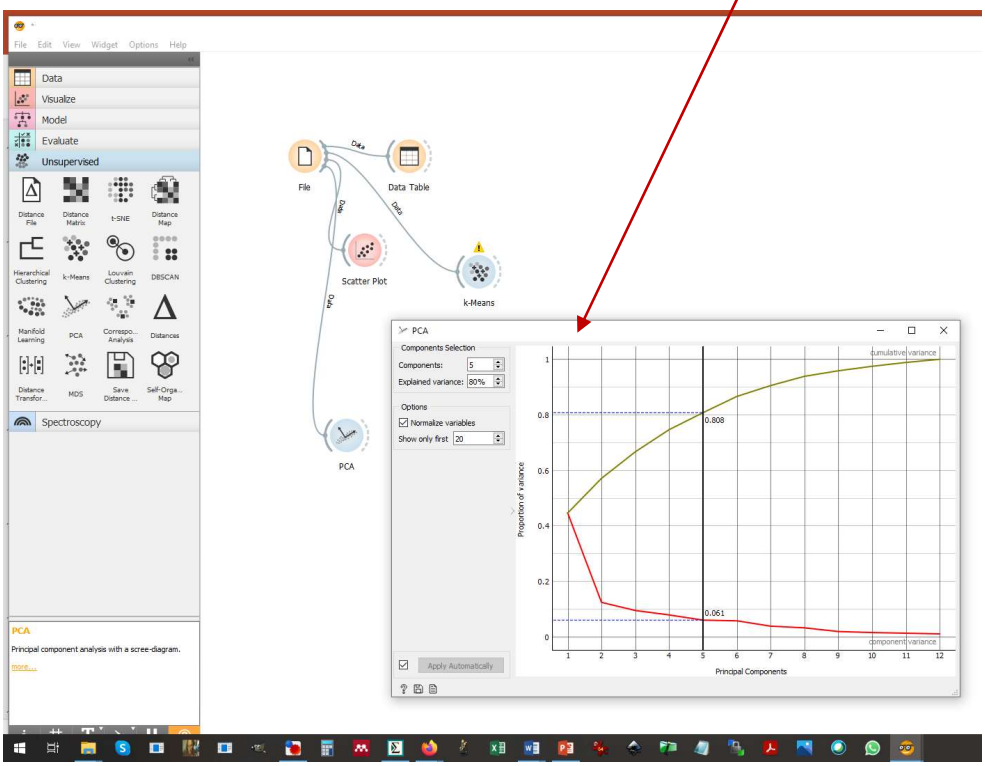
Another way of finding widgets is to drag the line from the File widget to an empty space on canvas and a list of recently used widgets will appear along with a search box. Type first letters of the widget you are looking for and you will be offered it.

Note: there is an exclamation mark above k-Means widget. This is a warning for k-Means inability to calculate silhouette values (they help us determining the number of clusters present) for n > 5000. Remember we have n= 65536.

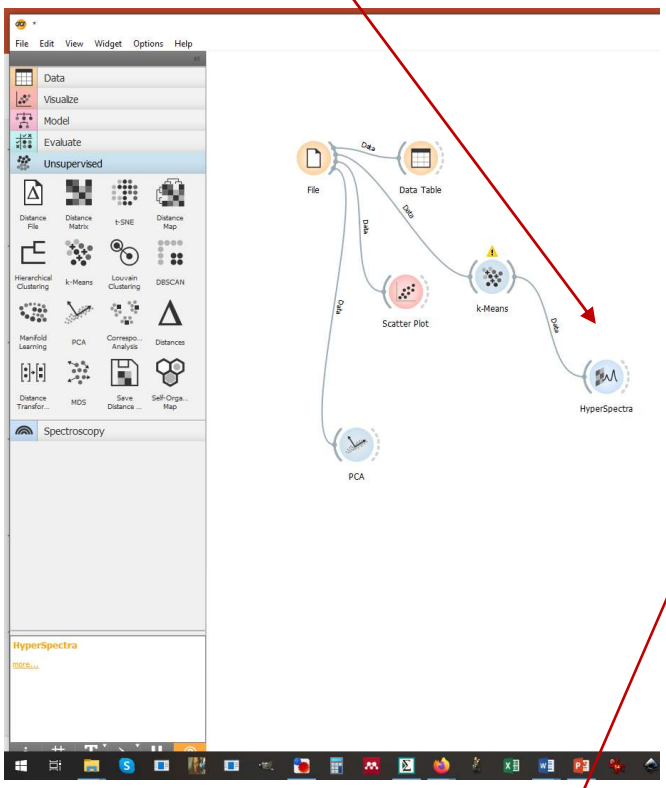
Double clicking on the k-Means widget opens a pop-up window where we can make selection for the parameters in the k-Means statistics.



