Manual. Shiny R App

Augusta University | 1120 15th Street, Augusta, GA 30912

Calcium imaging data analysis using shiny r app

Taraskin, Aleksandr

2023

Contents

[1. Opening the program for calcium imaging data analysis 2](#_Toc146807661)

[2. TAB: Preliminary analysis 2](#_Toc146807662)

[Uploading excel files 3](#_Toc146807663)

[How to handle several sheets in a file 4](#_Toc146807664)

[Custom sheet names 5](#_Toc146807665)

[Correct “Time” column 6](#_Toc146807666)

[Change column names 8](#_Toc146807667)

[Display option 12](#_Toc146807668)

[Calculate basic statistics 12](#_Toc146807669)

[Save descriptive statistics table as excel file 15](#_Toc146807670)

[Creating plots and excluding bad traces 15](#_Toc146807671)

[Excluding traces 17](#_Toc146807672)

[Discarding unnecessary traces 19](#_Toc146807673)

[3. TAB: Analyzing amplitude 21](#_Toc146807674)

[Uploading file with the “clean data” 21](#_Toc146807675)

[Plotting data 22](#_Toc146807676)

[Calculating amplitudes 23](#_Toc146807677)

[Amplitudes result table 24](#_Toc146807678)

[4. TAB: Shifting curves 26](#_Toc146807679)

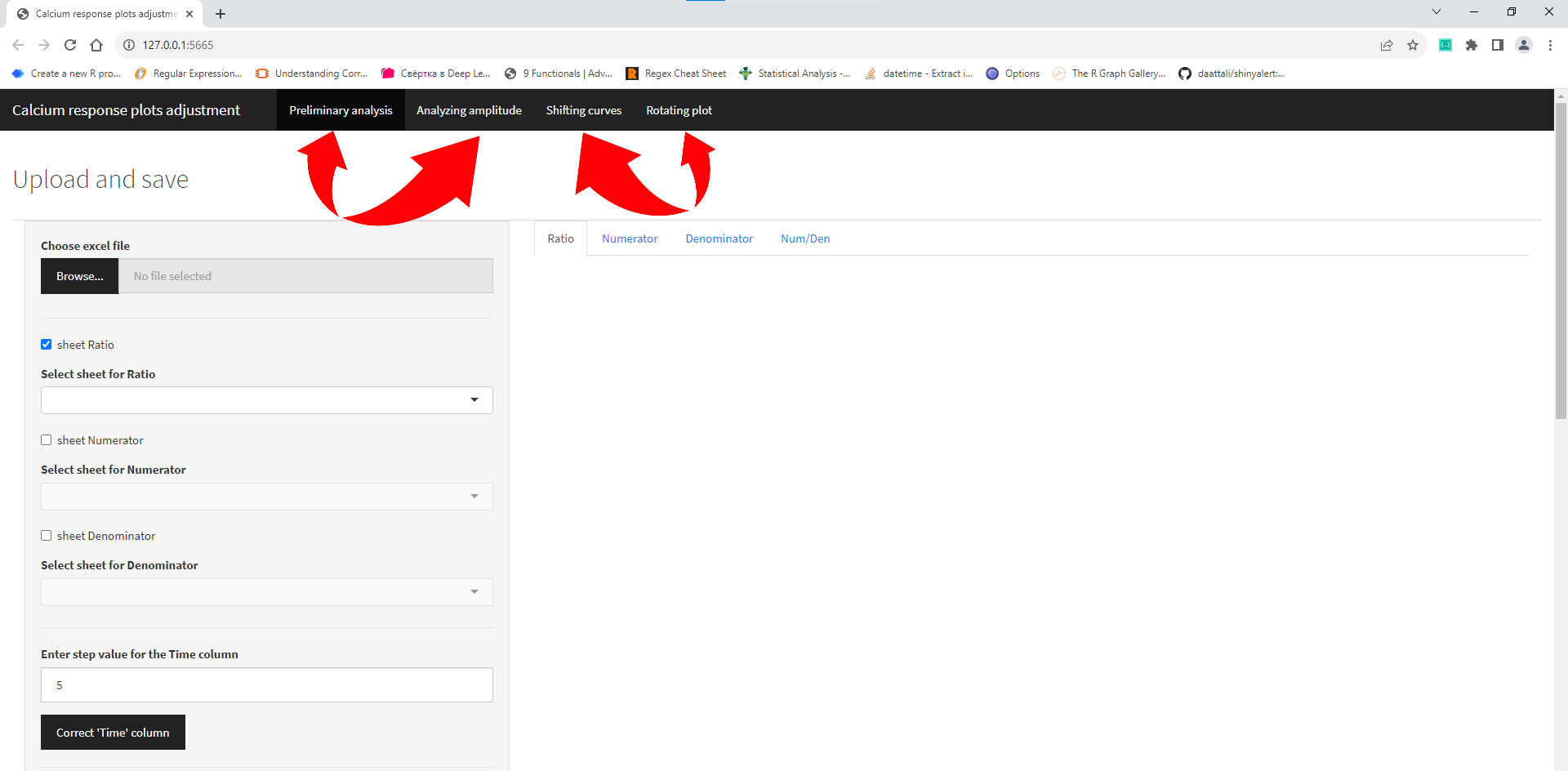
[5. TAB: Shifting curves 26](#_Toc146807680)

[Explaining algorithms 26](#_Toc146807681)

Calcium imaging

# Opening the program for calcium imaging data analysis

The upper part of the program’s start screen look like this:



Altogether there are 4 tabs:

* Preliminary analysis (loading raw excel files, correct Time column, change column names, exclude traces, etc.)
* Analyzing amplitude (finding amplitude for curves that have 1, 2 or 3 major maximums)
* Shifting curves (shift plots automatically to the left using two algorithms)
* Rotating plot (rotating plots and their parts)

You can start using program after choosing any of these 4 (Preliminary analysis is chosen by default).

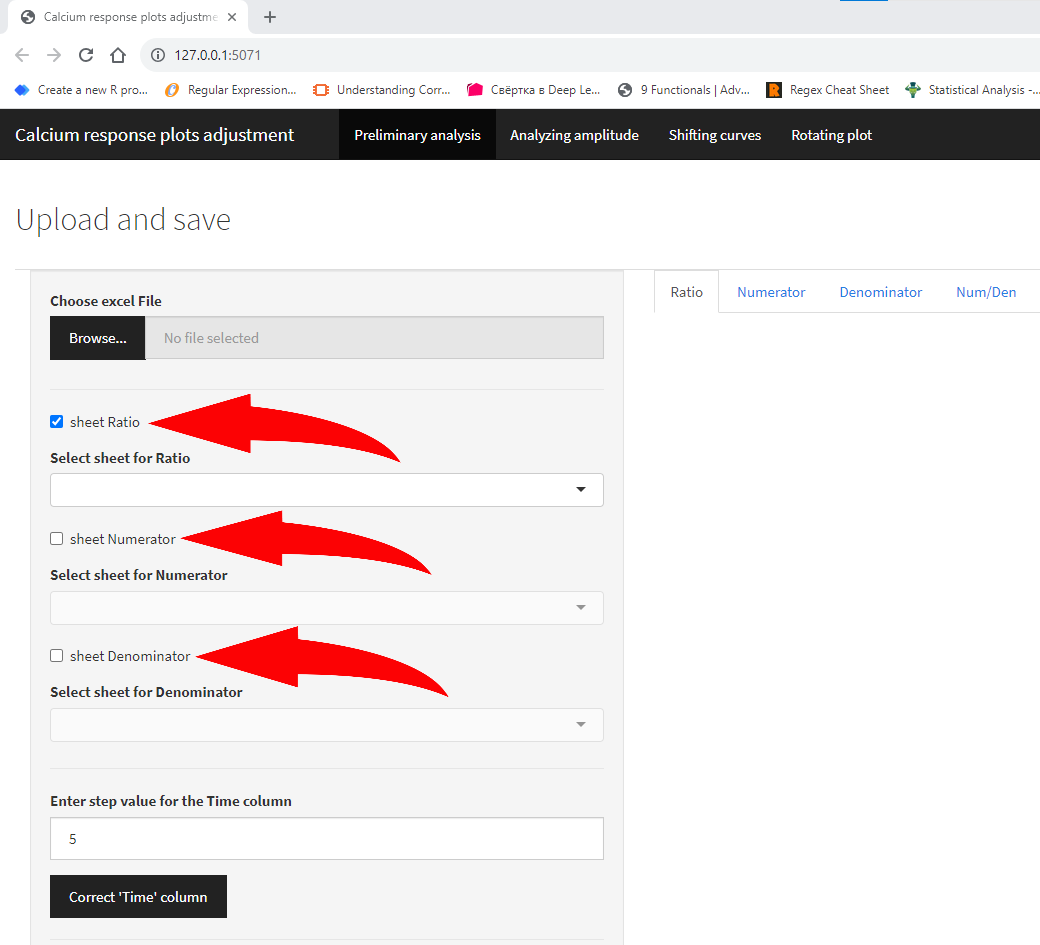
# TAB: Preliminary analysis

On this tab the main preview window (to the right from a sidebar) contains 4 internal tabs by default:

