

## Tutorial 2

### VARIABLE, TIME and PREPROCESS sections

This tutorial describes how to select a time series for subsequent nonlinear analysis (the VARIABLE section), how to specify the time interval (the TIME section), how to pre-process the data before applying the analysis methods (the PREPROCESS section).

INA adopts a strategy that drives the user to follow a linear progression from importing the data to their analysis:

**A. IMPORT -> B. VARIABLE -> C. TIME -> D. PREPROCESS -> E. Analysis**

This sequence is achieved by using the buttons located below the Next label in the “active” control panel displayed in all windows at right. The user can also use the fixed buttons on the left panel and “jump” between various sections of the library (data selection, time selection, analysis method selection). By default, the first variable appearing in the top of the right panel is chosen for the further processing, the default time interval is set as the full time interval of the uploaded data file, and the default preprocessing option is the “unmodified” (i.e. no pre-processing).

In this tutorial we continue the scenario initiated in the Tutorial number 1. The final outcome of the example introduced in the tutorial No. 1 was that the user selected a Cluster magnetic field data file whose origin is the Cluster Science Archive. Figure 1 illustrates the window opened after selection of the corresponding data file from the local directory.

The **VARIABLE** window depicted in Figure 1 opens when the user selects a data file. Figure 1 shows the 4 variables in the Cluster-FGM-CSA data file: the 3 components of the magnetic field (Bx, By and Bz) and the magnetic field magnitude (Bm). The upper part of central panel shows all the variables as small thumbnail plots, and the bottom part depicts the variable selected for the analysis. The selection of a variable can be made by using the corresponding buttons on the right panel. The button labeled “change variable names”, gives the user the option to easily change the names of one or more variables, and, correspondingly, the axis labels of the graphs, as well. Let us choose the Bx component and proceed to the **TIME** selection layer by pushing the “TIME” button below the label Next.

The **TIME** window is depicted in Figure 2. In the right part of the figure the user can apply a series of options: choose to analyze the full time interval, or extract a time interval using an interactive or a manual selection method.

