

# **XM - eXplainable Modeling**

## **How To Guide**



*Alberto Castellini, Alessandro Farinelli, Francesco Masillo*

Verona University



13 July 2018

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

# 1. How to install the software



# 1. How to install the software

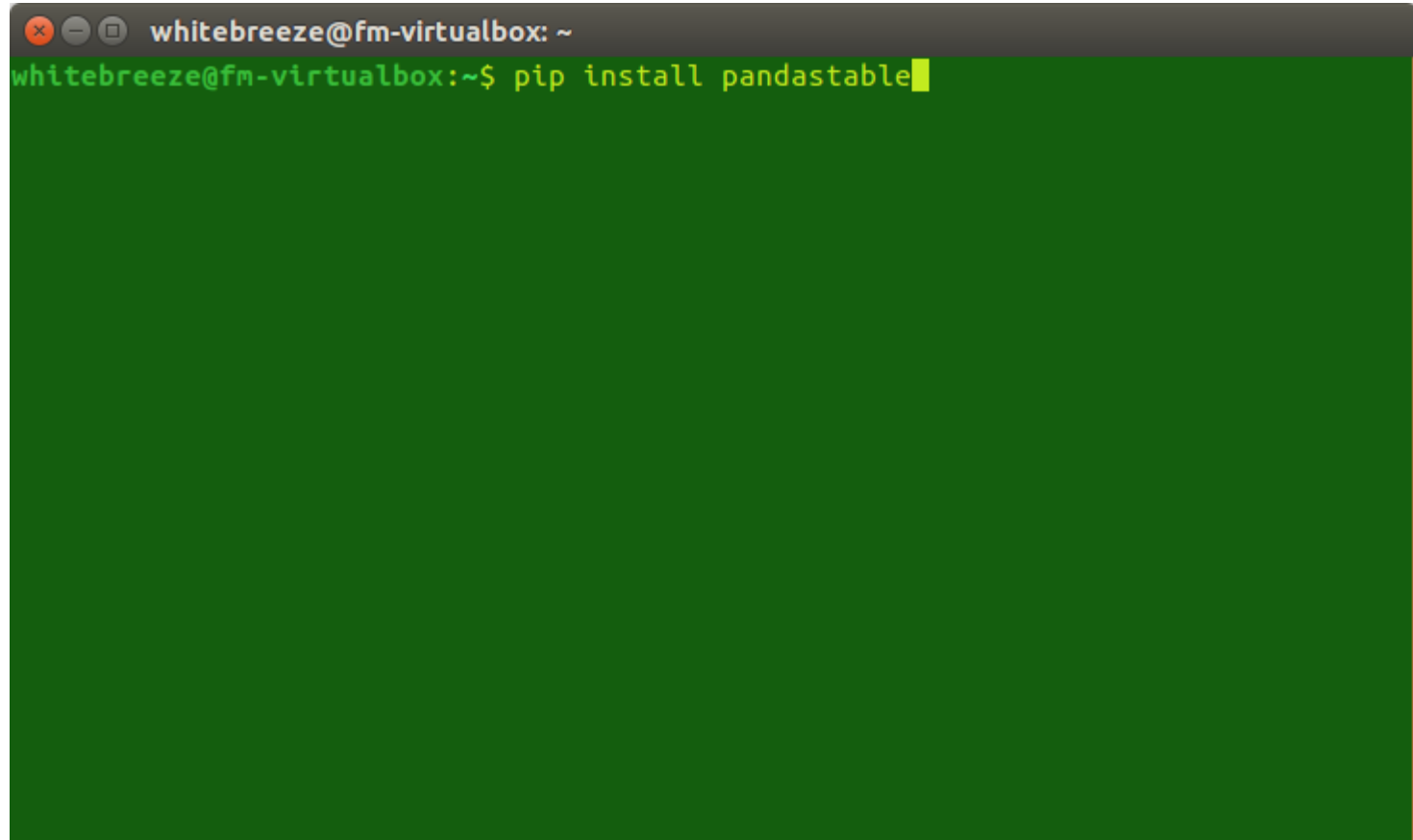
To use this software Python 2.7 has to be installed.

Some other libraries, also written in file “libraries.txt”, are required.

To install this libraries:

- pandastable, run “pip install pandastable”
- pyitlib, run “pip install pyitlib”
- xkcd, run “pip install xkcd”
- gsl, run “sudo apt-get install libgsl0-dev”
- lblas, run “sudo apt-get install libblas-dev liblapack-dev”

Example with  
pandastable:

A terminal window with a dark green background and a grey title bar. The title bar contains the text 'whitebreeze@fm-virtualbox: ~'. The terminal shows the command 'whitebreeze@fm-virtualbox:~\$ pip install pandastable' in green text, with a yellow cursor at the end of the line.

```
whitebreeze@fm-virtualbox: ~  
whitebreeze@fm-virtualbox:~$ pip install pandastable
```



## 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 “python XM.py” on a terminal opened in the sw folder.



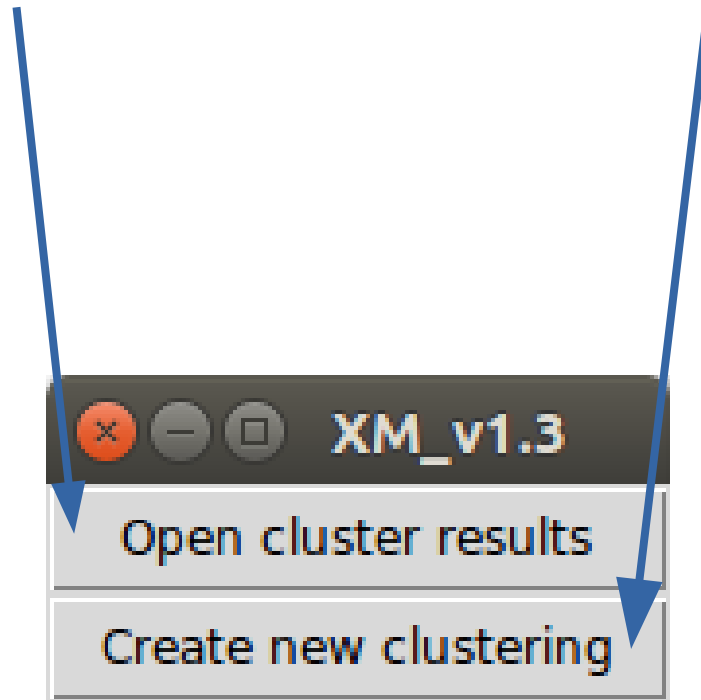
A terminal window with a dark green background and a grey title bar. The title bar contains the text "whitebreeze@fm-virtualbox: ~/XM\_v1.3". The terminal shows the command "python XM.py" being entered at the prompt "whitebreeze@fm-virtualbox:~/XM\_v1.3\$". A blue arrow points from the text "sw folder" in the paragraph above to the directory path in the terminal title bar. Another blue arrow points from the text "python XM.py" in the paragraph above to the command being entered in the terminal.

```
whitebreeze@fm-virtualbox: ~/XM_v1.3$ python XM.py
```



## How to launch the Model Visualizer (cont'd)

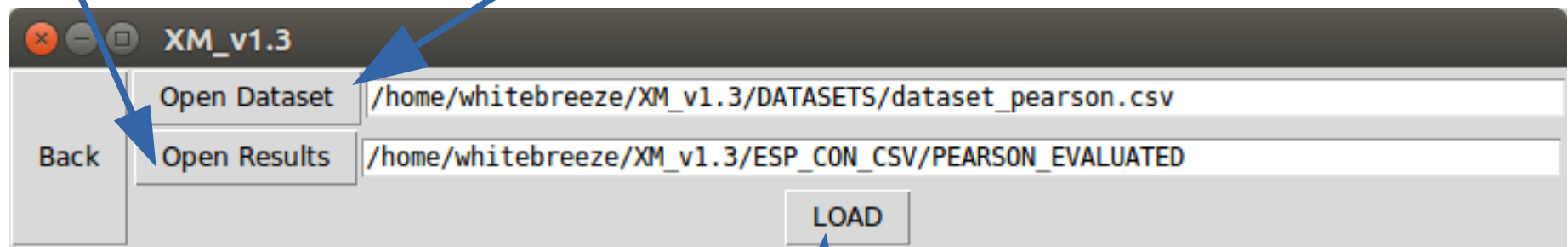
A new window should appear in the middle of the screen. It has two intuitive buttons, one for “Visualization” and one for “Generation”.