* Ratio – (fluorescence in channel 340/380)
* Numerator – (fluorescence in channel 340)
* Denominator – (fluorescence in channel 380)
* Num/Denominator – (custom ratio: values from Numerator are divided by values from Denominator)

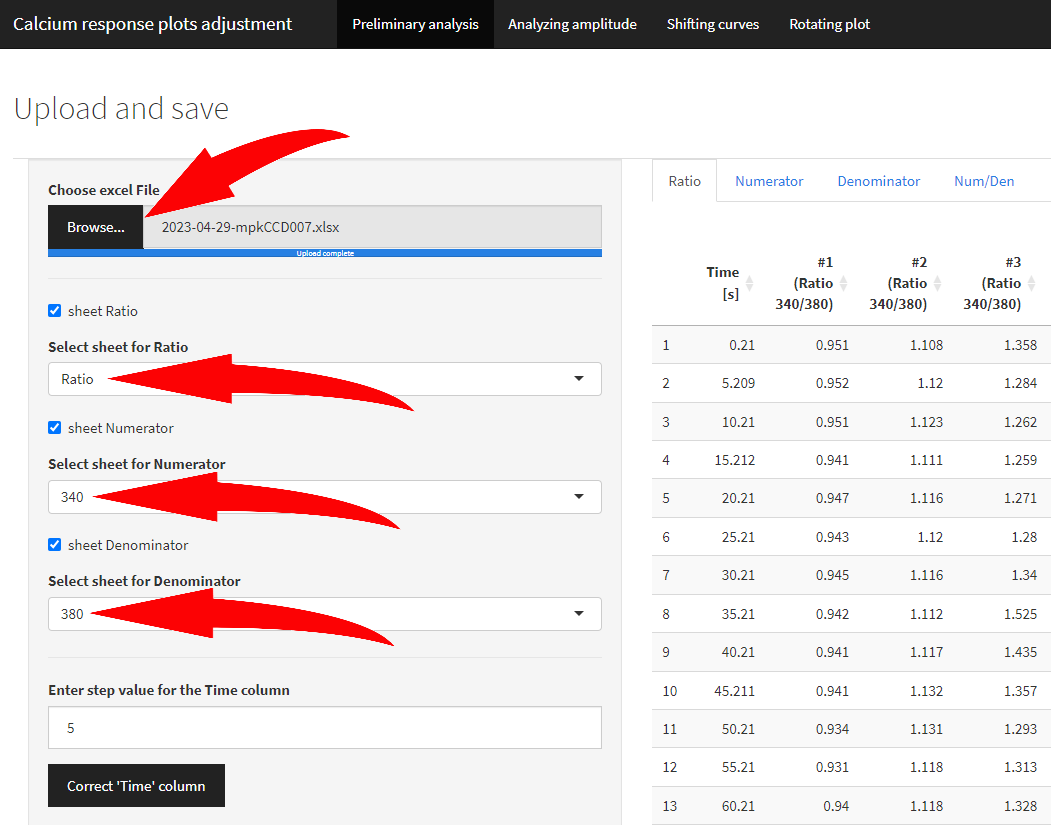
## Uploading excel files

In order to load your file and read sheets that contain information about Ratio, 340 or 380 channels you need to check necessary boxes on the left (user is allowed to do that either before the file was chosen and loaded or after):



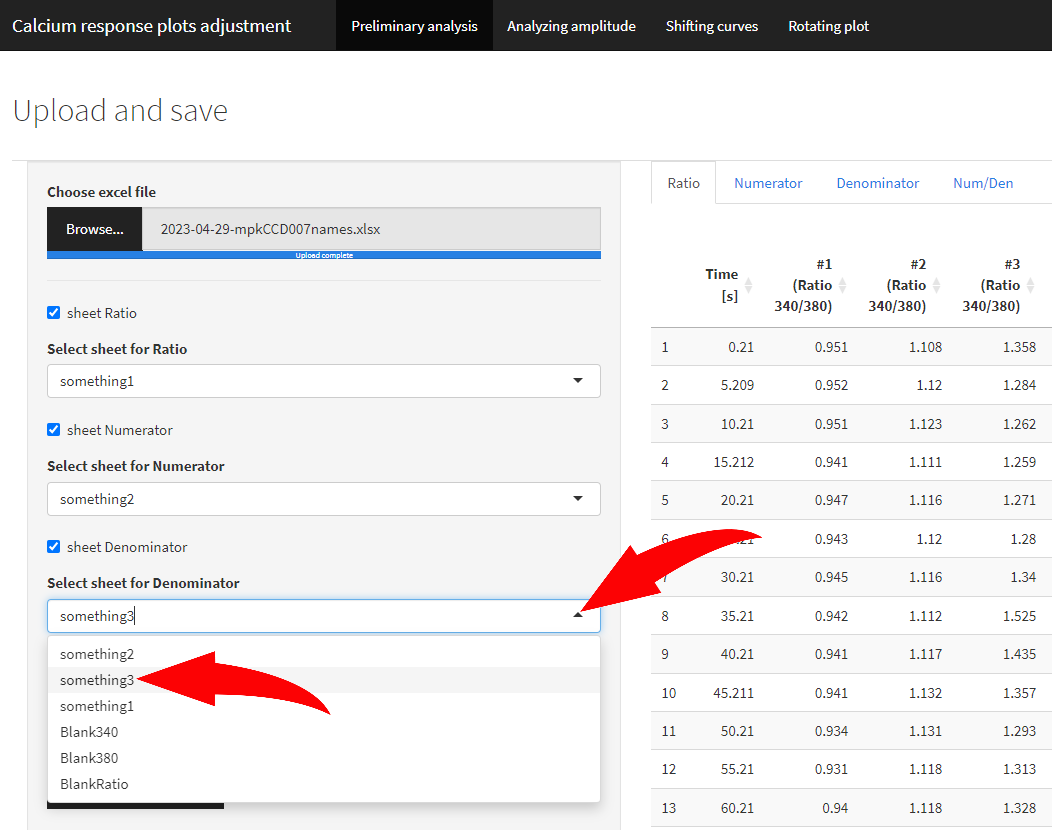
## How to handle several sheets in a file

An excel file should contain at least one sheet if you checked the one box or more. After pressing “Browse” button and choosing an excel file: if the file has sheets with names “Ratio” or “ratio”, “340” or “380” – they will be automatically recognized and reflected in “Select sheet for …” boxes:



## Custom sheet names

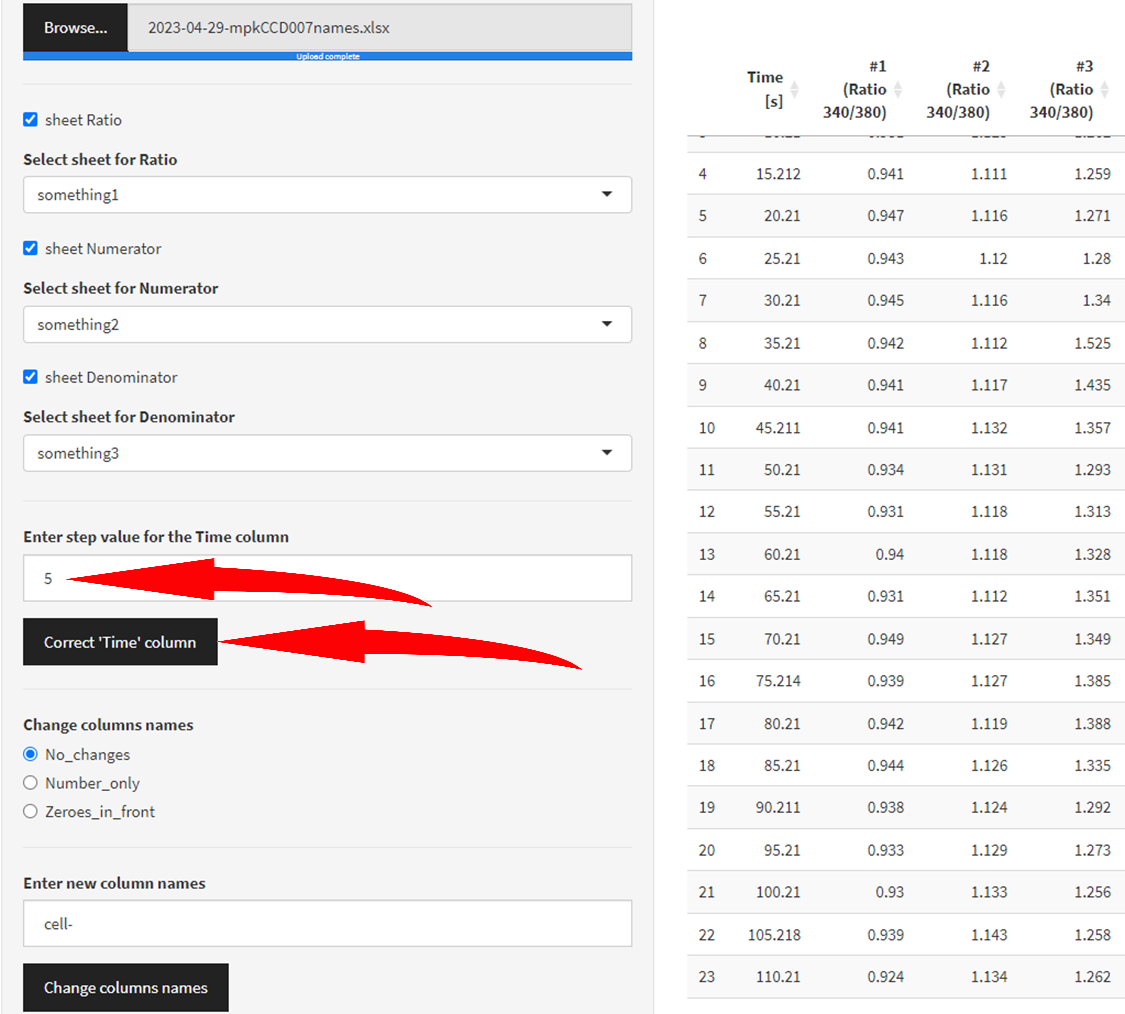
If names are different from reference, user can manually set any sheet name that corresponds to Ratio/Numerator/Denominator:



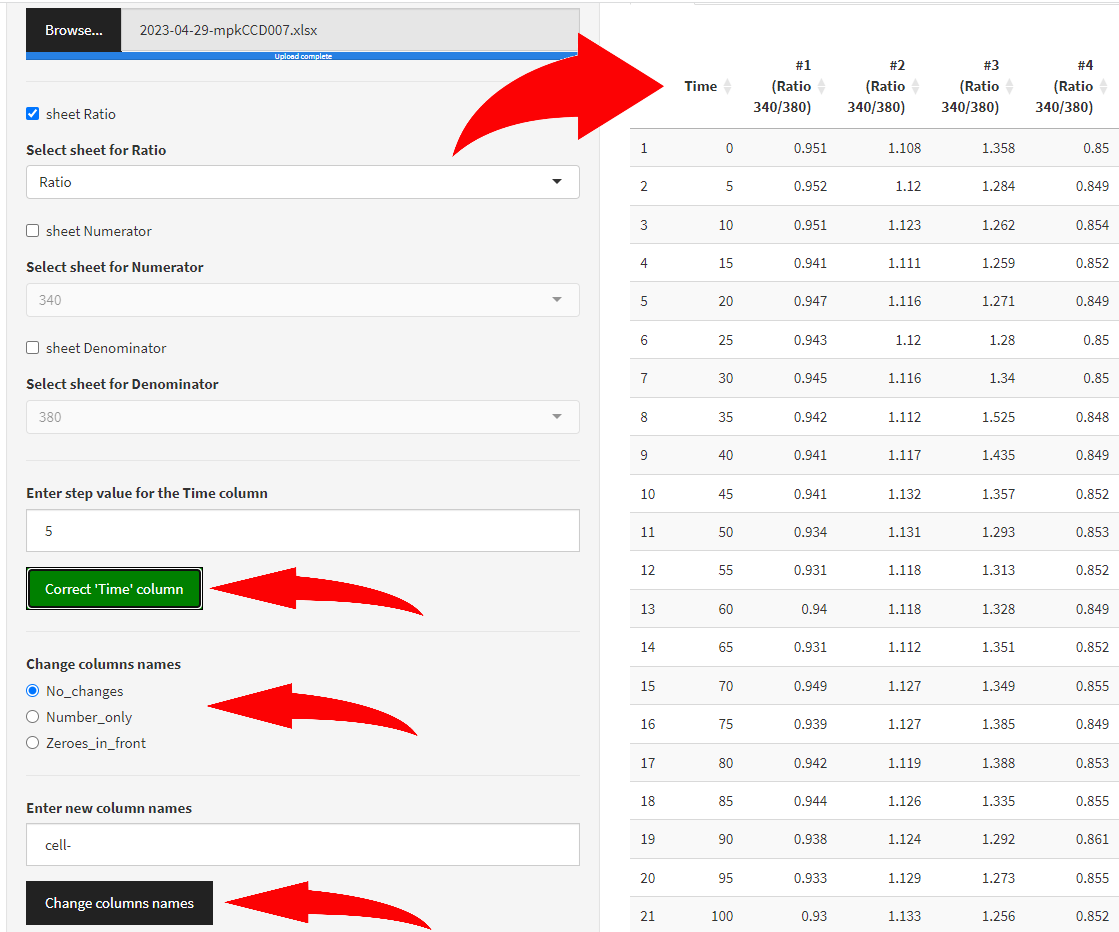
## Correct “Time” column

Next step allows you to correct the “Time” column and column names if necessary (but recommended).

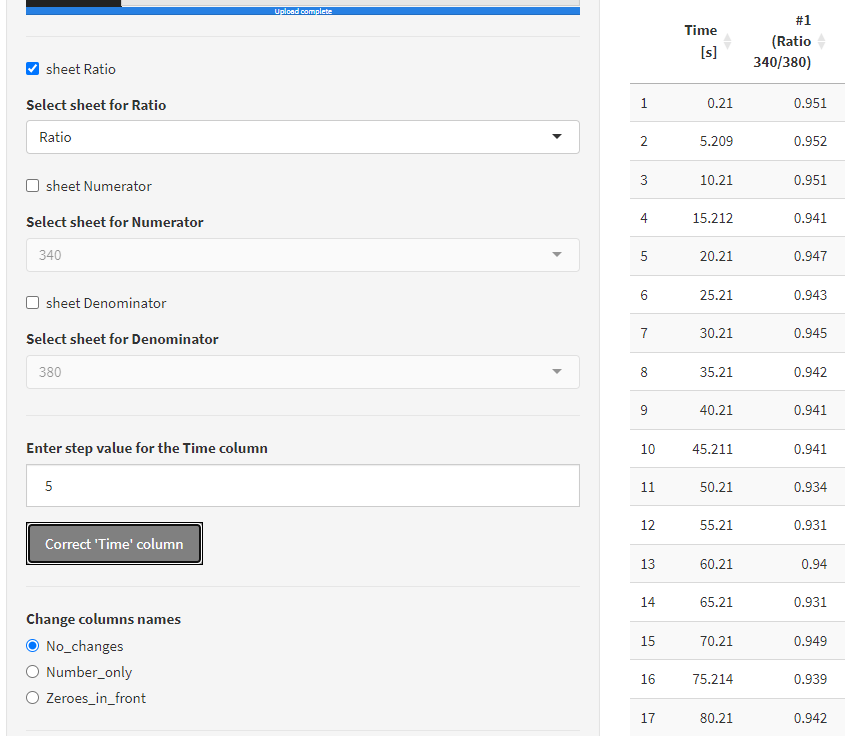
To replace “Time [s]” column to “Time” and fill it with values with custom step (by default 5 sec), enter the step value into the box below (if different from default) and press “Correct ‘Time’ column”.