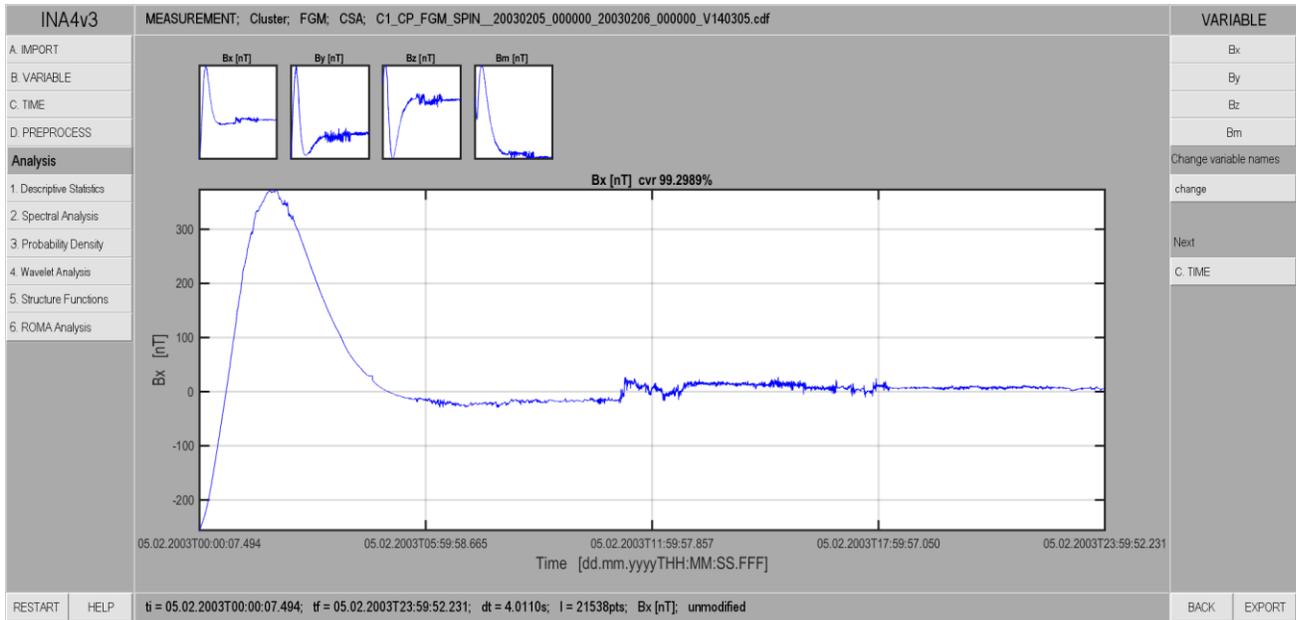


Figure 1. Illustration of the INA window opened after selection of a data file; this example corresponds to Cluster magnetic field data provided by the CSA database. The left panel is the main, permanent INA control and navigation panel. The central panel shows graphically the 4 variables of the Cluster FGM data file: Bx, By, Bz, and Bm; the bottom plot illustrates the actually selected variable. The right panel contains the buttons that allow the user to choose the variable for subsequent time selection and pre-processing, it also includes a window that allows changing the name of the variable and a button that automatically opens the next INA window, **TIME** selection.

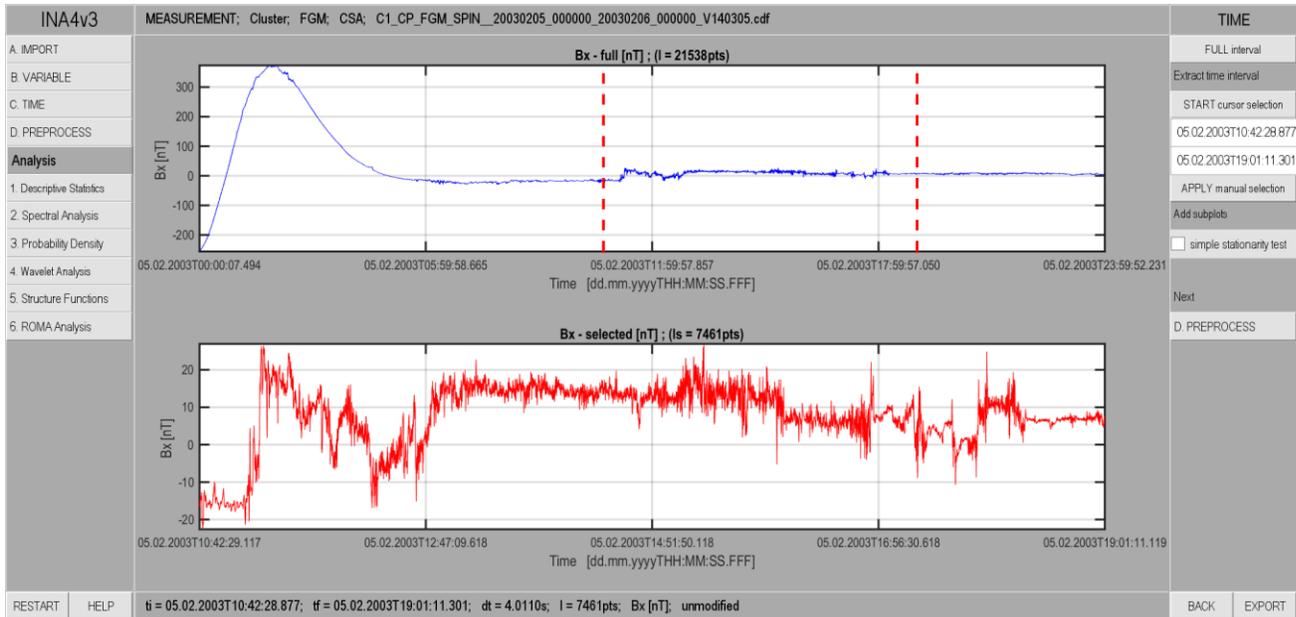


Figure 2. Illustration of the **TIME** window. The upper plot in the central panel depicts the full time series with marked limits of the sub-interval selected by the user; the bottom plot shows the extracted time sub-interval. The right panel contains the options available for this layer (see the text for details).

The interactive selection method lets the user choose the beginning and end points of the selected time interval with the cursor. To activate this method, the “START cursor selection” button must be pushed first, then click the two limits of the selectable interval by mouse.

The user can also extract a subinterval by specifying explicitly the time limits using the keyboard in the corresponding cases. Once the interval limits are set, the graph of the extracted time sub-interval of the original time series appears in the bottom part of the central panel, below the graph of the whole series.

