

XM - eXplainable Modeling

How To Guide



Alberto Castellini, Alessandro Farinelli, Francesco Masillo

Verona University



10 March 2019

1. How to install the software
2. How to launch the Model Visualizer
3. How to generate state models
4. Quick overview of folders and files



1. How to install the software



1. How to install the software

To use this software Python 3.X has to be installed.

Libraries:

```
sudo apt-get install libgsl0-dev
```

```
sudo apt-get install libblas-dev liblapack-dev
```

```
pip3 install --user pandas
```

```
pip3 install --user numpy
```

```
pip3 install --user pandastable
```

```
pip3 install --user pyitlib
```

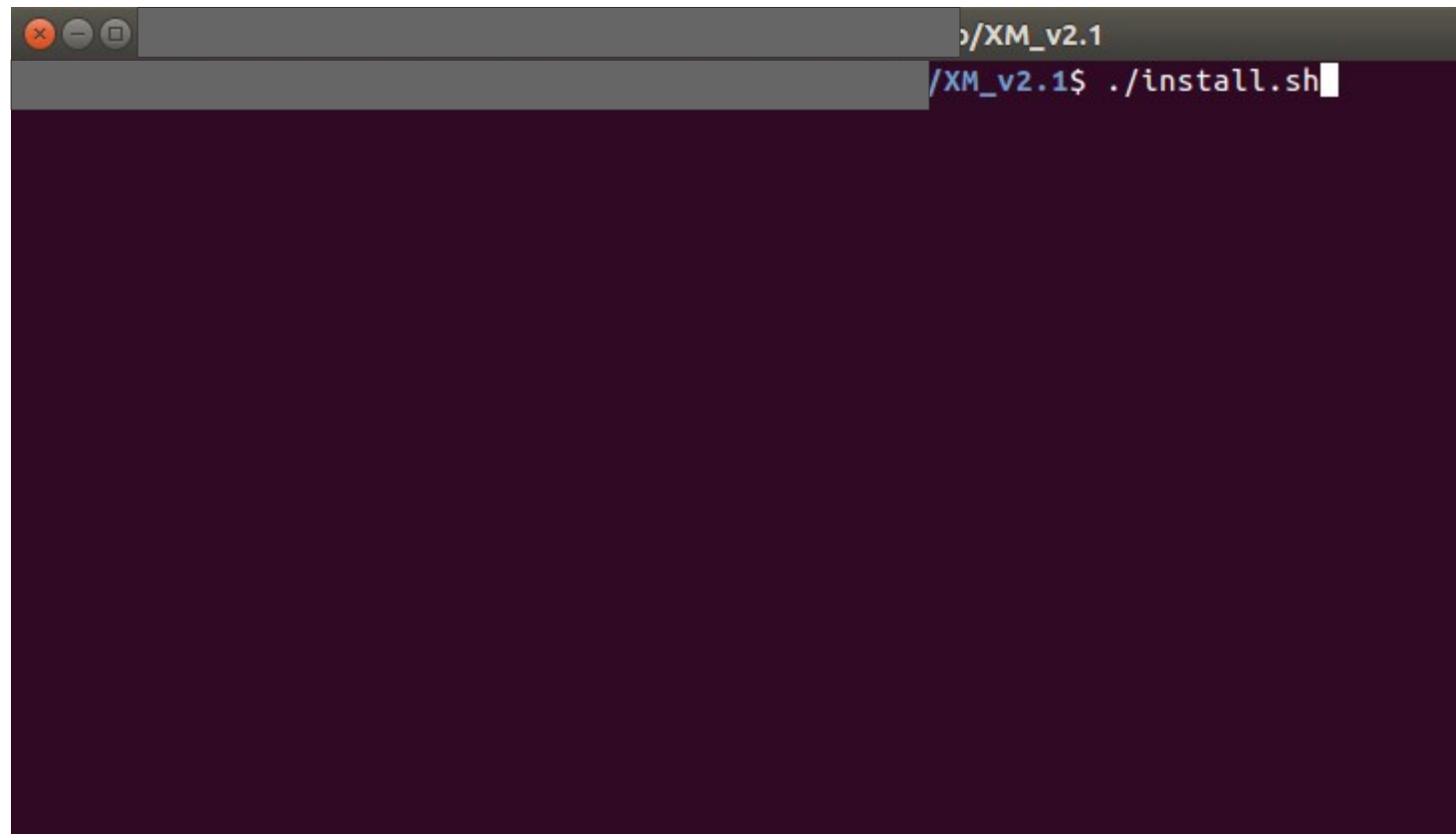
```
pip3 install --user matplotlib
```

```
pip3 install --user seaborn
```

```
pip3 install --user mplcursors
```

```
pip3 install --user pillow
```

Example:

A terminal window with a dark purple background. The title bar shows window control buttons and the text "/XM_v2.1". The prompt is "/XM_v2.1\$./install.sh" followed by a cursor. The rest of the terminal area is empty.

```
/XM_v2.1$ ./install.sh
```

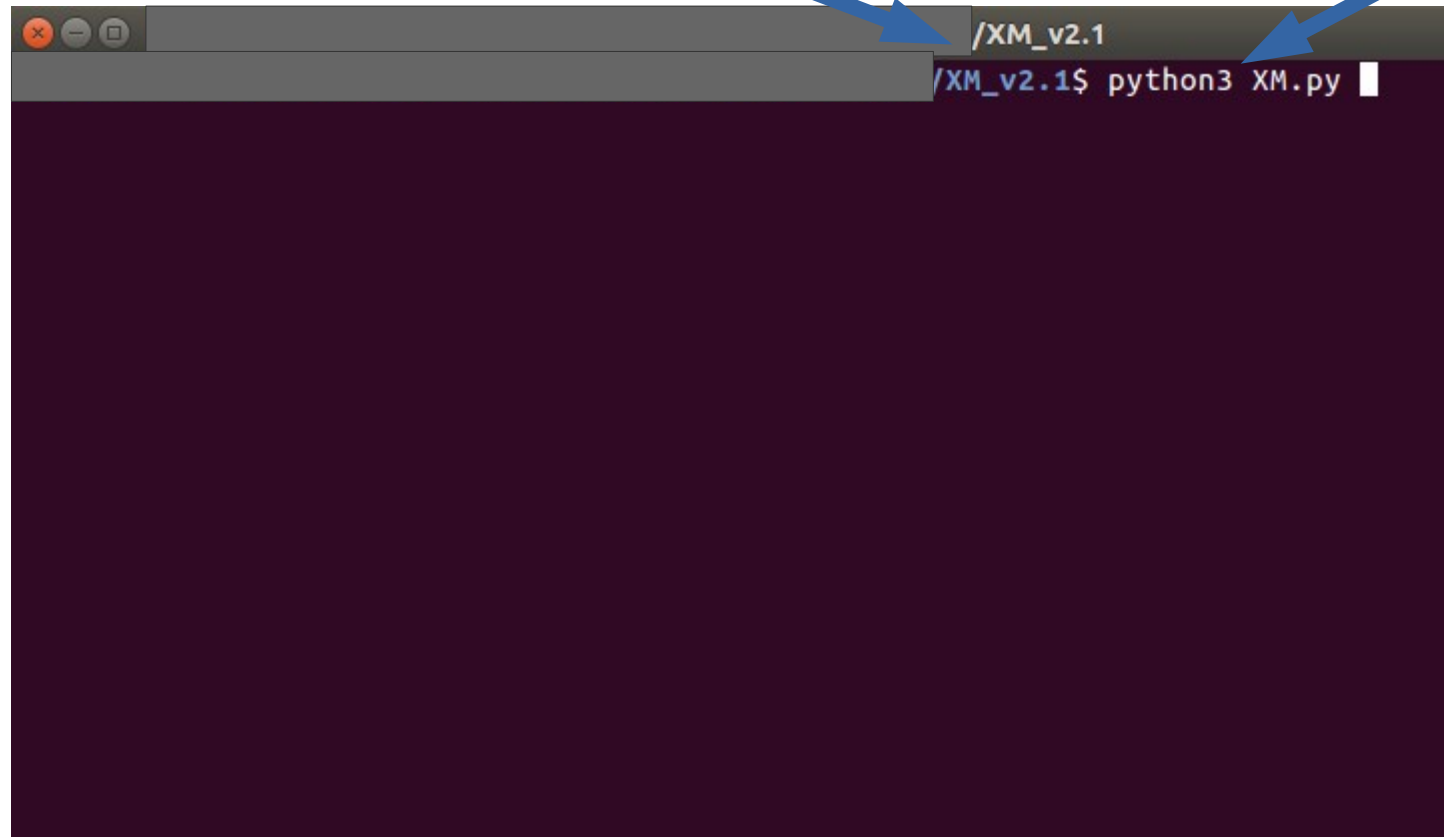


2. How to launch the Model Visualizer



How to launch the Model Visualizer

In order to launch this software, the user has to simply type the command “python3 XM.py” on a terminal opened in the sw folder.