For the visualization part “Open cluster results” must be clicked.

## 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: Symmetrical Uncertainty GO

	Clus	N. points	Silhouette
1	1.0	76	0.22032603
2	2.0	35	0.26842147
3	3.0	40	0.29407732
4	4.0	53	0.21609686
5	5.0	40	0.11154231
6	6.0	48	0.28272441
7	7.0	47	0.16868086
8	8.0	61	0.06328298

	Parameter	Values
1	Dataset_Path	/home/whitebreeze/XClust_v1.3/DATASETS/datas
2	Results_Path	/home/whitebreeze/XClust_v1.3/ESP_CON_C
3	Number_of_columns	3
4	Number_of_rows	400
5	Method	SubCMedians
6	NbExpCluster	4
7	SDmax	12
8	N	100
9	NbIter	480
10	Accuracy	0.655
11	CE	0.4175
12	F1	0.655817359510653
13	Entropy	0.5011221317327026

Select a type of visualization.

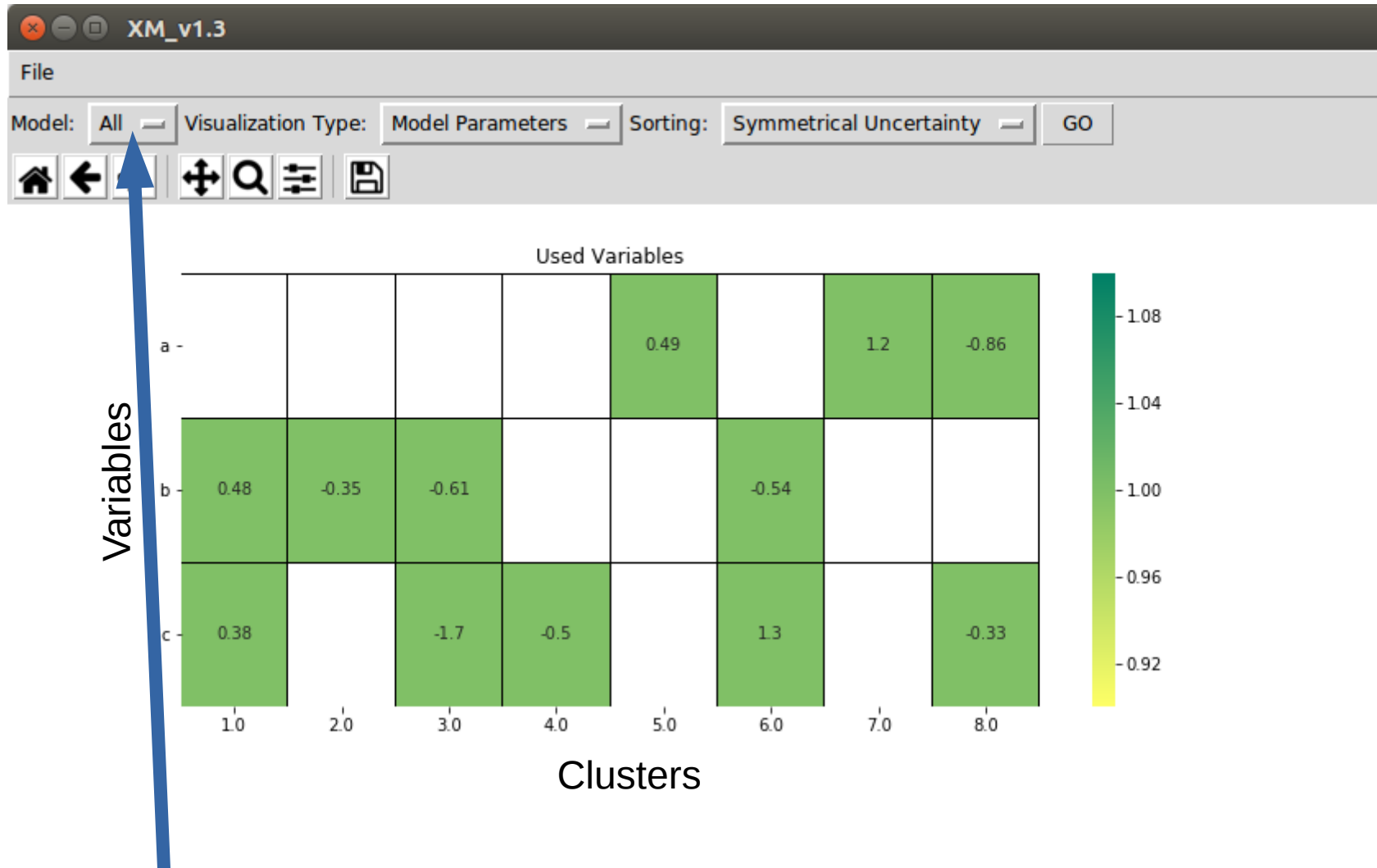
After selecting the type of visualization and/or cluster click GO



# Model Parameters Tab

“Model parameters” visualization.

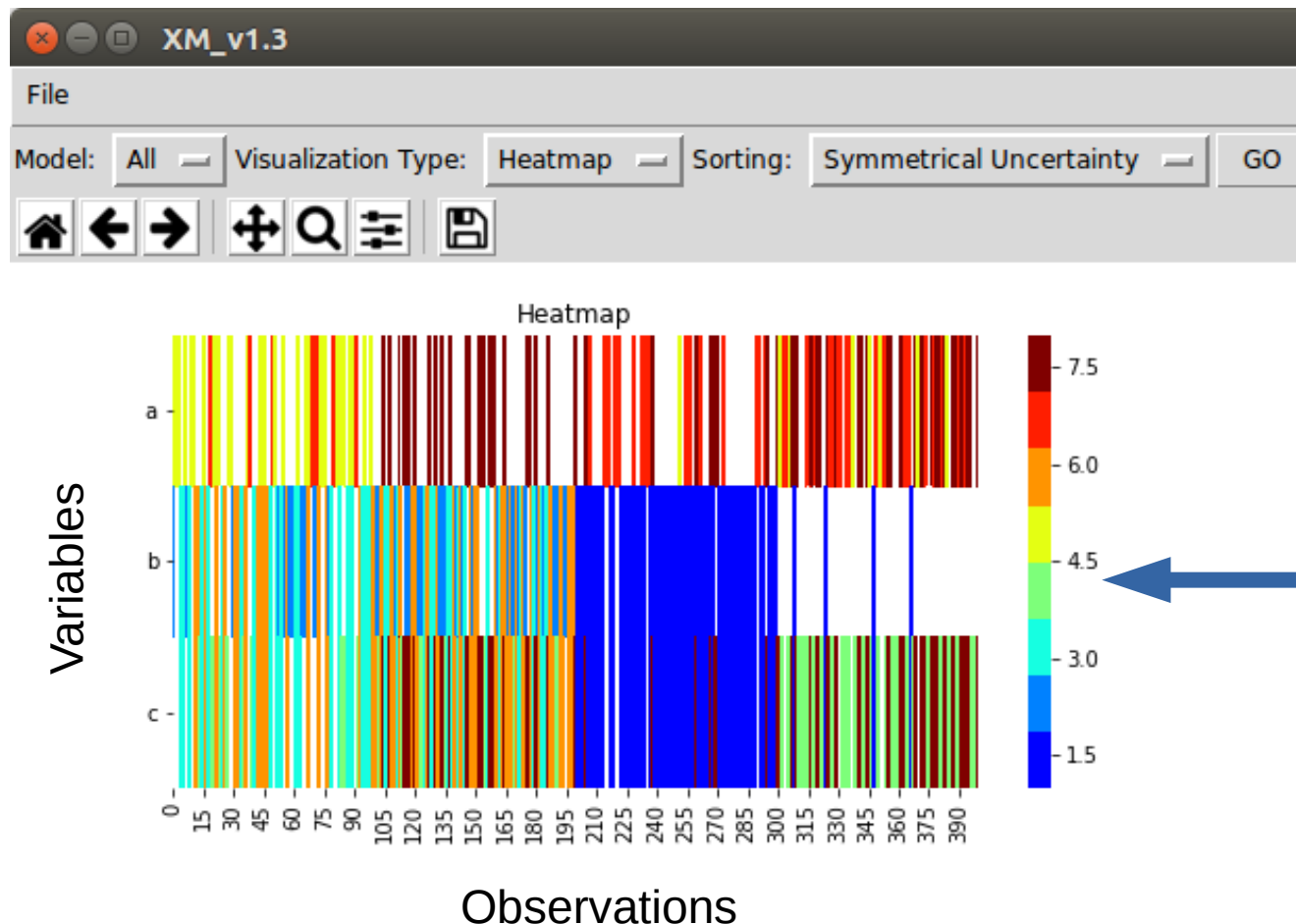
Each column represents a cluster centroid. Green cells are selected variables, white cells are non selected variables



