This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 776804 — H2020-SC5-2017.

DELIVERABLE 4.13

Appendix 2

User manual -SOM Toolbox

Horizon 2020 Project: **NEXT**

Author(s): **Andreas Kempe, Peggy Hielscher**

Institution: **Beak Consultants GmbH**

Date: **27.10.2020**

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Self-organizing maps and k-means method

Self-organizing maps (SOM) is an unsupervised artificial neural network that projects a set of n-dimensional vectors to a usually two-dimensional SOM lattice (Kohonen, 2001). The usability of SOM comes from its topology preserving nature: similar data vectors are assigned to SOM cells that are close together.

K-means clustering is a very basic clustering method where each data point is assigned to the cluster that best represents the data point.

2 INSTALLATION

2.1 Installation requirements

SOM toolbox requires ArcMap 10.6.1.

(ArcGis 10.6 requires Windows operating system 7, 8 or 10. For further recommendations see http://desktop.arcgis.com/en/system-requirements/10.6/arcgis-desktop-system-requirements.htm)

3 USING THE SOFTWARE

3.1 Implementation in ArcMap

First open the ArcToolbox window in ArcMap. Add the SOM toolbox by right-clicking at 'ArcToolbox' and choose the SOM toolbox from your file location. The SOM toolbox appears in the list of tools. In the SOM toolbox you will find the script SOM clustering (fig. 1).

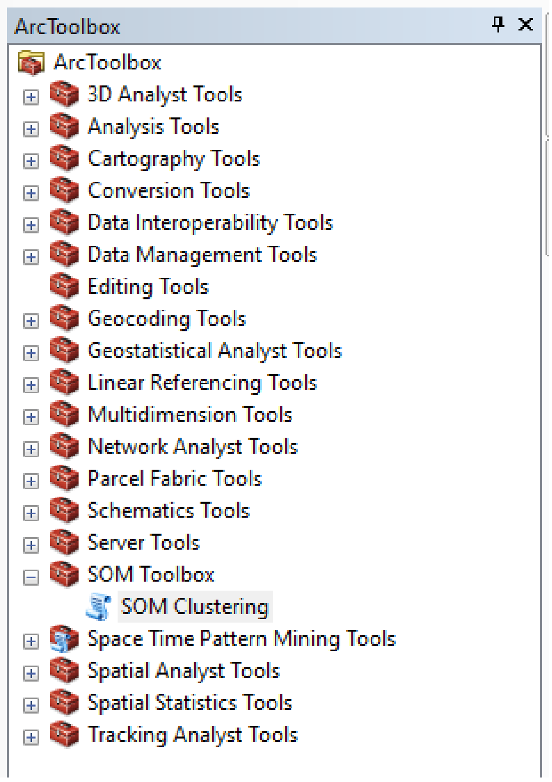

Figure 1. Arc Toolbox window.

3.2 Workflow

Help text can be shown or hidden by clicking the 'Show Help' / 'Hide Help' button on the bottom right of the toolbox (fig. 2). Every input line refers to an own explanation in the help window.

The appearance of the tool is easy and clear. After defining a workspace in the first input line, feeding the tool with input rasters in the next line, and defining parameters as for SOM, k-means, and others in the remaining lines the tool calculate results by clicking the 'Ok' button.

Input for SOM parameters are expected to complete, whereas all following parameters are set as default. Most relevant parameters as for SOM, map type, and grid shape appear directly on the GUI, and further parameters (initialization, neighborhood function, coefficient in the Gaussian neighborhood function, initial and final neighborhood, radiuscooling, initial and final training rate, scalecooling) can folded out by selecting 'advanced options' (fig. 3).

3.3 Defining Workspace

First define an empty workspace wherein all data will be processed. Defining workspace just means to create a new folder at a desired location. Therefore click on the 'New Folder' symbol and assign a name of your choice (fig. 4). In case of selecting an existing workspace containing already data from previous calculations, the toolbox will end with an error unable to overwrite present files. After creating your workspace click once to the folder so that the folder name appears in the line 'Name' and then click 'Add'. Do not double click to the workspace folder.

Another way is to create and assign a new folder in win explorer, copy the path and paste it into the workspace line of the SOM toolbox window.

