XM - eXplainable Modeling
How To Guide

Alberto Castellini, Alessandro Farinelli, Francesco Masillo

Verona University

13 July 2018

Overview

1. How to install the software

2. How to launch the Model Visualizer

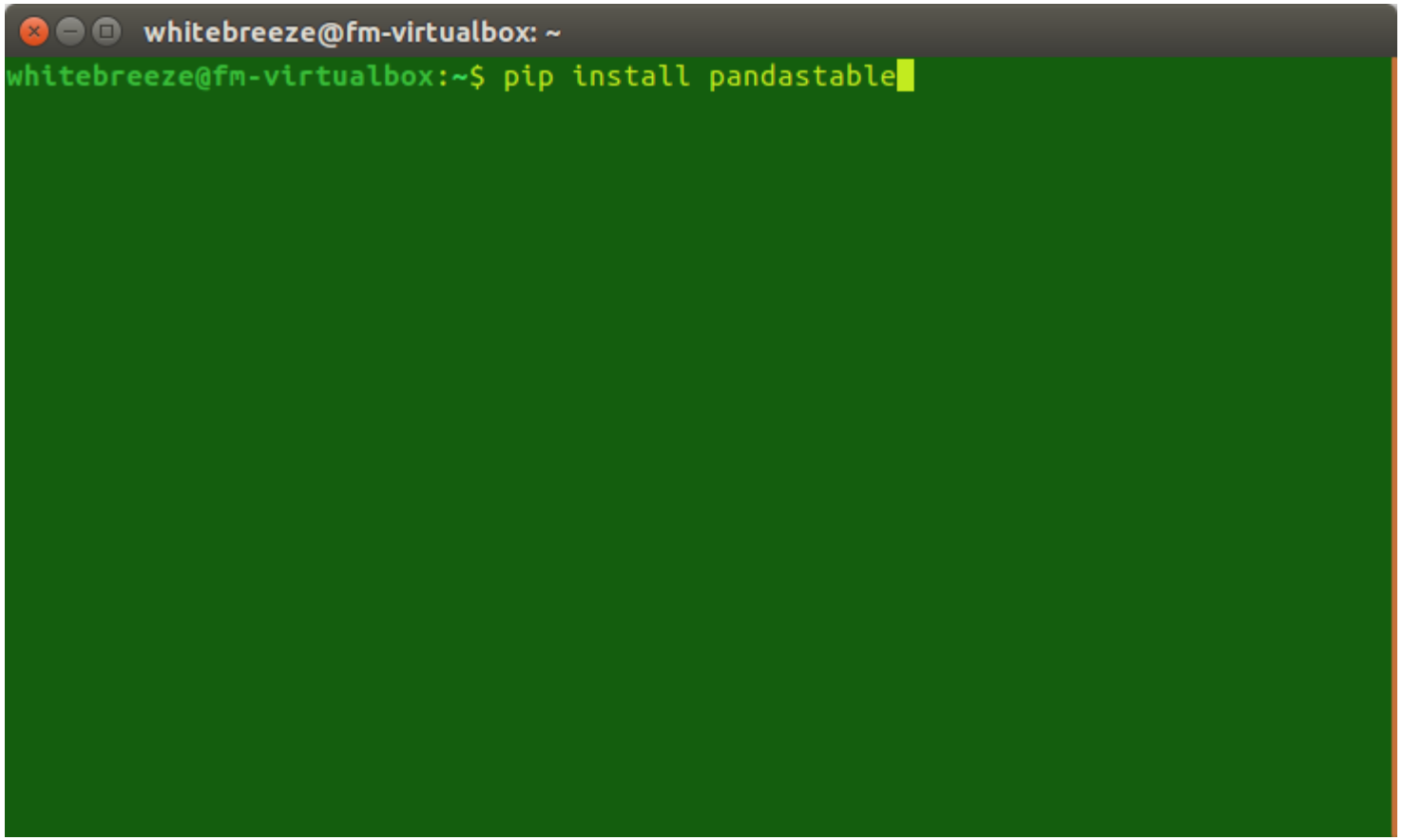
3. How to launch the Model Generator

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:



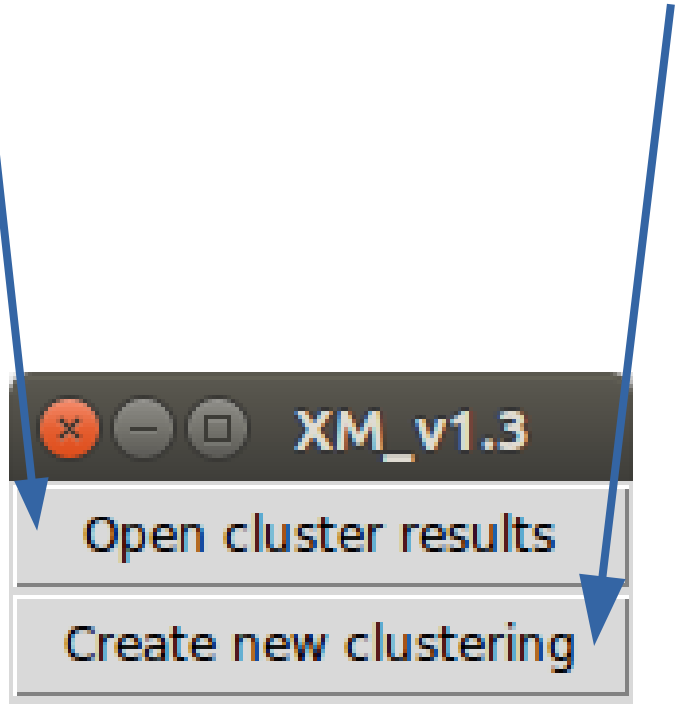
2. How to launch the Model Visualizer

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.

How to launch the Model Visualizer (cont'd)

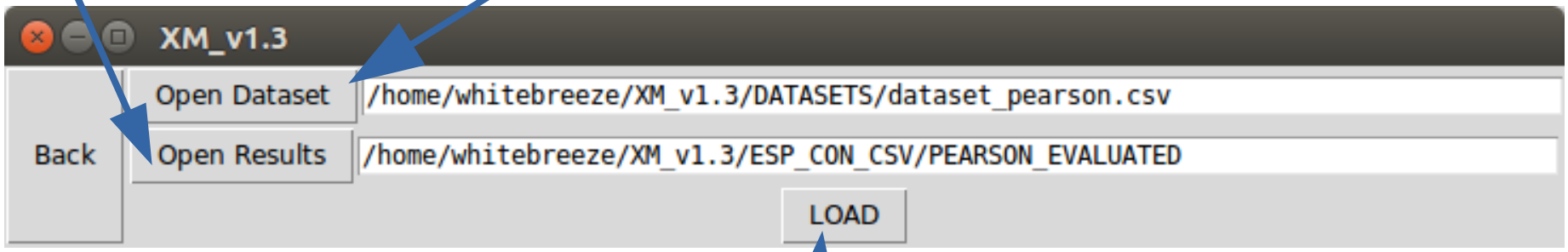
A new window should appear in the middle of the screen. It has two intuitive buttons, one for "Visualization" and one for "Generation".