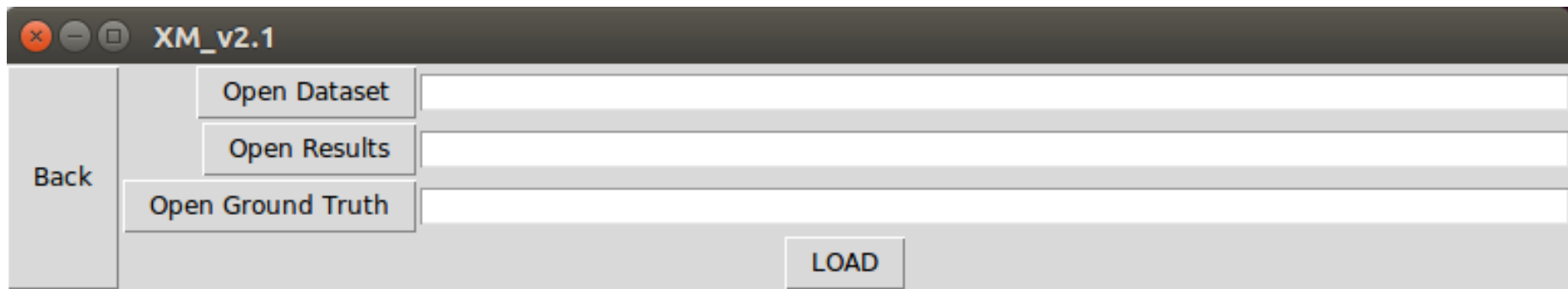


A terminal window with a dark purple background. The title bar at the top shows standard window controls (close, minimize, maximize) and the path `/XM_v2.1`. The terminal prompt is `/XM_v2.1$`, and the command `python3 XM.py` has been entered, with a white cursor at the end. Two blue arrows point from the text above to the terminal: one points to the title bar path, and the other points to the command text.

```
/XM_v2.1$ python3 XM.py
```

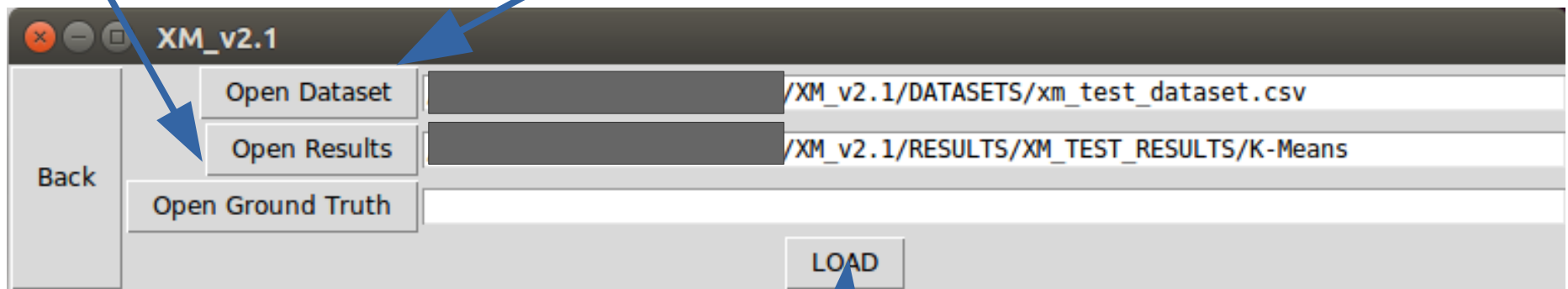
How to launch the Model Visualizer (cont'd)

A new window should appear in the middle of the screen.



Selection of dataset and modeling result

To select the dataset the user has to click on “Open Dataset”. Then a dialog will show up to allow the user to select the file. Same goes for the selection of the directory of the results. Here are shown those two paths already selected.



To finally visualize the results click on LOAD

General Info Tab

This is the first tab that will show up on a successful load.

It's called "General Info" and wraps up the list of clusters and related properties and general info about the dataset, clustering parameters and clustering performances.

Model: All Visualization Type: General Info Sorting: Silhouette Uncertainty Experiments: All GO

	Model	N. points	Silhouette	N. variables
1	4	9	0.595	11
2	2	101	0.457	11
3	3	105	0.444	11
4	1	12	0.267	11
5	5	74	0.203	11

	Parameter	Values
1	Dataset Path	/home/francesco/Tirocinio/XM_v2.1/DATASETS/x
2	Results Path	/home/francesco/Tirocinio/XM_v2.1/RESULTS
3	Number of columns	11
4	Number of rows	301
5	Method	K-Means
6	N. Cluster	5
7	Accuracy	NA
8	CE	NA
9	F1	NA
10	Entropy	NA
11	Purity	NA
12	NbClusterFound	5
13	Variables	['speed', 'voltage', 'm0_current', 'm1_current', 'ac
14	SSE	1.46e+03
15	Silhouette total	0.38677377651

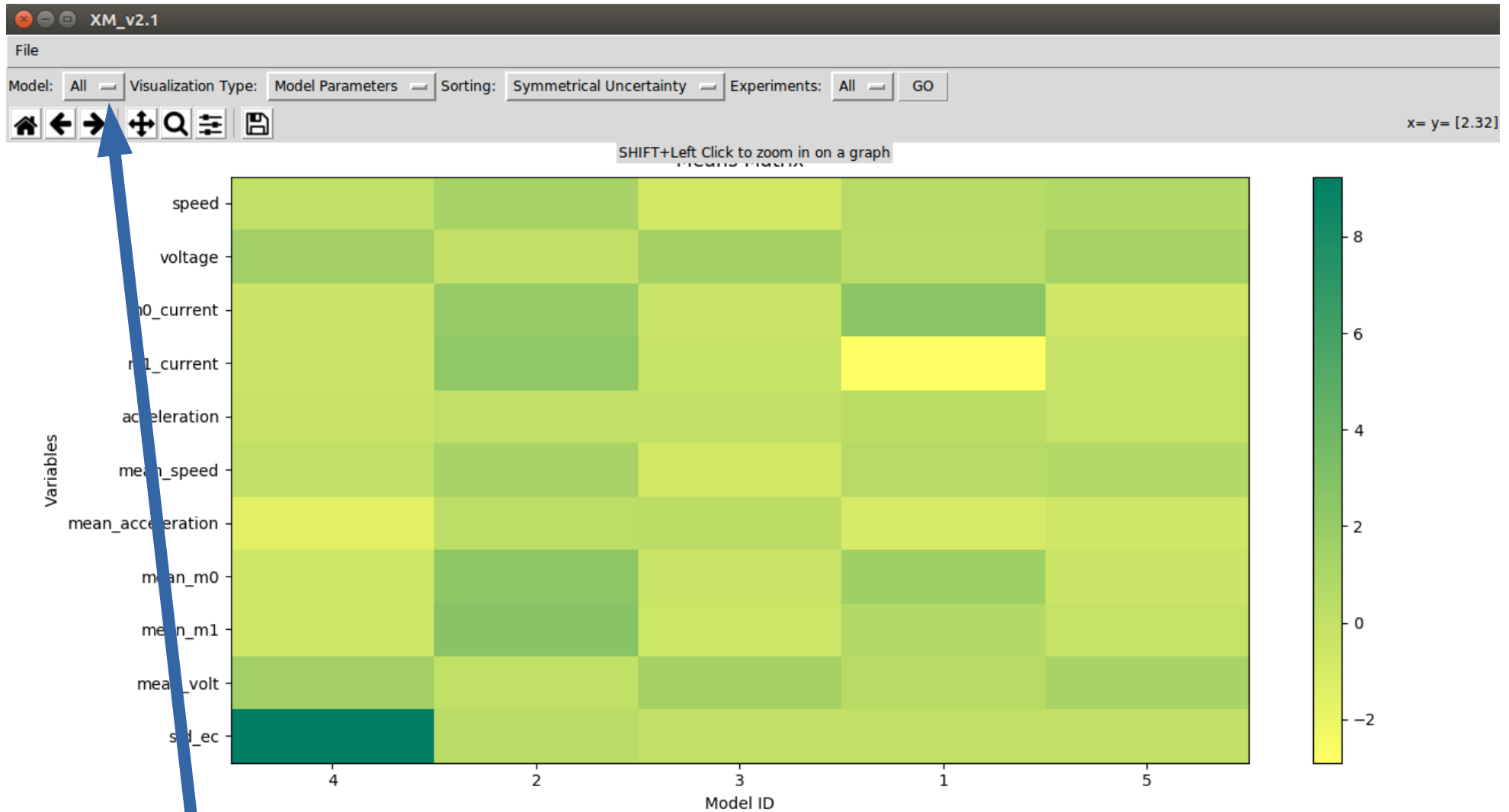
Select a type of visualization.

