1   Topic

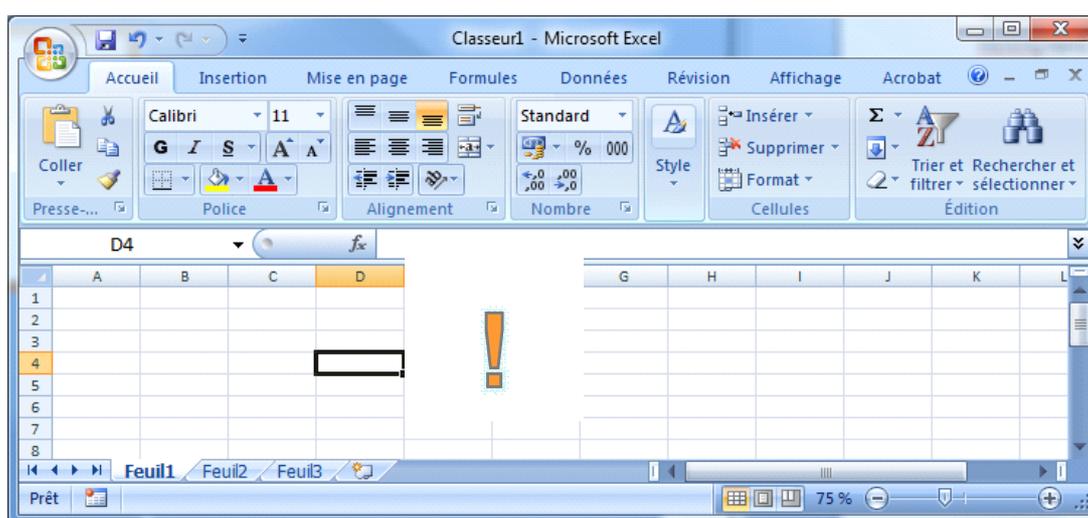**Installing and using the add-in "Tanagra.xla" in Office 2007 and 2010.**

The "tanagra.xla" add-in for Excel contributes to the wide diffusion of Tanagra. The principle is simple. It is to embed a Tanagra menu in Excel. Thus the user can run statistical calculations without having to leave the spreadsheet. It seems simplistic. But this feature facilitates immensely the work of data miner. Indeed, the spreadsheet is one of the most used tools for preparing dataset (see KDNuggets Polls: Tools / Languages for Data Cleaning - 2008). By embedding the data mining tool in the spreadsheet environment, it avoids to the practitioner the tedious and repetitive manipulations: importing the dataset, exporting the dataset, checking the compatibilities between data file formats, etc.

The installation and the use of the "tanagra.xla" add-in under the previous versions of Office are described elsewhere (Office 1997 to Office 2003 - see http://data-mining-tutorials.blogspot.com/2008/10/excel-file-handling-using-add-in.html). This description is obsolete for the latest version of Office because the organization of the menus is modified for these versions i.e. Office 2007 and Office 2010. And yet, the add-in is still operational. **In this tutorial, we show how to install and to use the Tanagra add-in under Office 2007 and 2010**.