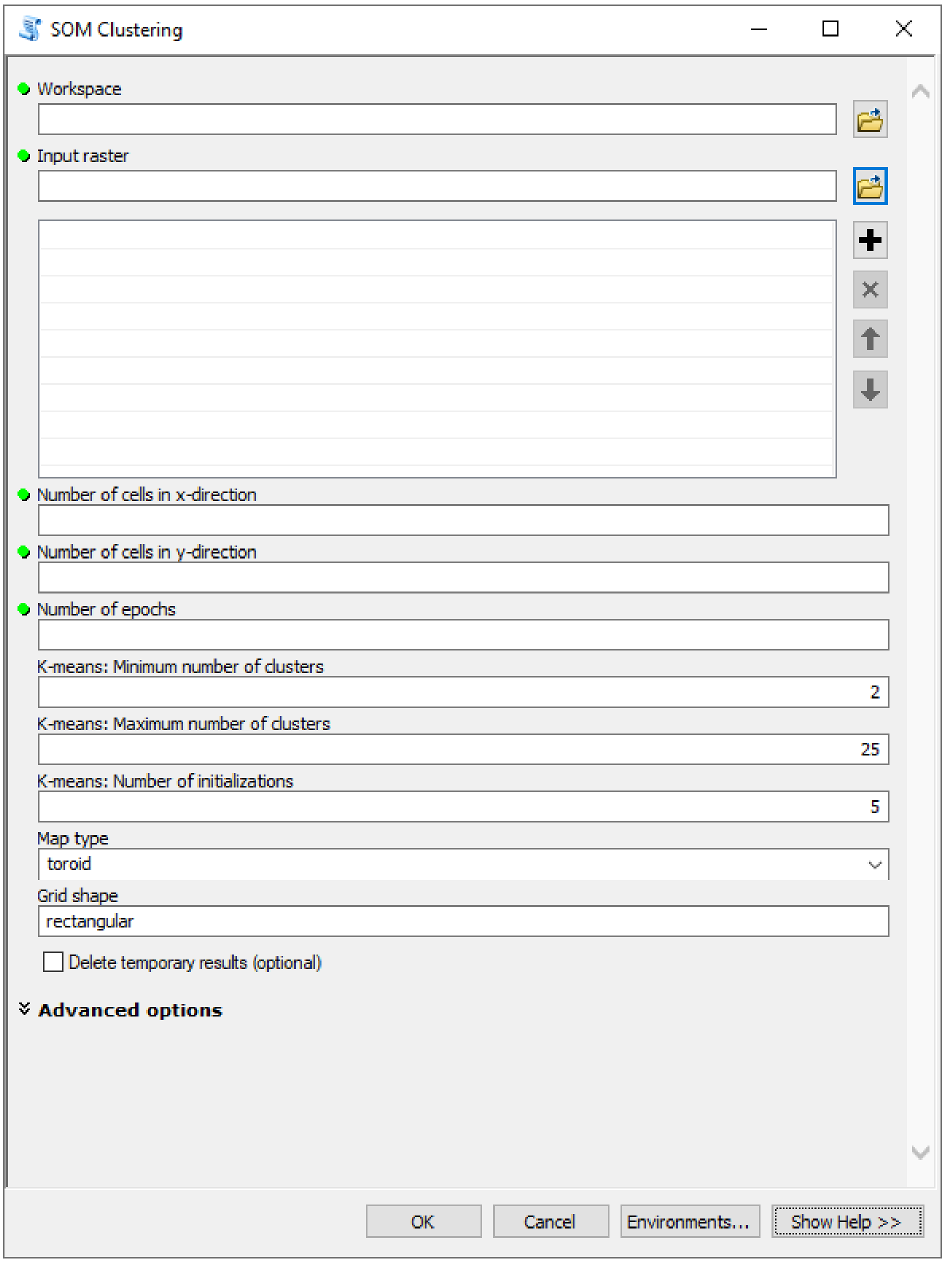

Figure 2. SOM toolbox window.