After selecting the type of visualization and/or cluster click GO



Model Parameters Tab

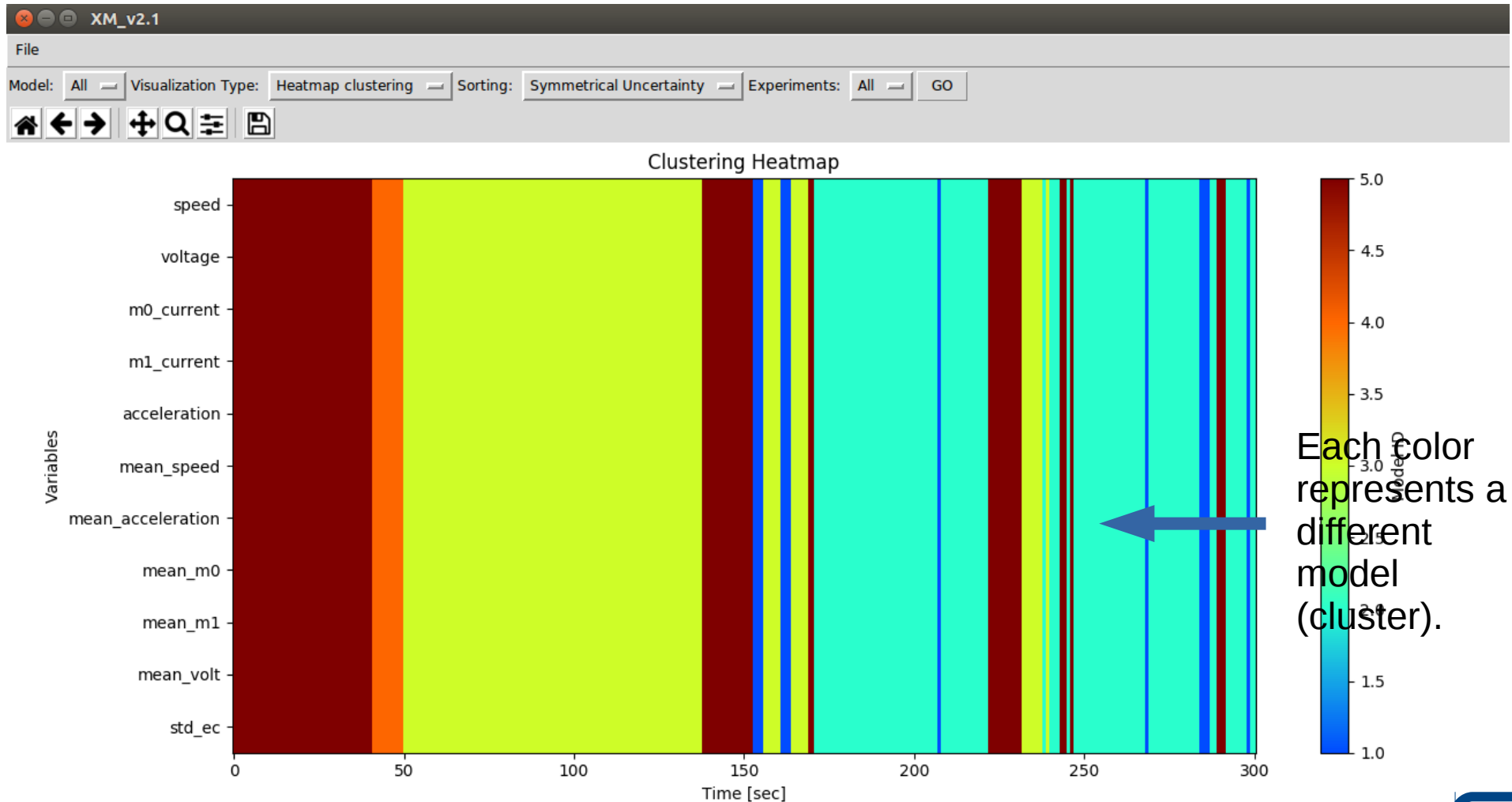
“Model parameters” visualization.
Each column represents a cluster centroid.



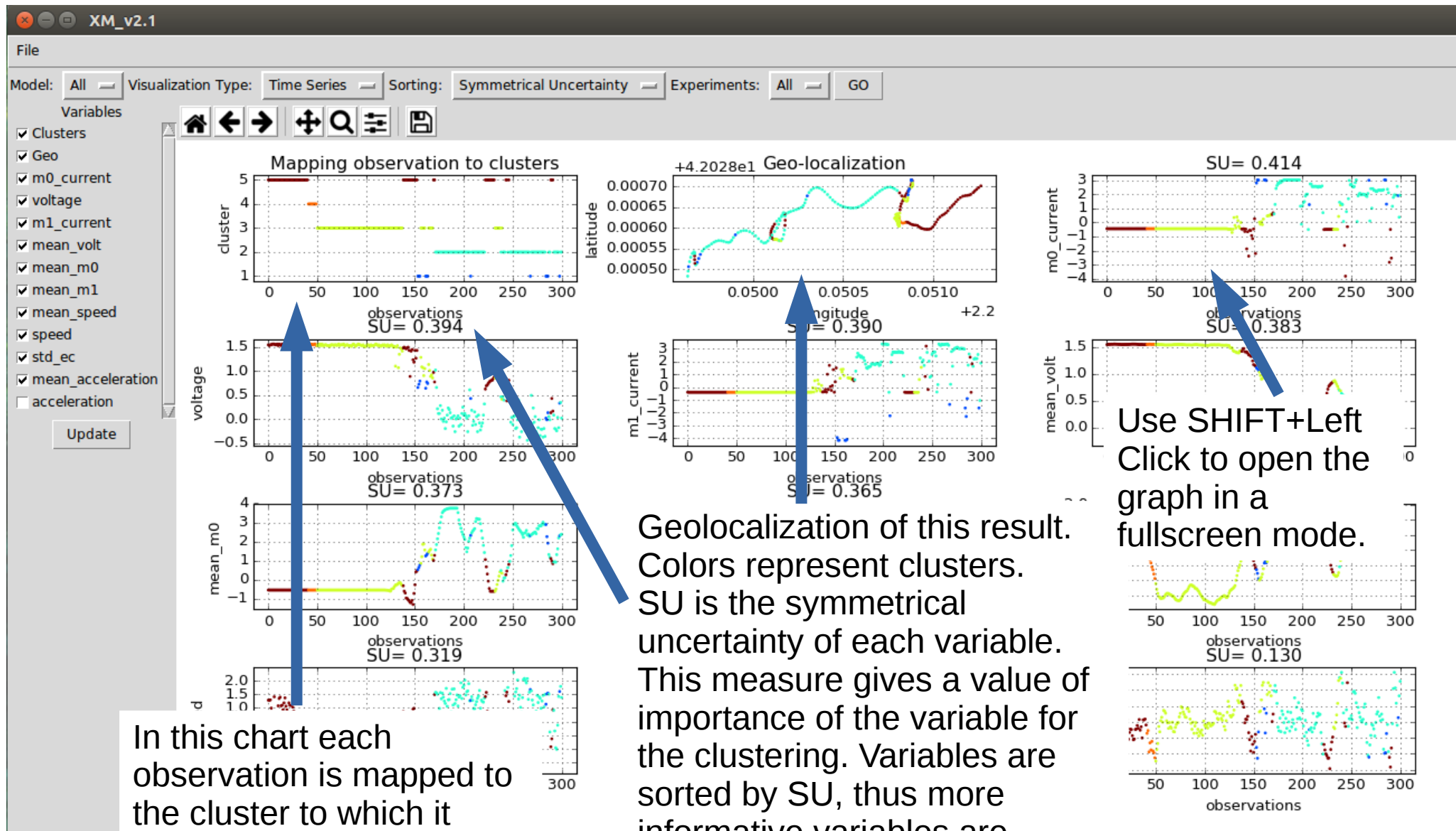
Select a specific model (cluster) to highlight the related properties.