For the visualization part "Open cluster results" must be clicked.

Selection of dataset and modeling result

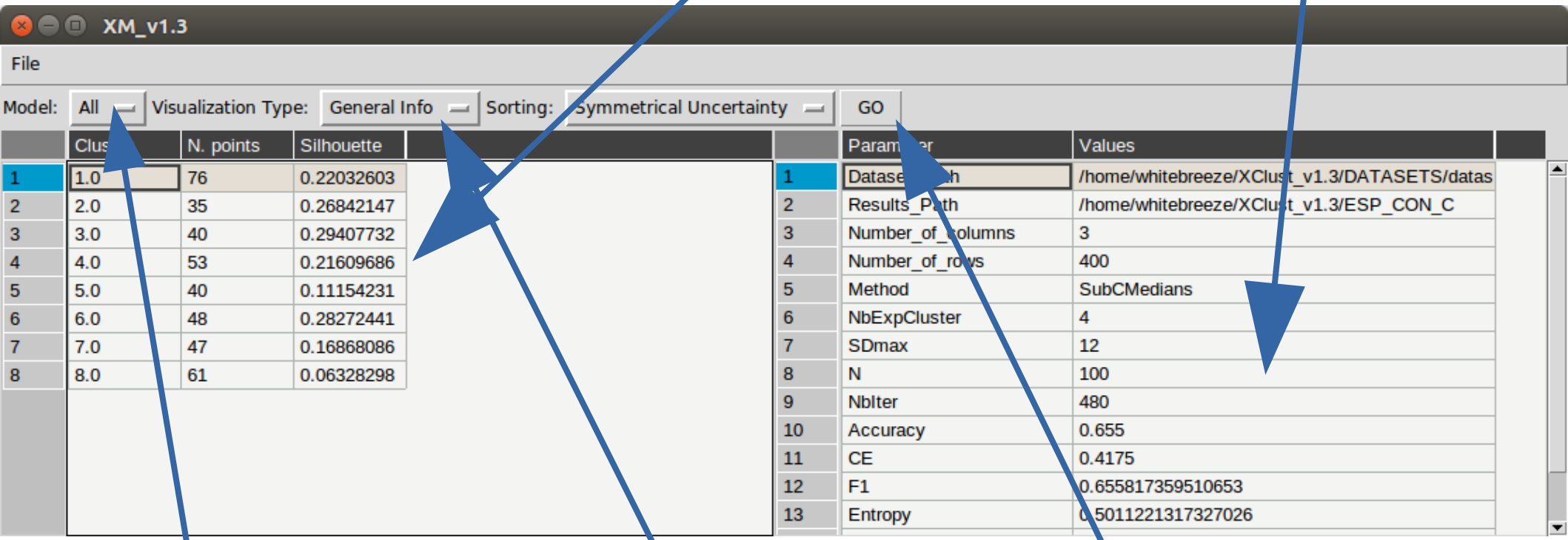
To select the dataset the user has to click on “Open Dataset”. Then a dialog will show up to allow the user to select the file. Same goes for the selection of the directory of the results. Here are shown those two paths already selected.



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.

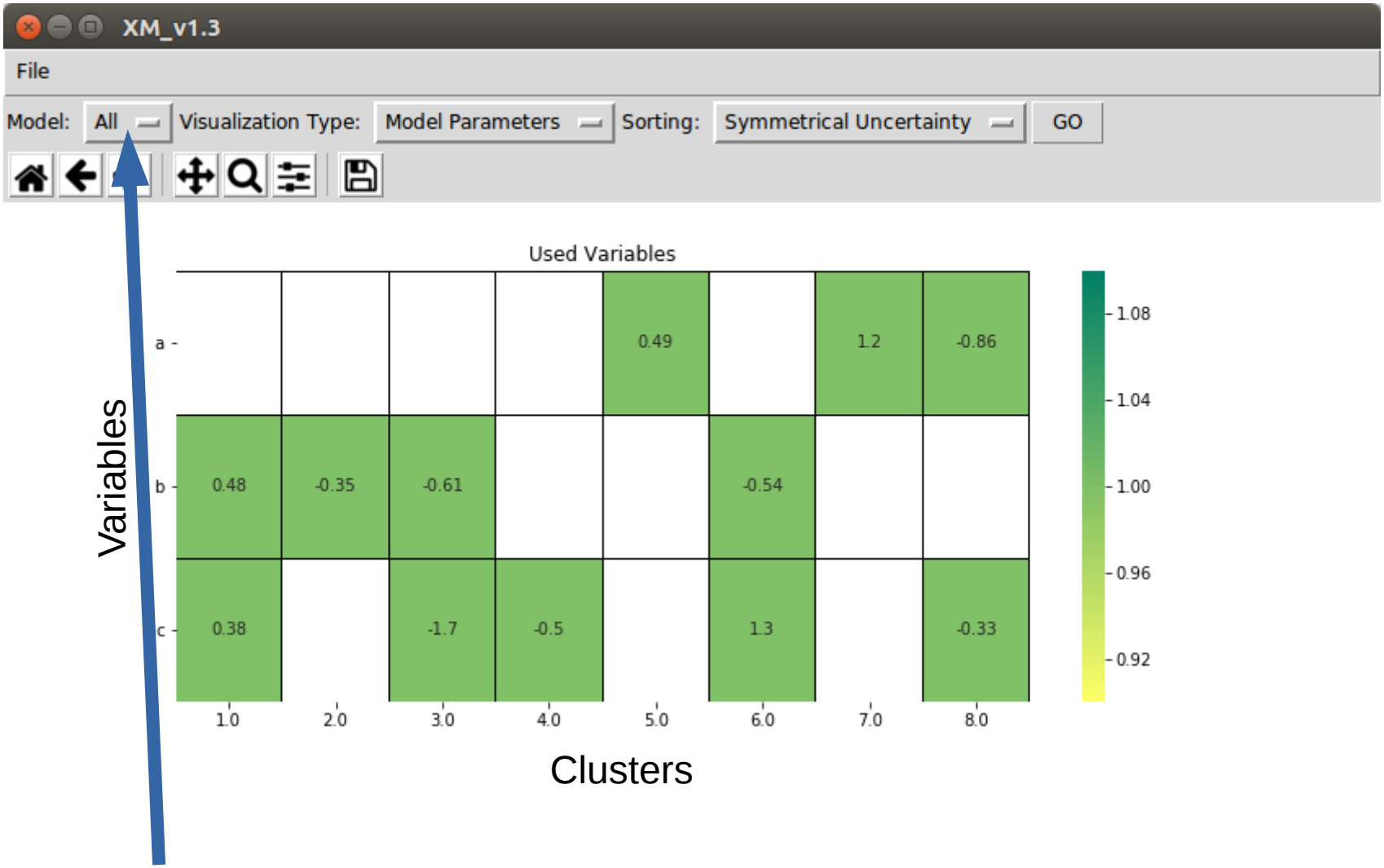


Select a type of visualization.