3.4 Loading input raster

Load input raster by clicking to the folder symbol at the right of the line and choose data from your data source. An input raster window opens where you can browse your file locations and select raster data (fig. 5). It is not possible to load the same input raster twice. At least two input raster has to be chosen for calculation (fig. 5). The supported file format is ESRI-Grid.

From input rasters with different extents only congruent regions were used for calculation. Furthermore, if one of many rasters contains NoData, the resulting raster exclude these regions from calculation and the final output returns equally NoData (fig. 6).

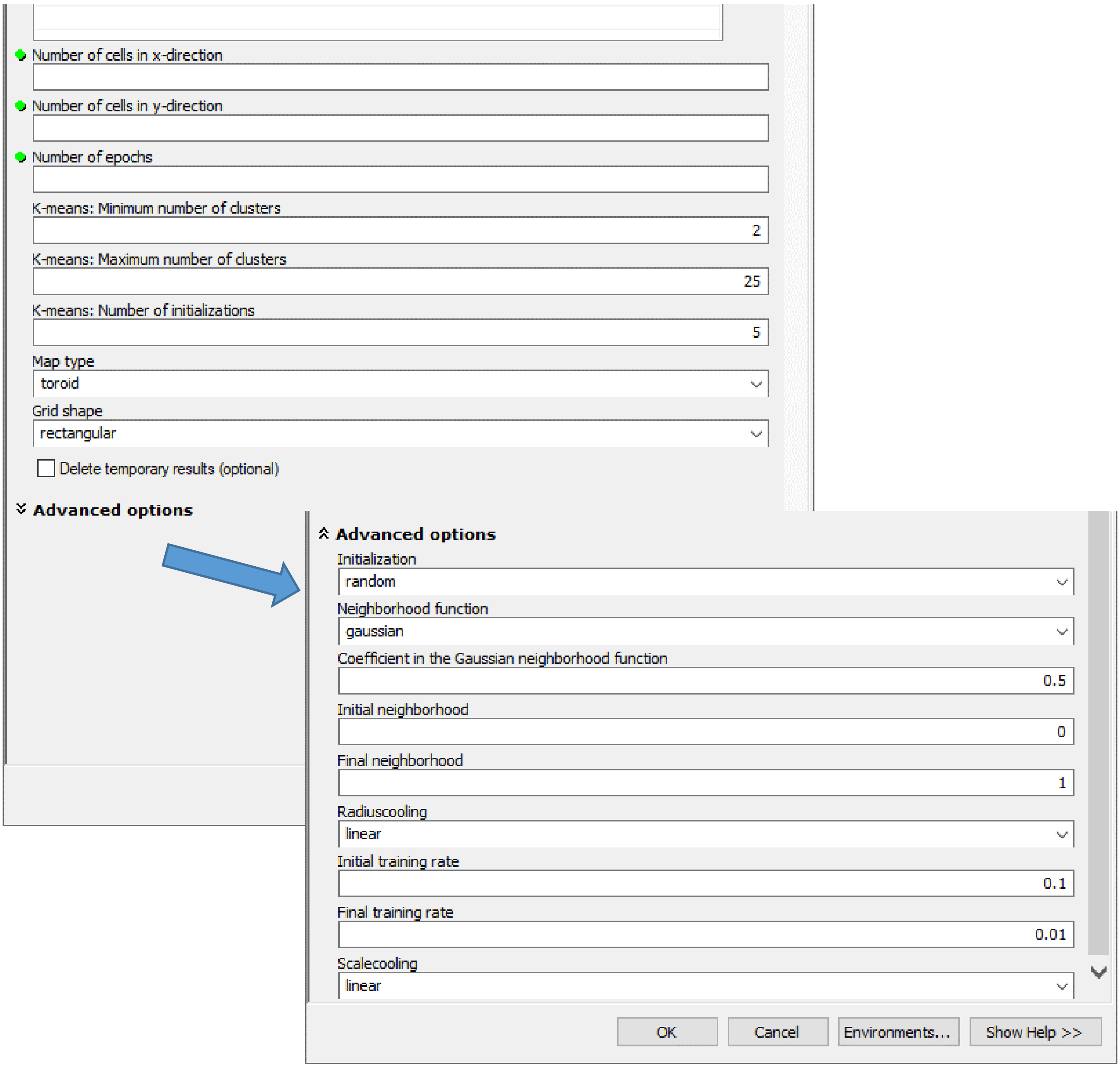

Figure 3. SOM toolbox window with advanced options folded out.