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

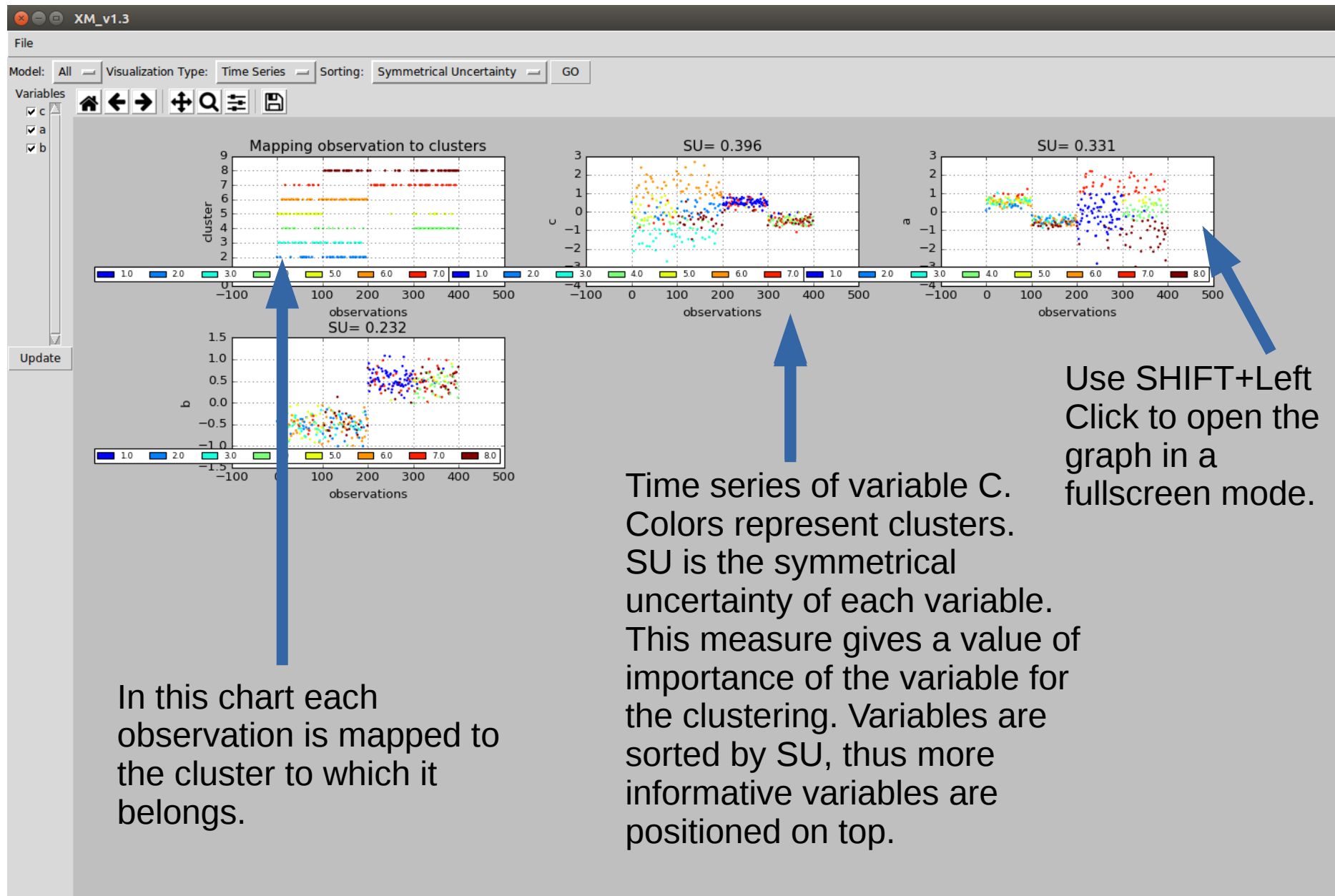
# Heatmap Tab

“Heatmap” visualization of the dataset and clustering.  
Every color represents a different cluster. For instance the blue cluster is mainly located between instant 200 and 300 and uses variables b and c.

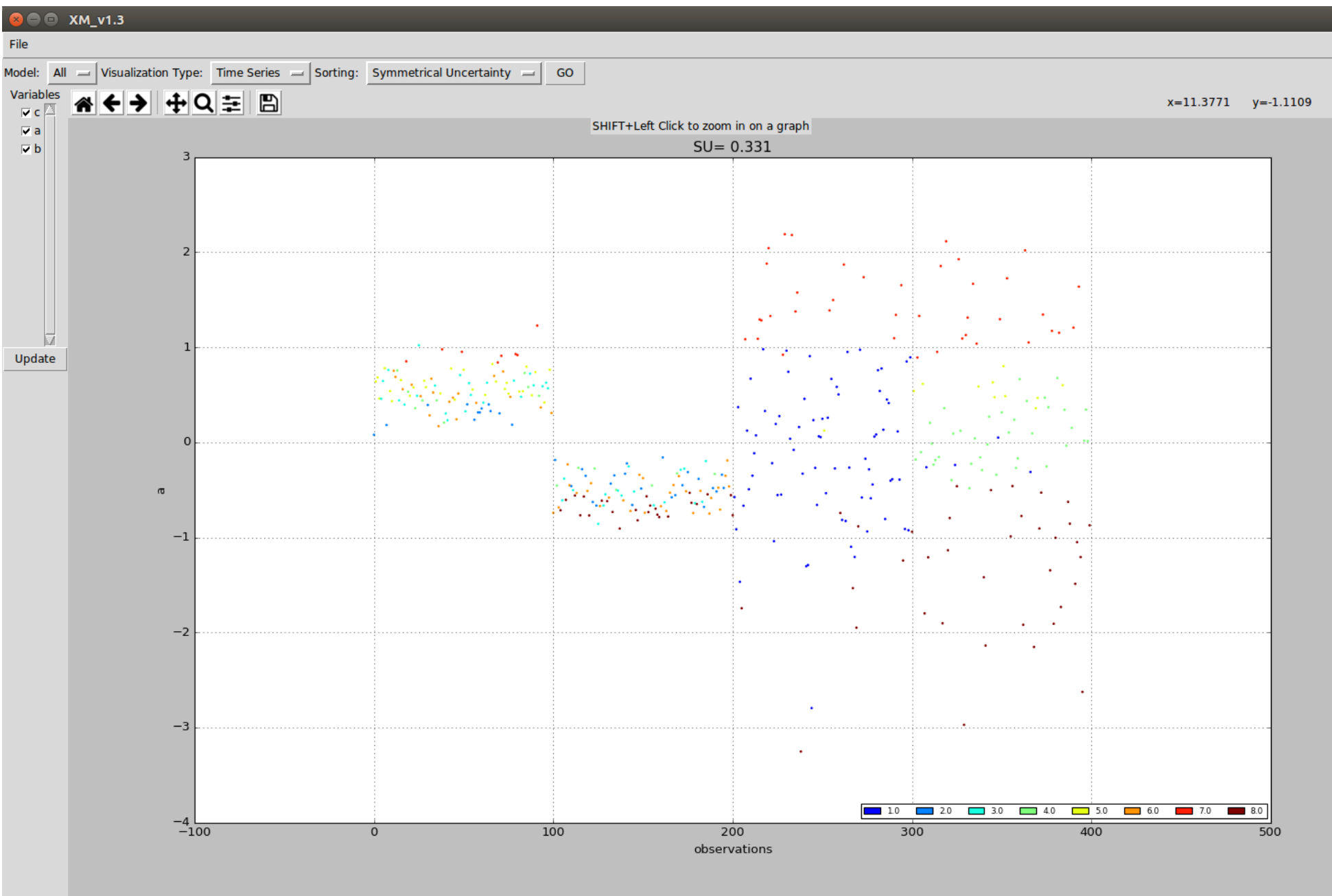


Each color represents a different model (cluster).