This transition to recent versions of Excel is absolutely not neutral. Indeed, compared to the previous Excel versions, Excel 2007 (and 2010) and can handle more important rows and columns. We can process a dataset up to **1,048,575 observations** (the first line corresponds to the variable names) and **16,384 variables**. In this tutorial, we will treat a database with 100,000 observations and 22 variables (wave100k.xlsx). This is a version of the famous waveform database. Note that this file, because of the number of rows, cannot be manipulated by earlier versions of Excel.
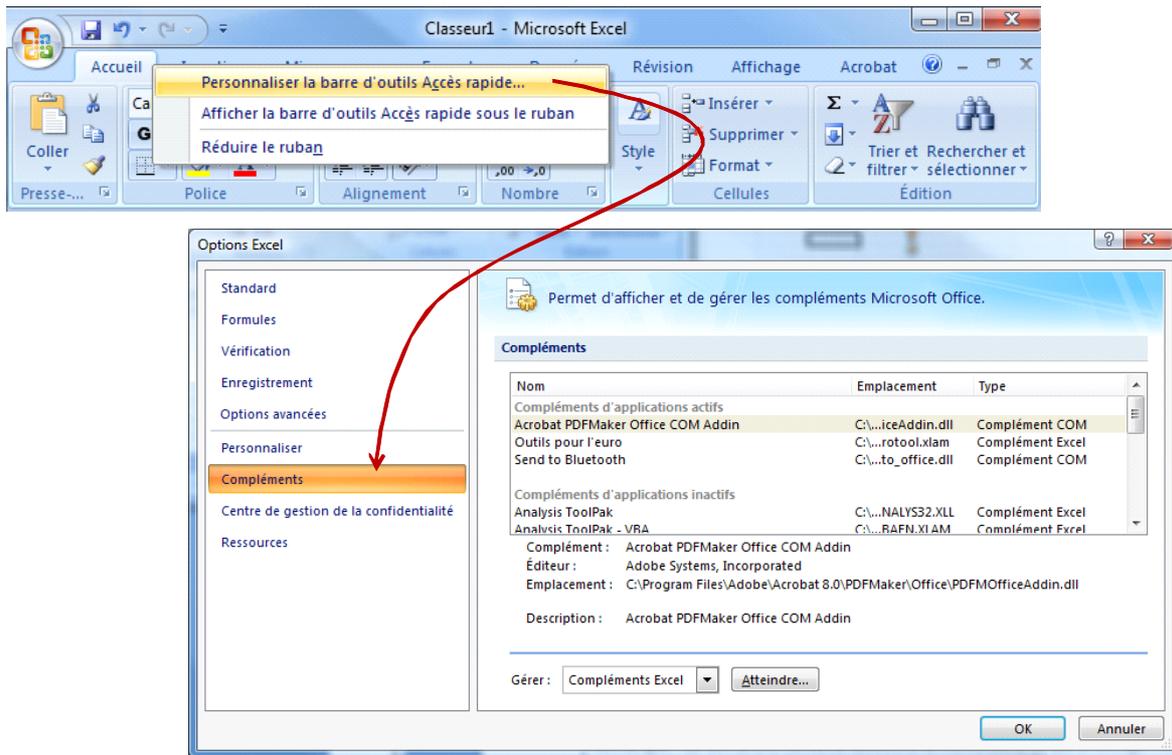
The process described in this document is also valid for the SIPINA add-in (sipina.xla).

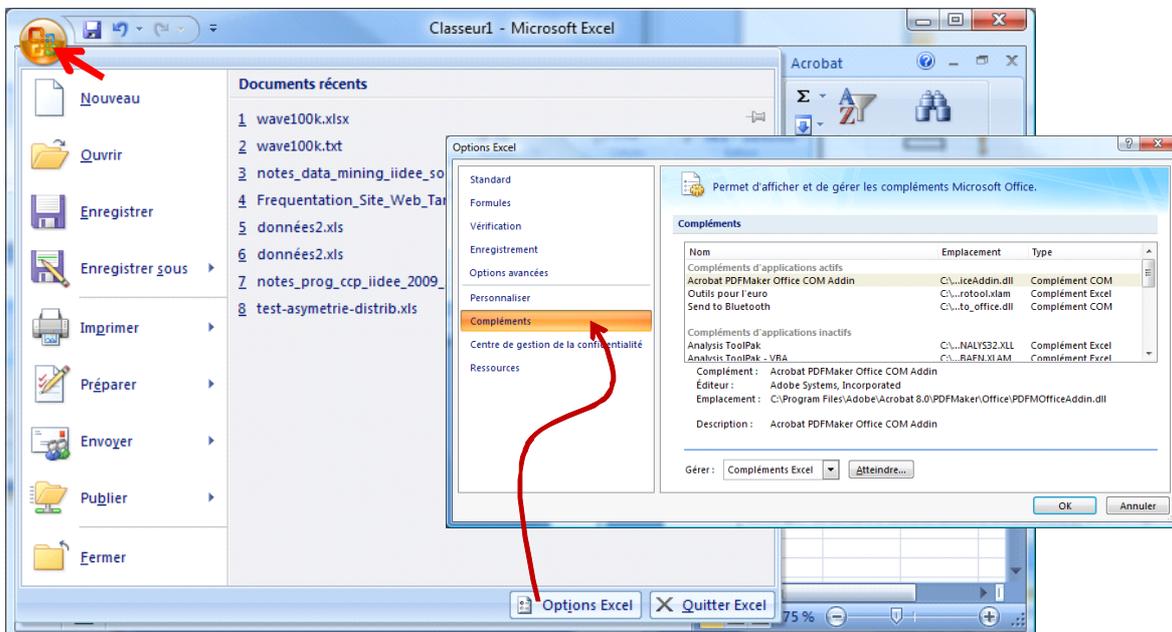# 2   Installing the add-in under Excel 2007