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

Model Parameters Tab

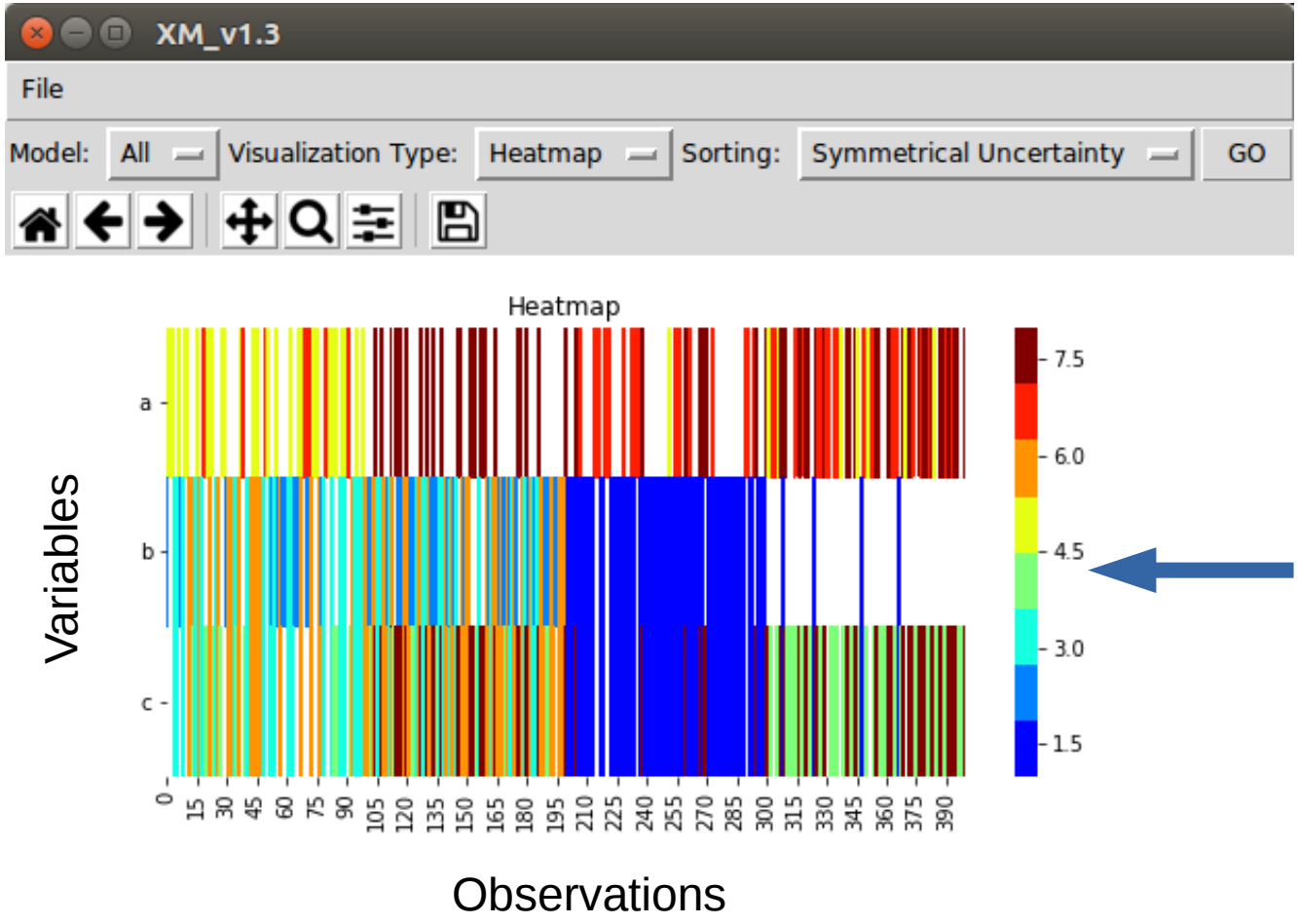
"Model parameters" visualization.
Each column represents a cluster centroid. Green cells are selected variables, white cells are non selected variables