# Time Series Tab



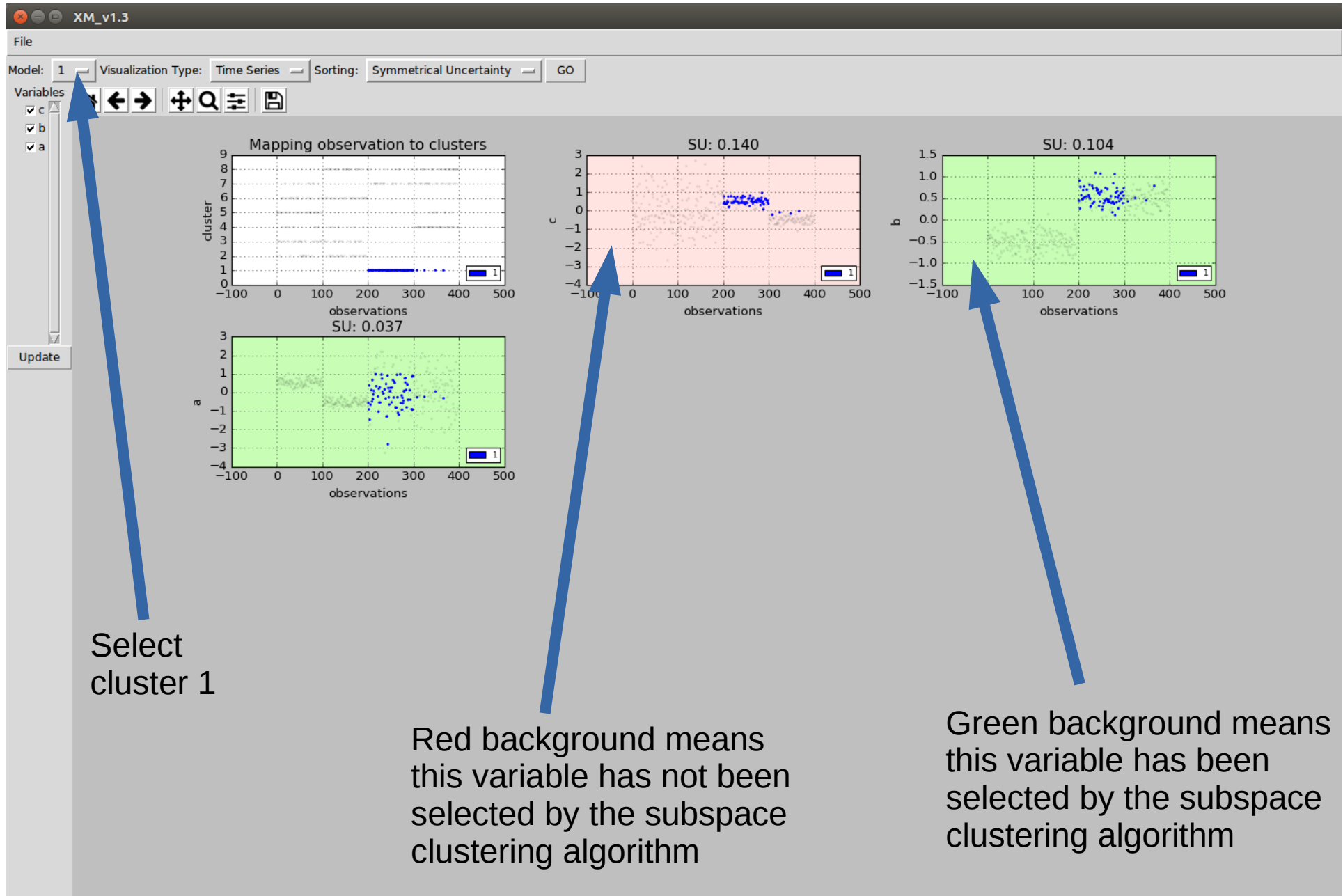
# Time Series Tab fullscreen mode



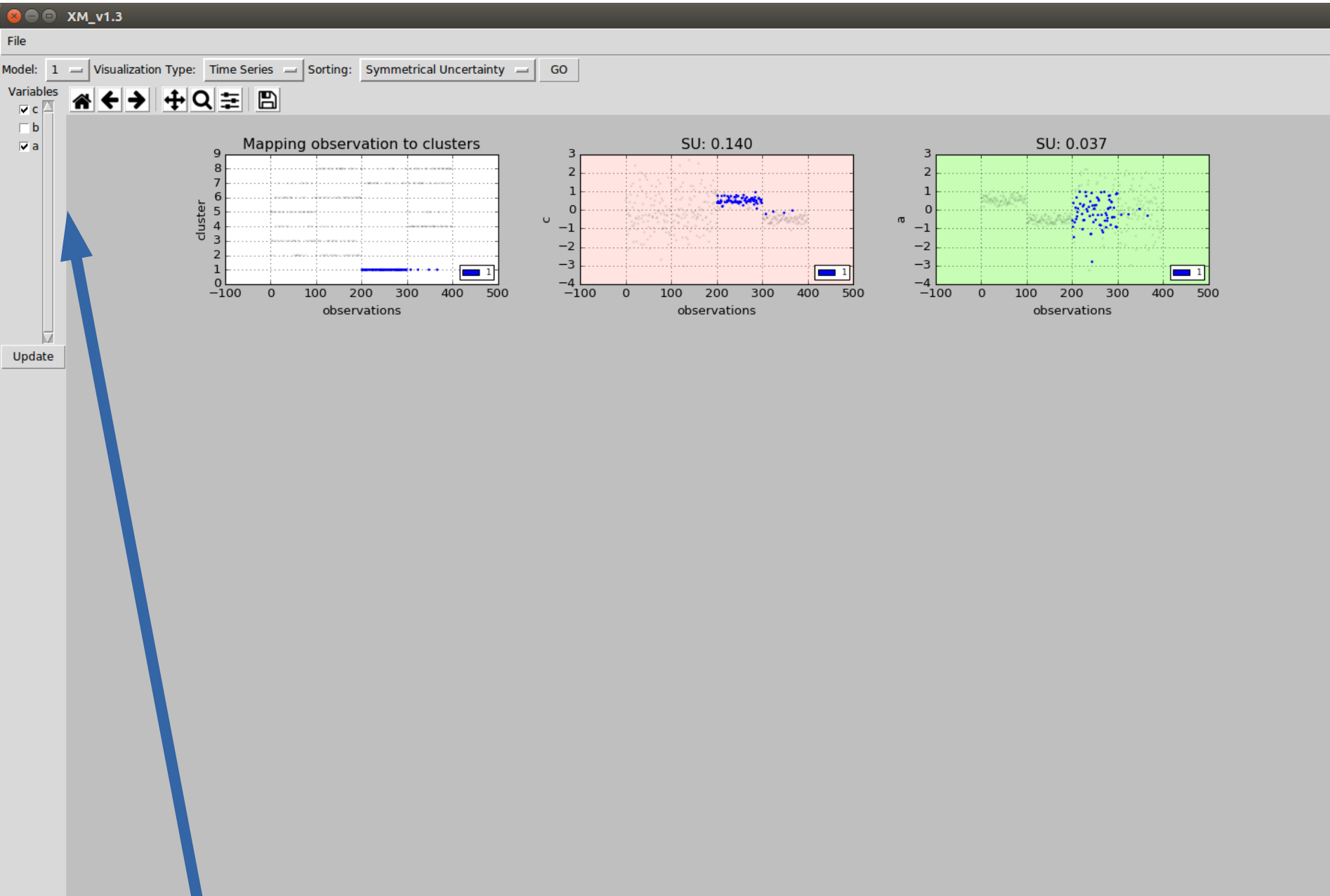
Press again SHIFT+Left click to go back to the standard view.