Heatmap Tab

“Heatmap” visualization of the dataset and clustering. Every color represents a different cluster. For instance the green cluster is mainly located between instant 50 and 130.



Time Series Tab

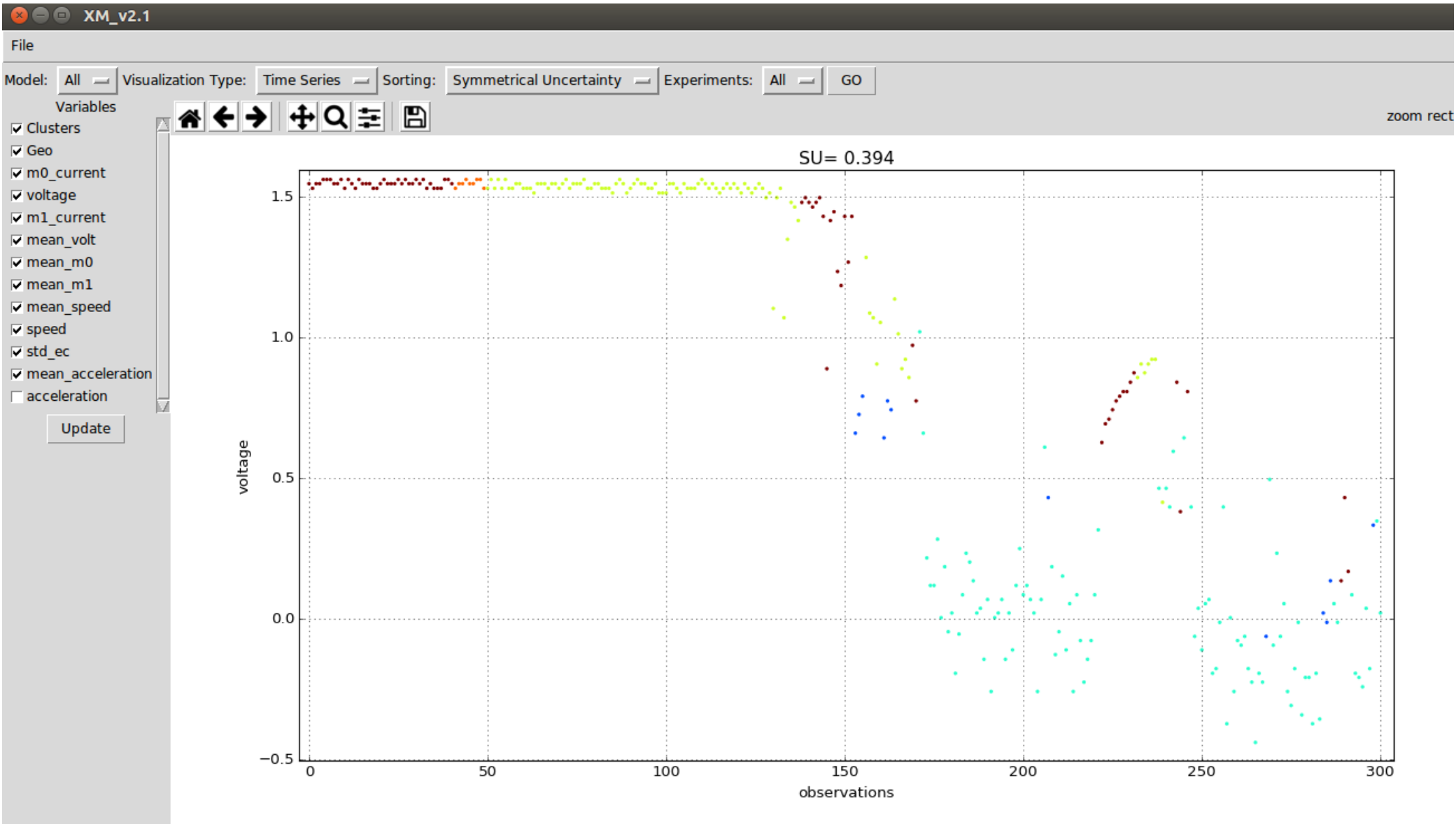


In this chart each observation is mapped to the cluster to which it belongs.

Geolocalization of this result. Colors represent clusters. SU is the symmetrical uncertainty of each variable. This measure gives a value of importance of the variable for the clustering. Variables are sorted by SU, thus more informative variables are positioned on top.

Use SHIFT+Left Click to open the graph in a fullscreen mode.