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

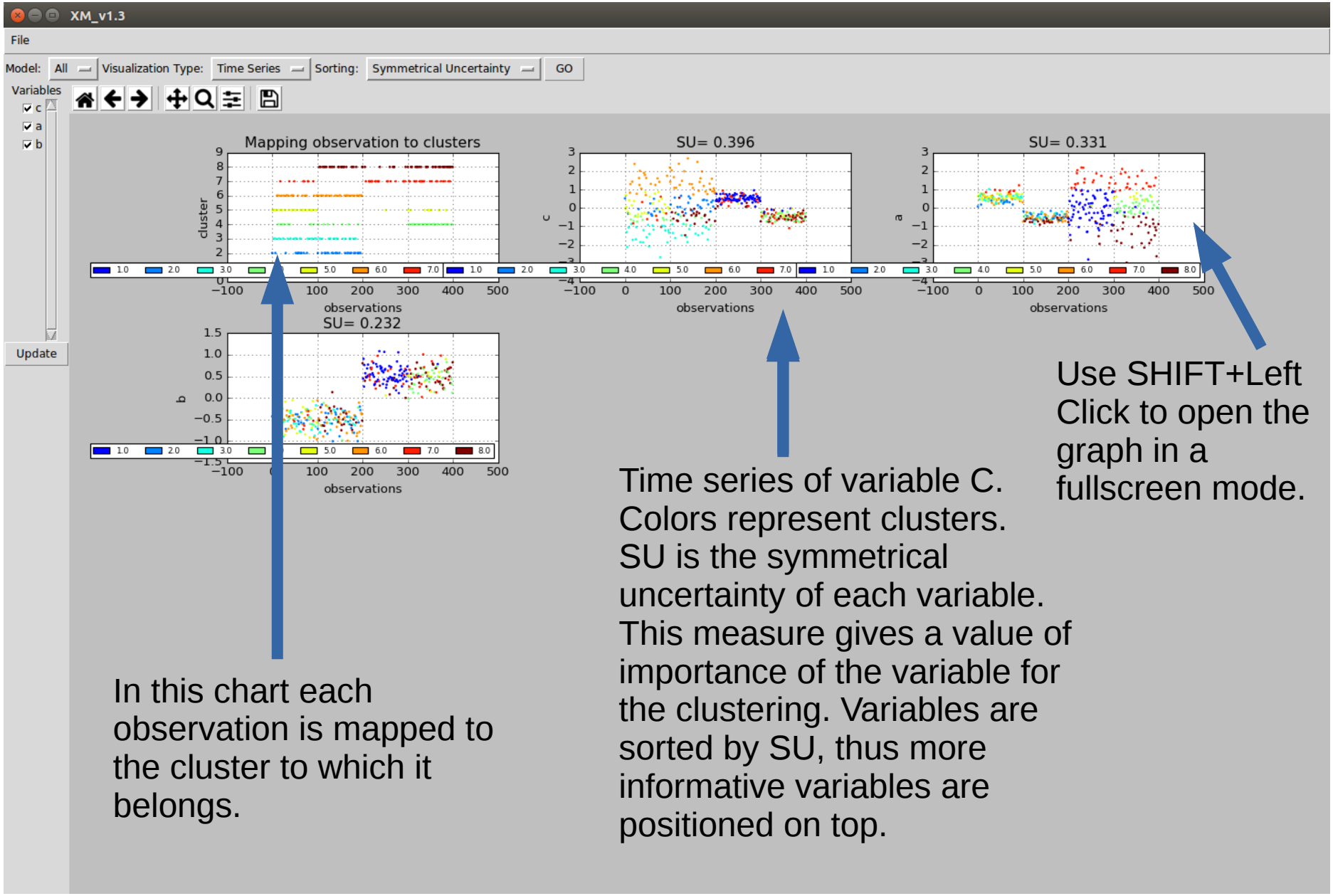
Heatmap Tab

"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

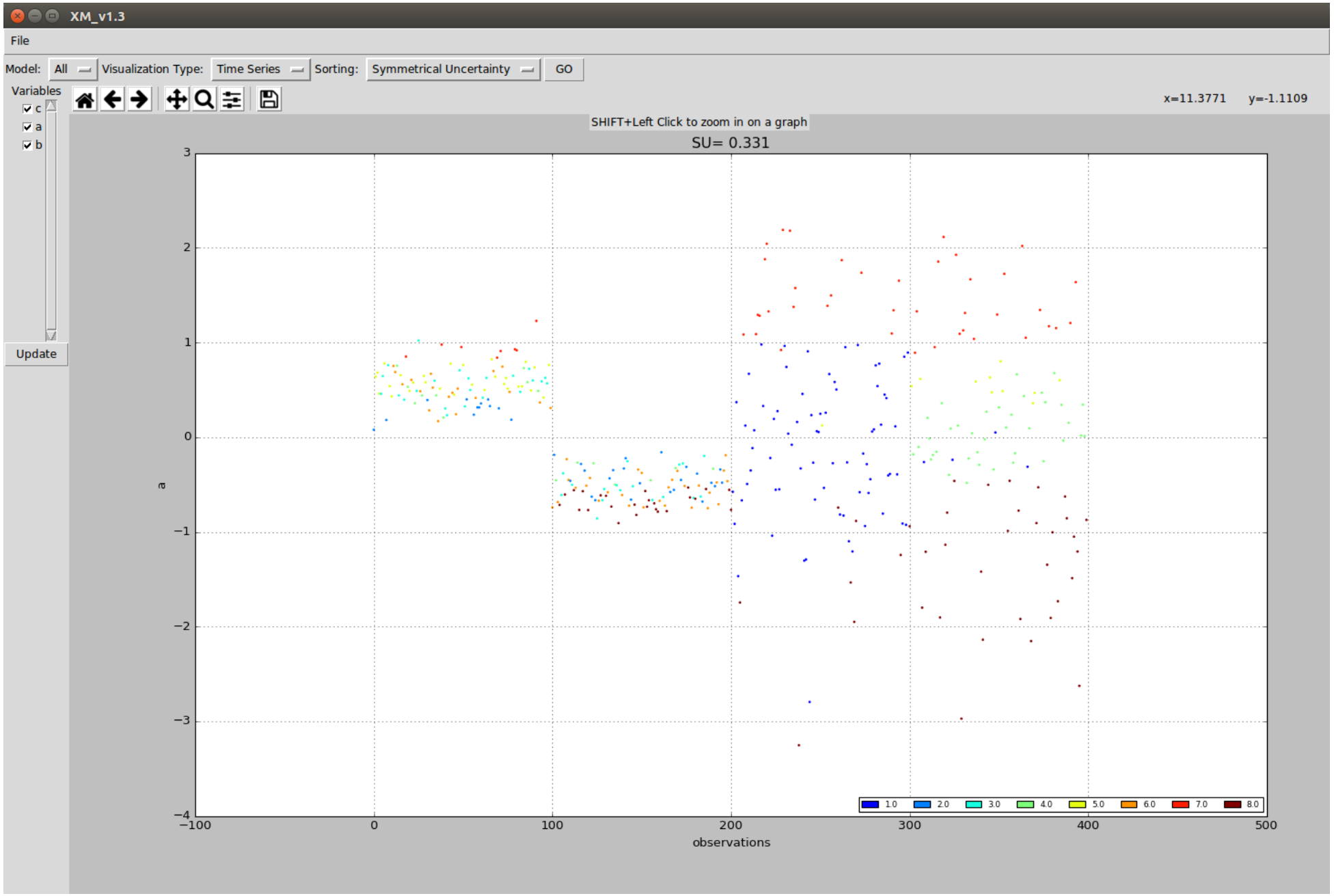


In this chart each observation is mapped to the cluster to which it belongs.

Time series of variable C. 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.