# Time Series Tab with selection of specific model (cluster)

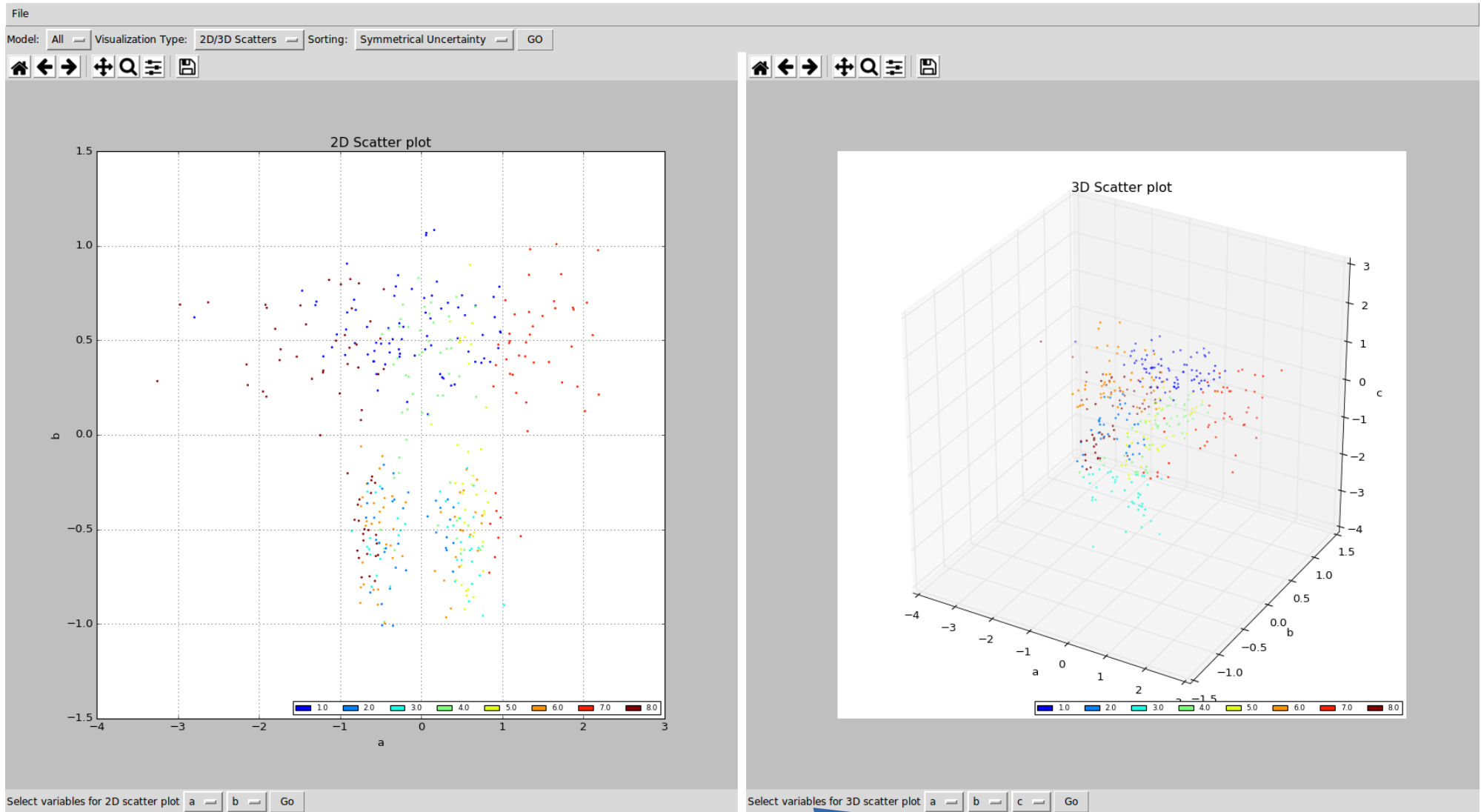


# Time Series with selection of specific variables



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

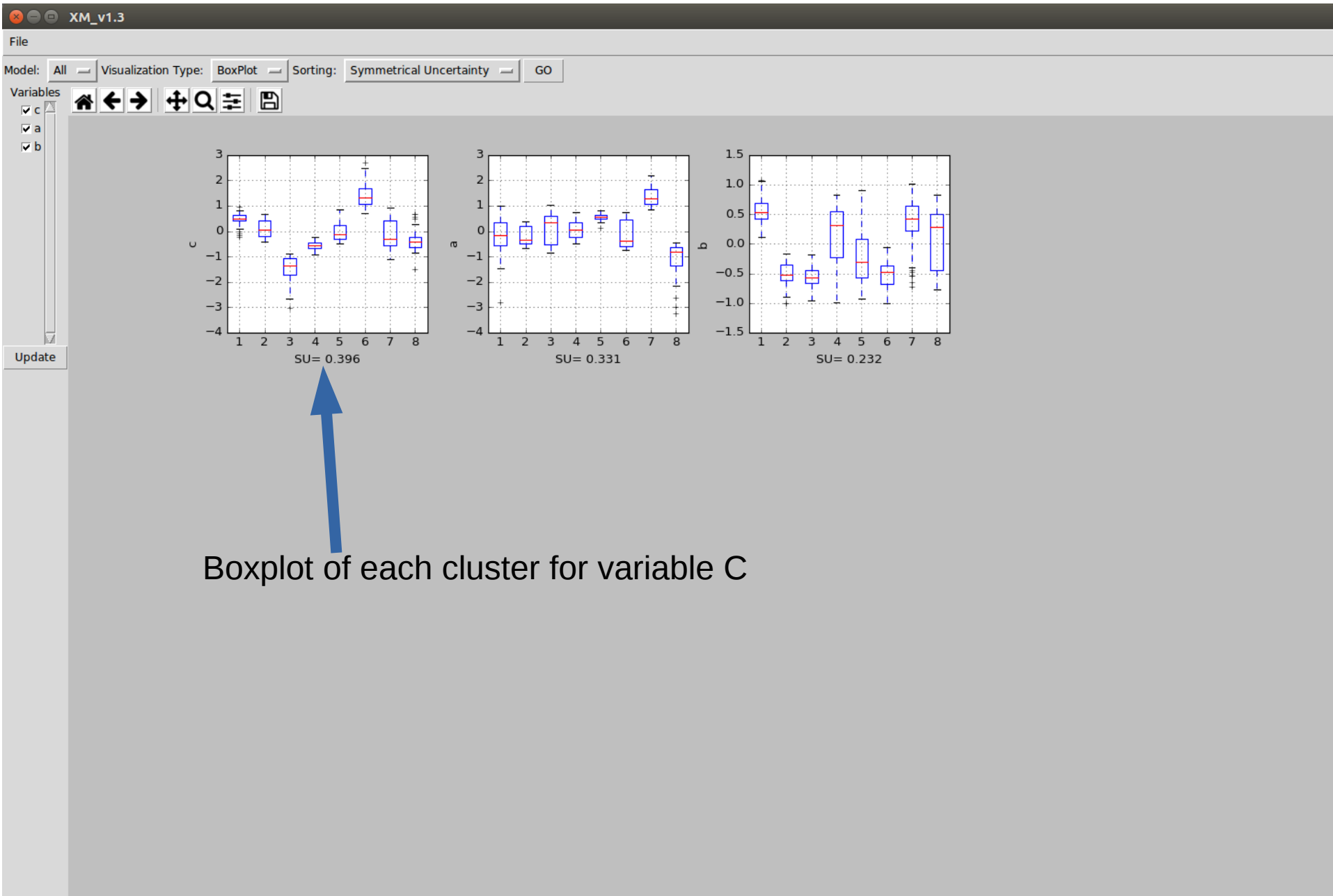
# 2D/3D Scatter plot Tab



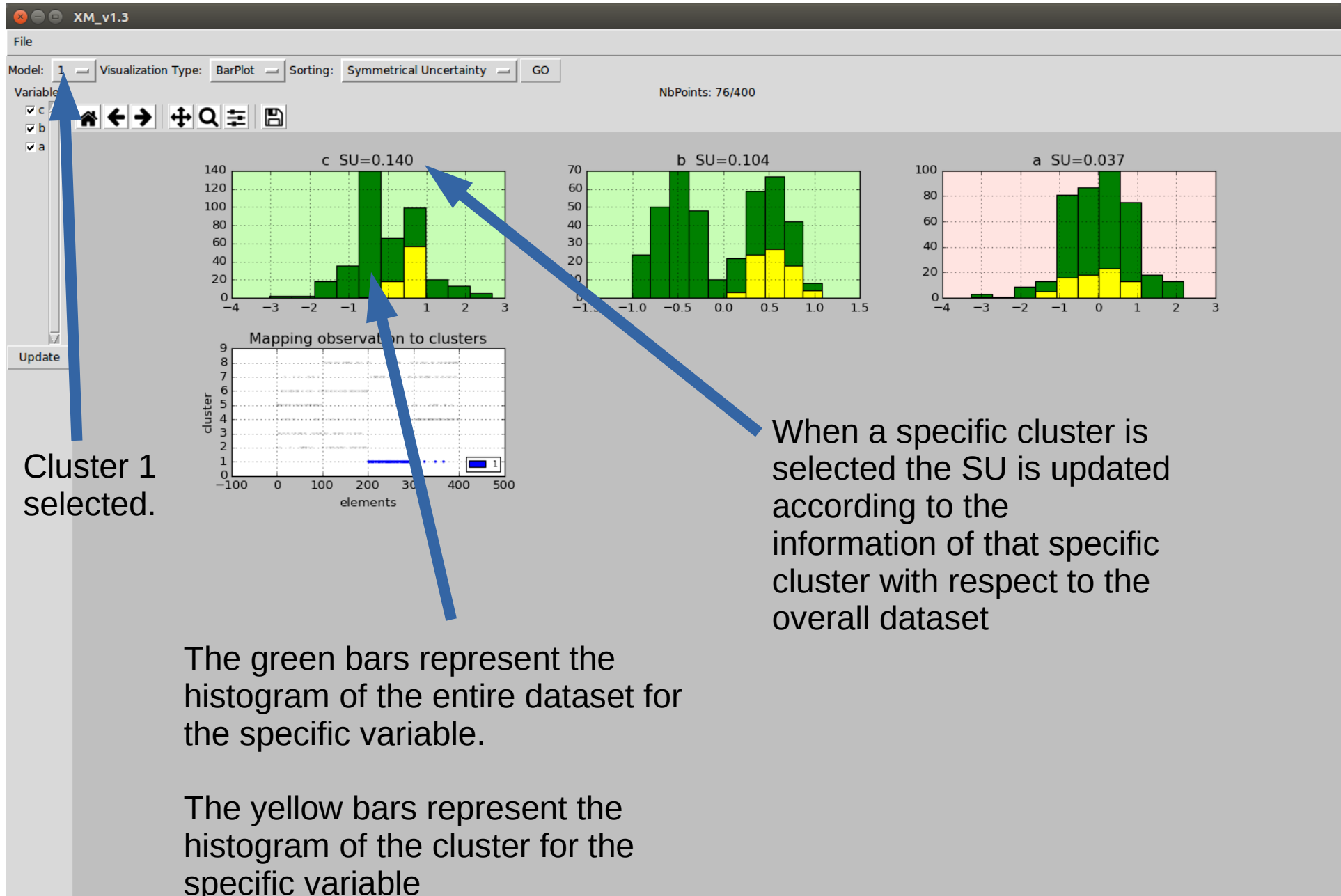
Select variables for each axes of the 2D/3D scatter plot



# Boxplot Tab



# Barplot Tab

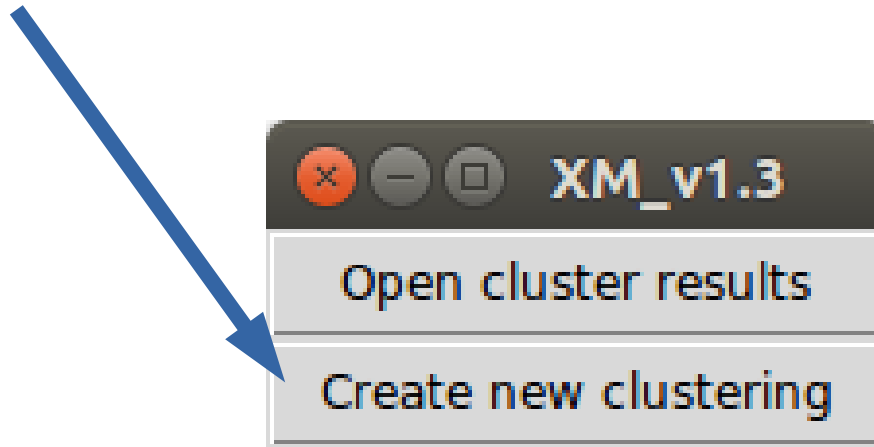


### 3. How to launch the Model Generator

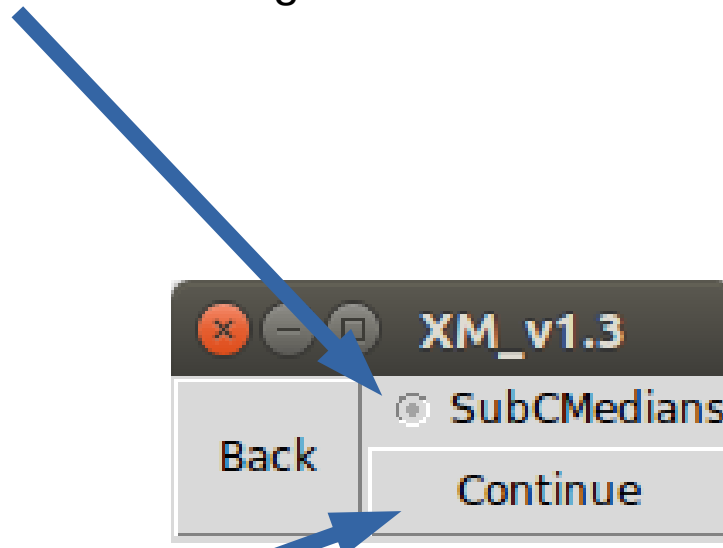


### 3. How to launch the Model Generator

Click on “Create new clustering” button.



Choose the method used to generate new results.



Then click continue.



# Default parameters setting

This window enables parameters setting for SubCMedians (*Peignier et al, 2018*). Similar windows enable parameters setting for other methods.

This file is optional, select only in presence of a synthetic dataset.

Select the folder where to save the results. If a folder is selected then its name will appear in the text area.

When ready click Start.