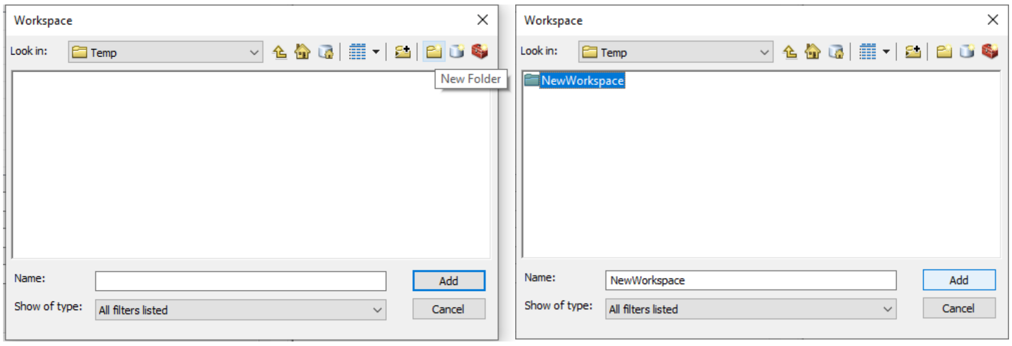

Figure 4. *Workspace window.*

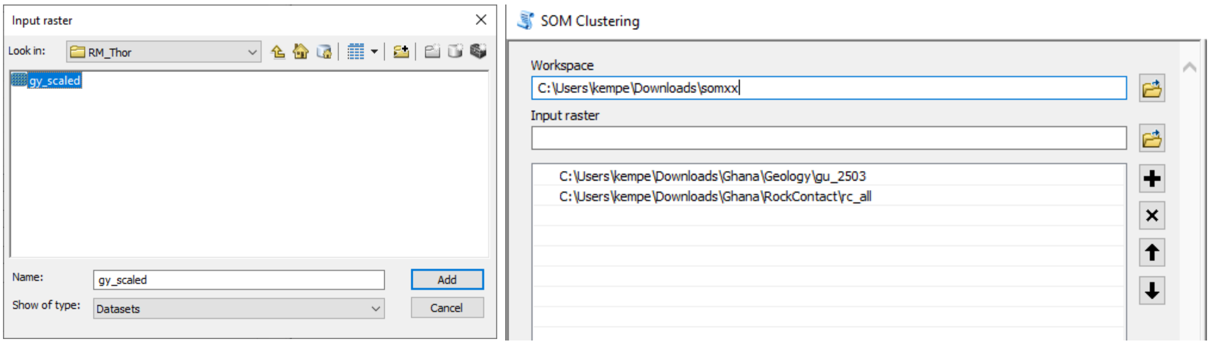

Figure 5. *Input window and exemplary filled input raster list.*

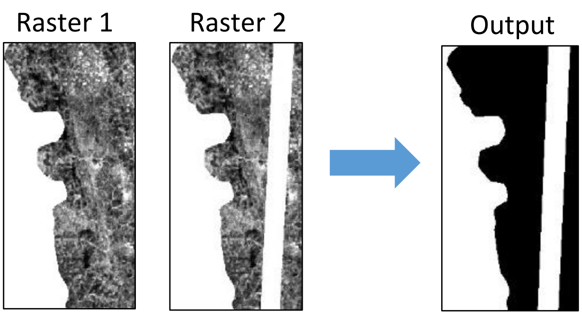

Figure 6. *NoData in one of the input rasters result in NoData in the output (calculated region in black).*

3.5 Parameters for SOM

Number of cells in x- and y-direction

In this step, you need to choose the parameters used in SOM clustering (fig. 7).

The size of the SOM is given by the number of SOM cells in x and y direction. Rule of thumb: the total number of cells is 5√n with n = number of data points. The larger the values, the more detailed the SOM, but it also takes more time to compute.