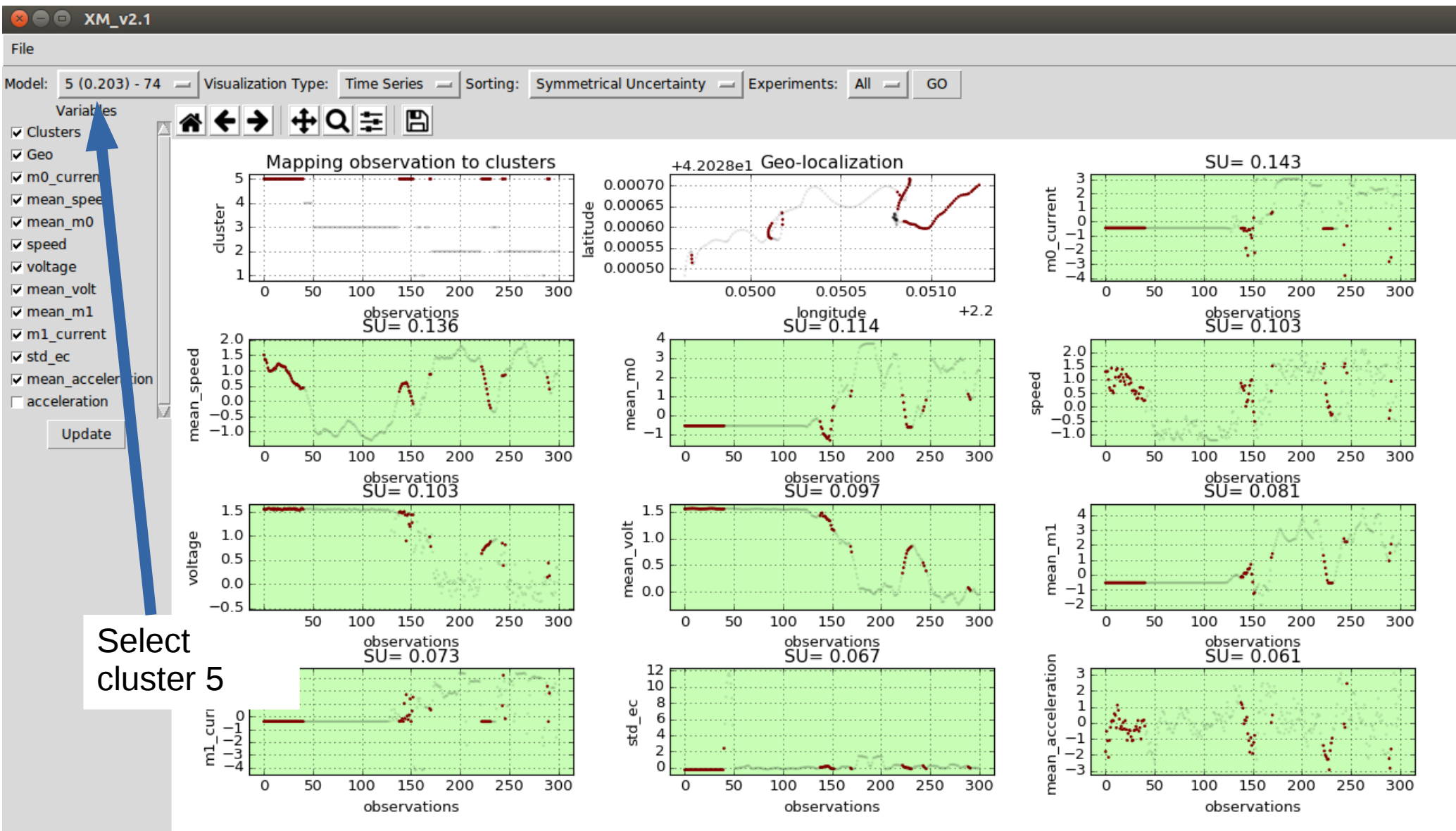
Time Series Tab fullscreen mode



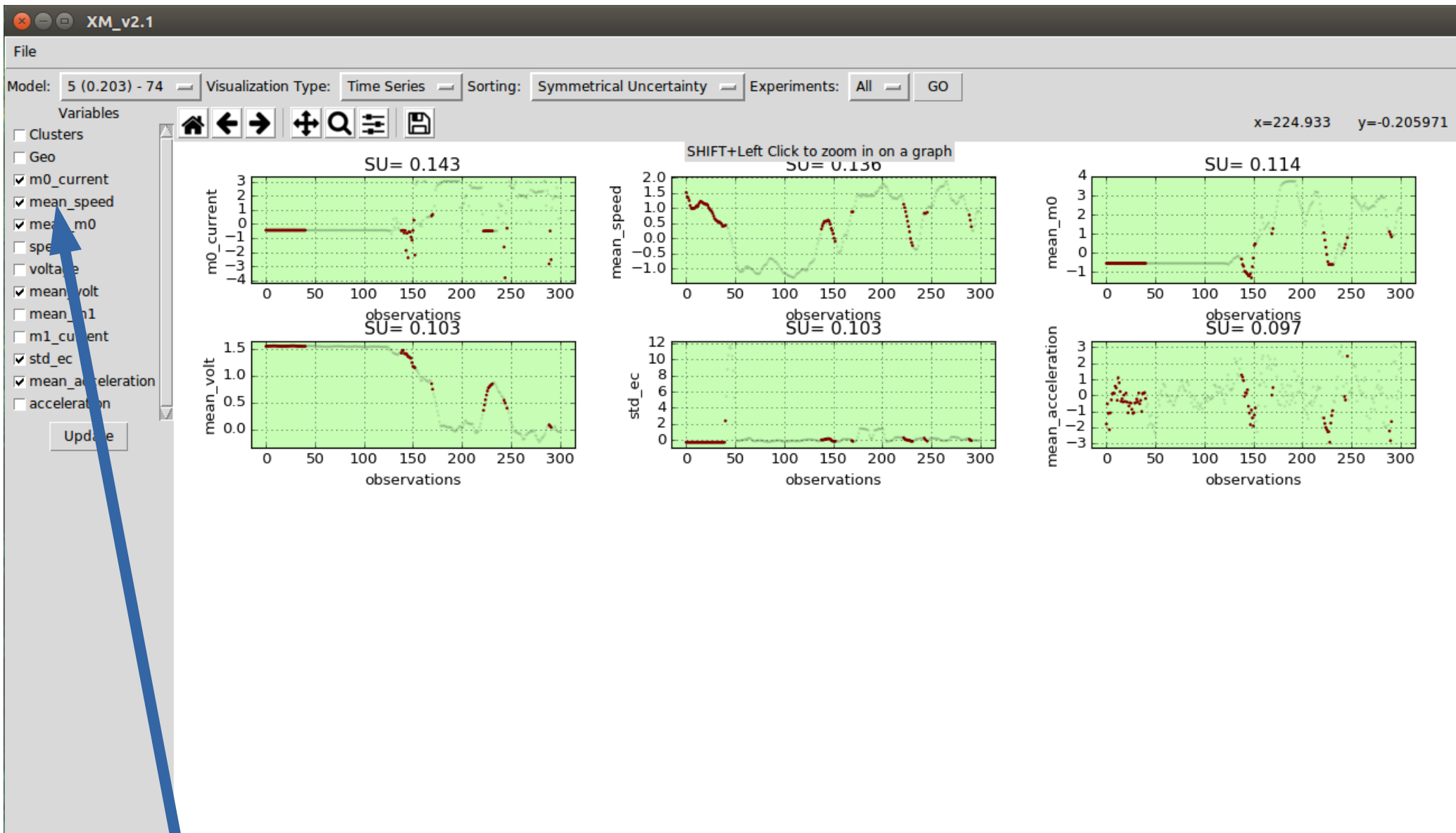
Press again SHIFT+Left click to go back to the standard view.



Time Series Tab with selection of specific model (cluster)



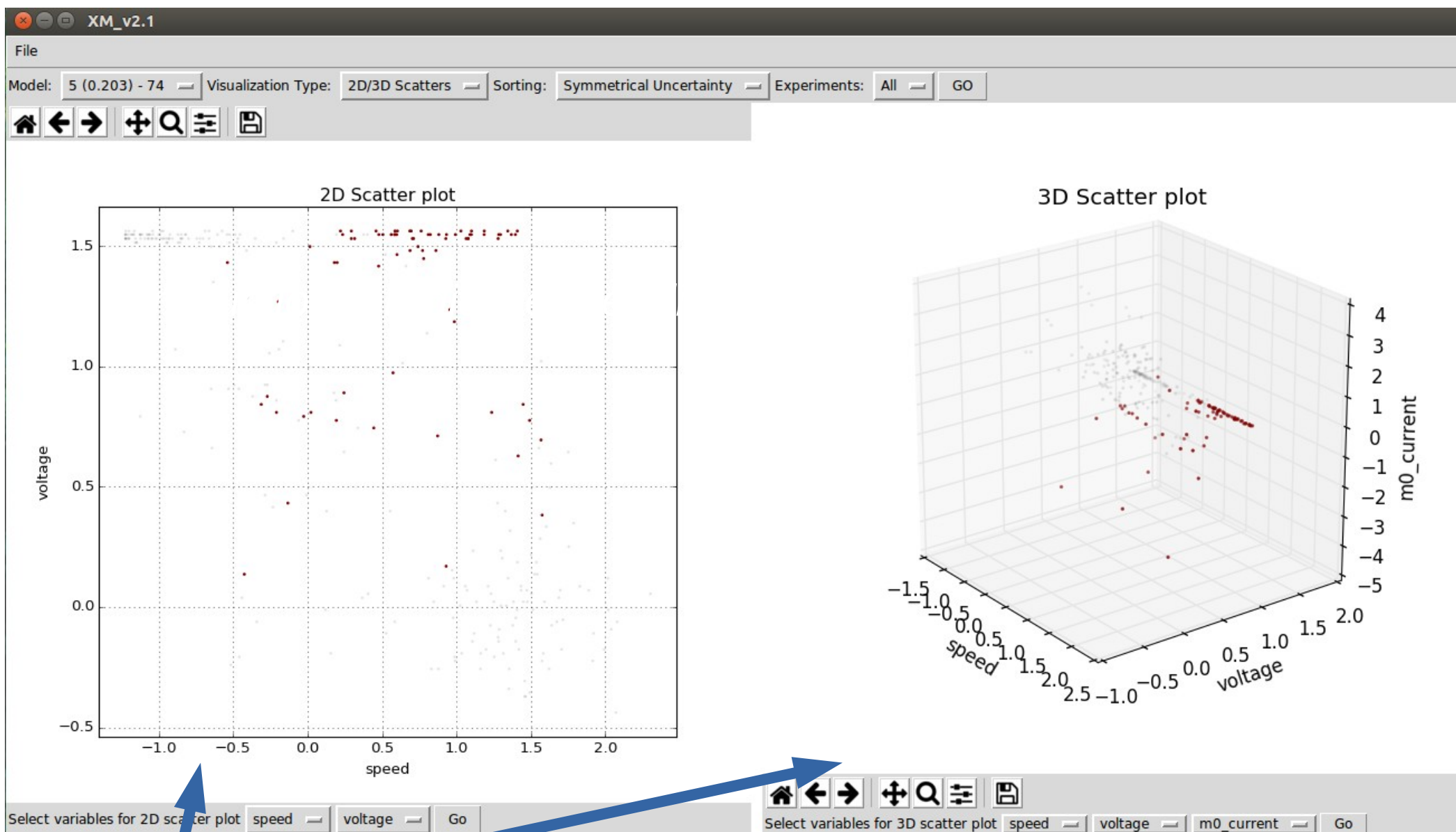
Time Series with selection of specific variables



