

# *temDM*

## **temDM MSA basic version**

Pavel Potapov

February 21, 2022

### **CONTENTS**

<b>1</b>	<b>Installation</b>	<b>2</b>
<b>2</b>	<b>Get familiar with temDM tools</b>	<b>2</b>
<b>3</b>	<b>Basic PCA treatment</b>	<b>3</b>
<b>4</b>	<b>Treat EDX data</b>	<b>8</b>
<b>5</b>	<b>Cut spectrum-images</b>	<b>10</b>
<b>6</b>	<b>View Results</b>	<b>10</b>
<b>7</b>	<b>More about Spectrum Picker</b>	<b>12</b>
<b>8</b>	<b>Rotation of PCA results</b>	<b>13</b>
<b>9</b>	<b>Clustering</b>	<b>16</b>
<b>10</b>	<b>View rotated and clustered results</b>	<b>19</b>
<b>11</b>	<b>Endmembers</b>	<b>20</b>
<b>12</b>	<b>Settings</b>	<b>23</b>
<b>13</b>	<b>Export to plotting programs</b>	<b>25</b>
<b>14</b>	<b>Troubleshoot</b>	<b>25</b>
	<b>References</b>	<b>25</b>

## 1 INSTALLATION

You should place the **temDM MSA.gtk** plugin into a Plugins folder of DigitalMicrograph. There are typically several such folders, for instance `C:\ProgramData\Gatan\Plugins`. If import of Velox XEDS spectrum-images is desired, you should also place there an appropriated HDF5-reading plugin. These open-source plugins were compiled by Tore Niermann, TU Berlin :

**hdf5\_GMS2X\_amd64.dll** for the 64-bit system,  
**hdf5\_GMS2X\_x86.dll** for the 32-bit system.

The script **find plugins folders.s** included in the distribution package will help you to localize the desired folders. Open **find plugins folders.s** in DigitalMicrograph and run it by pressing **execute** or by pressing **ENTER** while holding the **CNTR** key. Read the list of the available plugins folders. The first folder in the list is most appropriated for placing the temDM plugins.

Some folders can be hidden in Windows. If you do not see all folders, make them visible in the file explorer of Windows 10: **View** tab - click **hidden items** checkbox. In some network-based systems, **find plugins folders.s** could fail to localize the user-specific DigitalMicrograph plugins folder. You might try to find them manually checking for the paths like  
`.../users/user/AppData/Local/VirtualStore/Program Files (x86)/Gatan/DigitalMicrograph/Plugins`.

After you localize the appropriate plugins folder you should

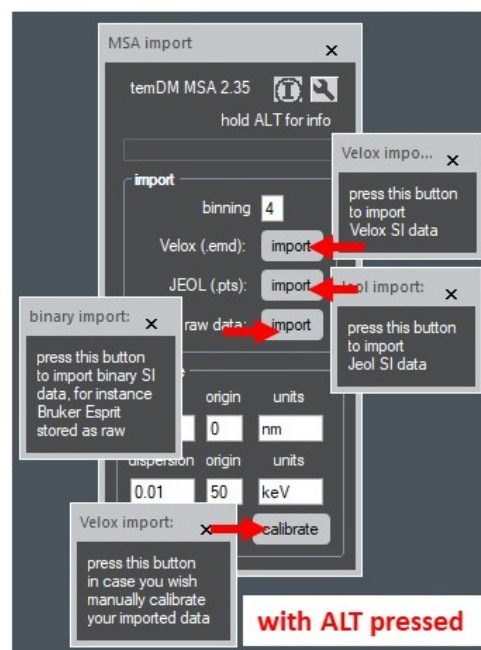
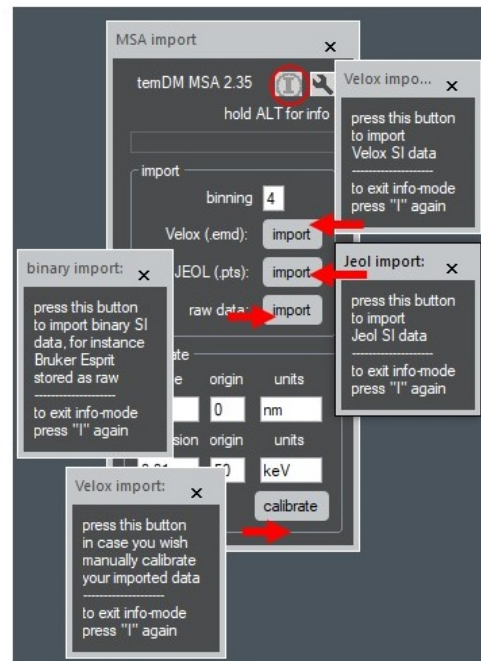
- drop **temDM MSA.gtk** and **hdf5\_GMS2X\_amd64.dll** into the chosen Plugins folder.
- restart DigitalMicrograph and find the MSA items in the Menu **temDM**.

If you are using the most recent versions of GMS (Gatan Microscopy Suite), you might get the warning message "You have incompatible plugin *hdf5\_GMS2X\_amd64.dll*" during the start of DigitalMicrograph. That is because this dll was compiled for the older versions of GMS. In fact, the plugin works with any version. This message is just over-security and you can safely ignore it. Alternatively you can get rid of this boring message by removing *hdf5\_GMS2X\_amd64.dll* from the plugin folder. However you then would not be able to import Velox spectrum-images. The problem with the warning message will be fixed in the near-

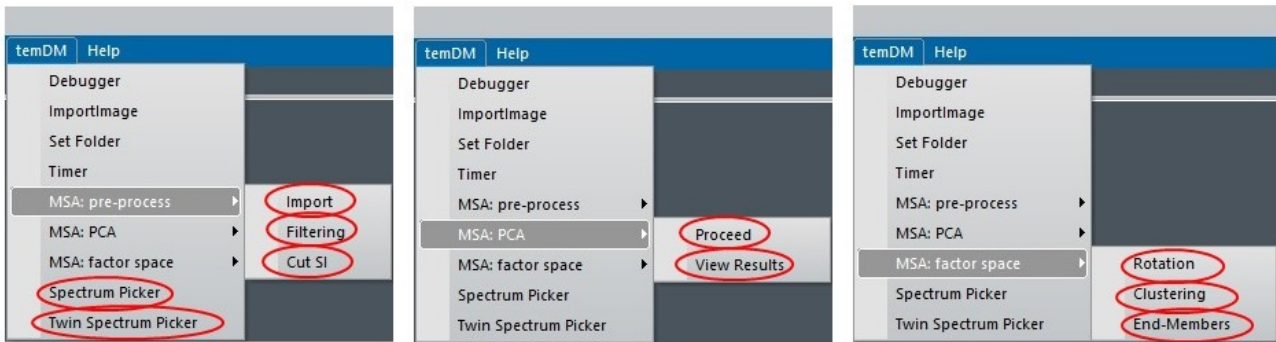
est future.

To update the version, just overwrite the plugin of the previous version in the Plugins folder. All versions of temDM MSA have the same name to avoid confusion with loading ambiguous commands. If you have several versions, it is recommended to keep Plugins in individual folders with meaningful names like `temDM MSA basic version 2_XX`.

## 2 GET FAMILIAR WITH TEMDM TOOLS



Now you have a **temDM** menu with a range of items



(shown at the top of the next page).

Choosing an item in the temDM menu opens a given temDM tool - a small pallet to proceed certain treatment procedures. All **temDM** tools share the common features designed for smooth and comfortable operation. The standard DigitalMicrograph tools are sometimes not easy to find at the workplace. They might be accidentally shifted to the the very corner of the screen or shadowed by another tools. No such a problem for the **temDM** tools! These tools remember their positions. Try, for instance to open the **Import** tool, move it around the workspace, close and open again. Next time you open **Import**, it appears exactly at the place you left it previously. So you can situate a tool at the most convenient for you position and that will be remembered by the system.

If the position was accidentally set crazy such as you cannot find it, open the tool while holding the SHIFT key. The default position will be reset.

Another useful feature is a so-called information mode. You are not obliged to remember what each button is doing. Just click at the small **(I)** icon at the upper-right corner of the tool to activate the information mode. Now pressing any button would output a small text window explaining in short the action of the button.

Note that the **(I)** icon goes dim when the tool is in the information mode. To restore the normal mode just click the **(I)** icon again. You can also get information about a given button without activating the **(I)** icon. Just click the button of interest while holding the ALT key.

### 3 BASIC PCA TREATMENT

This chapter will introduce you in how to perform Principal Component Analysis (PCA) of STEM EELS (Electron Energy-Loss Spectroscopy) or EDX (Energy Dis-

persive X-ray Spectroscopy) spectrum-images.

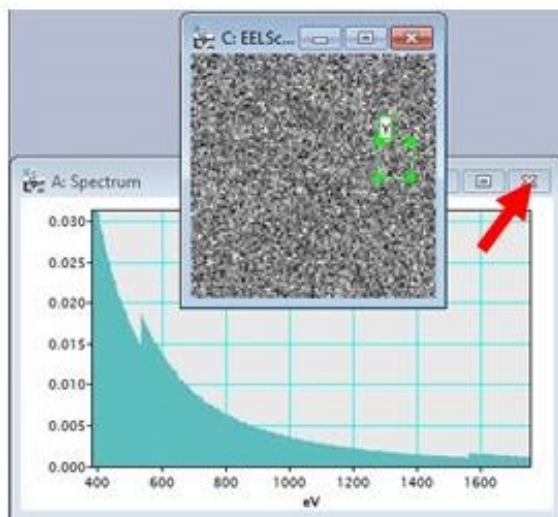
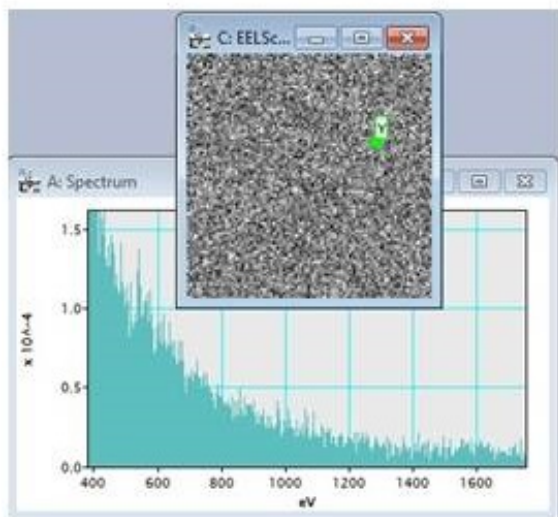
First of all you have to open the data cube where the results of your spectrum-imaging are stored. That might be an EELS data acquired by the Gatan software and stored in the standard format - with spatial dimensions along X and Y and the energy axis along Z (not directly visible). For the purpose of learning you might open the simulated EELS spectrum-image available in <http://temdm.com/web/msa/>. The simulated data cube **EELScube.dm3** mimics the format and hidden information tags of real experimental data cubes. You can inspect how the spectrum is changing from pixel to pixel in **EELScube.dm3**.

- Put **EELScube.dm3** in front and choose **Spectrum Picker** in the **temDM** menu.

A green marker appears inside the image while a spectrum is displayed in the separate window. You might enhance the vertical scale by mouse with holding the CNTR-key. Notice that the spectrum is always very noisy. This spectrum-image was generated noisy by purpose. You may drag the marker across the image - the spectrum will be live updated. By default, the marker points to the one-pixel area but you can enlarge the green rectangle and get the spectrum averaged over the larger area. If the area is too large for live update, the spectrum will be updated as soon as you release the mouse button. A spectrum from larger areas look much nicer, is not it? Well, the averaging removes random noise. However, you probably need the spatial information, not just a spectrum averaged over a large area. The question is how spectra change from pixel to pixel. With the PCA denoising, you can obtain such a nice spectrum at *each* pixel. Just continue, you get it.