After starting Excel, we obtain the window above. To install the add-in, we perform a right-click on one of the top Ribbon tabs. On the shortcut menu, we click on "Personnaliser la barre d'outils Accès rapide..." (*Note*: Unfortunately, the menus are in French, I hope the reader can find easily the
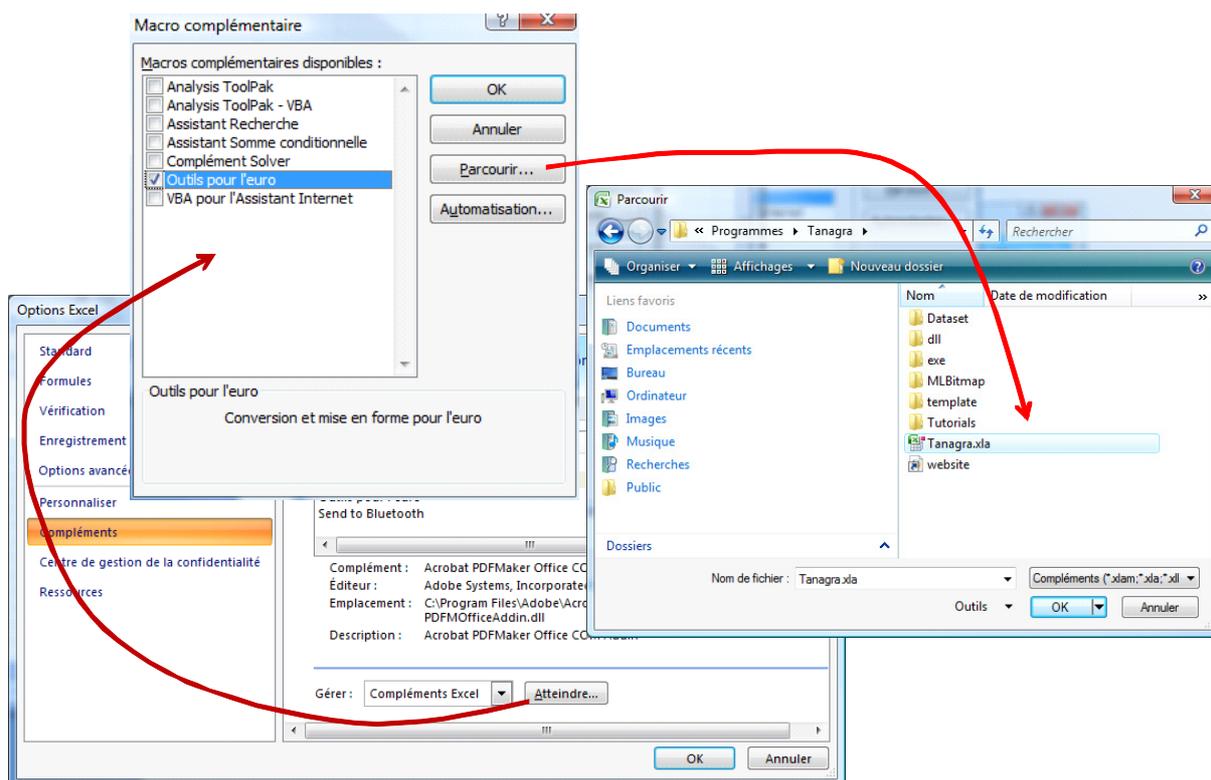