On the left user can select which variables are shown on the screen (min 1, max 10 variables)



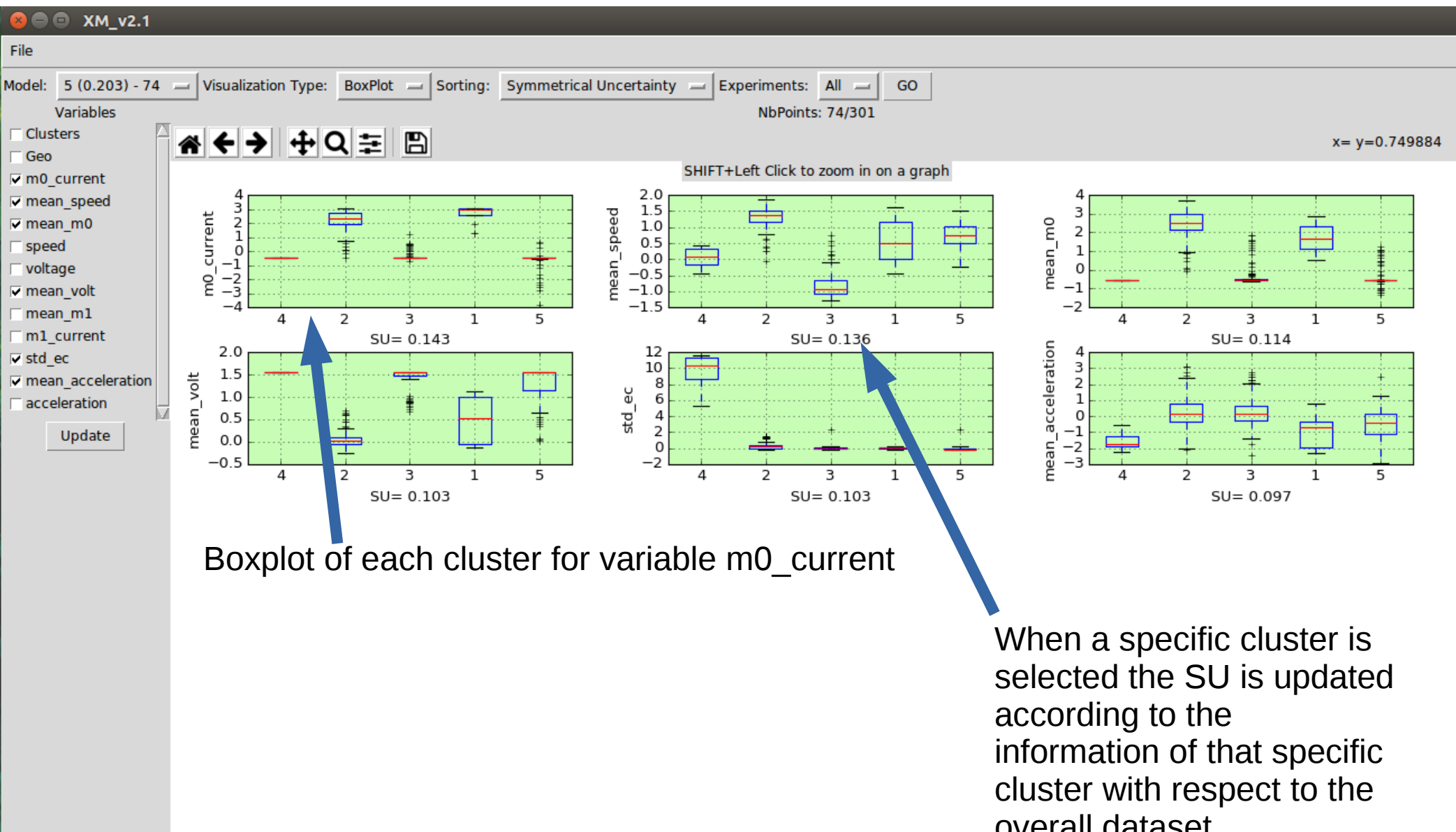
2D/3D Scatter plot Tab



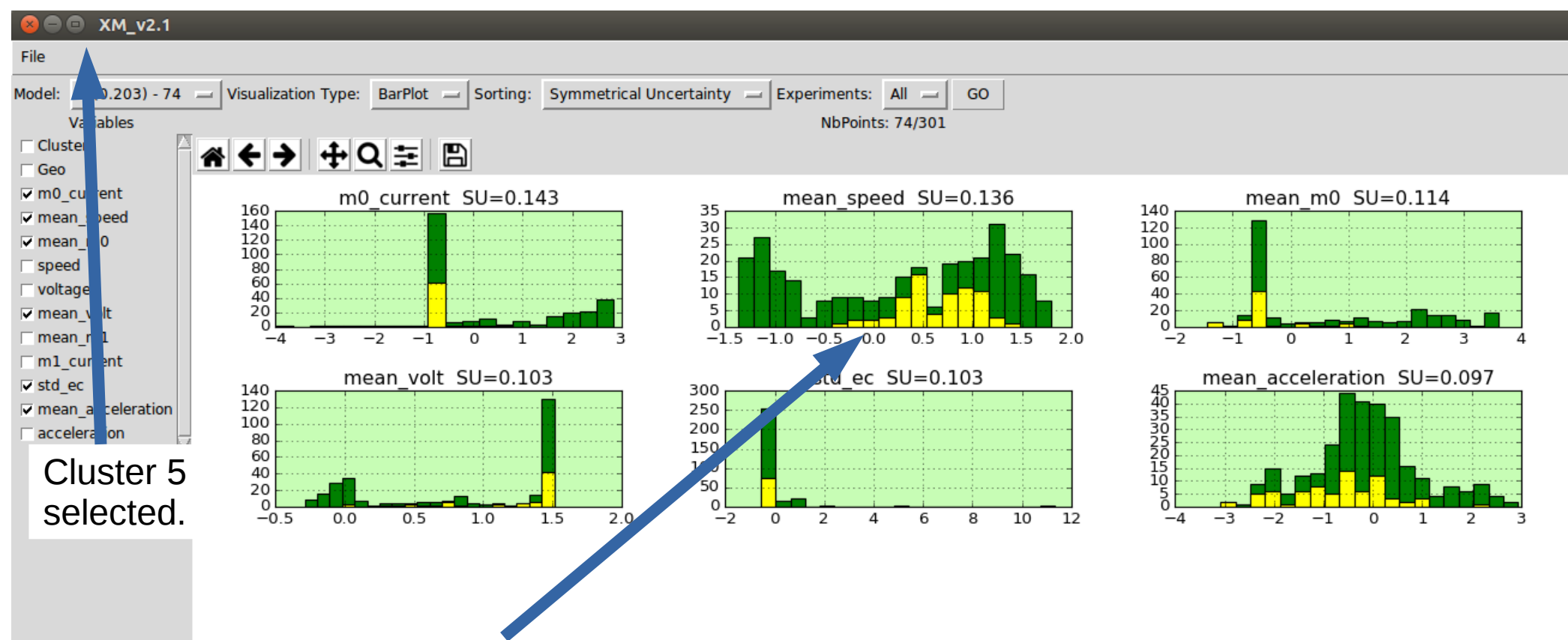
Select variables for each axes of the 2D/3D scatter plot