*How to stop viewing spectrum ?*

- Close live-spectrum display without saving (you might keep ALT-key pressed to omit the DigitalMicrograph saving prompt)



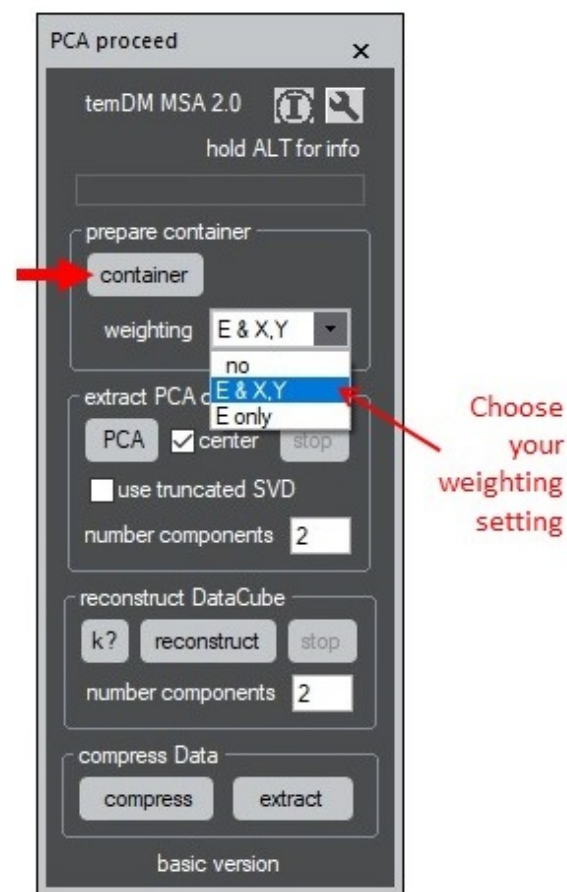
Lets treat **EELScube.dm3** and try to improve it.

All results of the statistical analysis will be kept in a special container **EELScube MSA.dm3**. Before you generate such a container, decide whether your Principal Component Analysis (PCA) will be weighted or not. The weighting treatment is needed to equalize the Poisson noise across the dataset and is strongly recommended. In particular, unweighted PCA will *not* pick up the relevant data variation in the example **EELScube.dm3**. Instead, the random variations of the low-energy background will be highlighted. Weighting really makes a good job!

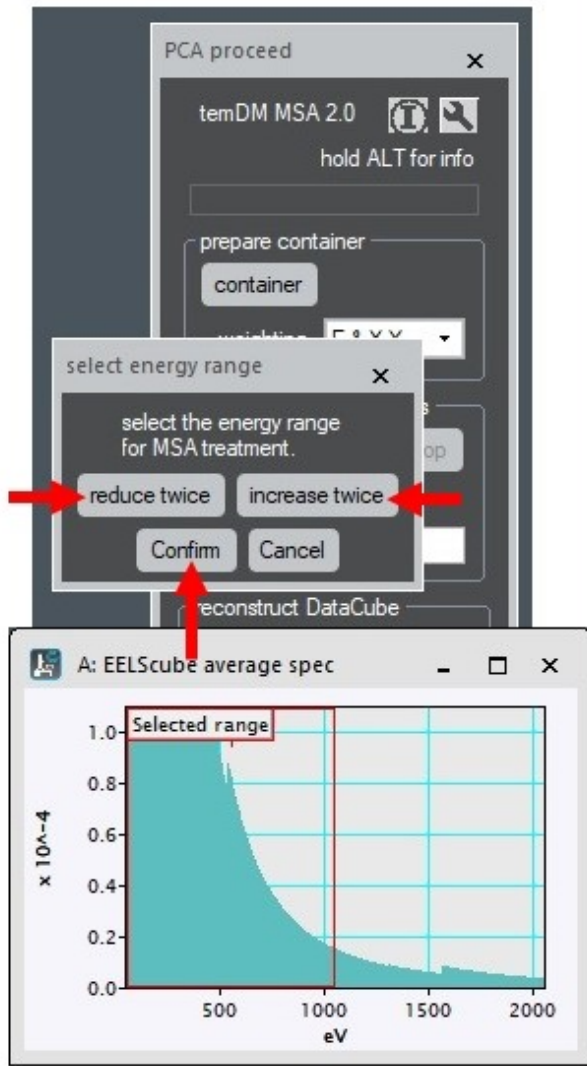
You may choose to weight over spatial and energy dimensions or over energy only. Weighting for only spatial dimensions makes no sense for typical STEM spectrum-images.

To generate **EELScube MSA.dm3**:

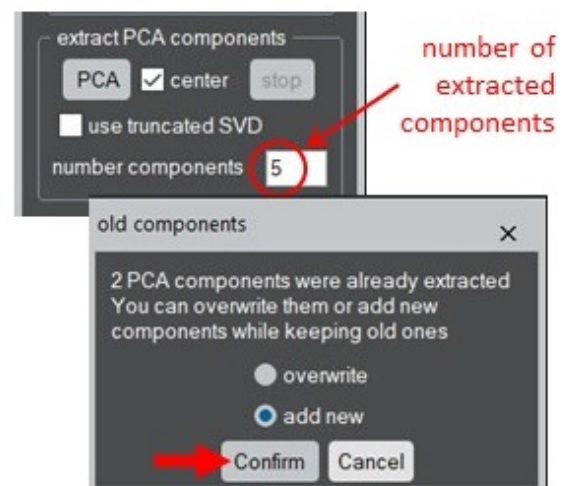
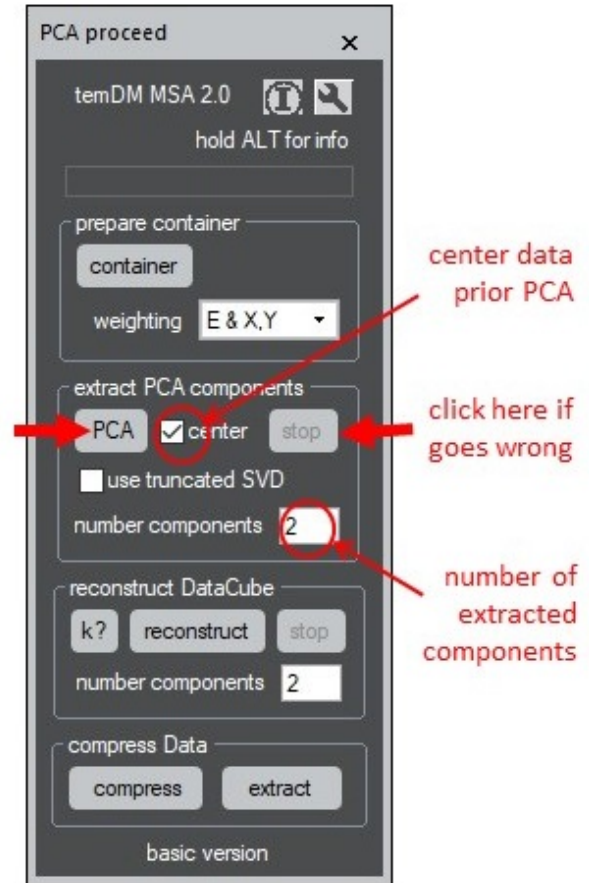
- Open the **Proceed** tool by choosing **temDM - MSA:PCA - Proceed**,
- press **container** button while having the original **EELScube.dm3** in front. The spectrum averaged over all pixels appears. You can choose the energy region you are interested in by dragging the red rectangular "Selected range". Clicking the button **<** or **>** would reduce or increase the region twice. When you are confident in your energy region press **OK**. Alternatively you might just close the spectrum window without storing it. That is the most convenient way if you wish to use the complete energy range. Just close the spectrum with holding the ALT-key pressed!



You have obtained a new entity called **EELScube MSA.dm3**. This 2D image displays nothing but values averaged over all energy channels, i.e. an “average image”. However the 3D data distribution is not lost ! This is attached to the image in the matrix form suitable for further processing. Such image will be called *MSA container* through this guide. Do not be surprised that *MSA containers* are quite large when stored. There is a lot of hidden information in it. All MSA results such as PCA loadings and scores will be



sition (like number of components, centering, clustering) and still use the same **EELScube MSA.dm3**. No need to start again with the original data cube. But you *do have* to generate a new MSA container if you wish to change the energy range or to evaluate weighted vs unweighted treatment.



kept in this image as tags. The key point is - you may play with different parameters of the PCA decompo-

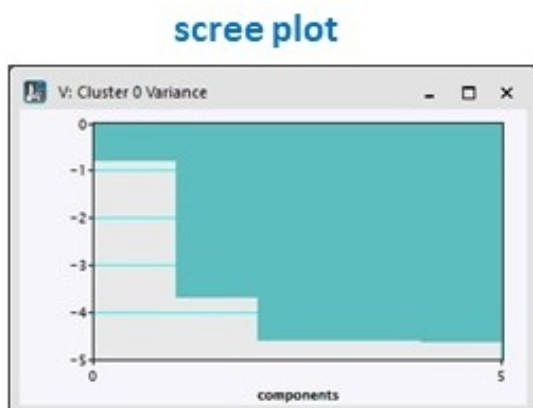
Now you are ready to extract PCA components. Before doing that, it is a good idea to check few parameters. One is *centering*, which means that all your data variations will be counted relative the average spectrum. Centering is recommended. Another important

parameter is the number of principal components to be extracted. Do not think much about that right now, it will be discussed later in detail. For the moment, just realize that the **temDM MSA** is very flexible and allows anytime to add more principal components *without overwriting* the previously extracted ones. Let's practice in it:

- press **PCA** using 2 (default) components while having **EELScube MSA.dm3** in front. The extraction proceeds followed by displaying the PCA results. We will explain later their meaning. Now just close the resulted images without storing.
- Input more components, for example **5**, and press **PCA** again. The dialog appears asking you if you want to overwrite the old components or add the new ones. Confirm adding more components. If you realize that too many components are requested and their extraction takes too much time, you can interrupt the procedure at any moment by pressing **stop**.

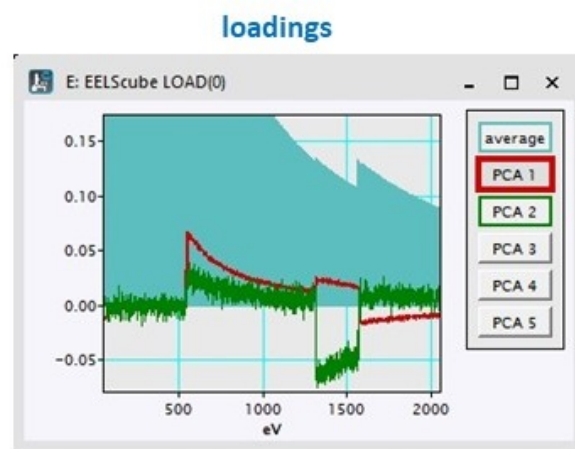
Once the PCA decomposition was executed successfully, several forms of presenting the results show up.

One is a *screepplot*, i.e. the variance of data within each principal component. The PCA components are sorted in such a way that the variance decreases with increasing the component index (this is actually the essence of PCA). Inspection of a screepplot might help you to find the number of meaningful components in the system. For instance, the **EELScube.dm3** data shows two PCA components with large variance while components of index **3-5** exhibit the lower variance that changes little from index to index. That is just a noise! It is clear that all forthcoming principal components would not be very different because you have approached the noise level already. Thus, we conclude that this data cube exhibits two principal components, i.e. two independent variation trends.



A screepplot is important but not the only hint to determine the number of non-noise components in a dataset. More tricks will be demonstrated below.

Another essential PCA results are spectra of principal components, or *loadings*. In this example, the only two first components show the meaningful spectra. PCA components **3-5** (hidden in the shown display) are simply noise. The average spectrum is also displayed in the same graph. All PCA components should be considered as the deviations from this average spectrum. Do not try to interpret what each component actually means at this stage of treatment. The interpretation aspects will be touched later.

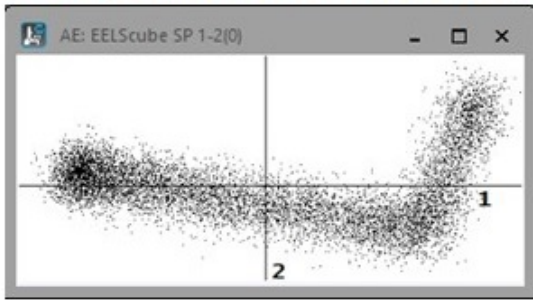


This all would look differently if you choose to do uncentered PCA. Then, the first component is effectively the average spectrum, while the total number of components is increased by one. To do or not to do centering is in a large extent a matter of flavor. In most cases the final results e.g. denoised spectrum-images will be identical. The **temDM** package assumes the centered treatment as default.

A very important characteristic is a *scatter plot*, i.e. the joint distribution of two selected component scores. After the extraction of the principal components, the scatterplot of 1st vs 2nd score will be automatically displayed. Later you may inspect the scatter plots between any couple of components. The scatter plot of the two first components is typically most informative, that is why it is displayed automatically. In the present example you see that **EELScube.dm3** shows two independent trends of the data variation (two branches in the scatter figure).