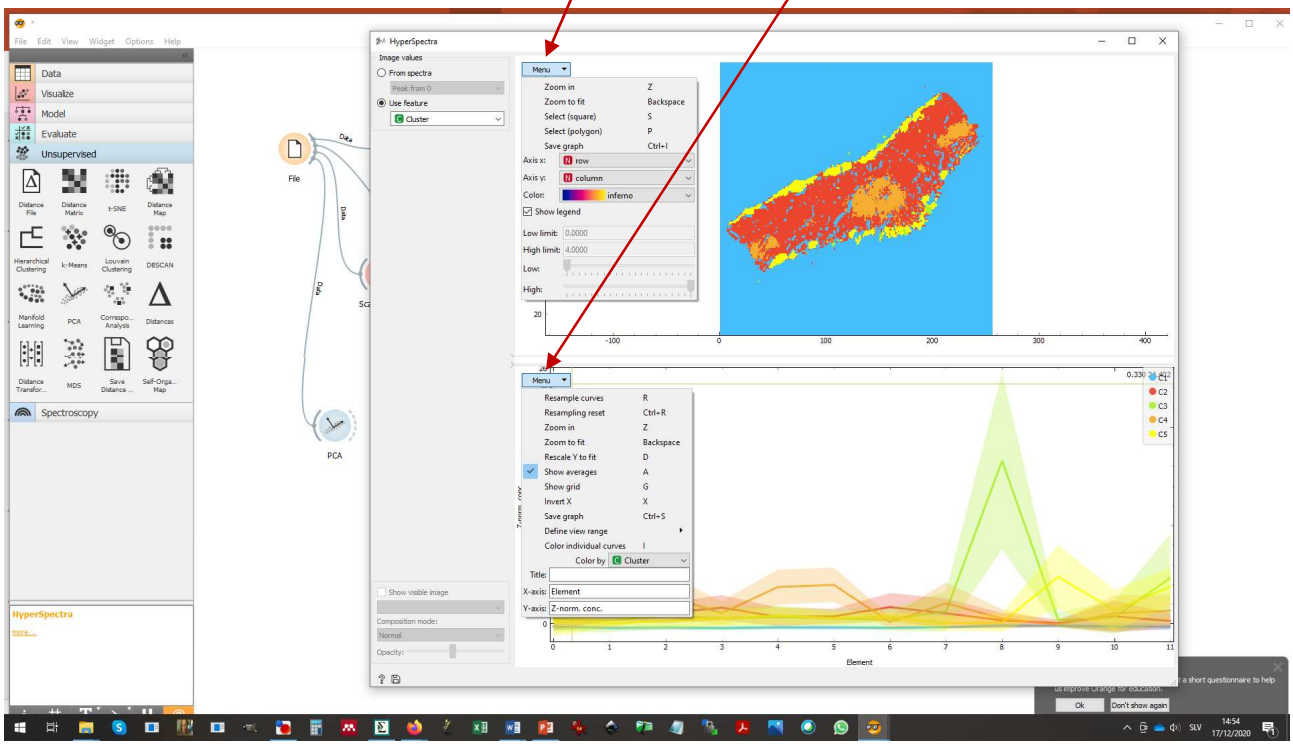
Because we cannot calculate silhouette values (our n>5000), we can use PCA do determine the number of principal components (PC) describing the variability in our dataset. Connect the File and the PCA widget and inspect the graph: it indicates that with 5 PC we describe >80% variability. In the k-Means statistics we select 5 as Fixed number of clusters.