After pressing the “Correct ‘Time’ column” button, it turns green indicating that action has been taken.

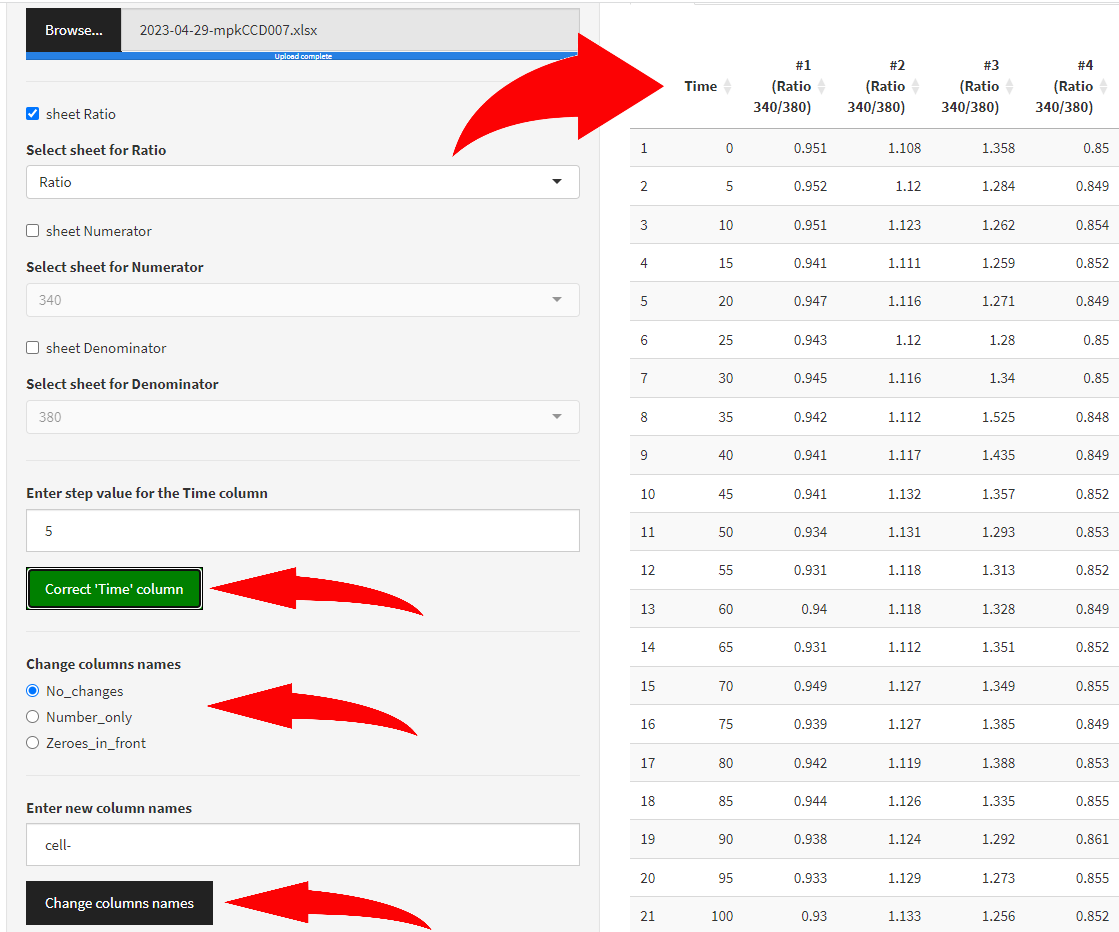


By pressing it again it turns grey and deny changes that have been made to the “Time” column.

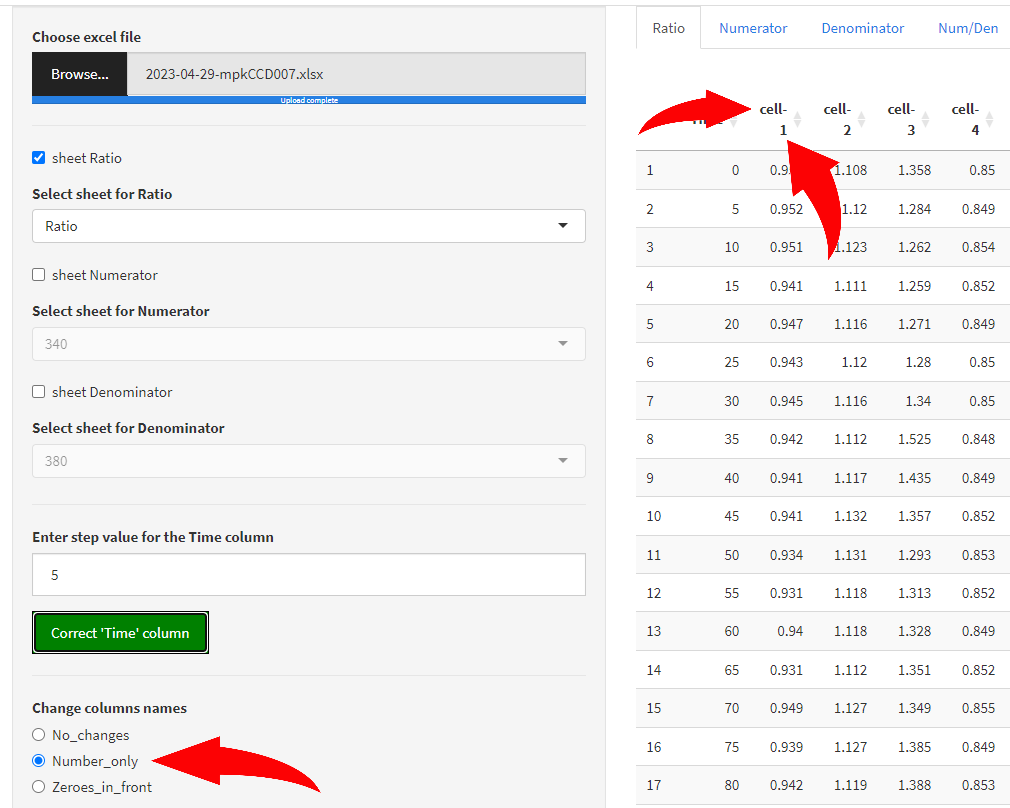


## Change *column names*

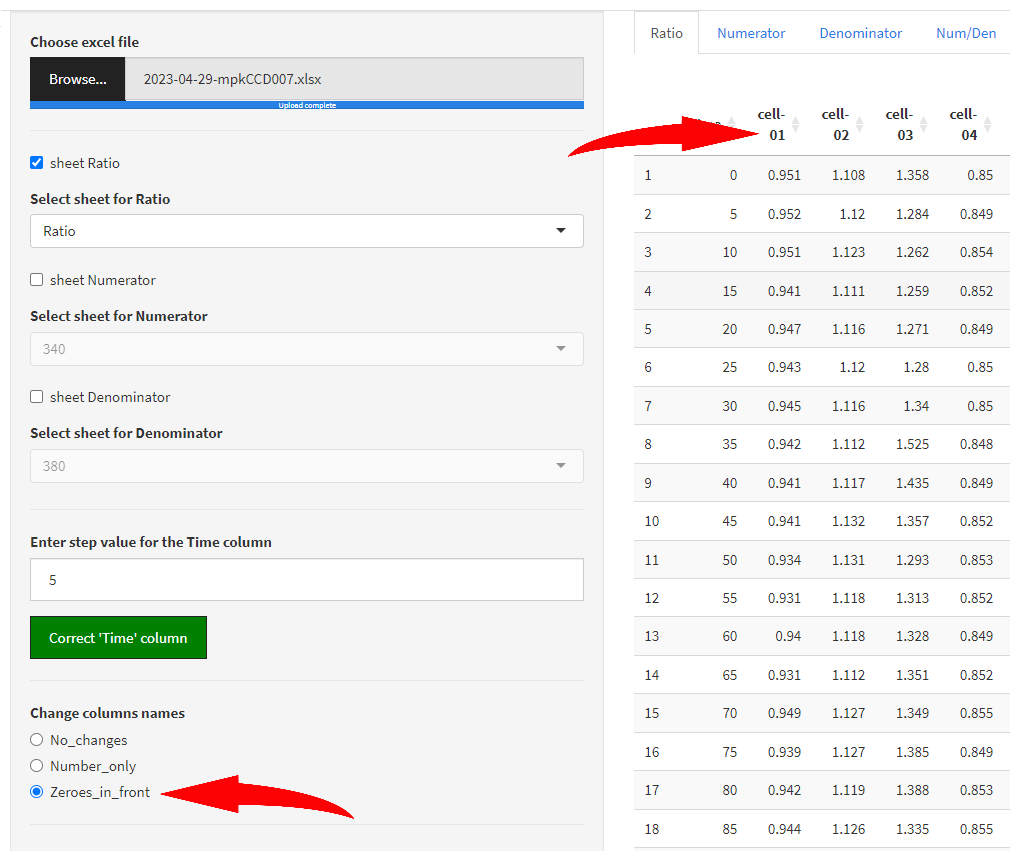
To change other column names there are two options for number conversion and option for custom prefix.