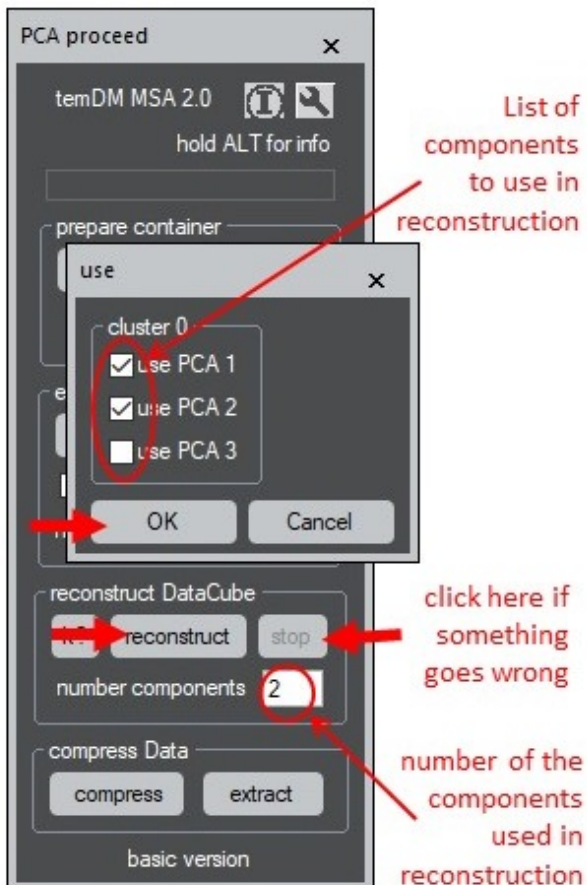
All these three pictures are for your information only. After you have a look, just close them without saving. They are automatically stored in **EELScube MSA.dm3**. You can easily display them any time when necessary.

### scatterplot of 1st vs 2nd component



If you agree that **EELScube.dm3** consists of only two non-noise components, you can reconstruct the data cube using these two components only and get rid of noise.

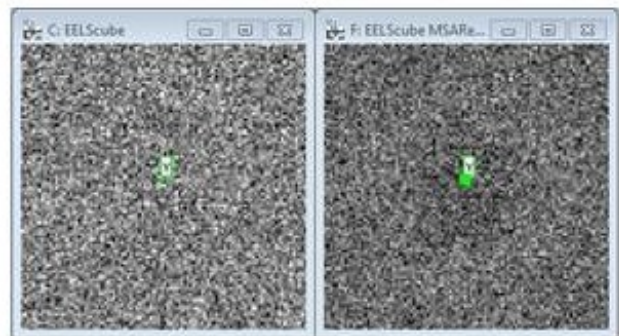
- Look at the **reconstruct DataCube** box in the **Proceed** tool. Choose the number of components for reconstruction. As a little exercise input 3, not 2 for **EELScube MSA.dm3**.
- Click **reconstruct**. You get the list of components ready to include in reconstruction. That is your last chance to reduce the number of components. Uncheck PCA 3 and press **OK**.



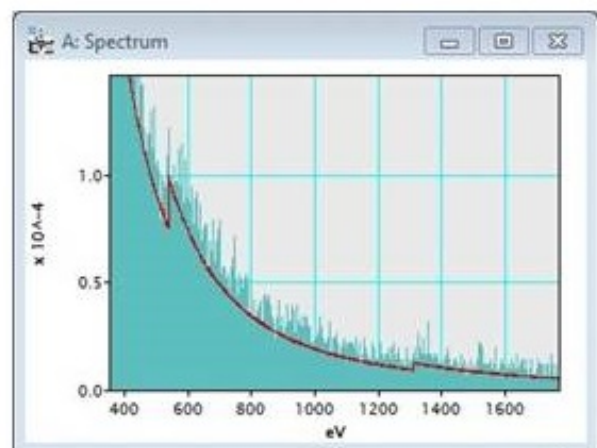
in between the meaningful ones, for example remove PCA 2 while retain PCA 3. This however is *not* recommended to do unless you see clear artefacts in the intermediate component.

Finally the reconstructed data cube **EELScube REC(3).dm3** appears. You might wish to compare it with the original data cube.

- Place **EELScube REC(3).dm3** front-most and **EELScube.dm3** second front-most; choose **temDM - Twin Spectrum Picker** in the DigitalMicrograph menu. Enhance the vertical scale.



Green markers will appear in both the images while the extracted spectra will be displayed as an overlay in the separate window. You may drag or resize the marker at any image. The companion marker will move synchronously.



Do you see the difference between the original and the reconstructed data cubes? Well, this is the simulated data. Real examples can be not as straightforward... Nevertheless if you tune all steps carefully, the results might be astonishing as a plenty of unwanted noise is gone.

Theoretically you can remove a PCA component

- To stop twin-spectrum picker close spectrum win-

dow without saving.

The choice of the number of relevant principal components is not always evident. Inspection of a screeplot, loadings or score maps might help you to make a decision. The advanced version of the **temDM MSA** package offers also the advice on how many components to retain for reconstruction (the button **k ?**).

Now you probably wish to store the results of your PCA decomposition. Try to store **EELScube MSA.dm3**.

Ooops... Why this 2D image is so large ? It is even a bit larger in size then the original spectrum-image. This is because the original 3D data is kept as a hidden matrix in **EELScube MSA.dm3**. But you can get rid of that.



- Click **compress** while having **EELScube MSA.dm3** in front.

A new 2D image - **EELScube COMP.dm3** - seemingly identical to **EELScube MSA.dm3** - appears. It consists of all the results of your analysis *except of* the original noisy data. Store it and find out that it is *very compact*.

What is important - for 95% of further operations,

**EELScube COMP.dm3** is identical to **EELScube MSA.dm3**. You may do the further MSA (Multivariate Statistical Analysis) : rotation, clustering, endmembering and others. Enjoy the wide range of treatment possibilities with the very compactly compressed format ! You do not need necessarily to store the large **EELScube MSA.dm3** file. Just find the right number of PCA components, ensure that the reconstruction goes correctly and convert **EELScube MSA.dm3** into **EELScube COMP.dm3**. Then, you may close **EELScube MSA.dm3** without saving.

You can anytime get the reconstructed spectrum-image from your compressed image.

- Click **extract** while having **EELScube COMP.dm3** in front.

This would take a while because the reconstruction process has to be repeated. Actually, **extract** is equivalent to pressing the **reconstruct** button with the only difference that you need not specify the list of components for reconstruction. The list of components from the previous reconstruction will be utilized to build the denoised spectrum-image.

*At this point, job is basically done. However, you are advised to explore more features of the **temDM MSA** package.*

## 4 TREAT EDX DATA

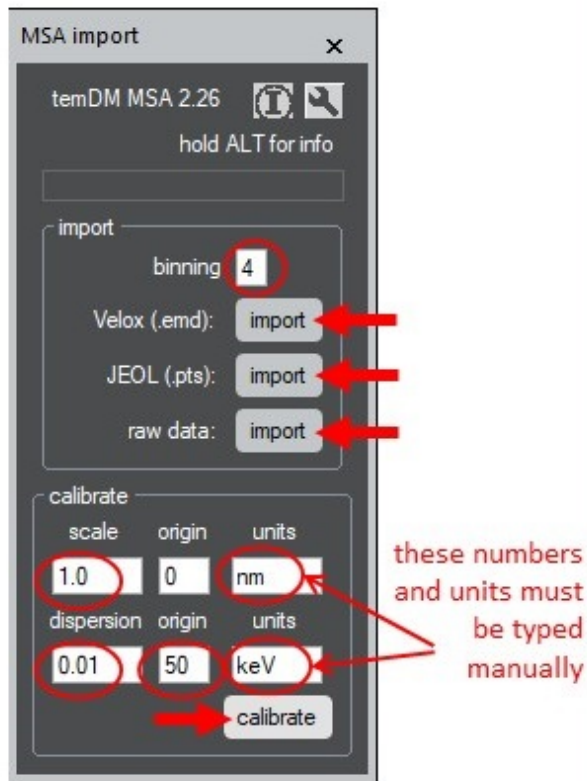
EDX (Energy Dispersive X-ray Spectroscopy, sometimes called EDS) data cubes can be treated in the same way. In case your EDX data were collected with a software other than DigitalMicrograph, the **temDM MSA** package allows to convert it into the .dm3 format with an **Import** tool. In particular, you can import EDX spectrum-images stored by the Thermo Fisher Velox software (version 1.3 and higher). Use the **superalloy 1510** example free downloadable at <http://temdm.com/web/msa/> to learn the importing procedure:

- Open the **import** tool by choosing **temDM - MSA:pre-process - Import** in the DigitalMicrograph menu.
- Click **import** button in **import FEI Velox** box, browse and choose the **superalloy 1510.emd** file.
- The range of available frames to integrate in the spectrum-image appears. If you are sure all frames



are appropriated for integration, just click **OK**.

The resulted .dm3 spectrum-image will be already properly calibrated.



Similarly, you can import a JEOL EDX spectrum-image. In that case you should choose the .pts file where the EDX stream is stored in the JEOL system. For the moment, Bruker, EDAX and Oxford native formats are not recognized by **temDM MSA**, but you can store data as raw binary files and then convert it into .dm3. Just for training, play with the **MagiCal EDX** example free down loadable at <http://temdm.com/web/msa/>. It was collected with the Bruker ESPRIT software and stored as binary **h-4.raw** file.

- Click **import** button in **import binary (Bruker, EDAX)** box, browse and choose the desired **h-4.raw** file. The ripple file **h-4.rpl** must also be in the same folder. Conversion take a while. You can stop it anytime by pressing **stop** if you clicked the wrong file.

Unfortunately the calibration is not kept in the **h-4.raw** or in **h-4.rpl** files, thus you should calibrate the resulted data cube manually by setting the proper numbers in the fields “scale” (how many nanometers are in one pixel) and “dispersion” (how many keV are in one energy channel). The value “origin” is usually 0 for spatial calibration while must be set at some pos-

itive number when calibrating energy. It varies from system to system but is fixed for a given spectrometer and a given spectrometer dispersion. Just check which channel is assigned to ZERO in your original EDX software. For the **MagiCal EDX** example, the calibration should be:

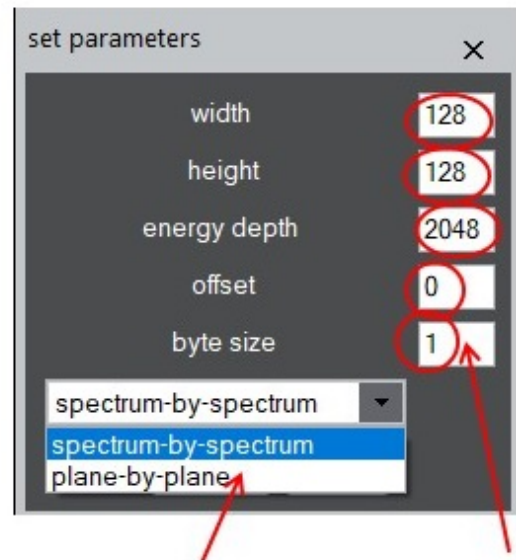
scale: 0.67

dispersion: 0.01

origin: 48.

- Click **calibrate** while having the converted spectrum-image front most.

In the considered example, the key parameters of the data cube are read from the ripple file **h-4.rpl**. If the system does not find such a file, you will be prompt to input the required parameters (width, height, energy depth ...) manually in the appearing dialog window.



check the sequence of data storage in a stream

check these numbers

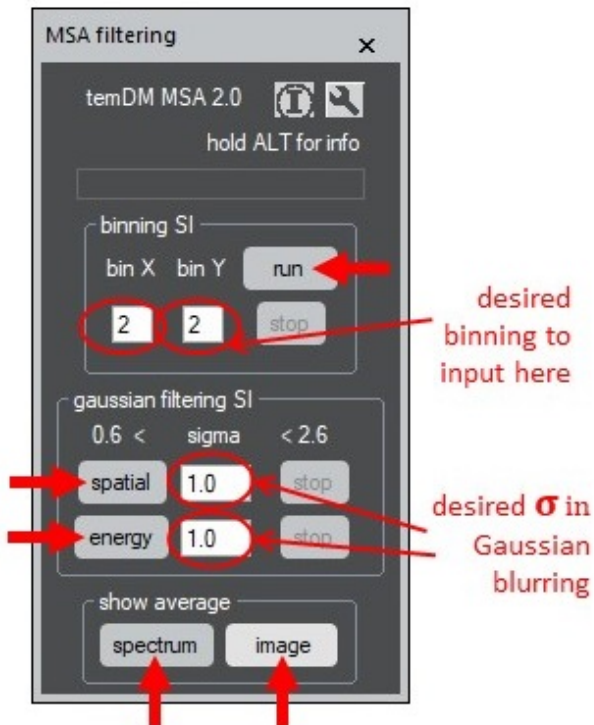
*Very important note:* EDX data are typically quite sparse. *This makes severe problems in PCA treatment.* Therefore, it is strongly recommended to down-sample data *in the course of importing*. Do not generate data cubes of huge size if it is not strongly needed! Keep the **binning** number in the **import** tool reasonably high, like 2 or 4.