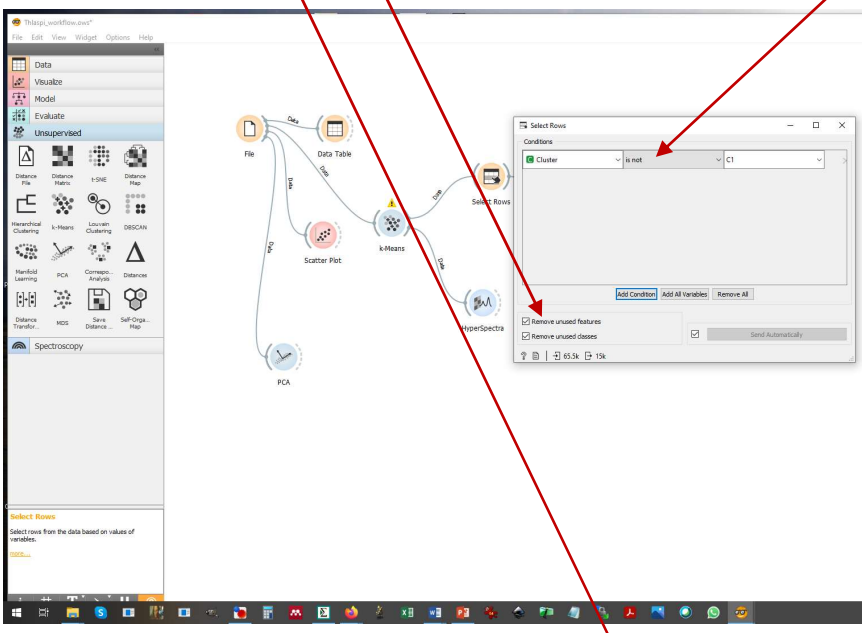
Results of k-Means statistics can be visualised through Hyperspectral widget, which you connect to the k-Means widget.

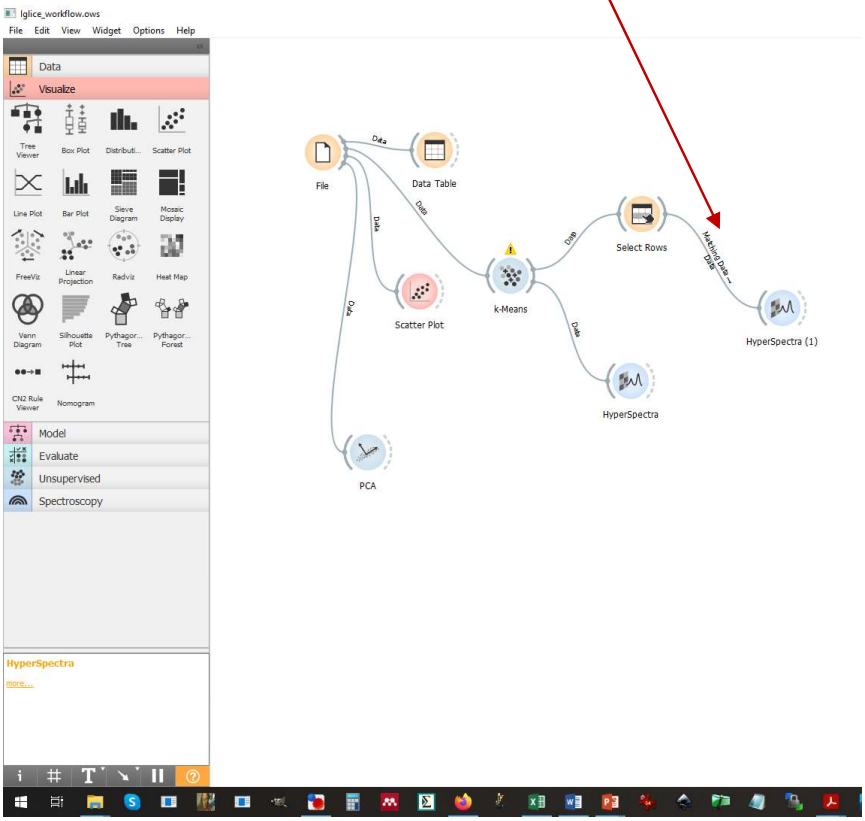


Double click on the Hyperspectral widget and inspect the pop-up window. In Menu (at the top): for x and y select „row“ and „column“ and in Menu (at the bottom): Show averages & Color by Cluster.