By choosing “Number\_only” column, any number in a column name will be captured and column will be renamed using the following pattern: column name prefix (by default “cell-”) + captured number.

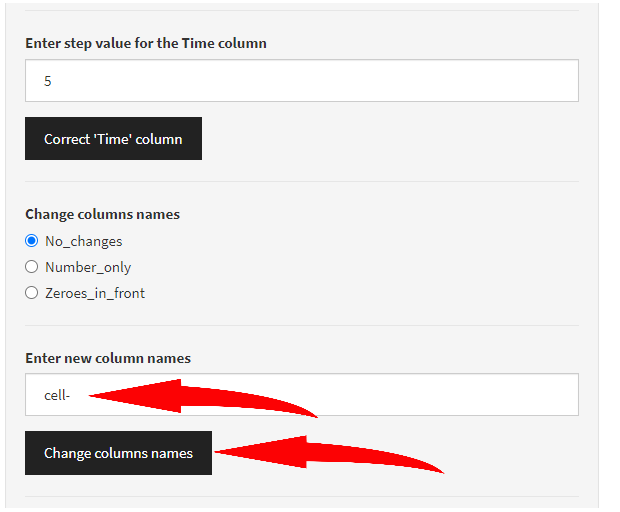


By choosing “Zeroes\_in\_front” option to any captured number a certain amount of zeroes will be added so the result number has fixed amount of all digits (this can help with sorting column names).



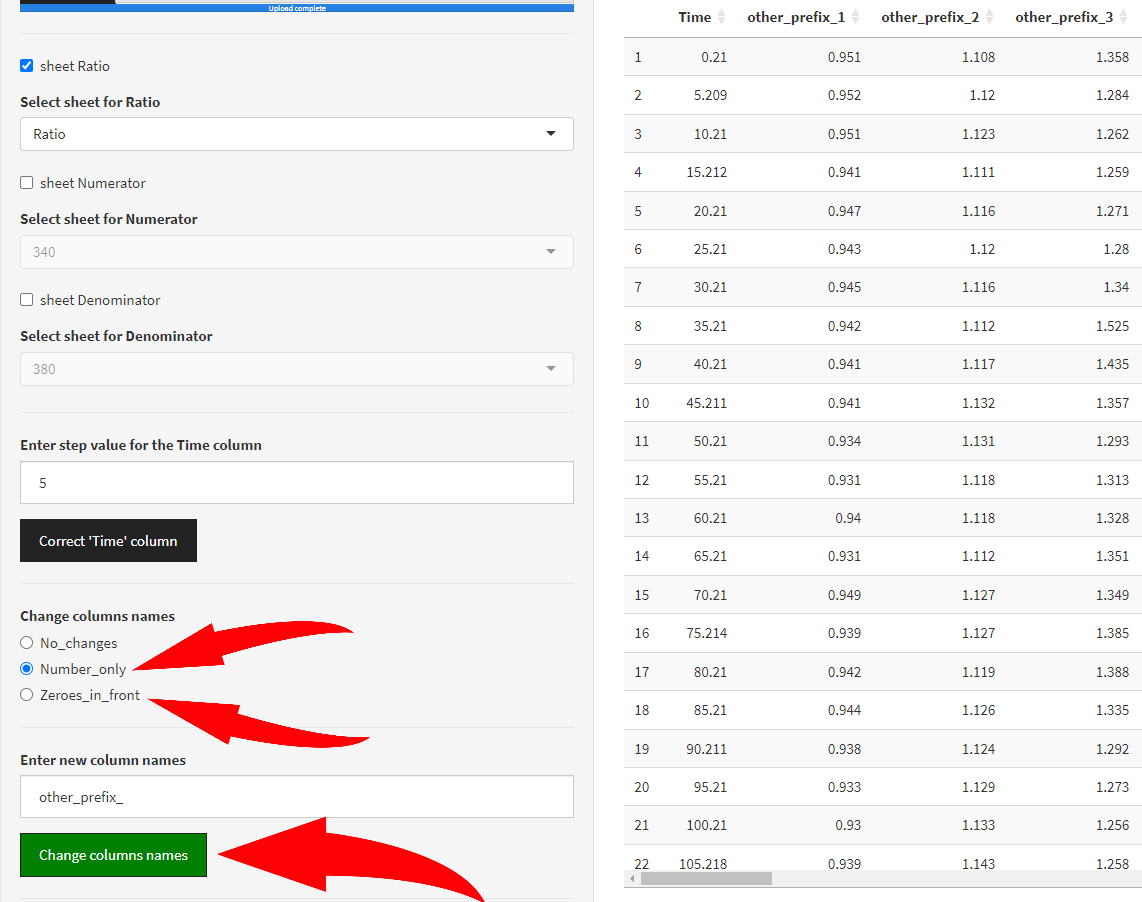
The amount of zeroes depends on the maximum number captured (maximum 9999, i.e. cell-9999). If maximum number has 4 digits and minimum number – only one, therefore the pattern is the following: “cell-0001, cell-0013, cell-0020, cell-0149, cell-1328”. The 2 digits maximum case is presented above.

If user wants to change the prefix from “cell-” to custom prefix, it can be done by entering the prefix in the “Enter new column names” box:



In this case if “No\_changes” radio button is chosen “Change column names” button won’t cause any changes and just will turn green one time indicating that the button was pressed at least one time. As long as this is an action button there are no “grey” state for it, “green” only. User can discard changes by choosing “No\_changes” radio button any time.

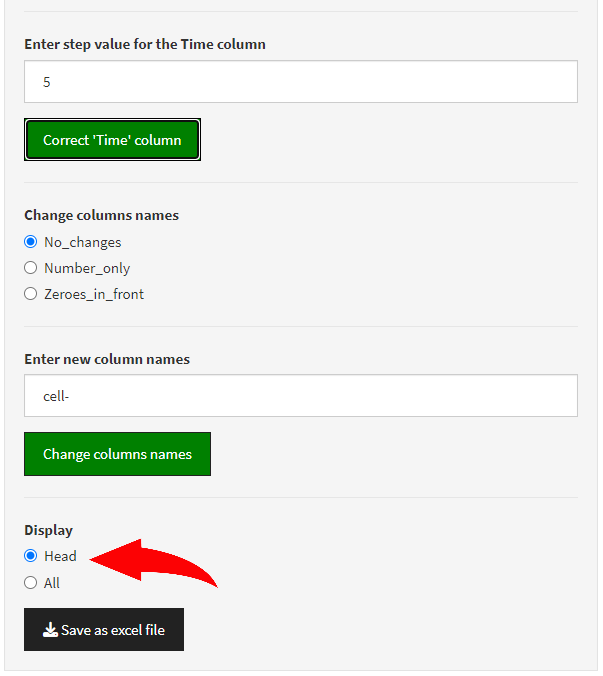
For changes to take effect user needs to choose first the column names style above (“Number\_only” or “Zeroes\_in\_front”) – entered prefix will be used momentarily to change column names. For the custom prefix AFTER this action has taken place user can enter custom prefix value and press “Change column names” button to change the prefix in column names.