If you did not care about down-sampling and found data too large after importing, you still have a possibility to bin it with another tool. Try this with **h-4.dm3** datacube:

- Open the **filtering** tool by choosing **temDM - MSA:pre-process - Filtering** in the DigitalMi-

crograph menu.

- Choose the binning factors **2** on both X and Y.
- Click **run** button.



This would reduce the sparseness in the data and greatly reduce its spatial dimensions although the number of energy channels will be unchanged (for the moment, there is no energy binning in the **temDM MSA** package) .

In case your PC has no problem to handle large spectrum-images, you might smooth the very noisy data by application of the “gaussian” filter. This filter makes the weighted averaging according the Gaussian function kernel. The smoothing parameter is the standard deviation *sigma* (expressed in pixels) in the Gaussian distribution. Both spatial and energy filtering is available. Currently, the allowed range is  $0.6 < \sigma < 2.6$ . Just try to smooth further the previously binned **h-4\_2x2.dm3** datacube:

- Having the binned **h-4\_2x2.dm3** spectrum-image in front and leaving the default  $\sigma=1$  in **Gaussian filtering SI** box, press **spatial** or **energy** to proceed filtering.

The Gaussian filtering is rather long procedure in the basic version, thus reserve some time for it. This process is significantly speed up in the advanced version.

Finally you might wish to check how your average image (image averaged over all energy channels) or average spectrum (spectrum averaged over all probes) look like.

- In the **show average** pan, click **spectrum** or **image**.

Then, you may denoise the resulted **h-4\_2x2\_xGss** cube like it was described in the previous section.

## 5 CUT SPECTRUM-IMAGES

Sometimes, you realize that the only part of a raw spectrum-image is useful for you. Then it is better to reduce your data *before* the MSA treatment. There is no sense to spend memory and computation time for treating useless pieces of data.

- Choose **Cut SI** in **temDM - MSA:pre-process** Menu having the target spectrum-image front-most.
- The average image appears. Select useful spatial area by dragging a green rectangular in the average image. Press **Cut selected**
- The average spectrum appears. Select useful energy region. Press **Cut selected**
- Finally, you have a possibility to rescale the intensity of your spectrum-image, i.e. multiply it by a given parameter. If necessary, change the default scaling **1** and press **OK**

## 6 VIEW RESULTS

- Open the **view results** tool by choosing **temDM - MSA:PCA - View Results** in the DigitalMicrograph menu.

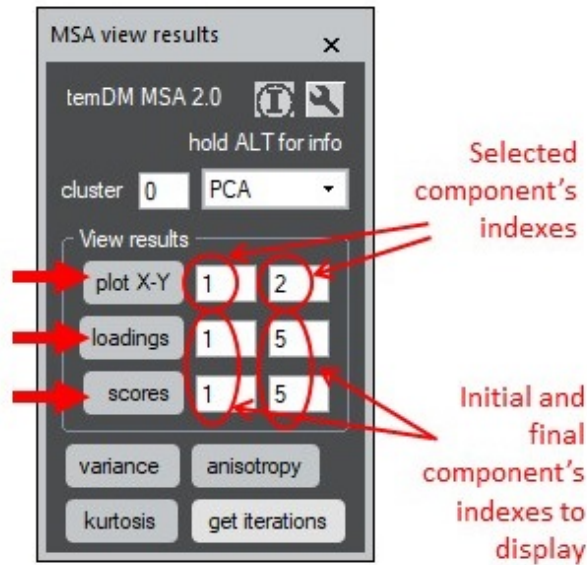
With this tool you can any time retrieve the stored information from **EELScube MSA/COMP.dm3**. For instance you can

- type the two component's indexes in the input fields and plot the corresponding scatterplot by clicking **plot X-Y**.

The sequence of the indexes plays a role! The index which stays in the left-hand field will be your horizon-

tal axis while the right-hand index will be the vertical one.

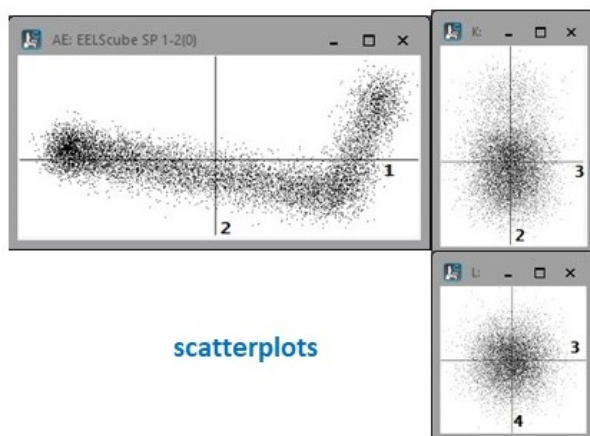
Now you can investigate the data distribution in the factor space with the desired deepness. Actually, you will quickly discover that all scatterplots with the principal components higher than 3 are essentially same - they form a round distribution of the uncorrelated noise nature. This is another way to determine the number of the non-noise PCA components in your dataset



You can also display the extracted PCA loadings in the desired range of indexes.

- Type the initial and final components and click **loadings**.

You have seen already such a display immediately after the PCA decomposition was finished. You can display these spectra at any moment without storing it in individual files.

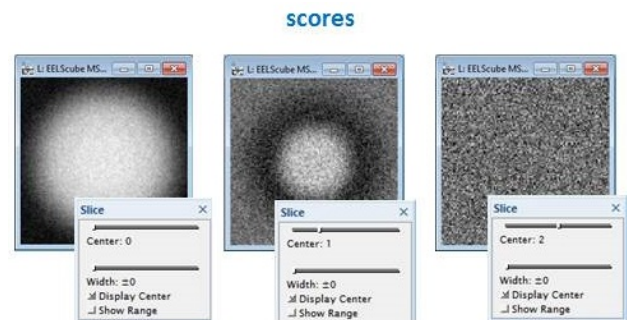


scatterplots

More information is retrieved by

- clicking **scores** with specifying the desired component ranges.

Then you get the maps showing how each PCA component is distributed across the object. The scores can be delivered as a number of 2D images or may be compacted into a data cube (choose your preference through the **spanner** button). In the latter case, you can scroll among different PCA components using the standard DigitalMicrograph tool - **Slice**. Again you see that the scores of the PCA components with the index 3 and greater show nothing but noise.



Explore now the buttons at the bottom of the **view results** tool.

- Clicking **variance** will plot the standard scree plot that you have seen already after finishing your PCA decomposition.
- Clicking **kurtosis** will calculate and plot the kurtosis for each extracted component.

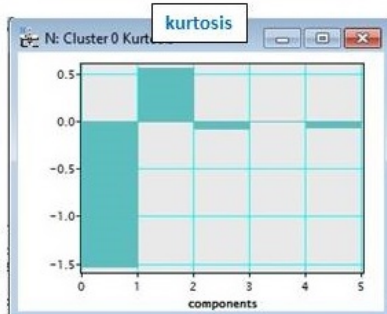
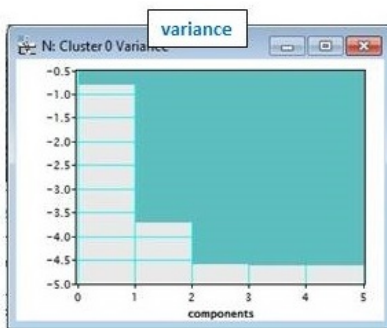
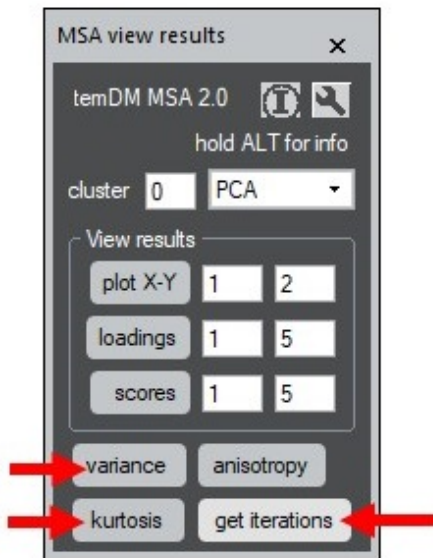
Kurtosis is a measure of non-Gaussianity in the data distribution. It may be negative or positive for an arbitrary distribution but it must be exactly zero for an ideally Gaussian-distributed data.

You might notice that the components with the index 3 and greater show almost zero kurtosis. Such a Gaussian-distributed component is very likely just a noise. This is another hint how to determine the number of non-noise components. *Caution:* a very large kurtosis is also not good - this is a fingerprint of heavy outliers in the component.

Finally you can inspect how the PCA decomposition was proceeded.

- Click **iterations**.

You have now access to the number of iterations required for finding each component. You can also see how the variance and precision (difference between two subsequent iterations) were progressing during the



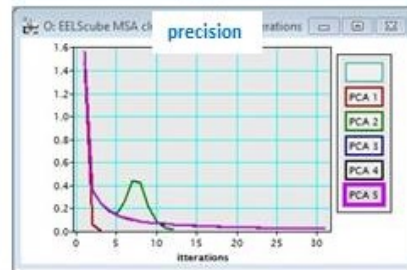
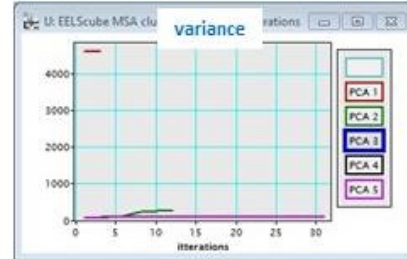
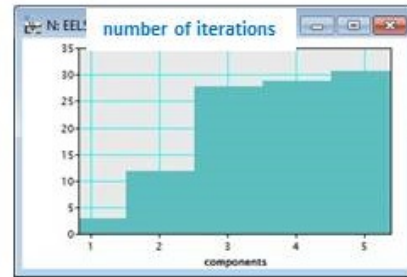
iterations. The inspection helps to identify the cases when the PCA algorithm did not converge for any reasons.

The button **anisotropy** is used in the advanced version only and will be explained in the corresponding manual.

## 7 MORE ABOUT SPECTRUM PICKER

With the tool **view spectrum** you can pick a spectrum not only from data cubes but also from MSA containers such as **XXX MSA.dm3**.

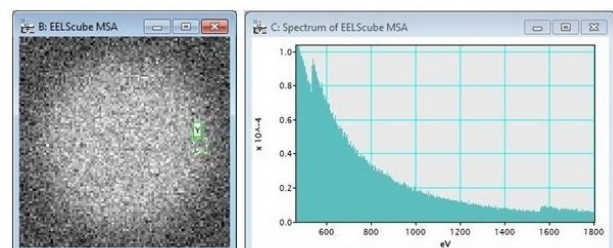
### iterations



- Put **EELScube MSA.dm3** front most and choose **temDM - Spectrum Picker**.

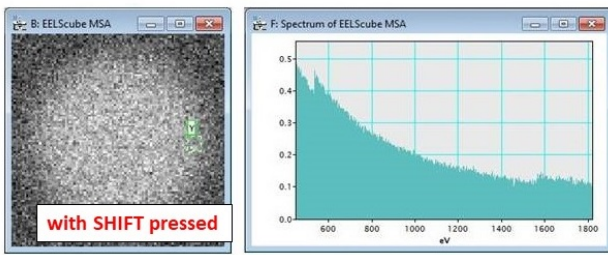
It works exactly same as that for data cubes although your spectra are now hidden inside the internal tag structure of **EELScube MSA.dm3**. You can also see how spectra have changed with the weighting procedure.

- Put **EELScube MSA.dm3** frontmost and choose **temDM - Spectrum Picker** while pressing SHIFT.



The same tool allows to extract a spectrum from scatter plots.