Figure 3 illustrates a functionality called “simple stationarity test”. The corresponding plot shows the difference between the mean value of the time series and the mean value obtained for the first 1% of the length of the times series (the first point), for the first 2% (the second point), and so on up to 100% of the time series (giving the value 0 for the last point). This test is performed for both time series (full and extracted), and can be used as a rough estimate of the time stationarity of the two series. Time stationarity can be inferred visually, from the rate of decay to 0. Figure 3, for example, shows that the selected time interval (red line) is more stationary than the full time interval (blue). INA also provides a numerical value to describe the “degree” of time stationarity, defined as the sum of the absolute values of the stationarity plot (shown in the legend of the corresponding plot). In figure 3, the stationarity parameter, “sp”, has a value of ~ 403 for the selected time interval, compared to the value of ~ 4500 for the full time interval.

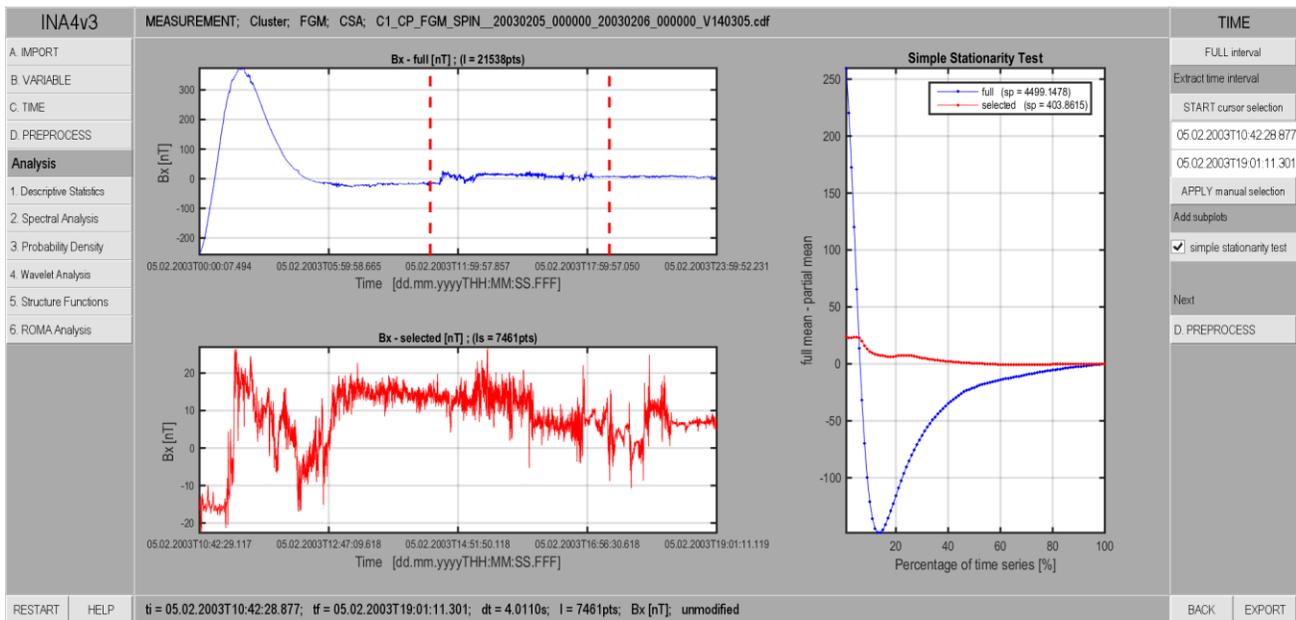


Figure3. Illustration of the **TIME** window where the user can select a sub-interval of the entire time series either by using the cursor or by specifying the exact start and end time of the sub-interval. The time selection buttons are displayed in the right hand side panel of the window. A simple Stationarity Test is also included. The left part of the

central panel depicts the full and extracted time series (as in the previous figure), and the right part shows a plot of the simple stationarity test (see the text for details).

The **TIME** and **PREPROCESS** layers function as data masks. For example, the user can choose one variable, then extract a subinterval of time, choose a preprocessing option and proceed with the analysis. If the user would like to analyze another variable contained in the same file, the user can switch back to the **VARIABLE** layer, select another variable, and then click the **Analysis** button. In this way, the time subinterval and preprocessing option chosen for the first selected variable will remain active, and are automatically applied to the new variable.

As a general remark, the “stationarity plot” is a persistent feature of the **TIME** and **PREPROCESS** layers. This means that if the user activated this option in the **TIME** window, it will be automatically activated also in **PREPROCESS**. Moreover, the stationarity test will remain active for the whole INA session, and the user will see the stationarity plot each time these two layers are visited.