After choosing column names style or changing column names prefix “Time [s]” column also changes its name automatically to “Time” only (if you haven’t done it earlier) but no automatic changes to its values are applied though.

## Display option

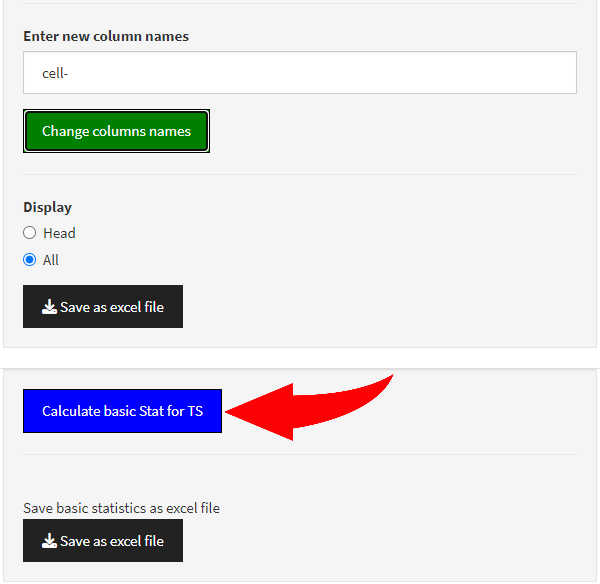
The last “Display” option in this section allows you to just show several rows only in a preview table. It is relevant only for amount of traces > 500 to speed up the process of loading and previewing data.



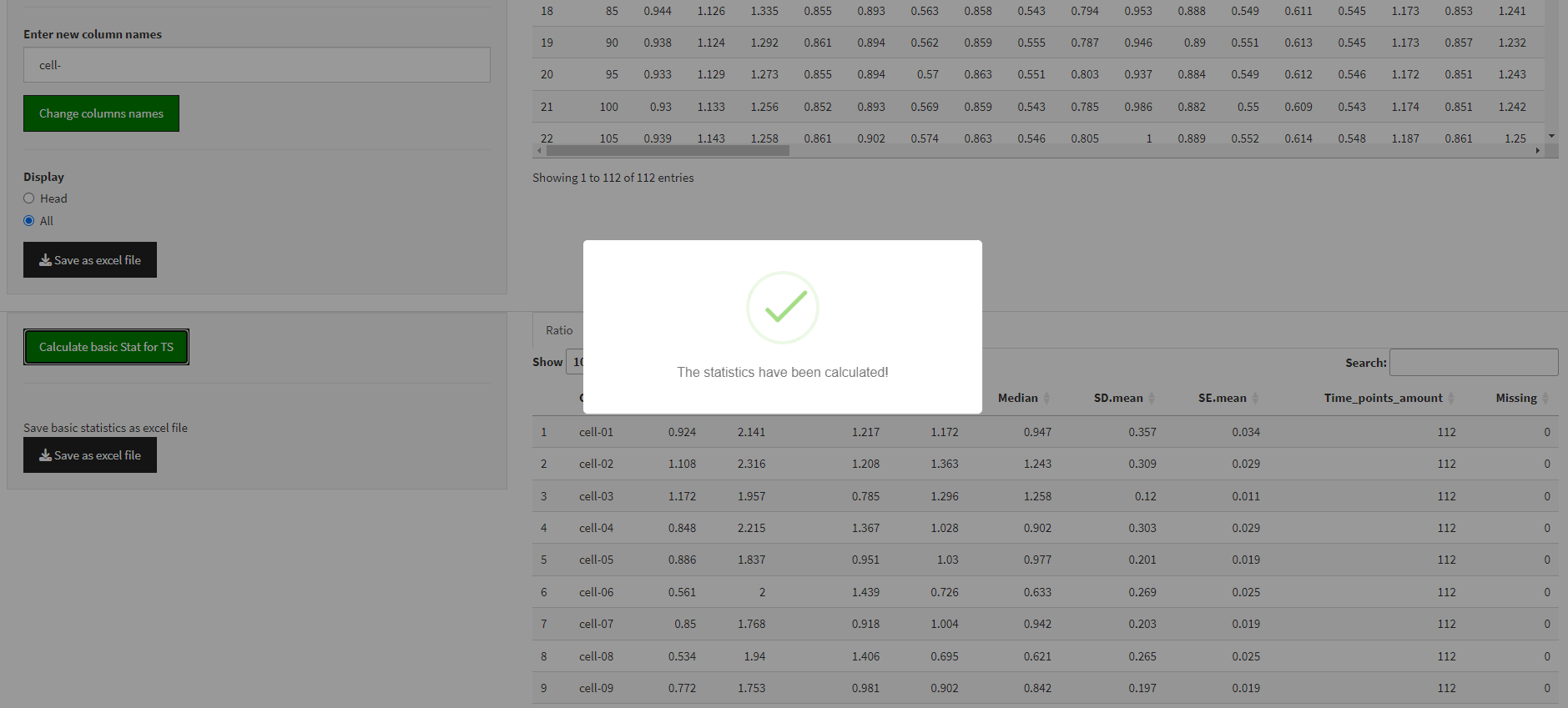
Finally, “Save as excel file” button allows to save an excel file (**{*initial name of the file}-*ProcessedTable.xlsx**) with data that match the preview’s format.

## Calculate basic statistics

This step is mandatory to proceed further with the preliminary data analysis (tab: Preliminary analysis). All the mandatory buttons that are required to be pressed in order to move further are in blue. After pressing it, button turns green. If data was processed a descriptive statistic table will appear at the right to the sidebar.

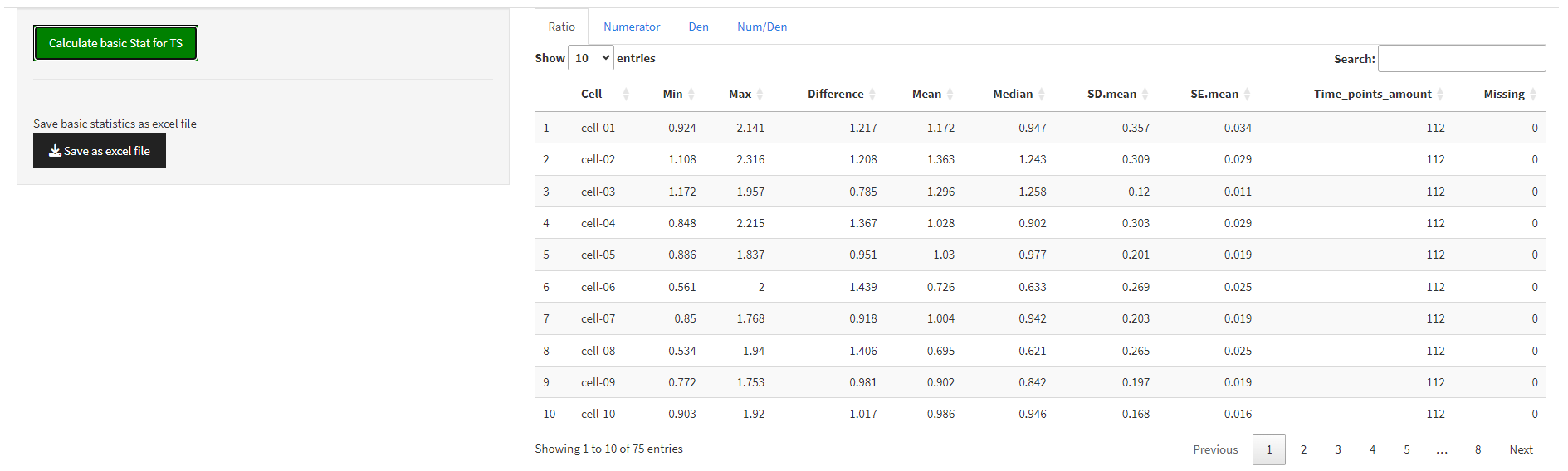


After pressing the button “Calculate basic Stat for TS” user needs to wait until the notification with the green check mark appears:



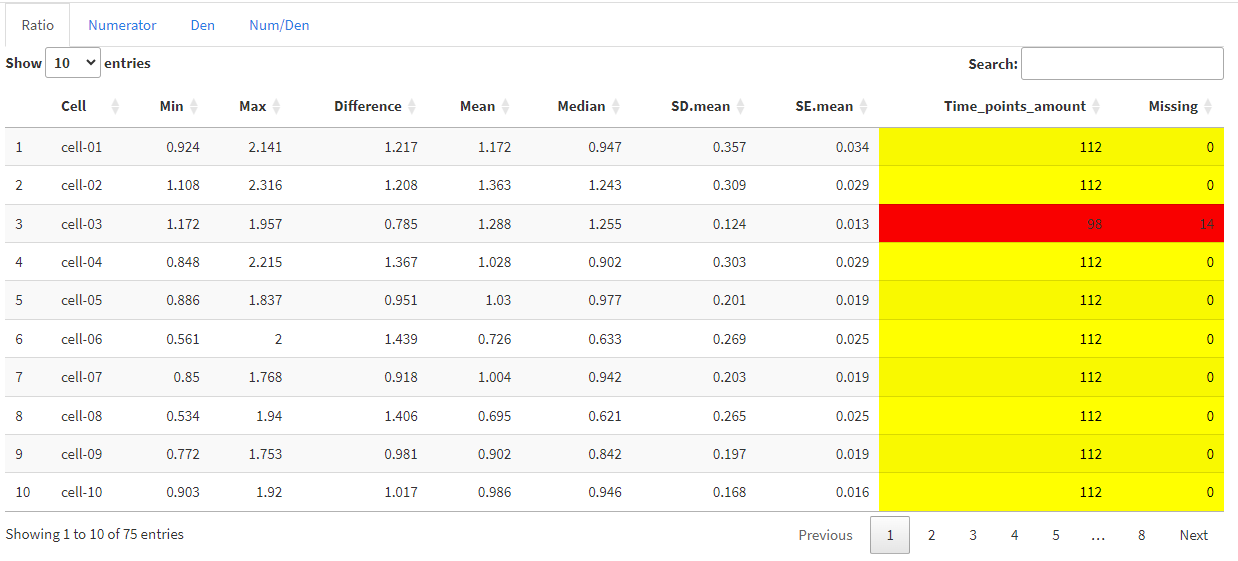
The obtained table with major descriptive statistics parameters have the following columns:

“Cell”, “Min”, “Max”, “Difference”, “Mean”, “Median”, “SD.mean”, “SE.mean”, “Time\_points\_amount”, “Missing”.



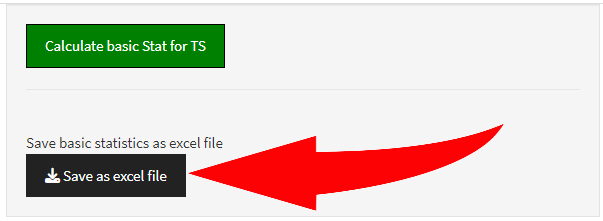
* “Cell”– the name of the initial column. All the parameters below are calculated for all the values in each column of the initial table.
* “Min” – minimum value in the specific column
* “Max” – maximum value in the specific column
* “Difference” – difference between Max and Min
* “Mean” – mean value in the specific column
* “Median” – median in the specific column
* “SD.mean” – standard deviation for values in the specific column
* “SE.mean” – standard error of the mean in the specific column
* “Time\_points\_amount” – the amount of time points in each column
* “Missing” – if some values are absent in the specific column this value will be > 0 indicating their amount

**If there are some missing values last two columns of this table will be marked as yellow and the trace with missing values (row) will be in red (otherwise nothing is highlighted):**



## Save descriptive statistics table as excel file

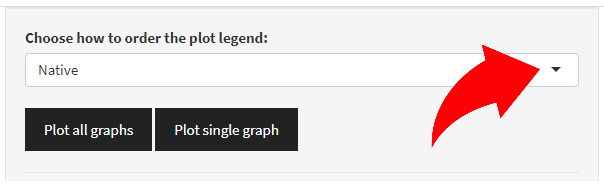
The table with descriptive statistics can be saved as an excel file (**{*initial name of the file}*-BasicStatisticsTable.xlsx**):



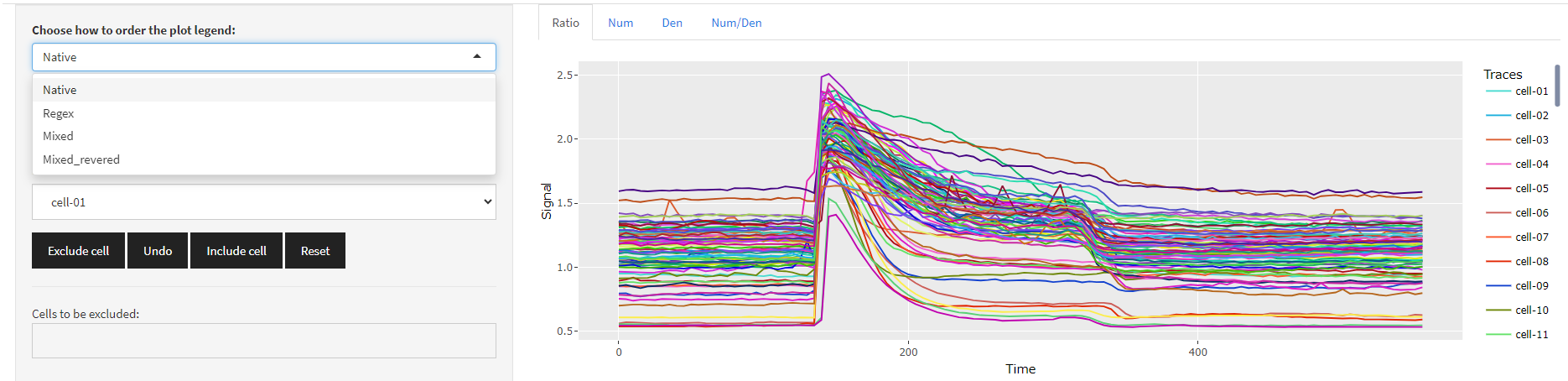
## Creating plots and excluding bad traces

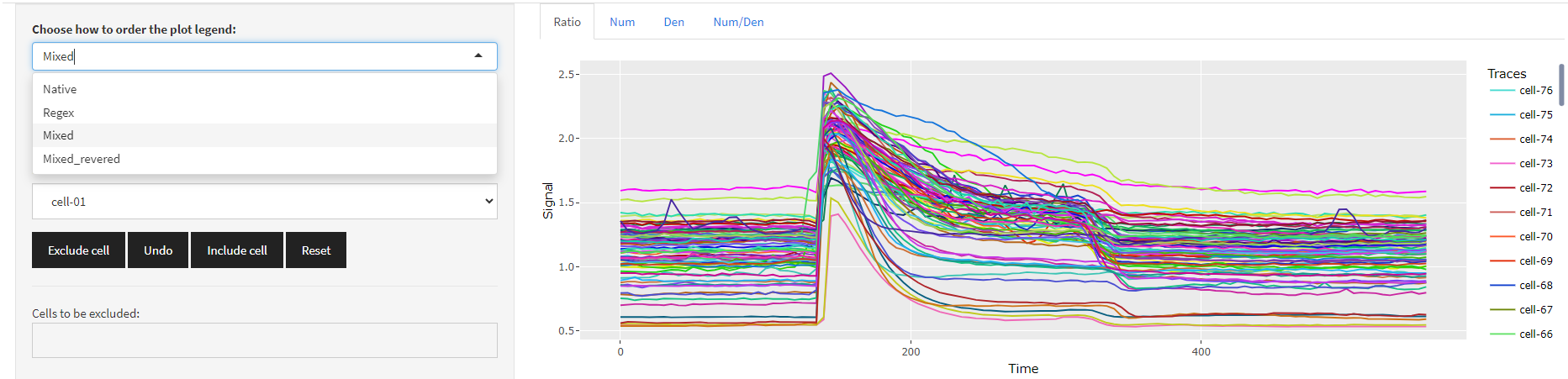
The final step for the “Preliminary analysis” tab allows user to plot all the graphs at once as well as each of them one by one to analyze them and exclude bad traces.