- Open or create a scatter plot as described in the previous section.
- Having the scatterplot front most choose **temDM**

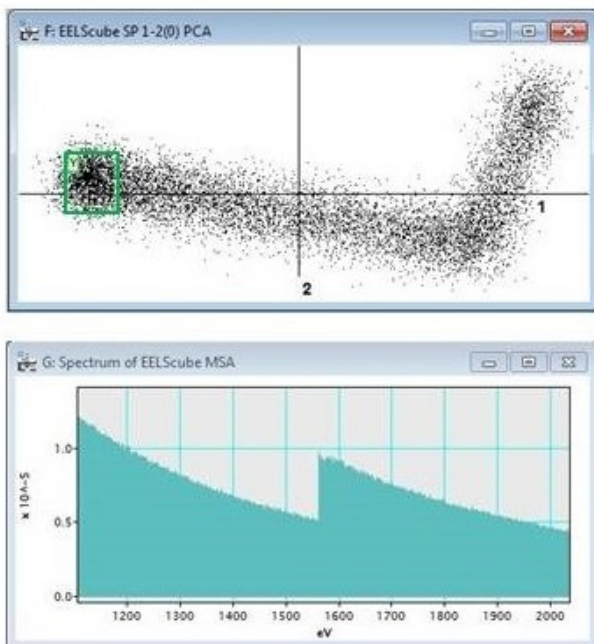


### - Spectrum Picker.

- Move the green rectangular to the place of interest in the scatterplot.

In this way you can get a clear idea about the distribution of your data in the factor space. This would help to identify the latent factors behind the data variation.

Again, you can get access to the weighted spectra when pressing SHIFT while loading **Spectrum Picker**.



## 8 ROTATION OF PCA RESULTS

You probably noticed that the obtained PCA loadings are not well interpretable. This is a typical feature of PCA - you can describe successfully the important data variation with few loadings but you cannot say what each loading actually means. You might try to improve interpretability by rotation of the PCA loadings in the factor space.

*Attention:* do not expect that the rotation will ulti-

mately improve the interpretability of the results. It might or might not. In fact, the *endmembering* treatment (explained below) is a much more reliable way to extract physically reasonable latent factors from your dataset. The rotation is included in the **temDM MSA** package rather for complementarity and for educational purposes.

Three rotation methods are incorporated in the **temDM MSA** package:

### Independent Component Analysis (ICA)

#### Maximal data simplicity (varimax)

#### Free rotation

How it works ?

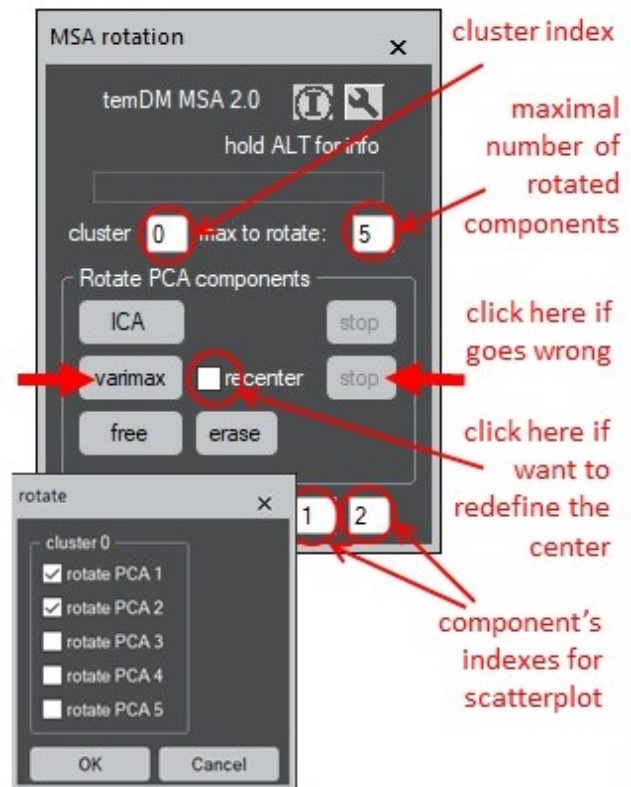
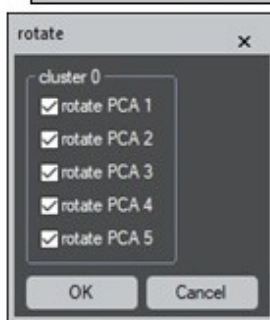
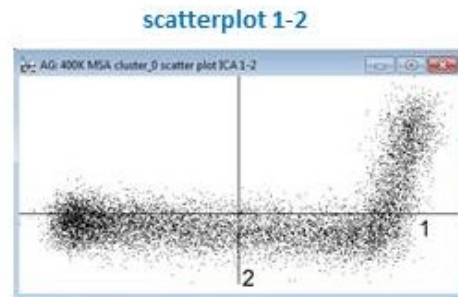
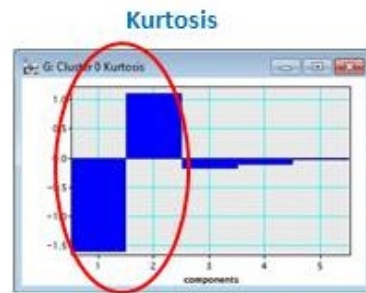
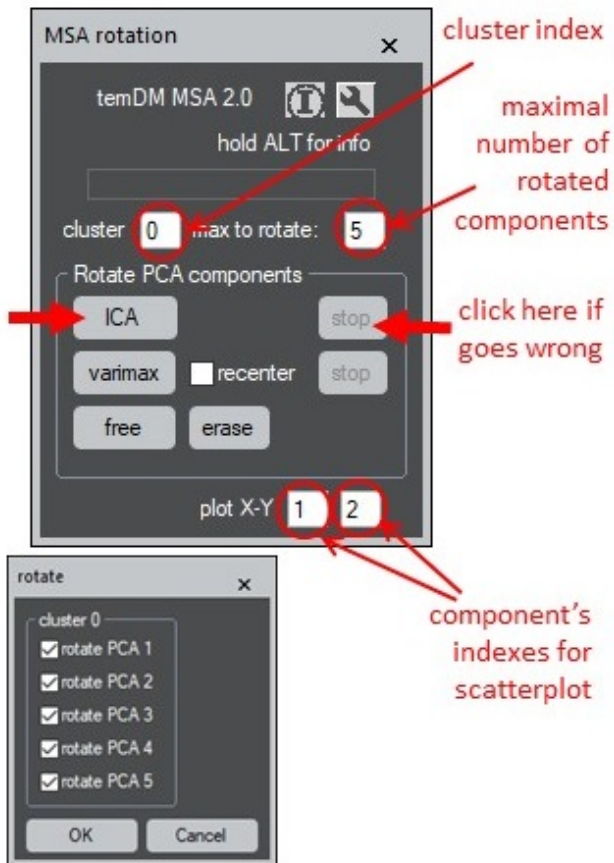
- Open the **MSA rotation** tool by choosing **temDM - MSA: factor space - Rotation**.

Lets start with ICA. ICA attempts to find the directions where the data distribution deviates most drastically from the Gaussian distribution. The criterion to be optimized is the kurtosis which is zero for the Gaussian distribution and positive or negative for the non-Gaussian one. Play with the example **EELScube MSA.dm3** (you can use also the compact **EELScube COMP.dm3** version).

- First choose which scatter plot (with which components) you want to see in the process of rotation. This is needed only for visualization of the rotation results, it does not affect the rotation itself. For **EELScube MSA.dm3** leave the default 1-2 scatter plot.
- Click **ICA**. You get the list of PCA components to be rotated. You may uncheck any of the components, then they will be fixed, i.e. not rotated. Do not do that for the moment. Press **OK**.

The rotation matrix appears showing how the new components are composed from the old ones. You might notice that components 1-2 are rotated moderately while components 3-5 are very significantly (and randomly) altered. This is because components 3-5 are pure noise, thus the algorithm is not able to find any optimal combination of them. The only rotation of components 1, 2 is meaningful.

You see the kurtosis of the new components and their loadings. The kurtosis of two first components is pretty large (although negative for component 1 and



positive for component 2) while it is close to zero for components 3-5. That is because they are noise. There is no need to rotate it. You can make the rotation faster and more economical.

- Click again **ICA**. Now unclick components 3-5 in the list of PCA components. Only components 1, 2 will be rotated. Press **OK**.

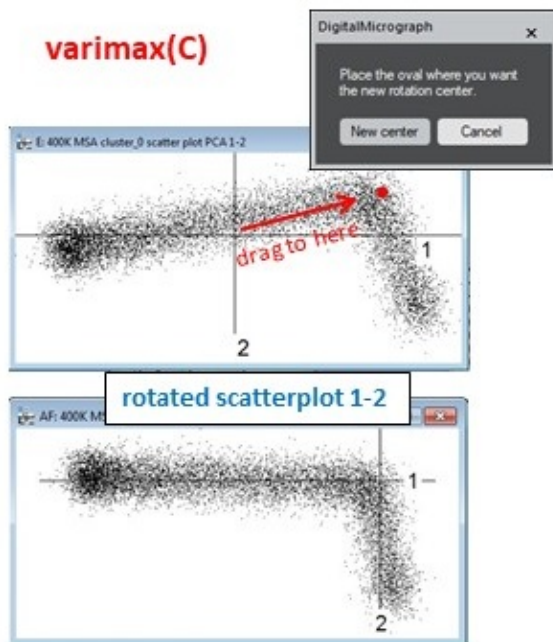
After the rotation is finished you see also the 1-2 scatterplot, which indicates that the major variation trend is almost along with the loading of ICA 1 component. This is the best job ICA can deliver for this set. In fact, ICA results do not differ very much from the PCA ones for this example. In other datasets, this might be very different.

Another rotation method - Varimax - attempts to obtain the best simplicity of data. That means, the scores must have preferentially the large coordinates along one axis while zero or small coordinates along the other axes. This is achieved by maximization of the squared variances of all scores. Varimax was designed for *uncentered* PCA decomposition while **temDM MSA** uses the *centered* treatment as default. To adapt for that, the slightly modified variant of Varimax is implemented in **temDM MSA**, namely Varimax with the

flexible definition of the rotation center. The package offers you the possibility to redefine it interactively.

- Click **varimax**. If the option **recenter** is chosen, the scatterplot appears with the red spot denoting the rotation center. Drag the spot to the desired place in the 1-2 component plane. Press **New center**.
- Then you get the list of PCA components to be rotated. You may uncheck any of the components and they will be fixed, i.e. not rotated. As in the case of the ICA rotation, it is worth to fix components 3-5. Press **OK**.

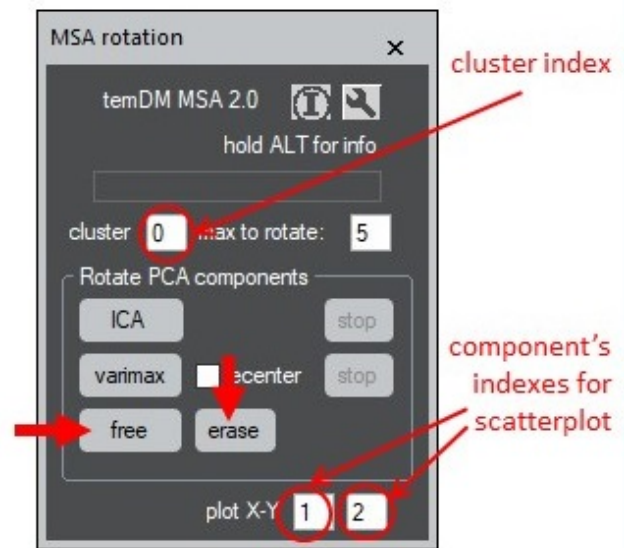
New component loadings and a new scatter plot appear. Now, the orientation of the trends relative the principal axis is much more logical and the loadings could be somehow interpreted.



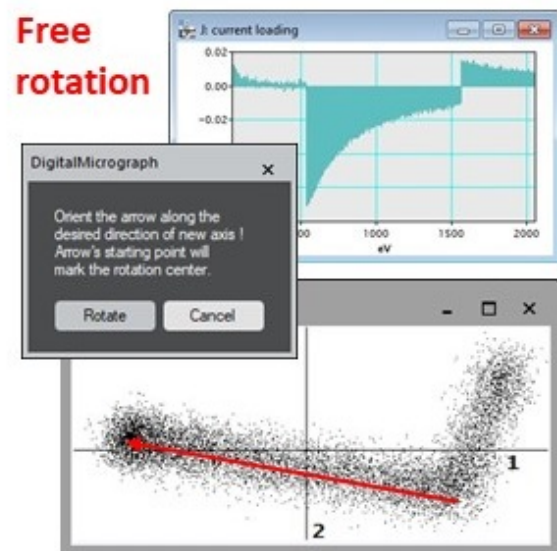
The trick with redefinition of the center can be naturally extended to the rotation fully defined by a user. This is called *free rotation* in the **temDM MSA** package.

- Click **free**. The scatter plot appears with the red arrow in the middle. The live spectrum profile shows the loading corresponding to the current direction of the arrow.
- Move around the red arrow on the scatter plot while watching the resulted loading spectrum. Orient the arrow such a way that its beginning denotes the rotation center while its end points to the desired direction for the best interpretable

loadings. Press **Rotate**.