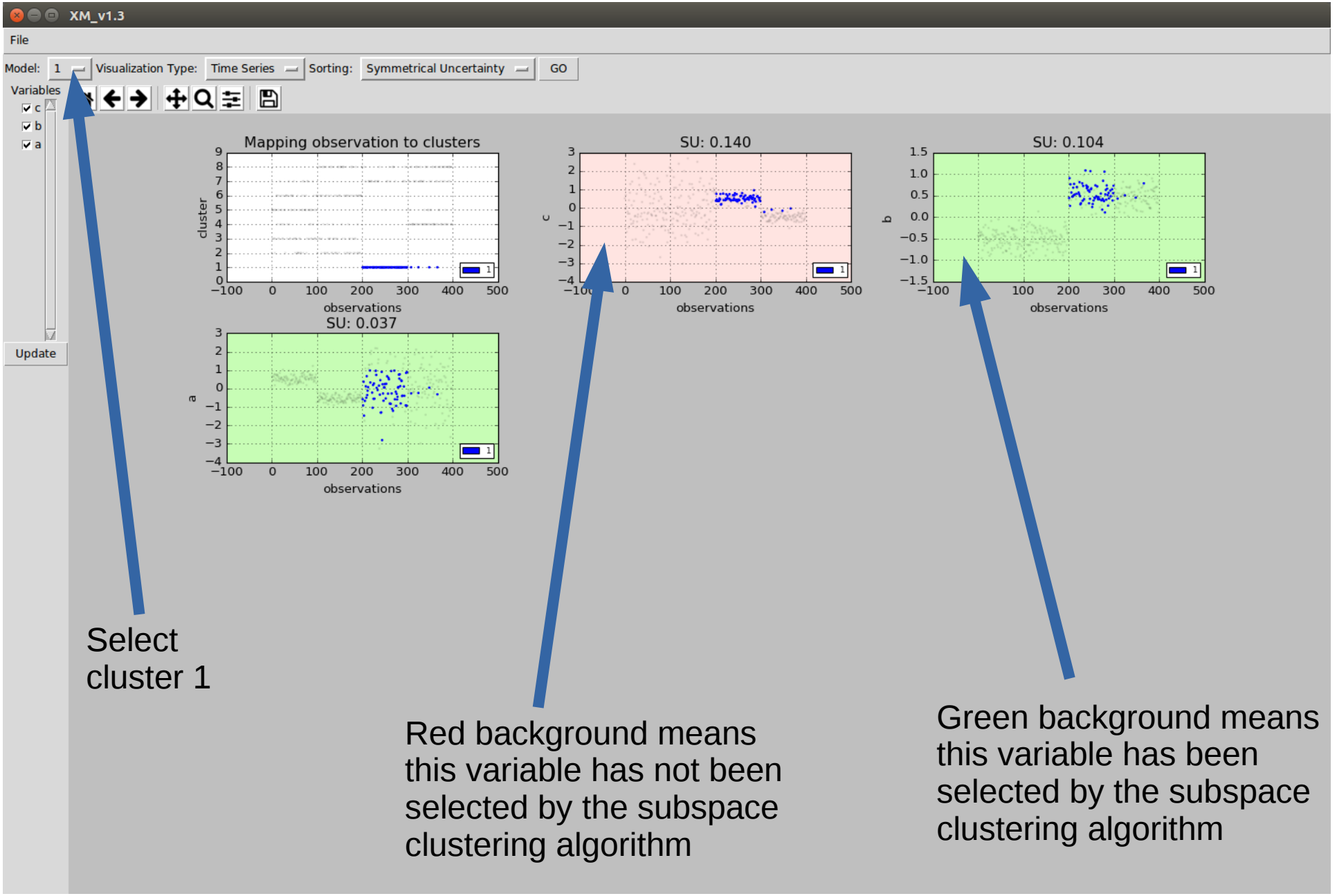
Use SHIFT+Left Click to open the graph in a fullscreen mode.

Time Series Tab fullscreen mode

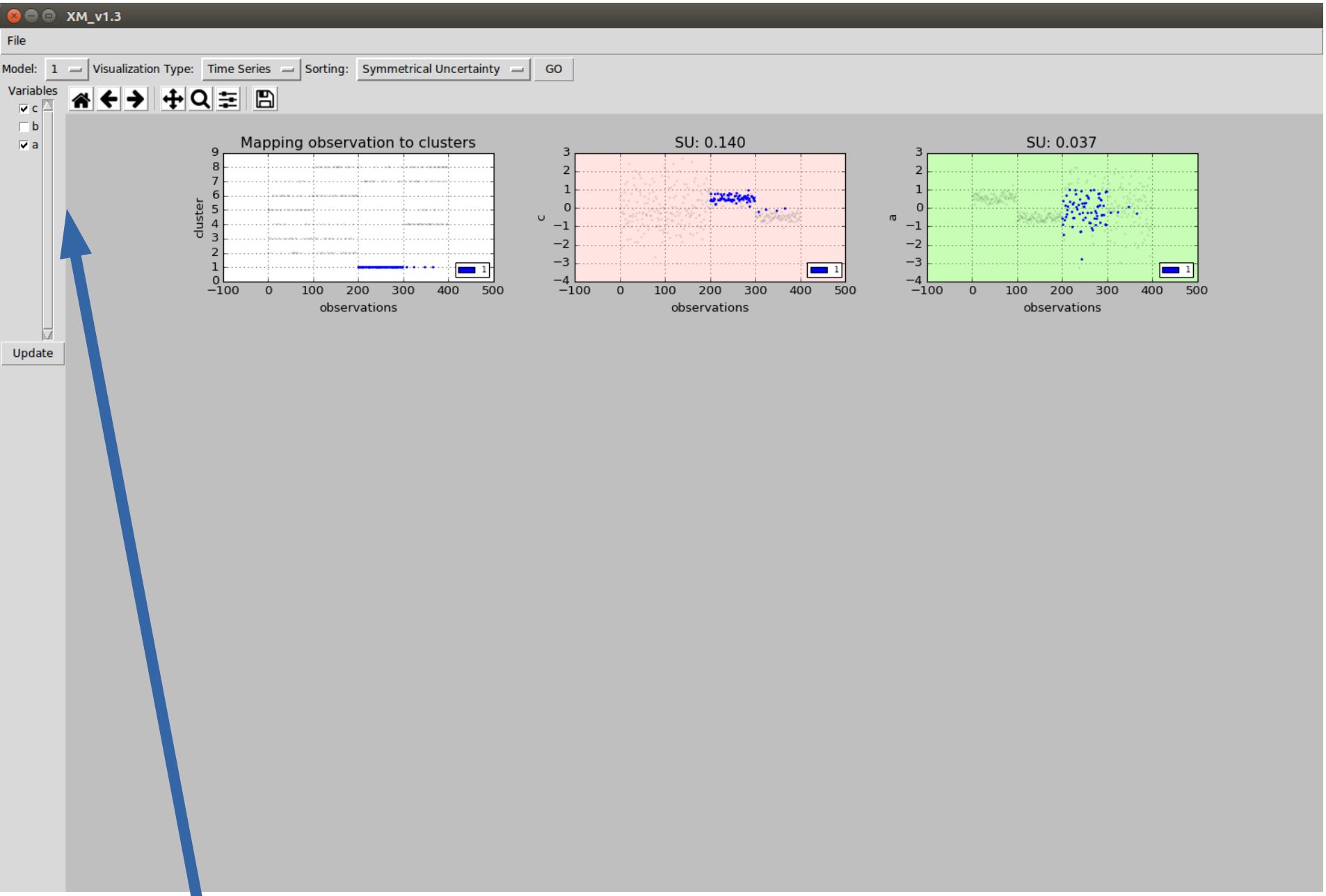


Press again SHIFT+Left click to go back to the standard view.