Number of epochs

'Number of epochs' is the number of times that the data set will be used when training the SOM. Possibly using 10 could be a good start. Small values result in faster computation, but possibly also in an inaccurate SOM. Larger values increase computation time, but might also improve the quality of the SOM. Usually the quality will not increase after certain amount of epochs.

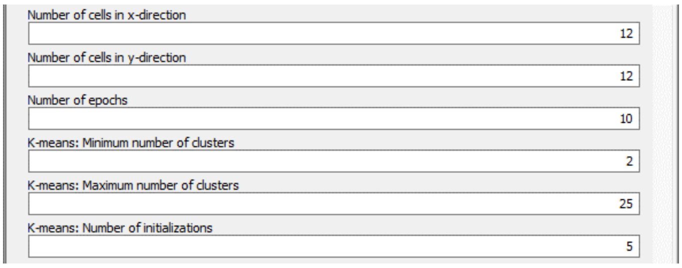

Figure 7. Parameters of SOM clustering and k-means.

All parameters of section 3.6 must not be zero or empty, which an error message warns against.

3.6 Parameters for k-means

K-means

In the next step, you can choose to apply k-means clustering in addition to SOM clustering by adding the tick in ‘k-means’ (fig. 7).

Number of clusters

If you checked k-means you need to select the minimum and maximum number of clusters. The default for minimum is 2 and maximum is 25. K-means requires a known number of clusters. This software applies k-means to the results of SOM using multiple values for the number of clusters and the best fit number is chosen based on the smallest Davies-Bouldin index.

Number of initial means

Choose the number of initial means. The default is 5. K-means utilizes random number generator in the algorithm and is sensitive to the initial means. Thus, this software runs k-means using different initial means and chooses the best fit based on the smallest Davies-Bouldin index.

All parameters of section 3.7 must not be zero or empty, which an error message warns against.

3.7 Further parameters

All parameters of section 3.8 and 3.9 has default settings and must not be zero or empty. Some parameters can be selected from drop-down list (fig. 8).

Map type

Parameter to specify the map topology. Planar (default) or toroid map.

Grid shape

Parameter to specify the grid shape of the nodes. Rectangular (default) or hexagonal shape.

3.8 Advanced options

Initialization

Parameter to specify the initialization. Random weights in the codebook (default) or pca, codebook is initialized from the first subspace spanned by the first two eigenvectors of the correlation matrix.

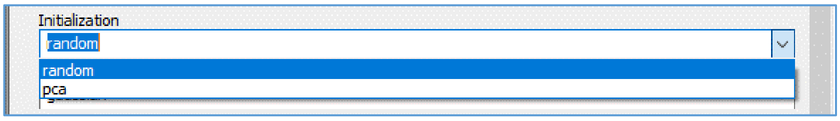

Figure 8. ***Parameter selection via drop-down list.***

Neighborhood function

Parameter to specify the neighborhood. Gaussian neighborhood (default) or bubble neighborhood function.

Coefficient in the Gaussian neighborhood function

Parameter to set the coefficient in the Gaussian neighborhood function. Default is 0.5.

Initial neighborhood

The initial radius on the map, where the update happens around a best matching unit. Default value of 0 will trigger a value of min(number of cells in x-direction, number of cells in y-direction)/2.

Final neighborhood

The radius on the map where the update happens around a best matching unit in the final epoch. Default value is 1.

Radiuscooling

The cooling strategy between initial and final neighborhood / radius. Linear interpolation (default) or exponential decay.

Initial training rate

The initial training rate. Default value is 0.1.

Final training rate

The training rate in the final epoch. Default value is 0.01.

Scalecooling

The cooling strategy between initial and final training rate. Linear interpolation (default) or exponential decay.

3.9 Deletion of temporary files

Temporary data can be deleted by enabling the intended checkbox before running the tool. These are SOM.lrn, SOM.xml, geospace.txt, somspace.txt and the mask.