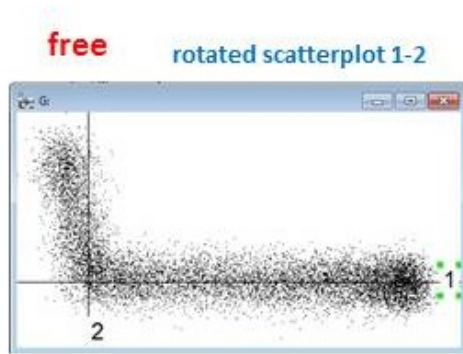


You get components 1 and 2 rotated as you specified. Of course, the rotation was proceed in the 1-2 plane only, while the rest components are unchanged. If you wish to rotate another axes, change the components indexes at the bottom of the **MSA rotation** tool. Free rotation affects only a couple of components at once. This is different from ICA and varimax that can rotate all available PCA components in one action.



This style of rotation is of course subjective, no quantitative criterion stays behind. However the real systems are often so complicated that any established rotation criteria fail. Still human brains have a chance to pick up a hidden structure in the data distribution. A flexible look-inside sometimes overperforms any fixed algorithm !

Note that you may use the free rotation tool to get an



idea about the correspondence between the given direction in the factor space and the shape of a spectrum *even if you are not going to rotate the components.*

- Click **free**. Move around the red arrow and learn what kind of spectrum appears in the live profile. After you have investigated all directions, press **Cancel**. The rotation will be NOT executed.

The free rotation may be proceed iteratively using the different couples of components. The results of the previous free rotation are not lost but suggested to you as the starting point for the next rotation operation. This is different from ICA and varimax which always start from the original PCA components.

If you feel your multiple free rotations has came to the configuration you don't understand anymore, you always can return to the original orientation of the PCA components.

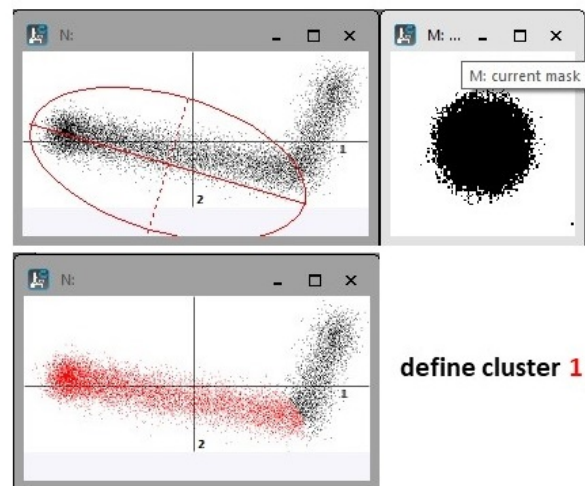
- Just click **erase**. The next free rotation will start from the original PCA components.



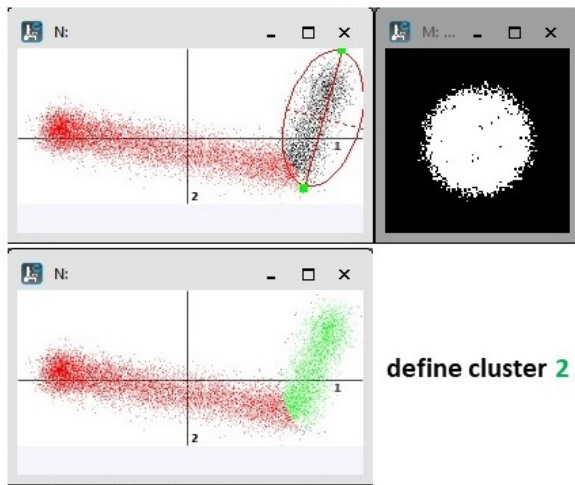
## 9 CLUSTERING

The example **EELScube MSA.dm3** shows the two trends of variations, more or less independent of each other. This is a typical feature of spectrum-imaging datasets; the real systems might show even more complicated structures. One way to deal with the complexity of datasets is to break it on several clusters in the factor space.

- Open the **clustering** tool by clicking **temDM - MSA: factor space - Clustering**.
- Having **EELScube MSA.dm3** in front, press **design clusters**. The default scatter plot of the two first PCA components will be generated.





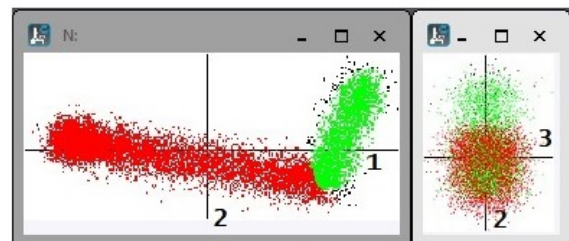


- Having the scatter plot in front, click **new** and notice an ellipse with the movable principal axes. This ellipse captures the cluster data points in the factor space. Also a live binary image (mask) appears that tells you how your cluster looks like in the real space.
- Orient the ellipse in the way it covers the points to be included in the cluster. You can drag the main axis of the ellipse and tune its proportion by stretching/expanding the minor axis. Click **confirm** when ready. You may also exit the process without storing the cluster by clicking **cancel**.
- Click again **new cluster** and manage the second cluster. Click **confirm**. If some pixels from the old cluster appear to lay within the new one, they will change their affiliation, i.e. will be attached to the new cluster.

Clusters are shown by points of different colors in a scatter plot. The default colors are - 1: red, 2: green, 3: blue, 4: pink, 5: yellow, 6: light-blue, For your convenience the pallet is always displayed on the tool. The colors of clusters with the index greater than 6 will be cyclically repeated as 7: red, 8: green and so on.

*One point should be stressed:* Although you designed your clusters in the 2D-plot, the actual structure is multi-dimensional. Thus, it is recommended to view it from the perspective of several scatter plots.

- Press **add plot X-Y**. A new scatter plot appears with the same clusters but seen from the other perspective. The default components for the second scatter plot are 3 vs 2 but that is something you can arbitrary change.



You may continue to design your clusters at any scatter plot you want. The changed clusters will be automatically repainted in the other scatterplots. Such multi-dimensional manipulation sounds challenging but after you get used to scatter plots, you will really have a fun exploring an n-dimensional space !

- You can anytime **delete** the last designed cluster. Then all its pixels will be marked as not belonging to any cluster and repainted in black.
- You can also correct any existing cluster by appending new data pixels. Press **append** and you will be prompted to choose the index or the color of the corrected cluster. Then the ellipse for choosing the new area will appear. Just drag it to the desired position and press **confirm**.

*Important:* the old pixels are not excluded from the corrected cluster but are *added* to those located in the new ellipse. This way you can design clusters of a very complicated geometry.

You can anytime visualize the clusters designed so far in the conventional (not factor) space.

- Press **maps** button at the bottom of the **clustering** tool. The binary masks and the average spectra for all existed clusters will be generated.



If you are happy with your job, you should store it.

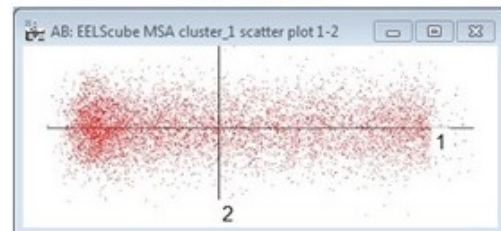
- Press the **store** button. Clustering is stored in the **EELScube MSA.dm3** container. This is however still not finally fixed because **EELScube MSA.dm3** is still in the operative memory only. Do not forget to store **EELScube MSA.dm3** on the disc at the end of your session.
- Alternatively, you can click **close** button. Then all your play with clustering will be erased without saving. Also you can do that after you stored your clusters and do not want the scatter plots to be displayed anymore.
- Later you can anytime load your previous clusters and continue the clustering job by clicking the **load old** button

You have now two clusters stored in **EELScube MSA.dm3**. You should proceed PCA again, this time in each cluster individually. To do that

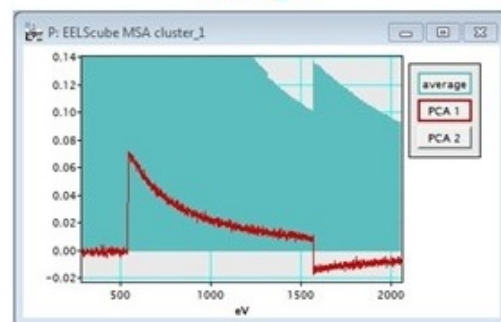
- Open the **proceed PCA** tool by choosing **temDM - MSA :PCA - Proceed**.
- Input **2** (it is sufficient) in **number components**. Click **PCA**. The package recognizes that data are clustered and ask you whether you wish to treat clustered.



scatterplot

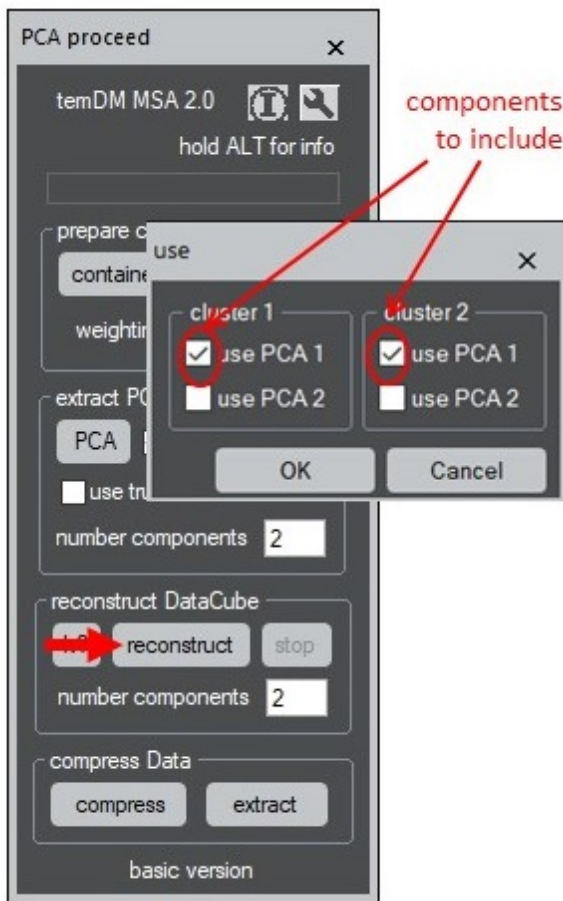


loadings



- Choose **treat all clusters** and press **OK**. Clusters will be treated non-stop one after the other.

Do you see the difference ? Each cluster is now essentially a one-componential object. It shows a simple scatterplot and a very transparent loading. As previously, close all displayed loadings, screeplots and



scatterplots without saving. Do not create mess in your DigitalMicrograph! All these results are kept in **EELScube MSA.dm3**.

So, clustering might be used to break the data onto the elementary pieces of simply interpretable structures.

Often, users are interested in one particular part of an investigated dataset. In that case you might design the only one cluster and then treat cluster 1 while the rest of the data will be ignored.

Finally, you can make reconstruction using the clustered data.

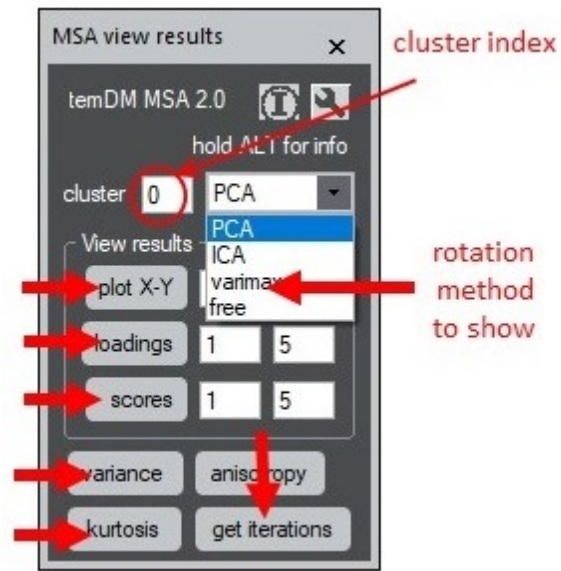
- Click **Reconstruct**. Confirm that you are going to use clusters. A list of available clusters will appear.
- Choose the desired number of components in each cluster. One component in each cluster is sufficient for **EELScube**. Press **OK**.

The reconstructed cube combines the pixels from all chosen clusters. The pixels which do not belong to any chosen clusters will miss the reconstruction and retain

their original noisy values.

## 10 VIEW ROTATED AND CLUSTERED RESULTS

The **view result** tool can be used to display loadings, scores and scatterplots for your rotated or clustered data.