The **PREPROCESS** window is depicted in Figure 4. There are 7 main preprocessing options: a) square; b) standardize; c) subtract mean; d) square + standardize and e) square + subtract mean, f) wavelet denoising and g) handling data gaps. Once a method is selected, the graph of the preprocessed time series appears in the **PREPROCESS** layer, next to the graph of the original signal. The “Square” button computes the square of the time series. The “standardize” button modifies the time series,  $y$ , according to:  $y_m = (y - \bar{y})/\sigma_y$ , where  $\bar{y}$  is the mean value and  $\sigma_y$  is the standard deviation of  $y$ . Wavelet denoising, based on the MATLAB denoising function **wden** performs an automatic denoising using wavelets. The denoising level is controlled by the parameter  $N$ : “weak” denoising for  $N < 5$  and respectively “harder” denoising for  $N > 5$ .

The “handling data gaps” functionality included in the **PREPROCESS** layer gives the user the possibility to choose between two methods to handle data gaps. By default, INA automatically detects and linearly interpolates all the data gaps. Another option is to fill-in the data gaps with NaN values. This approach can be useful when computing probability distribution functions where interpolation is not desirable in general. Note however that some analysis methods, like the **PSD-Welch**, for example, do not work when data gaps are replaced by NaN values.

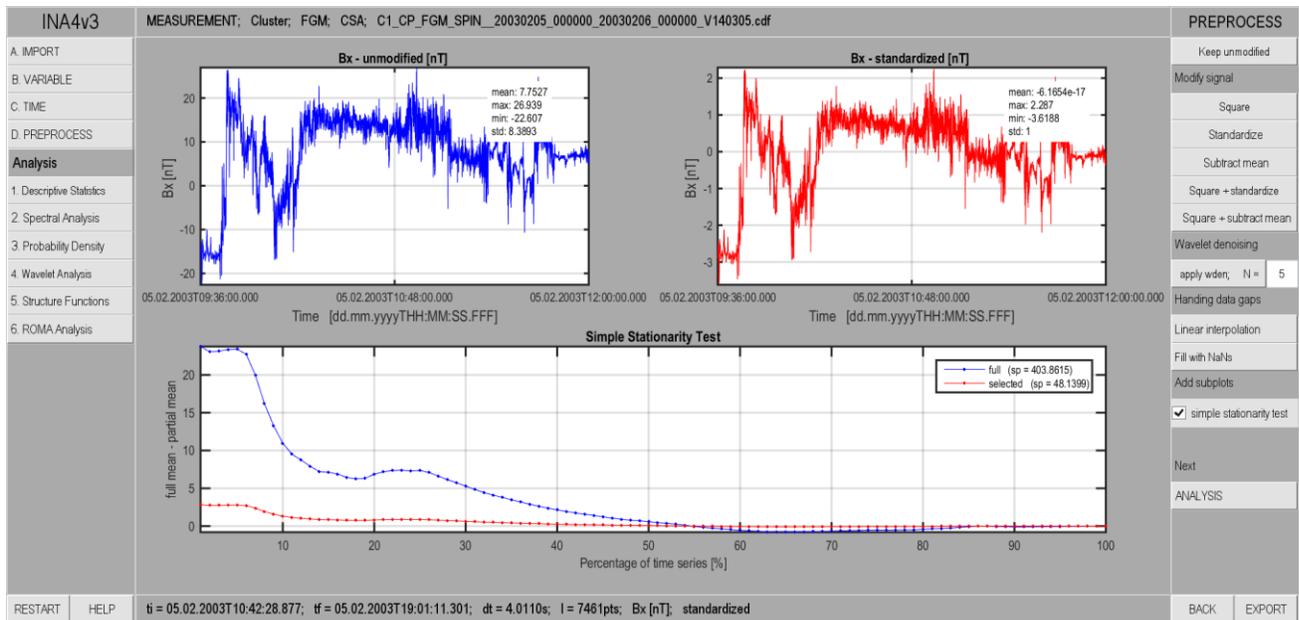


Figure 4. Illustration of the **PREPROCESS** window. The user can apply seven different preprocessing functions as described in text above. There are two different approaches for data gaps filling: linear interpolation or NaN values. The upper part of the central panel depicts the unmodified (left) and modified time series (right), and the bottom part shows a plot of the simple stationarity test for the two signals.