corresponding menu in English version). A dialog box appears. We select the item "Compléments" dedicated to the management, among other things, of add-ins.
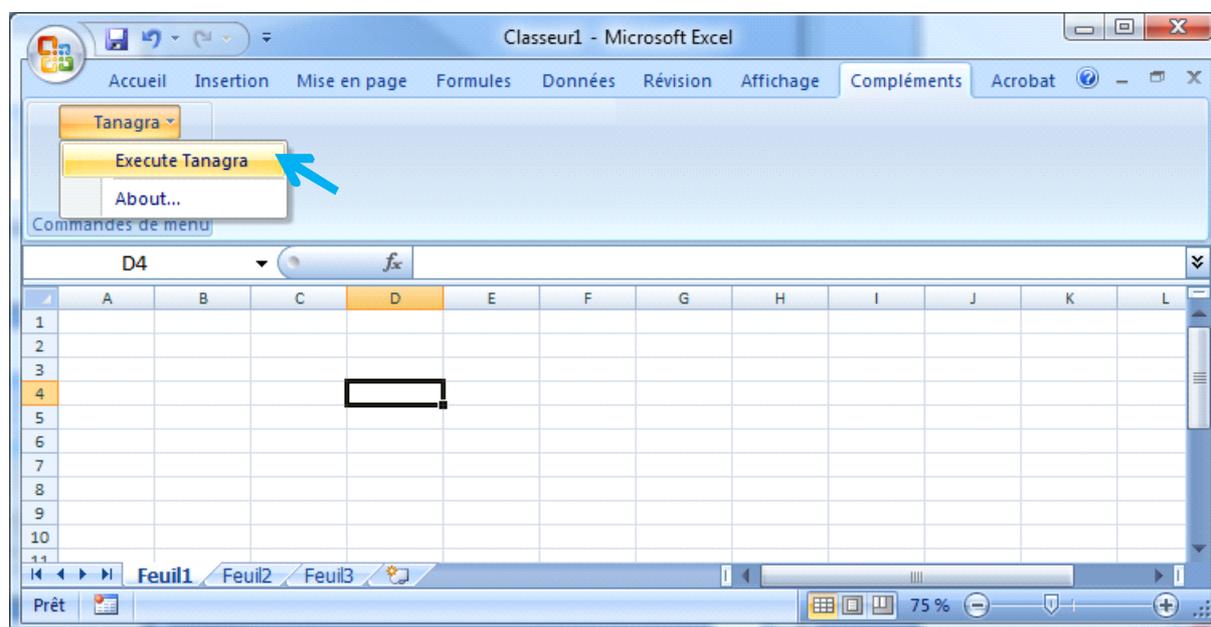
Note: *The following process leads to the same dialog box. We use the Office button at the left corner of the main window. Then we click on the "Options Excel" button.*