- Open the **view results** tool by clicking **temDM - MSA : PCA View Results**.
- Choose cluster index. Index **0** implies NO clustering.
- Choose the rotation method in the drop down menu.
- Use the appropriate buttons to view the desired stuff.

You can also check the **variance** or **kurtosis** of the obtained rotated components. **Get iteration** will show you the available information about how your rotated components were obtained. This is content-dependent. The free rotation has no iterations at all. The iteration in Varimax is done for all components at once.

We have included extra methods in our extended treatment - clustering and rotation. This means we are not anymore within Principal Component Analysis (PCA) but rather stepped into the more general area - Multivariate Statistical Analysis (MSA). As the next, we will explore another powerful MSA technique - endmembering.

## 11 ENDMEMBERS

You see that the MSA treatment can produce a lot of results to view and to analyze. Maybe too many...

Review of such output implies strong abstract thinking: loadings are differential, i.e. counted from some average spectra; scores may be both positive and negative, et cet. et cet.

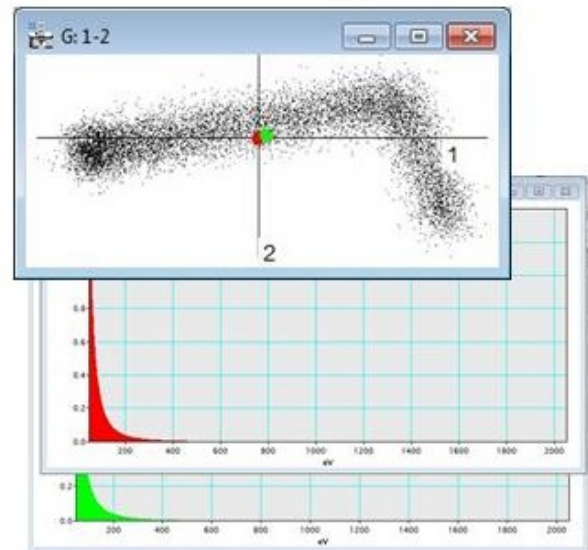
Can we come to some simpler, intuitively clear description of the results ? **Yes.**



Try to recast it in terms of *endmembers*. In such description you deal with few well interpretable spectra and few maps showing the contribution of these spectra across the object. Use again the example **EELScube MSA** where you have extracted at least 3 PCA components.

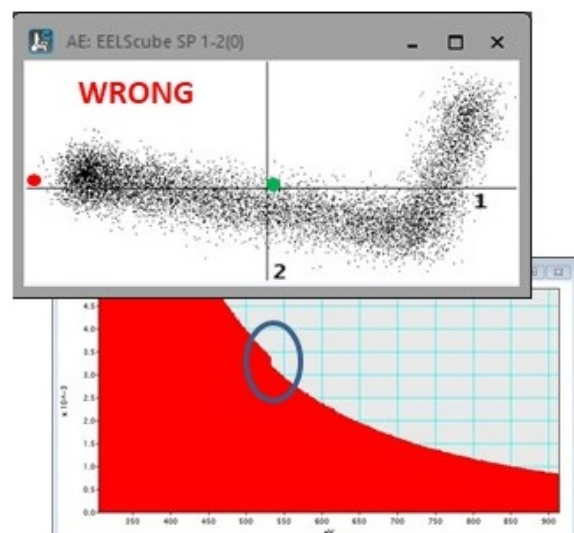
- Open the **end members** tool from **temDM - MSA : factor space - EndMembers** menu of DigitalMicrograph.
- Click the **new** button while having **EELScube MSA.dm3** or **EELScube COMP.dm3** in front.

A couple of new windows will appear. Find the 1-2 scatter plot. The red and green markers are located in the center of the scatter plot. They denote the positions of two endmembers spectra in the factor space while two other windows display the actual spectra. Try to drag the endmembers markers within the scatter plot and see how the spectra change.

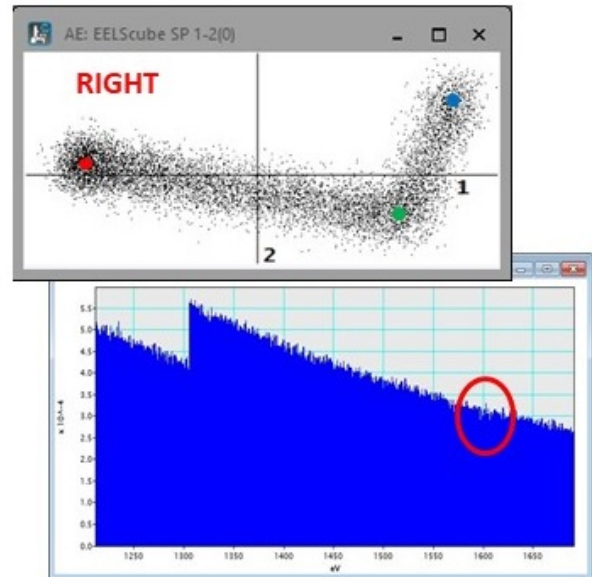
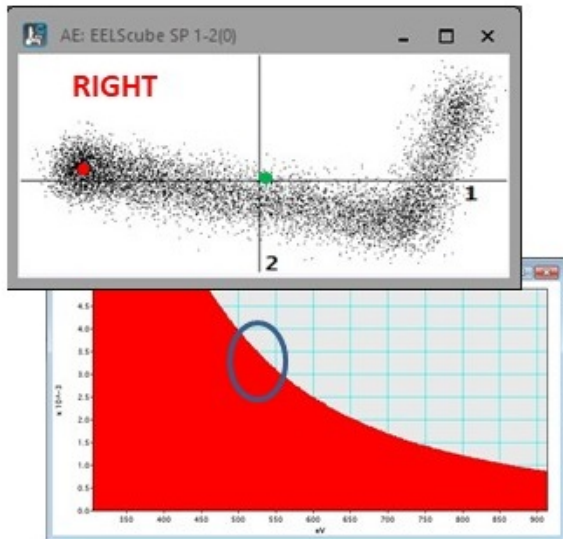


For your convenience, you may move the displayed spectra somewhere to the side of the DigitalMicrograph working space, zoom the intensity scale and the range of the displayed energy channels using the usual tricks of DigitalMicrograph (shift to left or right by dragging the scale bar and zooming the scale with the CNTR-key down).

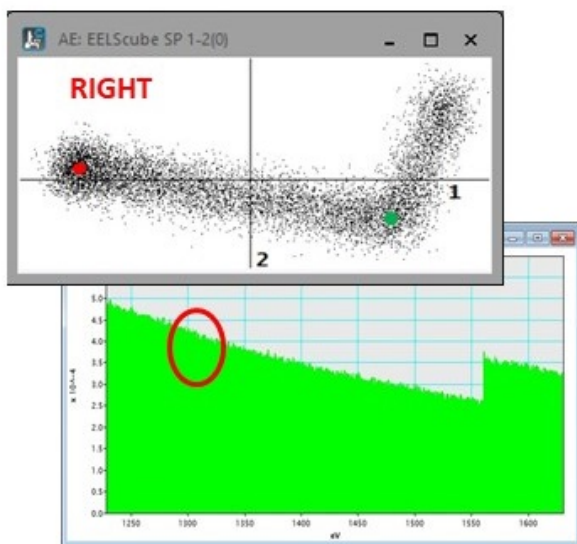
From the very name of the method you might guess that the reasonable endmembers positions are usually found at the periphery of your scatter plot. How to arrive at the right positions ?



By playing around you might discover some kind of a “saddle point” where the spectrum looks particularly proper. The deviation from that point results in the unphysical profile or too complicated spectrum. In the given example, place the red marker such as the kink at 530eV *disappears*.



Similarly, move the green marker a bit up-down and find the position where the peak at 1310eV is *not* observed at the corresponding spectrum.



Have you noticed that the significant portion of the scatter plot remains unattended? This is because the object requires at least 3 endmembers for the adequate description.

- Click **add member**. The third (blue) endmember marker will appear with the corresponding spectrum.

As previously, try to find an optimal “saddle point” for its position. In this case it would be the condition of the complete disappearance of the peak at 1530eV.

The markers and corresponding spectra are differentiated by color. The color sequence is displayed on the tool as a pallet of 6 colors. Starting endmember 7,

the colors will be cyclically repeated.

A number of images might create a mess in the DigitalMicrograph working space. A small tip: to operate more conveniently keep the endmembers spectra always in the right order:

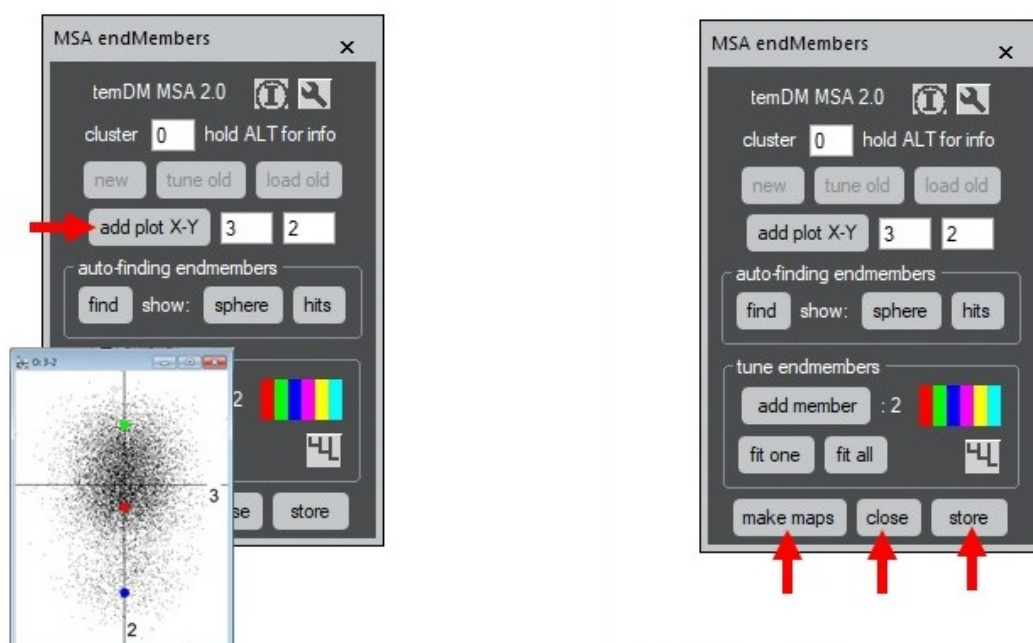
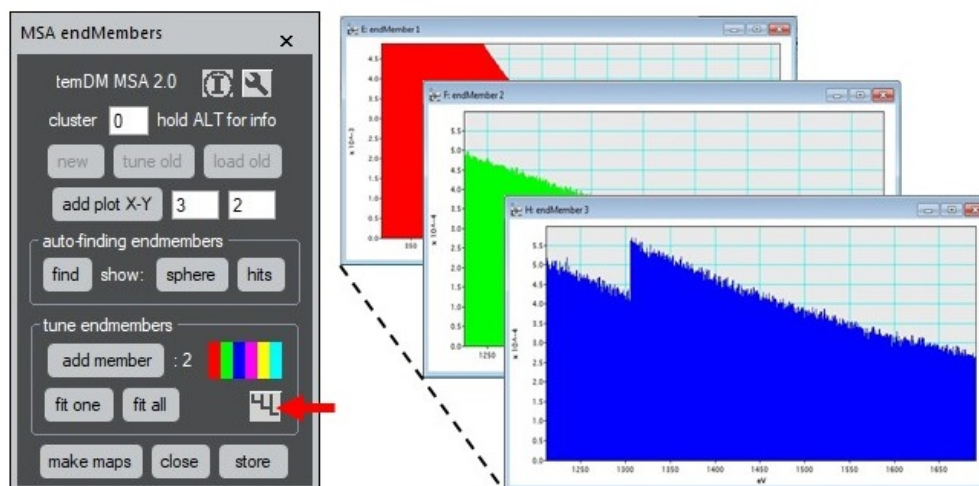
- Drag the the spectra of the first and the last endmembers to the most comfortable location for their inspection.
- Click the **graphical** button just below the color pallet. All intermediate member spectra will be ordered in between the first and the last ones in the cascade manner.

Up to now you explored the only one scatter plot (1st vs 2nd component). It is OK for this particular example but usually insufficient for real-life objects. You must look at your dataset from the perspective of *several* scatter plots.

- Click **plot X-Y**. The fields at the right denote the desired components to be plotted. A new scatter plot will appear while your tuned endmembers are already there.

Now you can have fun dragging the endmember markers at different scatter plots. They will move synchronously in both displays. Not bad exercise for n-dimensional thinking, is not it?