Time Series Tab with selection of specific model (cluster)

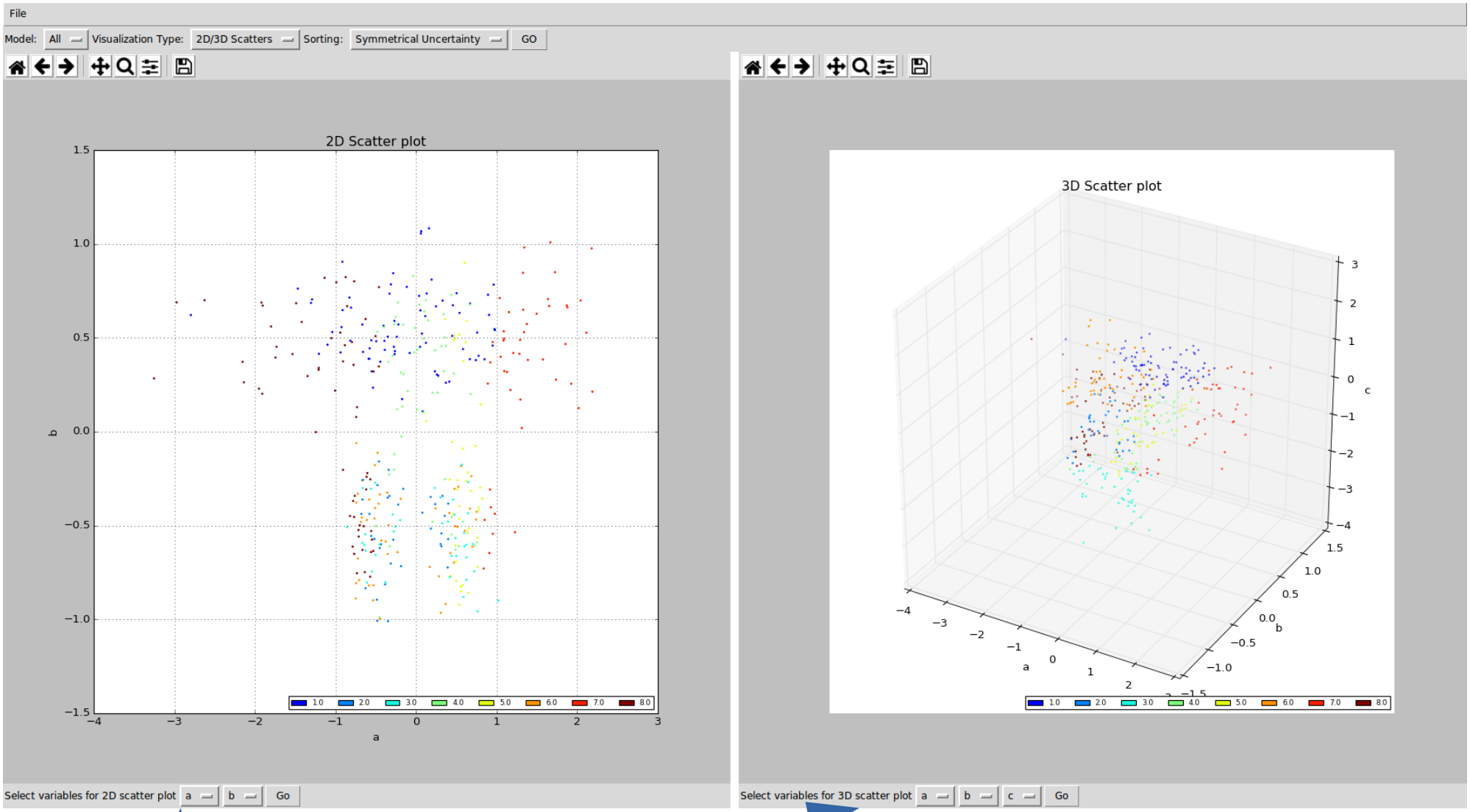


Time Series with selection of specific variables



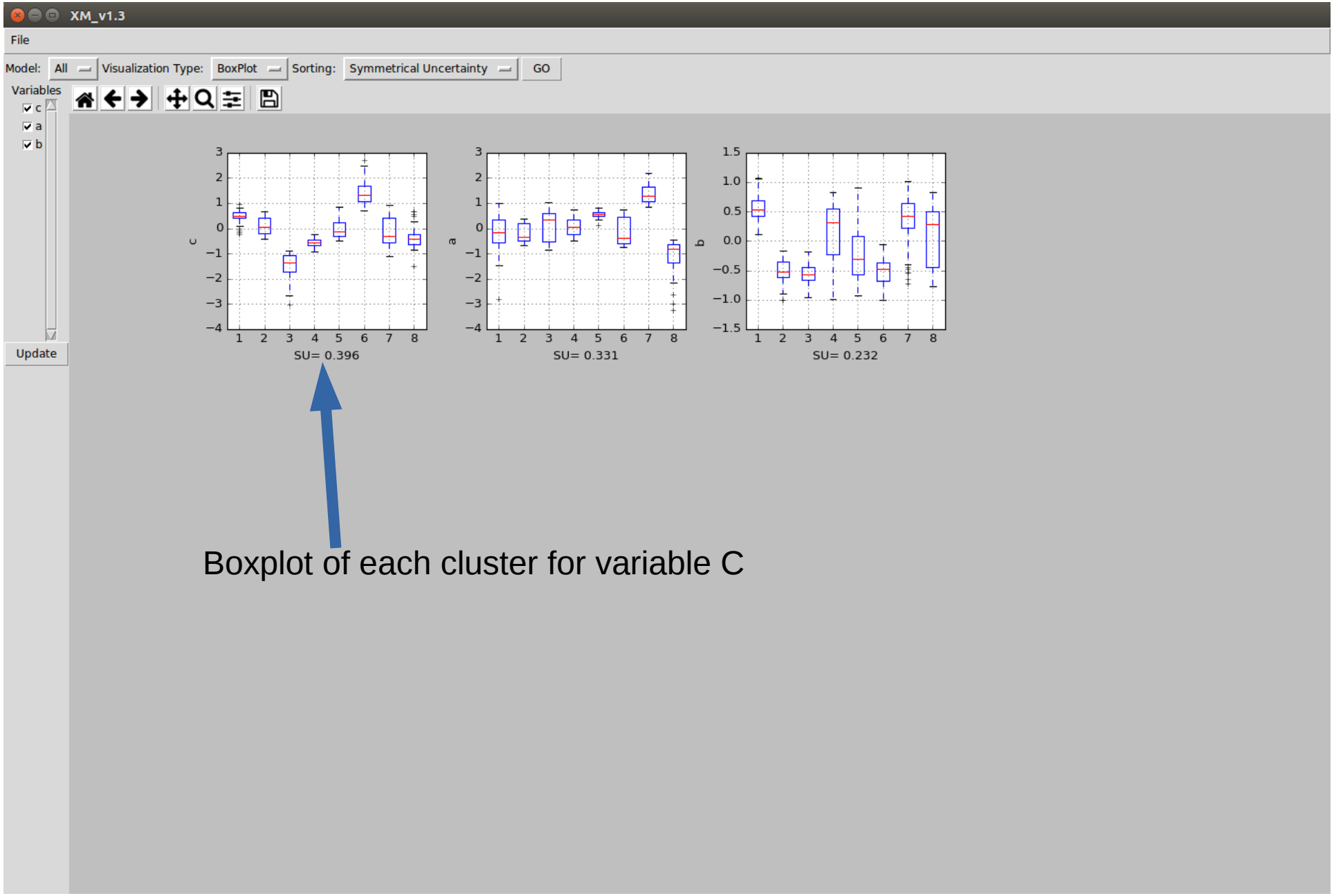
On the left user can select which variables are shown on the screen (min 1, max 10 variables)