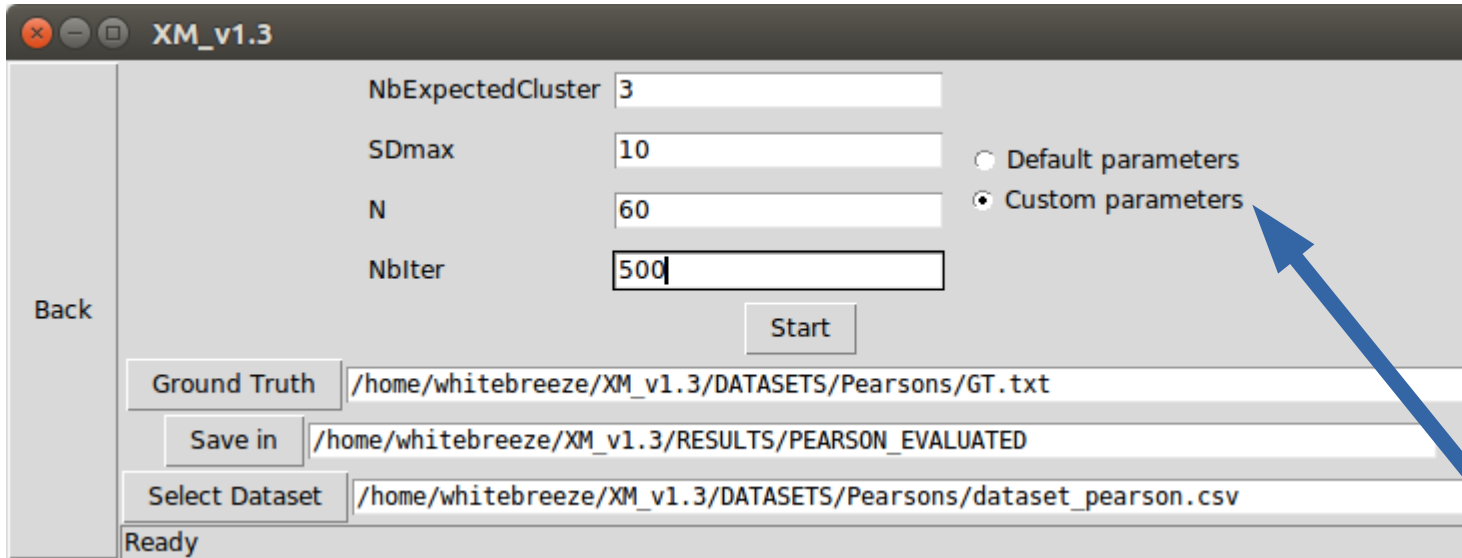
Select the dataset file. If a file is selected then its name will appear in the text area.

This option lets the user to input only a desired number of cluster that he would like to obtain. All the parameters will be automatically computed.

(*Peignier et al, 2018*): Sergio Peignier, Christophe Rigotti, Anthony Rossi, Guillaume Beslon, Weight-based search to find clusters around medians in subspaces. The 33rd ACM/SIGAPP Symposium On Applied Computing, pages 471-480, 2018.



# Custom parameters setting



The screenshot shows the XM\_v1.3 application window. On the left is a vertical sidebar with a 'Back' button. The main area contains four input fields for parameters: 'NbExpectedCluster' (3), 'SDmax' (10), 'N' (60), and 'NbIter' (500). To the right of these fields are two radio buttons: 'Default parameters' (unselected) and 'Custom parameters' (selected). A blue arrow points from the text below to the 'Custom parameters' radio button. Below the parameter fields is a 'Start' button. At the bottom, there are three text fields: 'Ground Truth' with the path '/home/whitebreeze/XM\_v1.3/DATASETS/Pearsons/GT.txt', 'Save in' with the path '/home/whitebreeze/XM\_v1.3/RESULTS/PEARSON\_EVALUATED', and 'Select Dataset' with the path '/home/whitebreeze/XM\_v1.3/DATASETS/Pearsons/dataset\_pearson.csv'. The status bar at the very bottom says 'Ready'.

Parameter	Value
NbExpectedCluster	3
SDmax	10
N	60
NbIter	500

Ground Truth: /home/whitebreeze/XM\_v1.3/DATASETS/Pearsons/GT.txt

Save in: /home/whitebreeze/XM\_v1.3/RESULTS/PEARSON\_EVALUATED

Select Dataset: /home/whitebreeze/XM\_v1.3/DATASETS/Pearsons/dataset\_pearson.csv

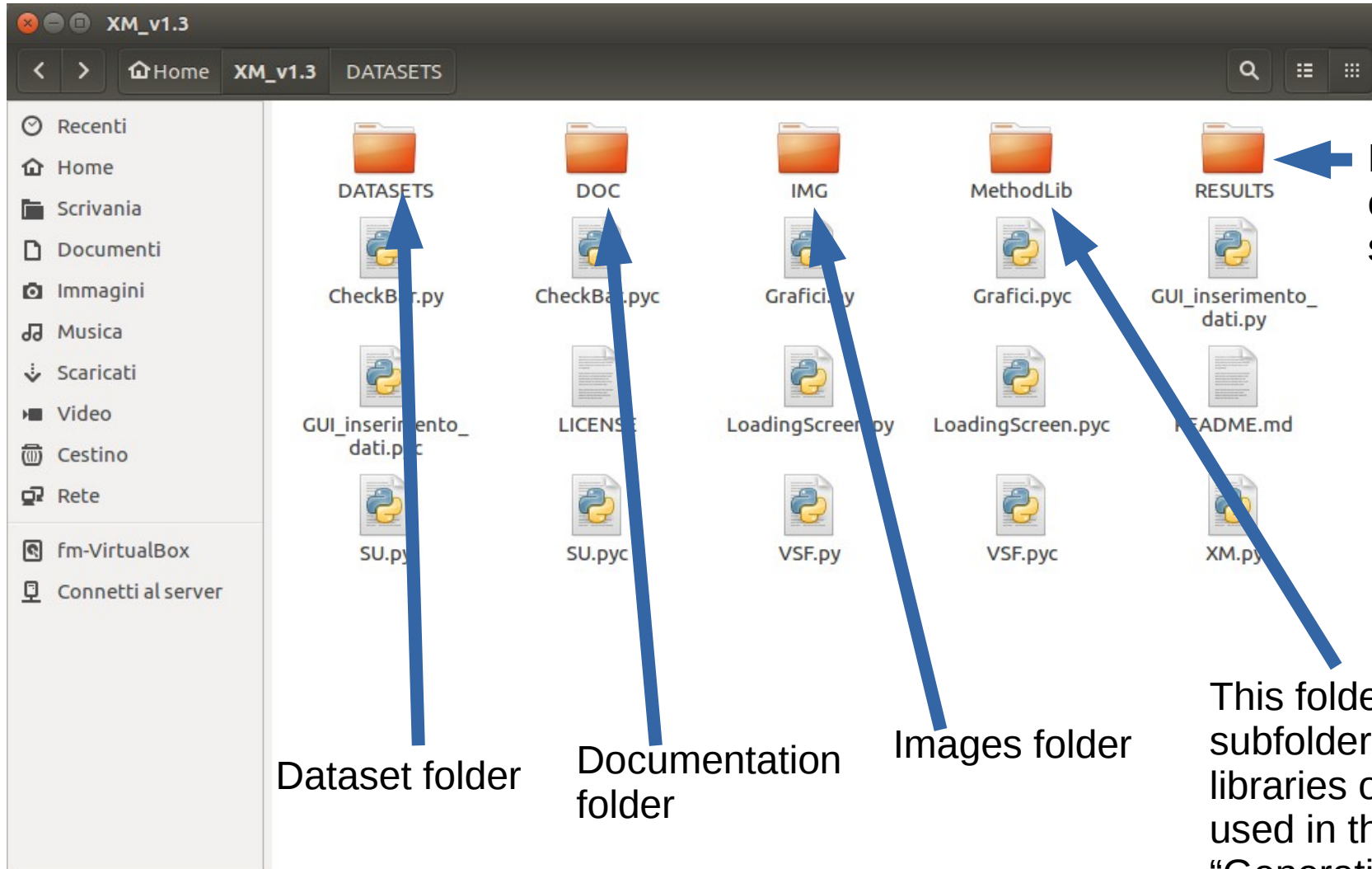
Status: Ready

If “Custom parameters” is selected this mask changes, allowing the user to input more specific values for the parameters.