The 3-2 scatter plot is not very useful for the given example. The theoretically reasonable endmembers should all sit at the line coinciding with the 2nd PCA component axis while the contribution along the 3rd



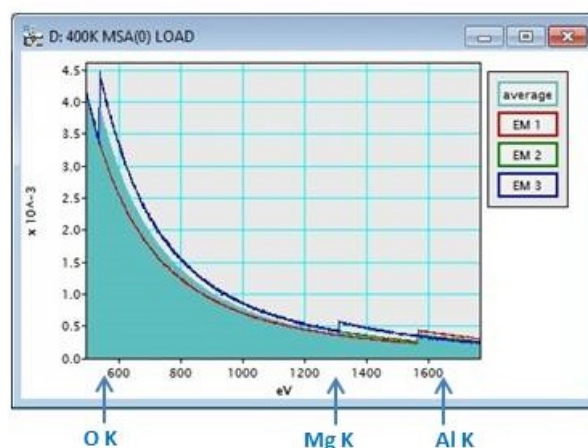
one must be zero.

Finally, plot the maps showing the fractions of all endmembers spectra in the dataset.

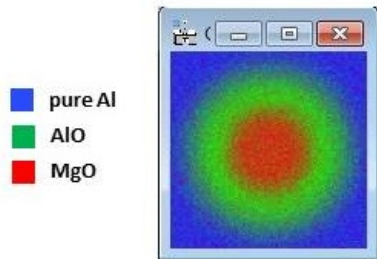
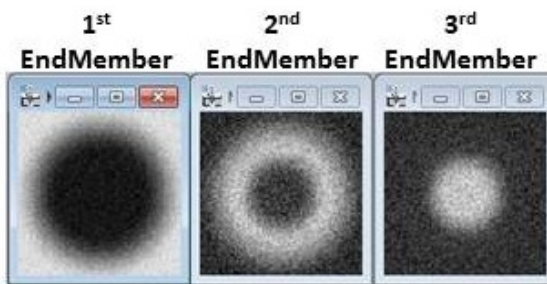
- Click the **maps** button. Three maps each corresponding the specific endmember will appear. The overlaid endmember spectra is displayed in the separate window.

You probably guess that the 1st endmember spectrum corresponds to pure Al, the 2nd one - to AlO and the 3rd one to MgO EELS spectra. The maps show nicely the distribution of these compounds. You might combine them into one colored map by using the standard DigitalMicrograph tool - **Color Mix**.

How to store your endmembering job ?



- Press the **store** button. Endmembers are stored in the **EELScube MSA.dm3** container.
- Then you can close all windows at once. Click **close** button and confirm that the endmembers are stored.



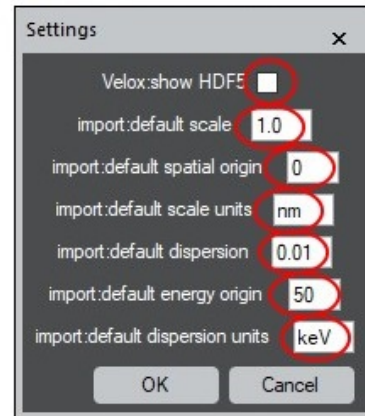
- Ensure that the endmembers are really in the container. Click **tune old** having **EELScube MSA.dm3** front most and the stored endmembers will be again displayed. You may continue to tune them.
- Alternatively you might wish to open only the endmembers maps and their average spectra. Click **load old** and you get it. So, all your work is indeed inside the **EELScube MSA.dm3** container. The container is however still in the operative memory only. Do not forget to store **EELScube MSA.dm3** on the disc at the end of your session.

The basic version of **temDM MSA** operates with the manual tune of endmembers. In many cases this is sufficient, however when the number of the dimensions in the meaningful PCA space exceeds 3, such manual adjustment becomes tricky. The automatic tune of the endmembers positions related with the buttons **find**, **fit one** and **fit all** is addressed in the advanced version of the program.

## 12 SETTINGS

Some tools have a **spanner** icon giving you access to the processing and displaying parameters. Here is a short description of tunable parameters. Note that different tools may have parallel access to the same parameter.

### „import“ tool



#### import tool:

**show HDF5:** The tag structure of the Velox HDF5 format may be displayed during import. This helps to retrieve the standard names for data pieces if they differ from the default ones.

**default scale:** Default scale for spatial calibration.

**default spatial origin:** Default origin pixel in spatial calibration.

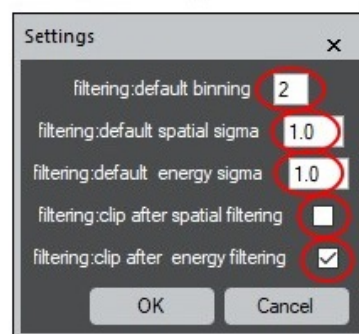
**default scale unit:** Default units for spatial calibration.

**default dispersion:** Default dispersion for energy calibration.

**default energy origin:** Default origin channel in energy calibration.

**default dispersion unit:** Default units for energy calibration.

### „filtering“ tool



#### filtering tool:

**default binning:** Default binning value.

**default spatial sigma:** Default standard deviation  $\sigma$  for spatial Gaussian smoothing (expressed in pixels).

**default energy sigma:** Default standard deviation  $\sigma$  for Gaussian smoothing along the energy axis (expressed in energy channels).

**clip after spatial filtering:** Filtering is not perfect at the border of spectrum image because there are less

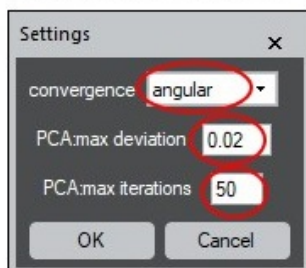
neighbors to average. You may choose to clip the border pixel by checking this box:

**clip after energy filtering:** The same for energy filtering - you may clip the border energy channels

loads (too many slices might make the display hard to handle).

**scores cubed:** Scores can be displayed as many individual images or as one datacube.

### „proceed PCA“ tool



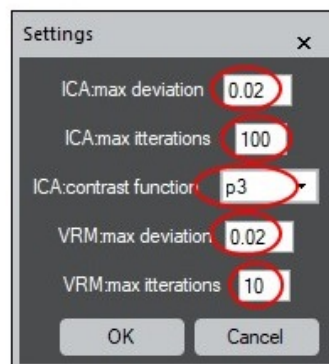
### proceed PCA tool:

**PCA convergence:** Convergence of the NIPALS algorithm for PCA can be estimated either from the angular difference between two subsequently found loading vectors (loading spectra are vectors in the n-dimensional energy space) or from the difference in their eigenvalues.

**PCA max deviation:** Criterion for stopping iterations in the NIPALS algorithm - maximal allowed difference between two subsequent iterations. Here: 0.02 rad (if angular convergence is chosen) or 0.02% (in the case of value convergence).

**PCA max iterations:** Maximal number of iterations in the NIPALS algorithm of PCA. If it is reached, the procedure will be stopped even if the stopping criterion is still not satisfied. The warning message “Precision ?” will be printed out in the results window. This usually happens when exploring the pure noise components.

### „rotation“ tool



### rotation tool:

**ICA: max deviation:** ICA convergence criterion: maximal angular difference between two subsequently iterated loadings, here: 0.02 rad.

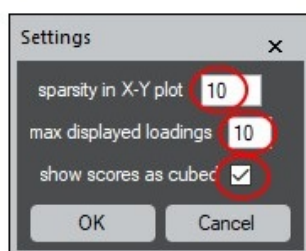
**ICA: max iterations** Maximal number of iterations in the fast ICA algorithm.

**ICA: contrast function:** fast ICA algorithm requires a certain “contrast” function for iterations. **temDM MSA** can use cubical polynomial (**p3**), hyperbolic-tangential (**tanh**) or exponential (**exp**) contrast functions.

**VRM: max deviation:** Varimax convergence criterion: maximal angular difference between two subsequently iterated loadings, here: 0.02 rad

**VRM: max iterations:** Maximal number of iterations in the Varimax algorithm.

### „view results“ tool

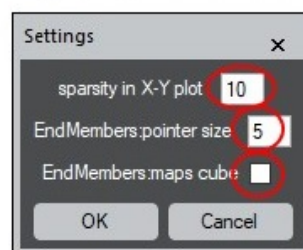


### view results tool:

**sparsity in X-Y plot:** Sparsity parameter defines a mesh to fill in an X-Y scatter plot. In the too sparse plot, tiny black pixels might be hardly visible. The too dense plot looks rough. Here the number of the empty (white) pixels will be 10 times the number of the occupied ones.

**max displayed loadings:** Maximal number of displayed

### „end members“ tool



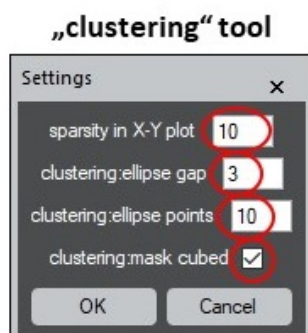
### end members tool:

**sparsity in X-Y plot:** Sparsity parameter defines a mesh to fill in an X-Y scatter plot. In the too sparse plot, tiny black pixels might be hardly visible. The too dense plot looks rough. Here the number of the empty (white) pixels will be 10 times the number of the occupied ones.



**endmember pointer size:** The radius of the marker pointing the position of loading in the X-Y scatter plot (expressed in pixels).

**maps cubed:** Maps can be displayed as many individual images or as one datacube.



### clustering tool:

**sparsity in X-Y plot:** Sparsity parameter defines a mesh to fill in an X-Y scatter plot. In the too sparse plot, tiny black pixels might be hardly visible. The too dense plot looks rough. Here the number of the empty (white) pixels will be 10 times the number of the occupied ones.

**ellipse points:** Ellipse for capturing clusters may be plotted with different number of points (here 10 points). The larger number of points makes the shape smoother but works slower.

**ellipse gap:** To move easier the ellipse its axes extend few pixels outside the ellipse itself (here 3 pixels outside).

**mask cubed:** Cluster masks can be displayed as many individual images or as one datacube

## 13 EXPORT TO PLOTTING PROGRAMS

An intrinsic DigitalMicrograph display of spectra and scatter plots is good for quick checking data but looks not nice in publications. Using scripting you can convert your PCA results into an ASCII text file that can be then imported by Origin, for instance:

### Loadings

Having the **EELScube MSA.dm3** image front most, run the script command:

```
getFrontImage().MSA112_getLoadings().
MSA112_SaveAsText(0)
```

### Scores

Having the **EELScube MSA.dm3** image front most, run the script command:

```
getFrontImage().MSA112_getScores().
```

```
MSA112_SaveAsText(0)
```

### Average spectrum

Having the **EELScube MSA.dm3** image front most, run the script command:

```
getFrontImage().MSA112_GetAverageSpectrum().
MSA112_SaveAsText(1)
```

## 14 TROUBLESHOOT

An error with the message “ambiguous function” is generated.

Most probably you have loaded **temDM MSA.gtk** twice. Localize all Plugins folders with script **find plugins folders.s** and check them. Only one **tem DM MSA.gtk** plugin must be there.

*You cannot load a **temDM MSA** tool from the menu. There is a failure message or the tool frame simply does not appear.*

Last time, the tool has finish functioning incorrectly. Open the tool again with holding the SHIFT key. The tool will be refreshed.

*You played too much with the **temDM MSA** settings. The program now fails or runs too long. You have no idea what have you changed.*

Choose **New Script** in the DigitalMicrograph menu. A new text window appears. Type there **MSA112\_delete()** and press **Execute**. The default settings will be restored.

## REFERENCES

- [1] P. Potapov. Why principal component analysis of STEM spectrum images results in abstract, uninterpretable loadings? *Ultramicroscopy*, 160:197–212, 2016.
- [2] P. Potapov, P. Longo, and E. Okunishi. Enhancement of noisy EDX HRSTEM spectrum-images by combination of filtering and PCA. *Micron*, 96:29–37, 2017.
- [3] P. Potapov. On the loss of information in PCA of spectrum-images. *Ultramicroscopy*, 182:191–194, 2017.
- [4] P. Potapov, P. Longo, and A. Lubk. A novel method for automatic determination of the number of meaningful components in the the PCA analysis of spectrum-

images. *Microsc. Microanal. Proceedings*, 24 S.1:572–573, 2018.

- [5] P. Potapov and A. Lubk. Optimal principal component analysis of STEM XEDS spectrum images. *Advanced Structural and Chemical Imaging*, 5:4, 2019.
- [6] P. Potapov and A. Lubk. Extraction of physically meaningful endmembers from STEM spectrum-images combining geometrical and statistical approaches. *Micron*, 145:103068, 2021.