We observe the list of installed add-ins. At the bottom of the window, we click the button "Atteindre" for "Compléments Excel". We obtain the management dialog box of add-ins. It is identical to the one of previous versions of Excel. To embed "tanagra.xla", we click on "Parcourir". We select the file in the directory setup of Tanagra.
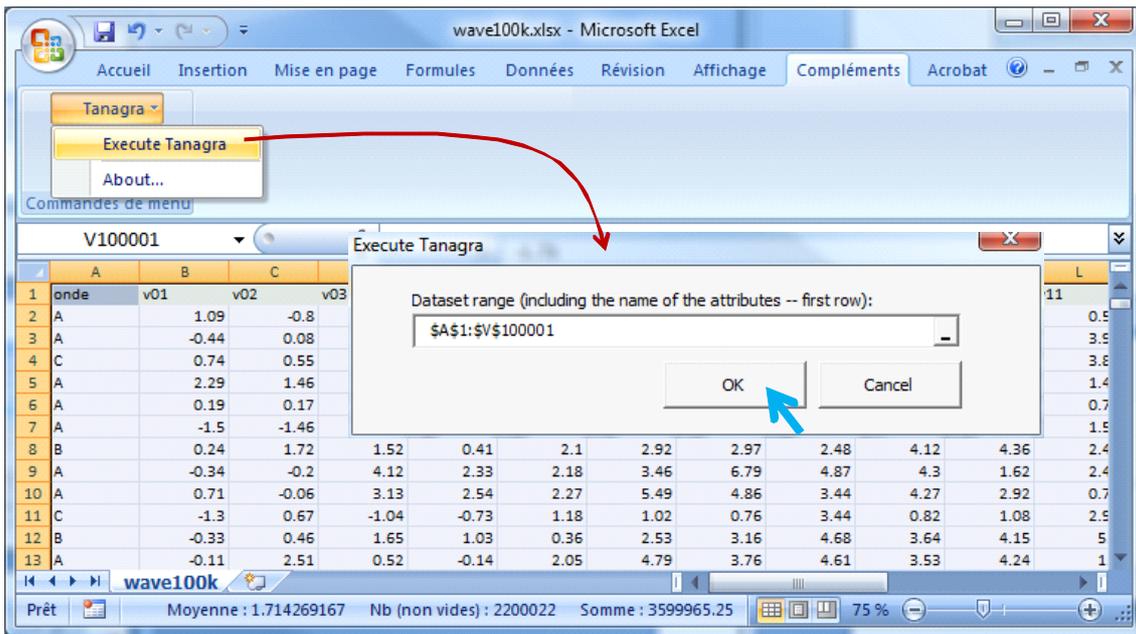
We validate this configuration. The new feature is now permanently installed in the "Compléments" Excel Ribbon tab i.e. it will be available even after restarting of Excel.
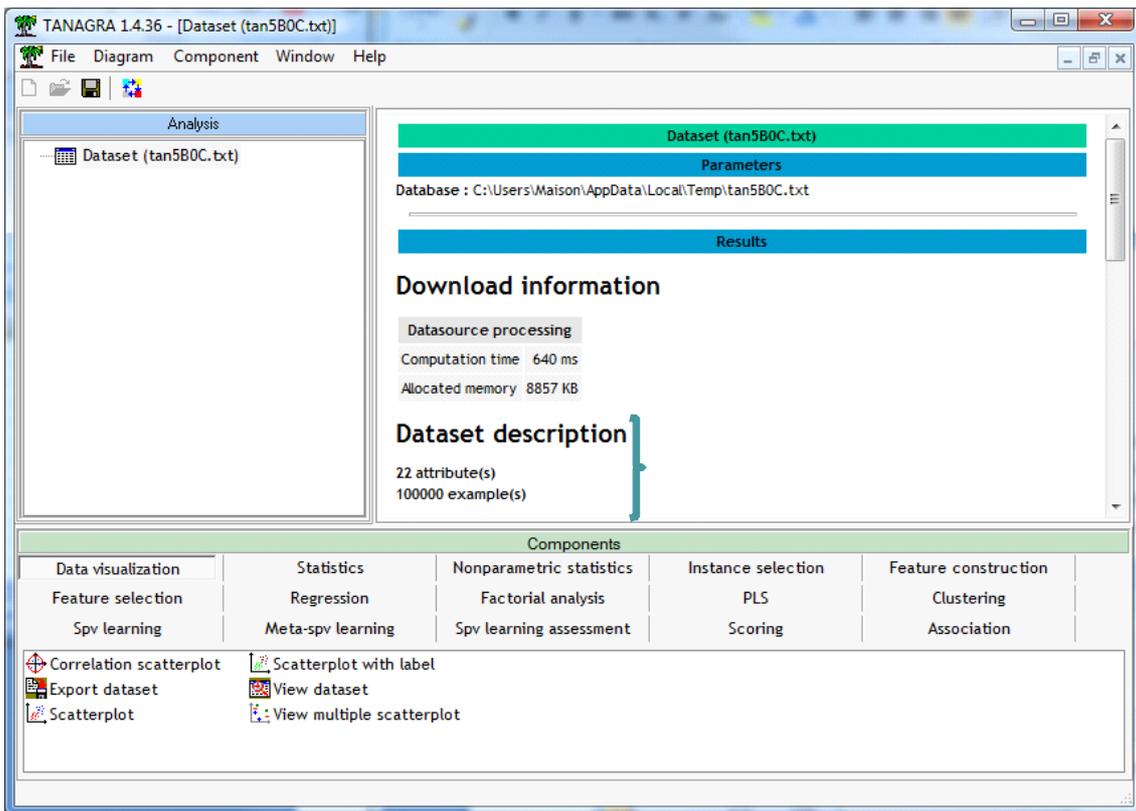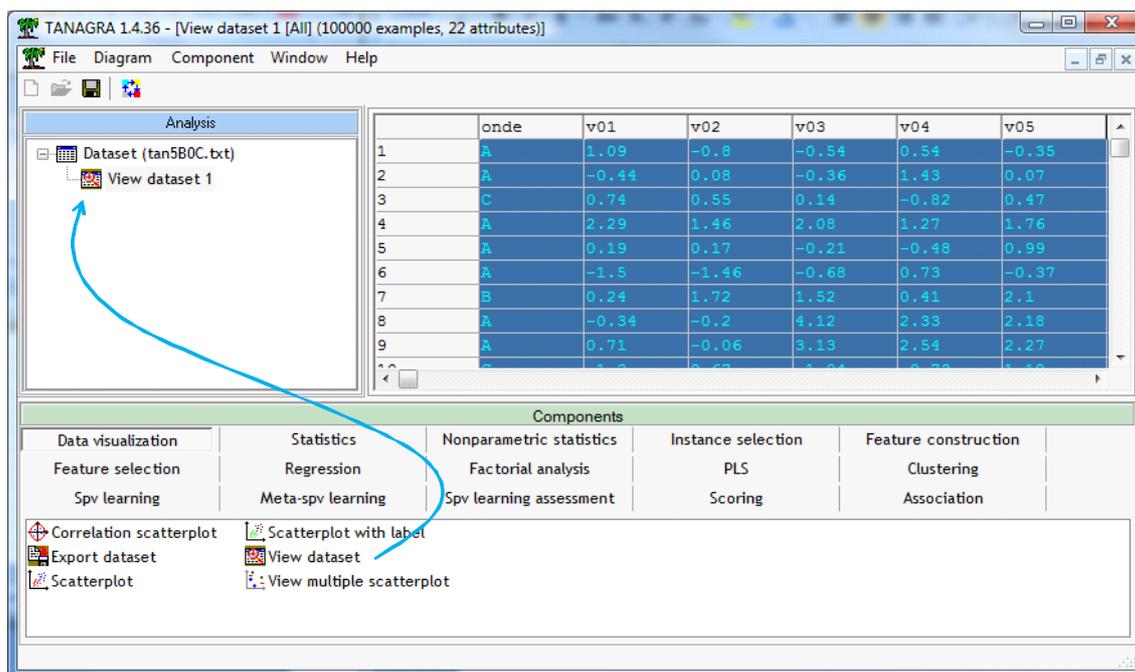