2D/3D Scatter plot Tab



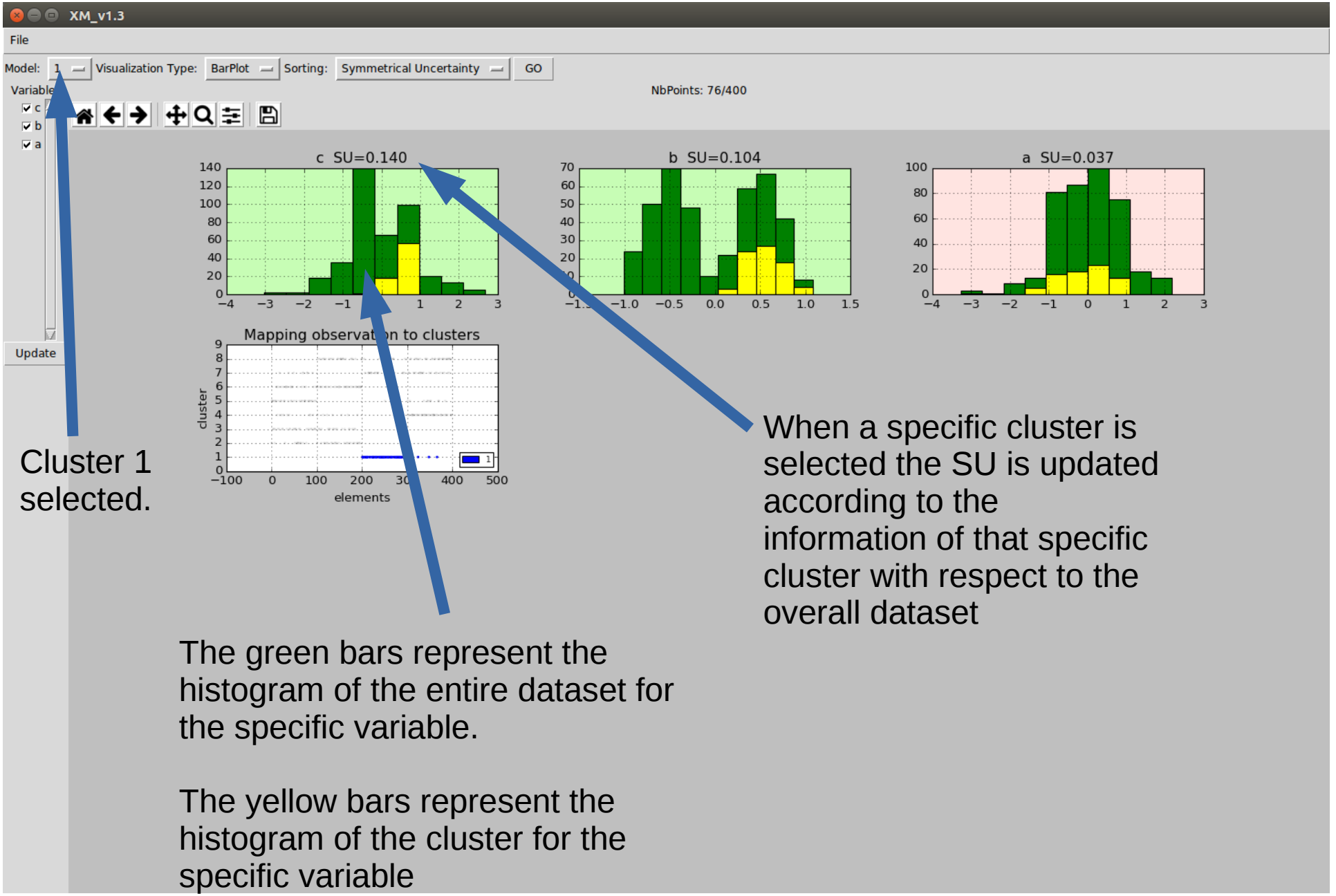
Select variables for each axes of the 2D/3D scatter plot

Boxplot Tab



Boxplot of each cluster for variable C

Barplot Tab



Cluster 1 selected.

When a specific cluster is selected the SU is updated according to the information of that specific cluster with respect to the overall dataset

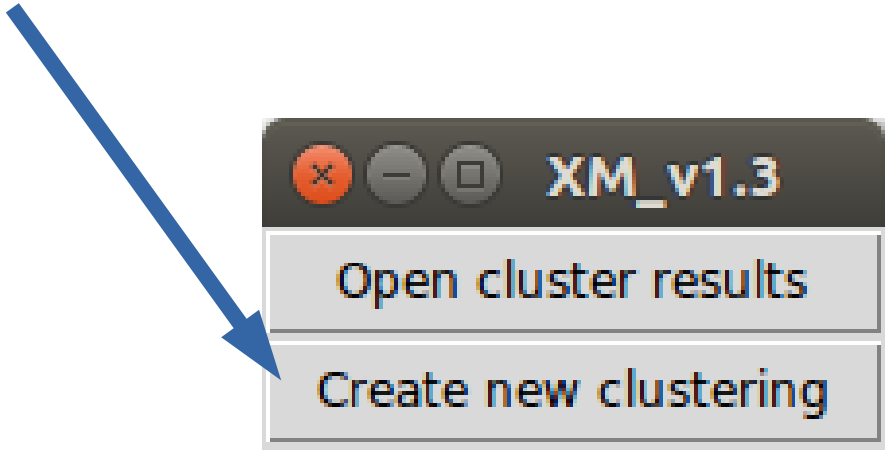
The green bars represent the histogram of the entire dataset for the specific variable.