Boxplot Tab



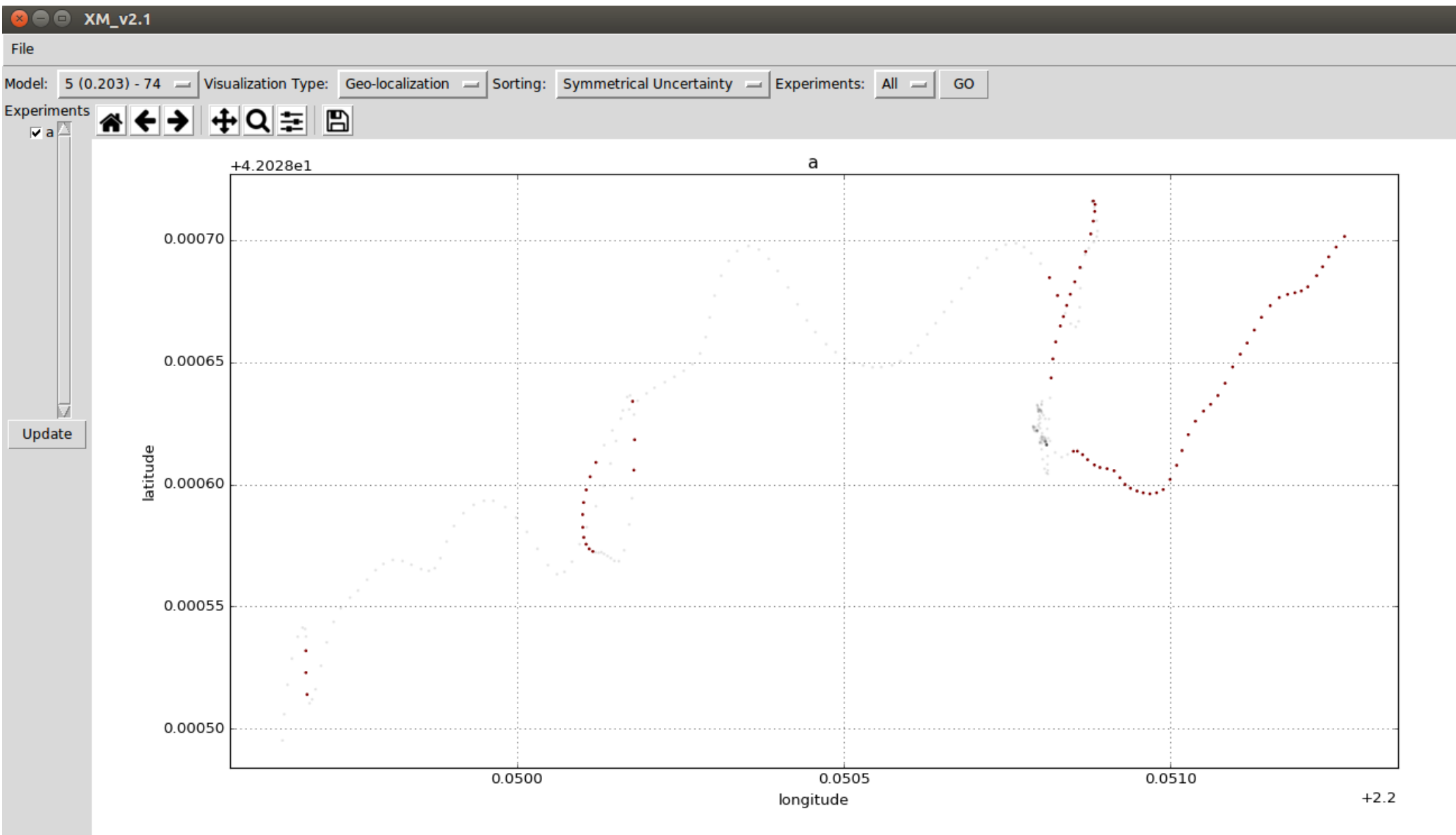
Barplot Tab



The green bars represent the histogram of the entire dataset for the specific variable.

The yellow bars represent the histogram of the cluster for the specific variable

Geo-localization Tab



3. How to generate state-models



3. How to generate state-models

```
from MethodLib.GMM import gmm
from MethodLib.KMeans import km
from MethodLib.SpectralClustering import spc
from MethodLib.SubCMedians import subc
from MethodLib.TICC import ticc
import os
```

```
actualPath = str(os.path.dirname(os.path.abspath(__file__)))
pathD = actualPath+"/DATASETS/xm_test_dataset.csv"
pathS = actualPath+"/RESULTS/XM_TEST_RESULTS/" #eventually change directory path
```

Setup path for dataset and result folder

```
print("Generating GMM results for test data")
gmm.gmm(5, "full", 10, pathD, pathS)
```

GMM generation part

```
print("Generating KM results for test data")
km.kmeans(5, pathD, pathS)
```

K-Means generation part

```
print("Generating Spectral clustering results for test data")
spc.spectral(5, pathD, pathS)
```

Spectral Clustering generation part

```
print("Generating SubCMedians results for test data")
subc.genera_cluster(5,0,0,0,pathD,pathS)
```

SubCmedians generation part

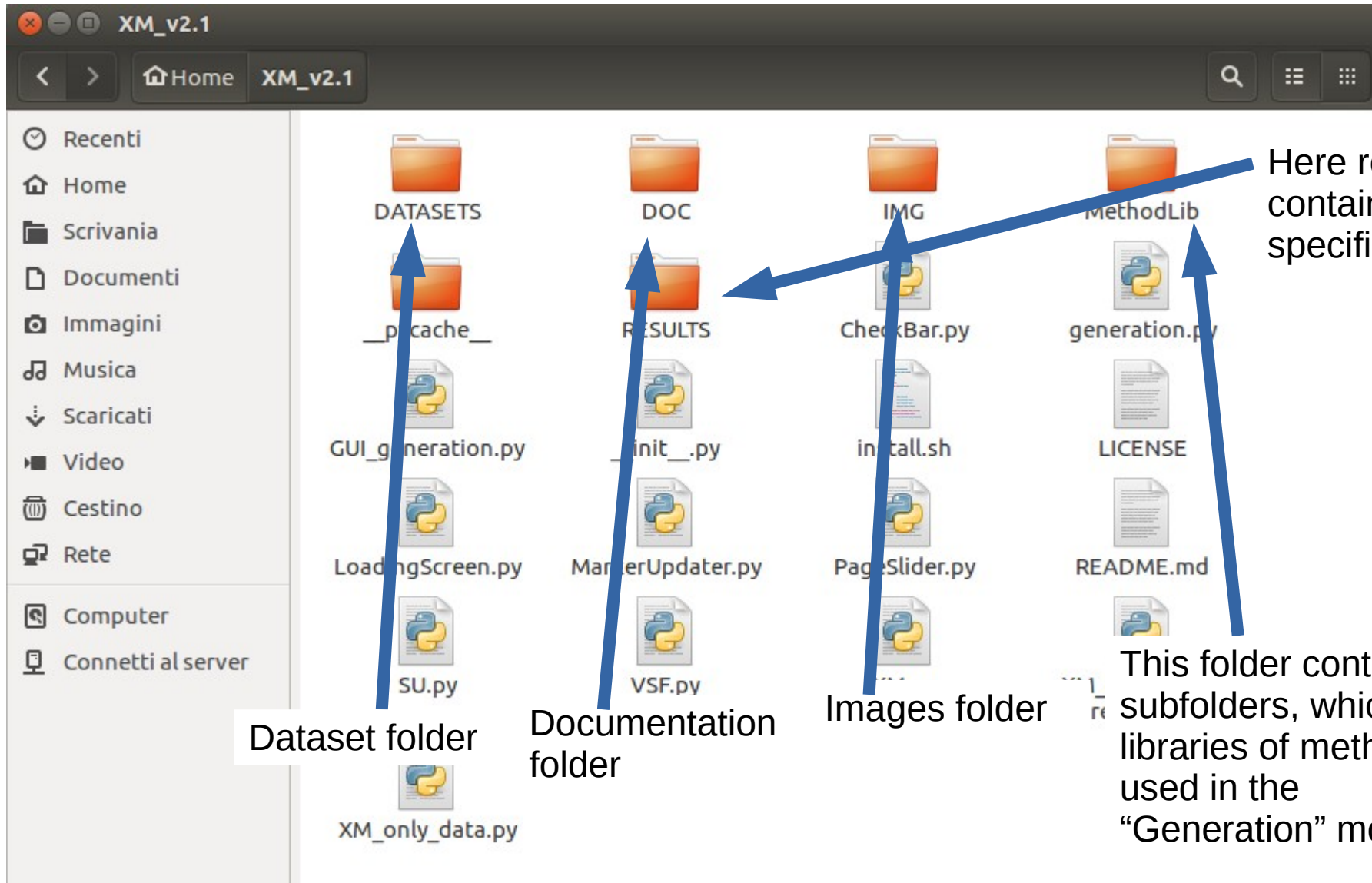
```
print("Generating TICC results for test data")
ticc.genera_cluster(5,1,0.5,150,pathD,pathS)
```