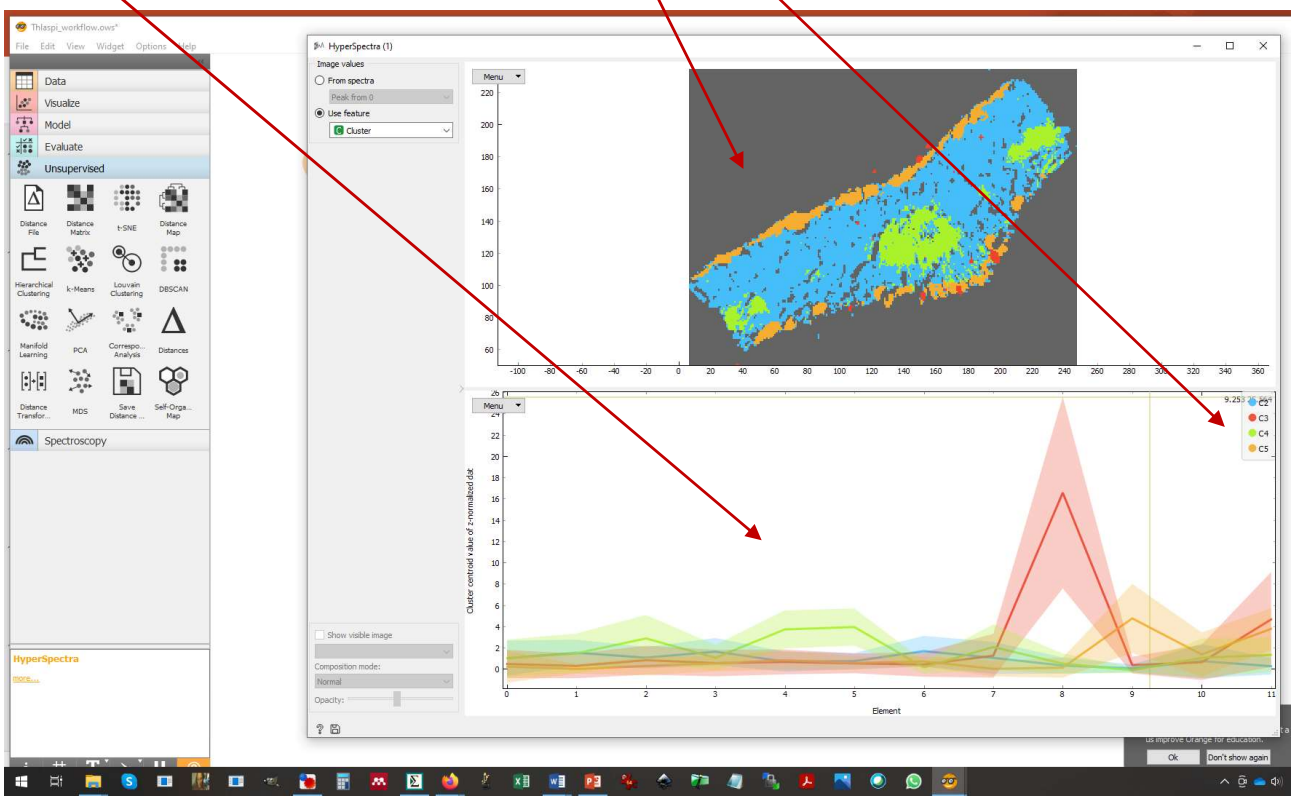


One of the clusters shown is actually the background (the C1 cluster, in blue) and because this in not a meaningful cluster, we can remove it through **Select Rows**, where you select: **Cluster is not C1**. Tick the boxes in front the Remove unused features & Remove unused classes. Repeat the visualisation of the result through the Hyperspectral widget, but with data originating from the Select Rows widget.





The final result: there are 4 clusters of tissues with unique element profiles: mesophyll (cluster in blue), vasculature (cluster in green), epidermal cluster (cluster in yellow) and hot-spots in epidermis (cluster in red). The colours on the image at the top match those in the graph at the bottom.



In the graph, x-axis only has numbers – their sequence represent the order by which elements are organised in the initial file. In this case: Mg, P, S, Cl, K, Ca, Mn, Fe, Zn, Cd, Pb. This graph indicates which element profiles describe certain cluster (tissue).

Few examples:

Yellow cluster (epidermis) is characterised by more-than-average Zn and Pb concentration.

Red cluster (the hotspots in epidermis) is characterised by more-than-average Fe in Pb concentration.

When interpreting this graphs do not forget it operates on z-score and not concentrations (thus: more-than-average or less-than-average).

Have fun with the tissue-specific distribution maps!