## *4. Quick overview of folders and files*

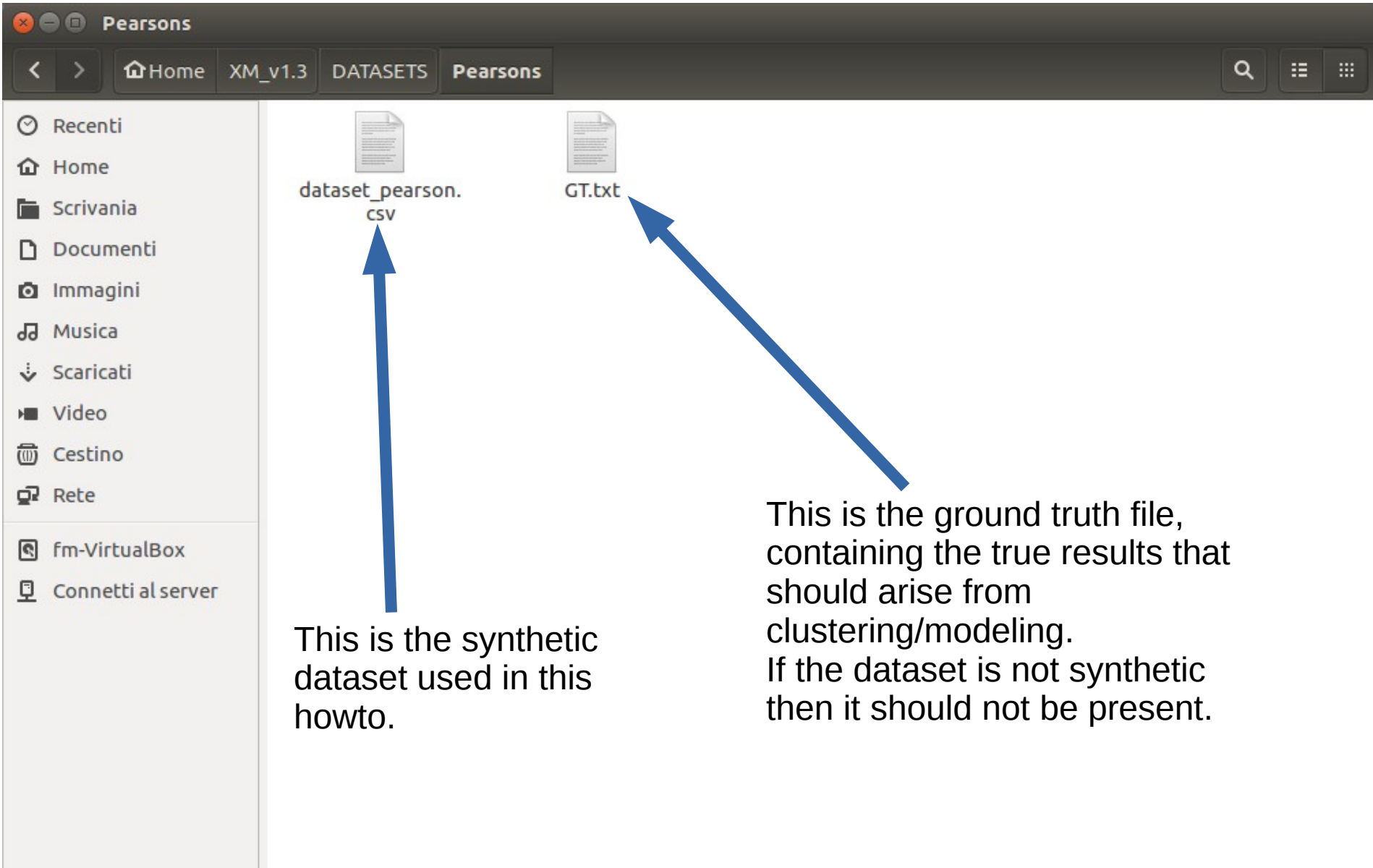


# Folders

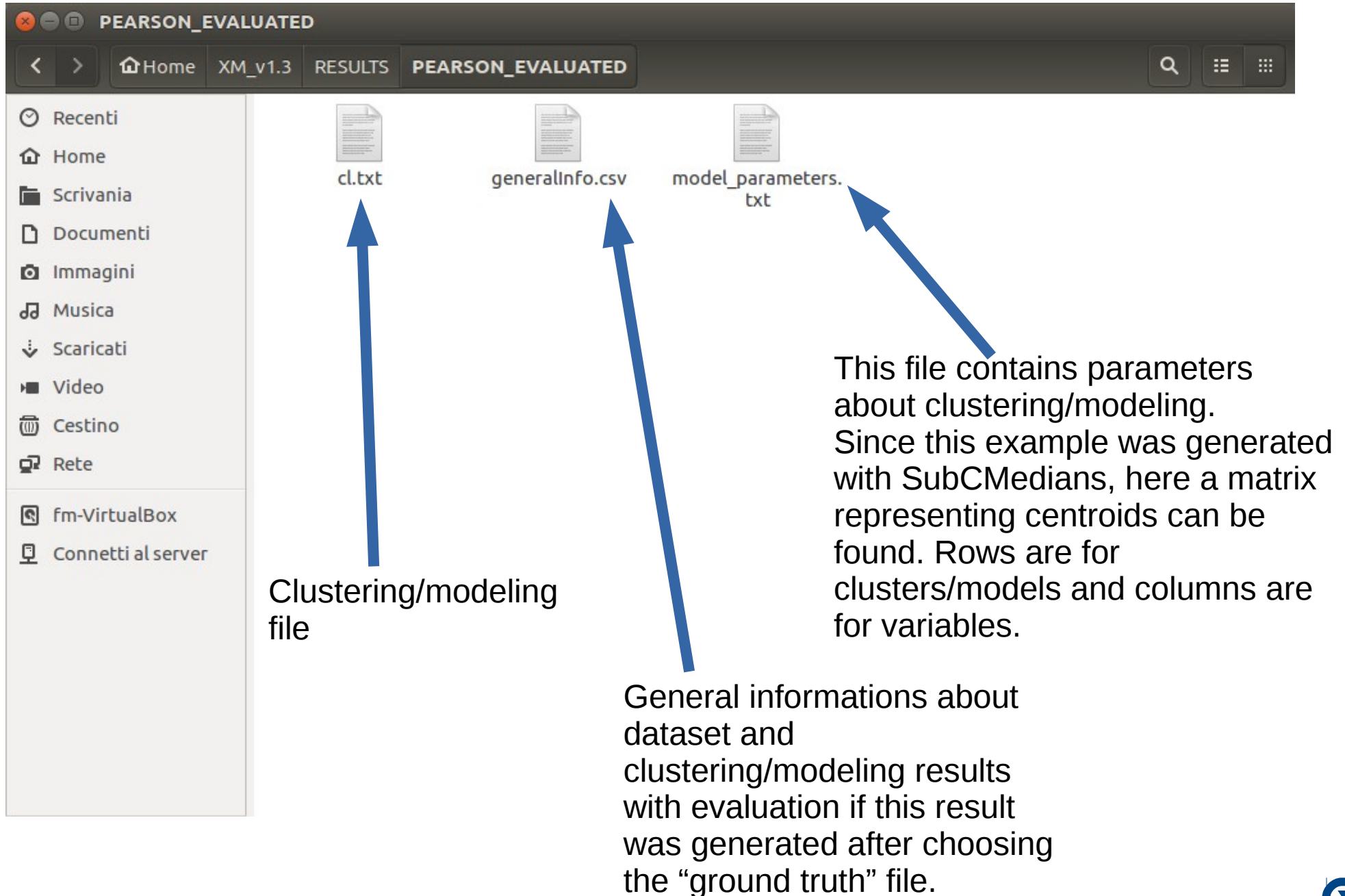




# View of a particular dataset folder



# View of a particular result folder



The screenshot shows a file explorer window titled "PEARSON\_EVALUATED". The breadcrumb path is "Home > XM\_v1.3 > RESULTS > PEARSON\_EVALUATED". The left sidebar lists various system folders. The main area displays three files: "cl.txt", "generalInfo.csv", and "model\_parameters.txt".

Clustering/modeling file

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

This file contains parameters about clustering/modeling. Since this example was generated with SubCMedians, here a matrix representing centroids can be found. Rows are for clusters/models and columns are for variables.