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

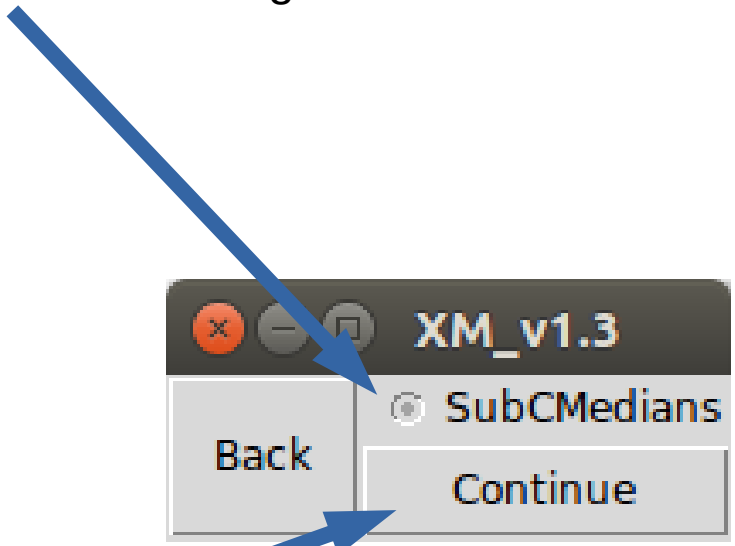
3. How to launch the Model Generator

3. How to launch the Model Generator

Click on “Create new clustering” button.



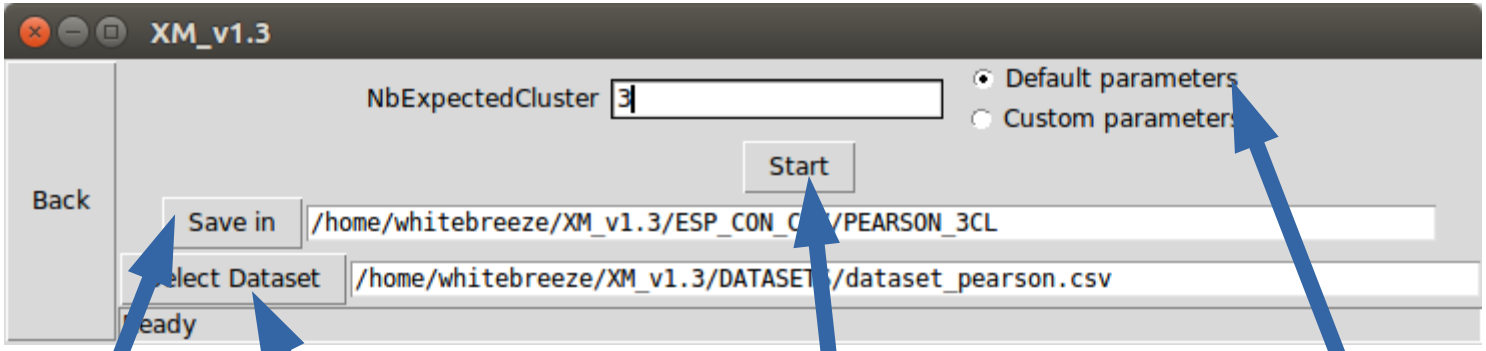
Choose the method used to generate new results.



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.



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

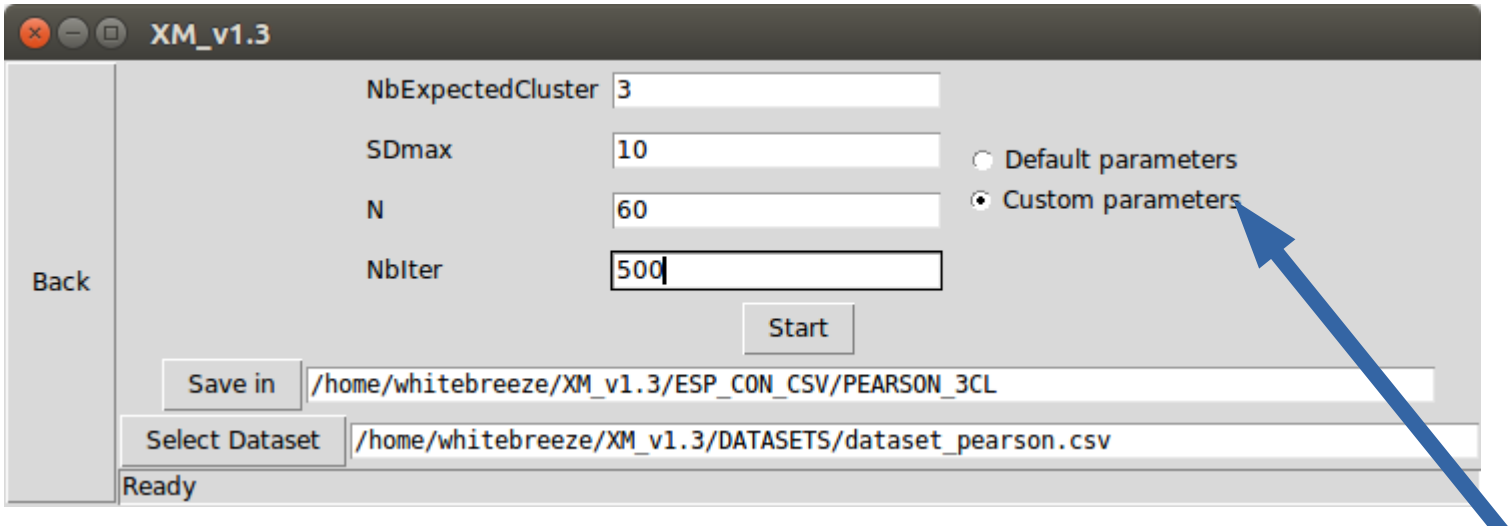
Select the dataset file.
If a file is selected then its name will appear in the text area.

When ready click Start.

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



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