3   Using the add-in - Excel 2007

The utilization of the add-in into Excel 2007 is the same as for previous versions of Excel. We load the "wave100k.xlsx" data file. We select the range of dataset. Then, we activate the menu "Compléments / Tanagra / Execute Tanagra…". A dialog box appears. We check the cells range. Last, we click on the OK button.

Tanagra appears very quickly. A new diagram is created and the data is loaded.
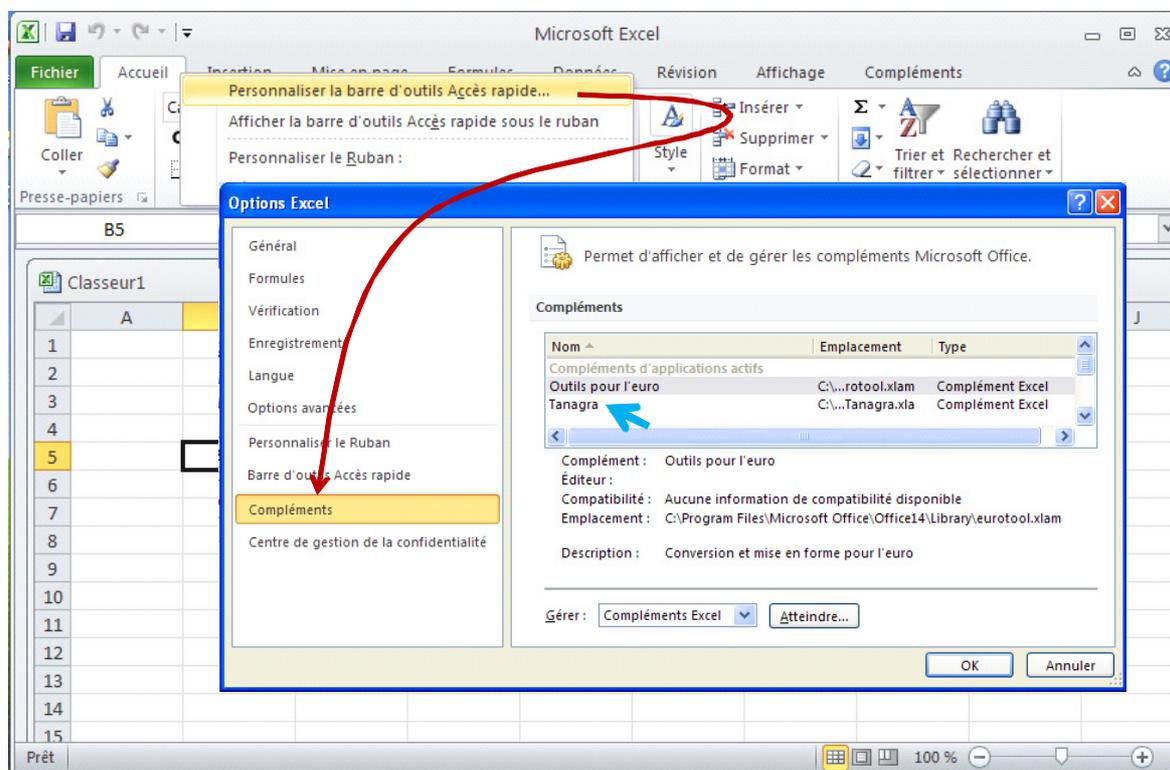

The VIEW DATASET component (DATA VISUALIZATION tab) enables to visualize the dataset.

4   Installation of the Add-in into Excel 2010

It was important to check if the procedure described in this document is valid for the latest version of Excel (Office 2010). We have the "beta" version of Office, but there should not be any notable changes with the final version. To install the add-in, we set the right click on one of the tabs of the ribbon. We select the "Personnaliser la barre d'outils d'Accès rapide" menu. We come to the settings dialog of add-ins.

The subsequent operations are the same than under Office 2007.

5   Conclusion

The add-in "tanagra.xla", defined originally for Office 2000, remains operational on Office 2007, and most recently, on the latest version Office 2010. We note that the transfer time is still reasonable, even when we handle large files (e.g. for the 500,000 rows version of the waveform dataset, the importation time is about 4 seconds).

The same add-in system is available for Open Office. It works under Windows and under Linux.