3.10 Execution

A processing window appears after clicking the ok button (fig. 9).

```
Executing SOM Clustering...

Close this dialog when completed successfully

Executing: Script \\vs-daten\Projekte\2018\0051-0100\20180096_Praktikum_Softwaree
\\vs-daten\Projekte\2018\0051-0100\20180096_Praktikum_Softwareentwicklung\Andreas
\Projekte\2018\0051-0100\20180096_Praktikum_Softwareentwicklung\Andreas\NEXT\ArcG
Start Time: Wed Oct  2 13:06:22 2019
Running script Script...
Folder 'Temp' is created.
Folder 'GeoSpace' is created.
Folder 'SomSpace' is created.
File 'SOM.xml' is written.
Creating mask.
Processing 'SOM.lrn' and mask.
Processing 'SOM.xml' and 'nextsom_wrap.exe'.
Creating results. This can take a few minutes.
```

Figure 9. ***Detail of processing window.***

If the calculation completed successfully the processing window displays if all files were successfully created and additionally the processing time (fig. 10).

```
File 'Geo_cluster' successfully created.
File 'quant_error' successfully created.
File 'SOM_cluster' successfully created.
File 'umatrix' successfully created.
Completed script Script...
Succeeded at Wed Oct  2 11:47:40 2019 (Elapsed Time: 17 minutes 10 seconds)
```

Figure 10. ***Detail of processing window after completion.***

4 RESULTS

In the last step, results will be loaded automatically to the current data frame. These are divided between two group layers containing geospace results and SOM space results.

5 RESULTS IN GEOSPACE

Clustering results from k-means clustering or SOM clustering, original attributes and quantization errors in geospace are shown.

5.1 Results in SOM space

These images show the resulting SOM using the attributes (also known as codebook vectors) and k-means clusters.

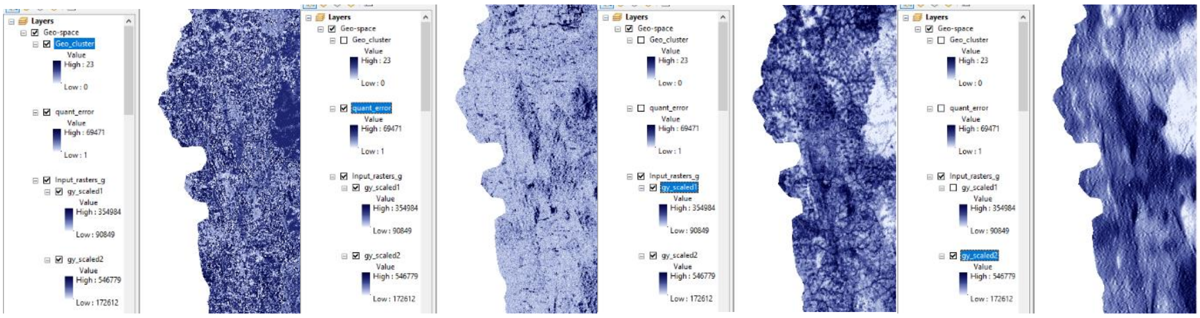

Figure 11. *Geospace results: geocluster, quantization error, raster 1, raster 2.*

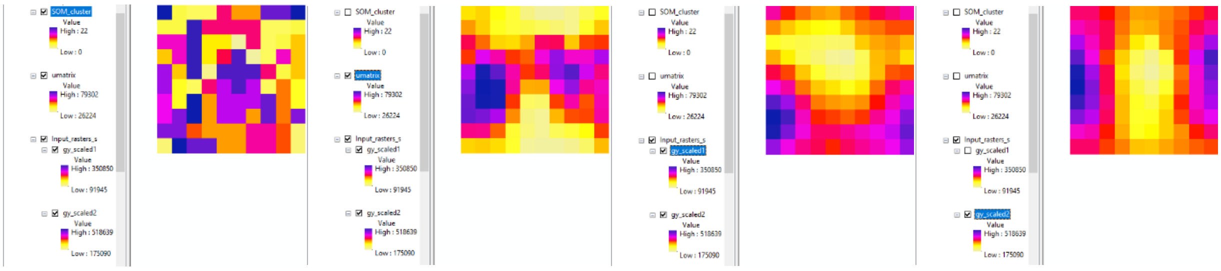

Figure 12. *SOM space results: SOM-cluster, U-matrix, raster 1, raster 2.*

6 REFERENCES

Kohonen T., 2001. Self-organizing maps, Third Extended Edition, *Springer Series in Information Sciences*, 30.