TICC generation part

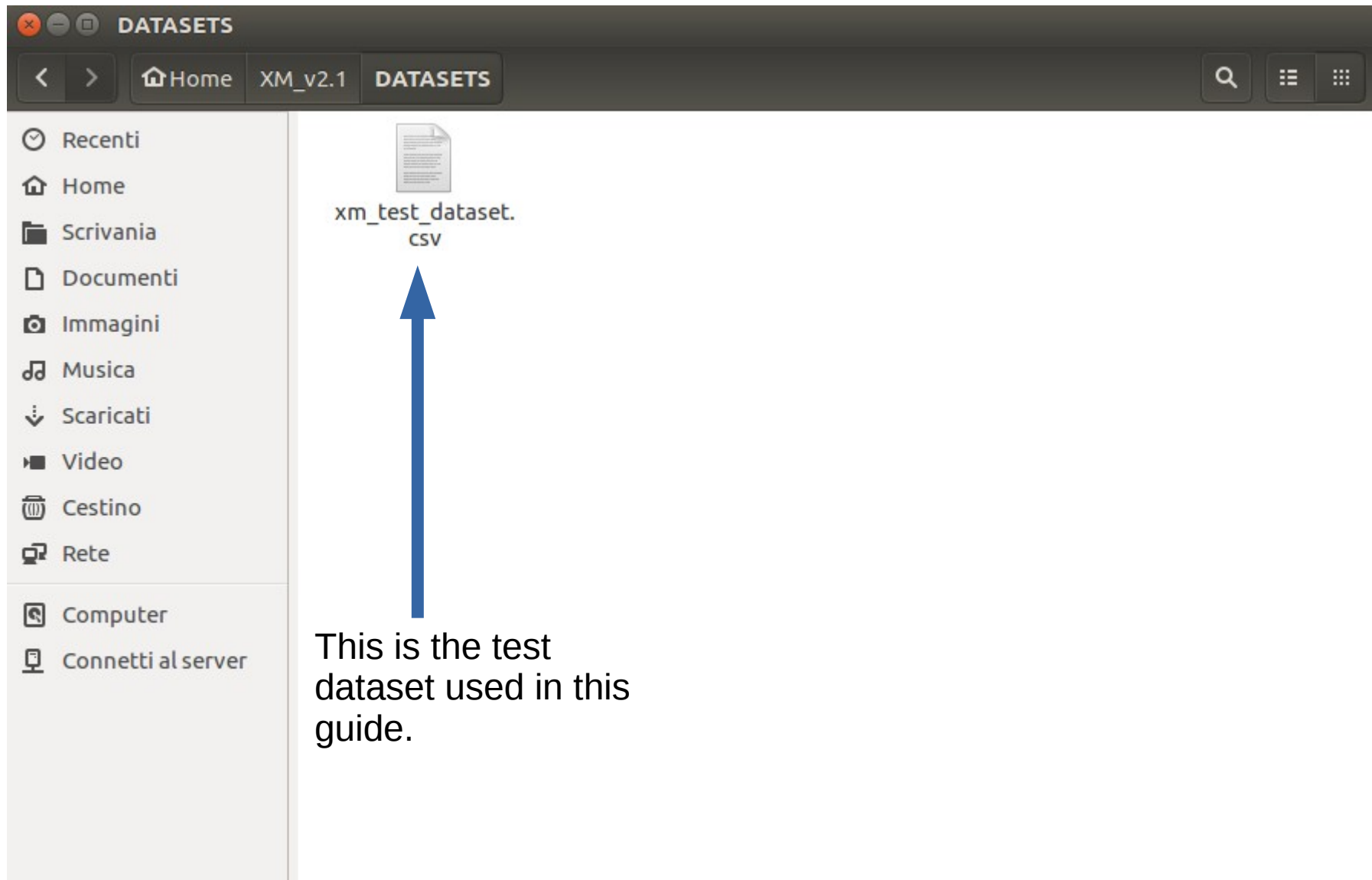
4. Quick overview of folders and files



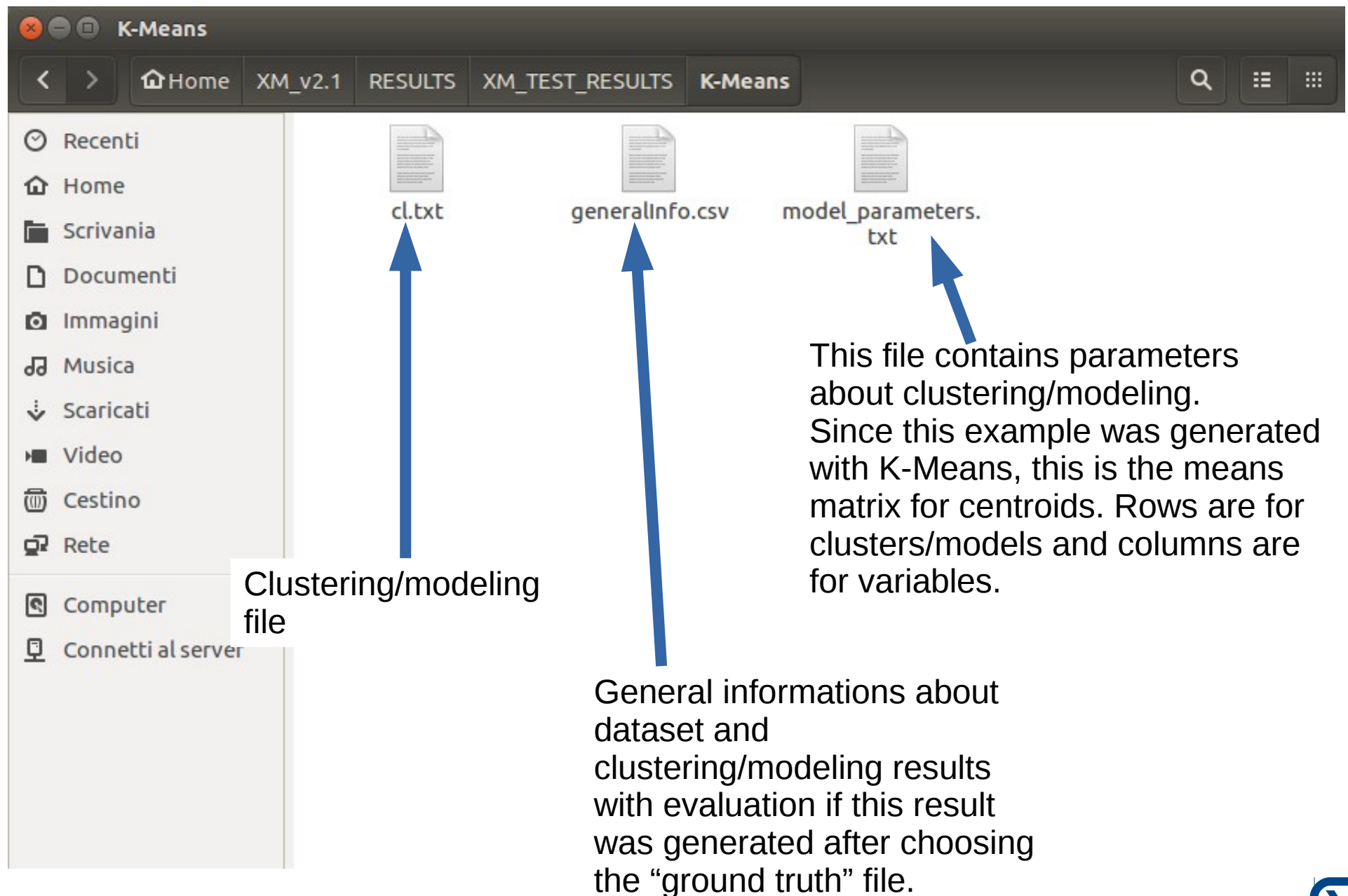
Folders



View of a particular dataset folder



View of a particular result folder



The screenshot shows a file explorer window titled "K-Means". The breadcrumb path is: Home > XM_v2.1 > RESULTS > XM_TEST_RESULTS > K-Means. The left sidebar shows the file system structure. The main area displays three files: `cl.txt`, `generalInfo.csv`, and `model_parameters.txt`. Blue arrows point from descriptive text blocks to each file.

cl.txt
Clustering/modeling file

generalInfo.csv
General informations about dataset and clustering/modeling results with evaluation if this result was generated after choosing the "ground truth" file.

model_parameters.txt
This file contains parameters about clustering/modeling. Since this example was generated with K-Means, this is the means matrix for centroids. Rows are for clusters/models and columns are for variables.