At first user can choose the way to order column names in the table uploaded from the excel file on the first step (the initial one).



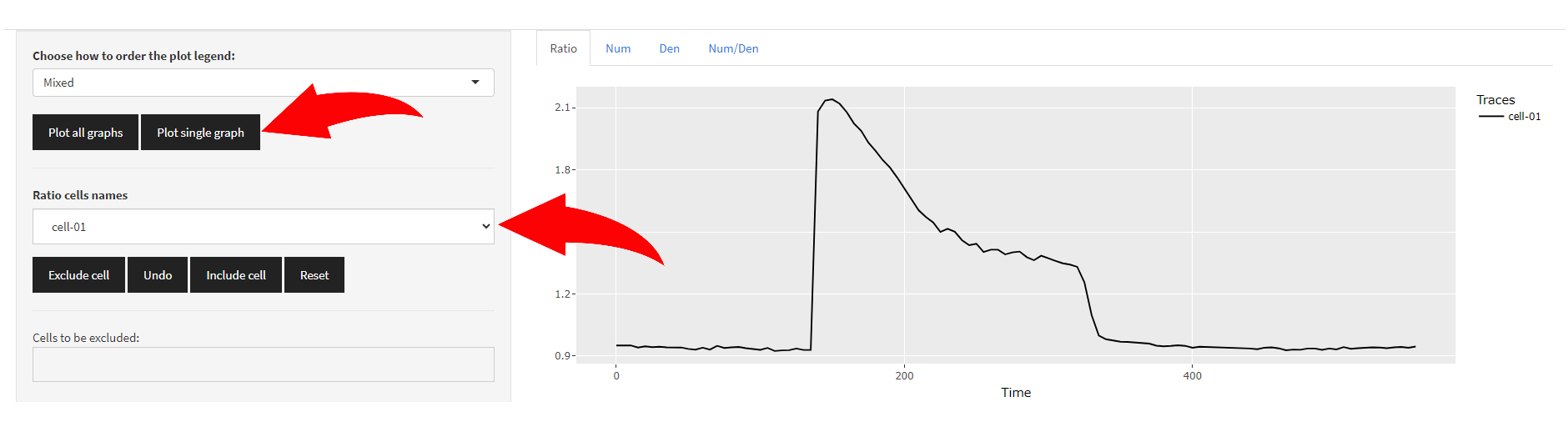
By default the sorting algorithm is “Native” which means that columns are be ordered the same way like in the initial excel file. If for some reasons they were unordered or mixed it is possible to reorder columns which will affect the legend of the plot (when pressing “Plot all graphs”). By looking at the legend when all traces are plotted at the same graph user can define the sorting method of choice:





The algorithm to choose depends on a prefix or a column name style and should be manually adjusted if columns are unordered or a custom prefix was chosen.

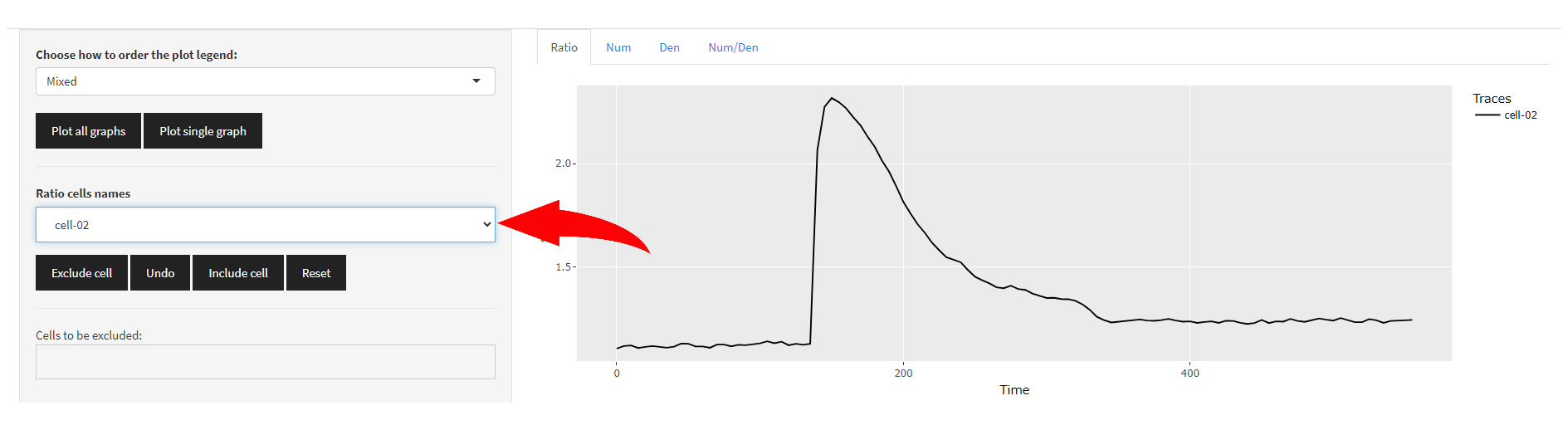
By pressing “Plot single graph” a single plot can be rendered. To choose a specific name of the traces use the “Ratio cells names” box:



To switch traces using keyboard’s arrows (up and down) it is necessary to press ones on the “Ratio cells names” box and choose any trace at first:

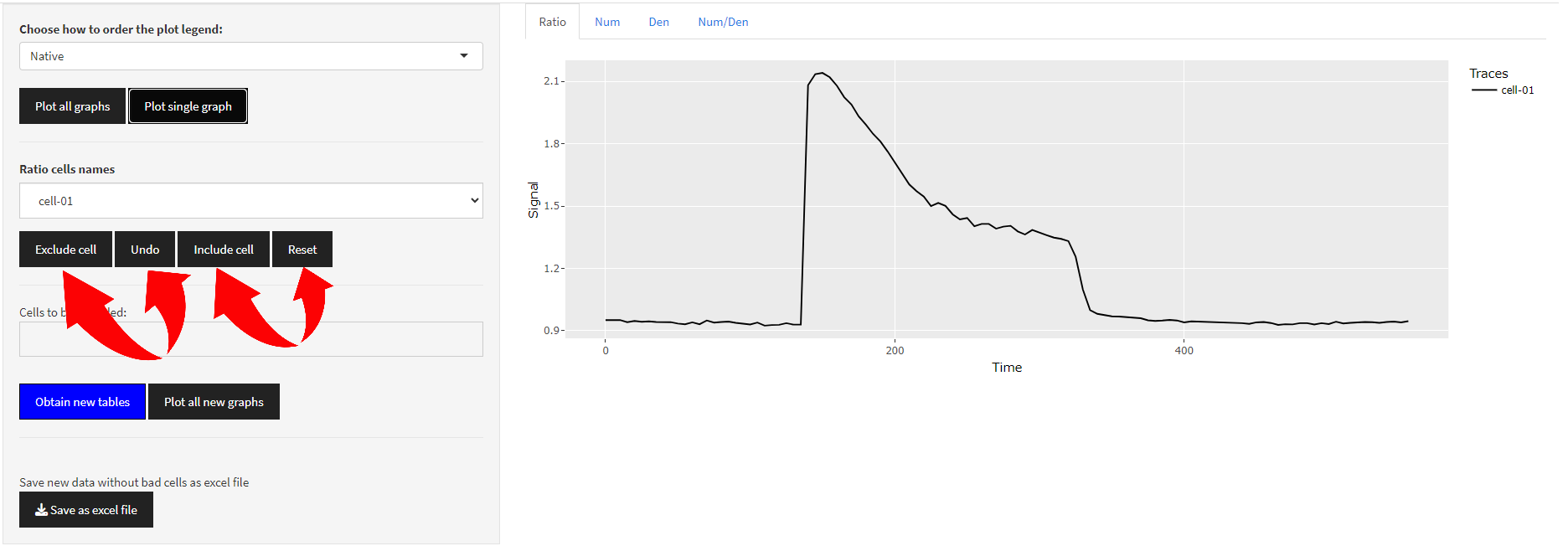


After choosing any trace it will be displayed to the right from the sidebar and user can switch traces (their plots) by pressing arrows buttons (up and down only) for convenience. When this mode is active (switching mode) the box is glowing in blue:

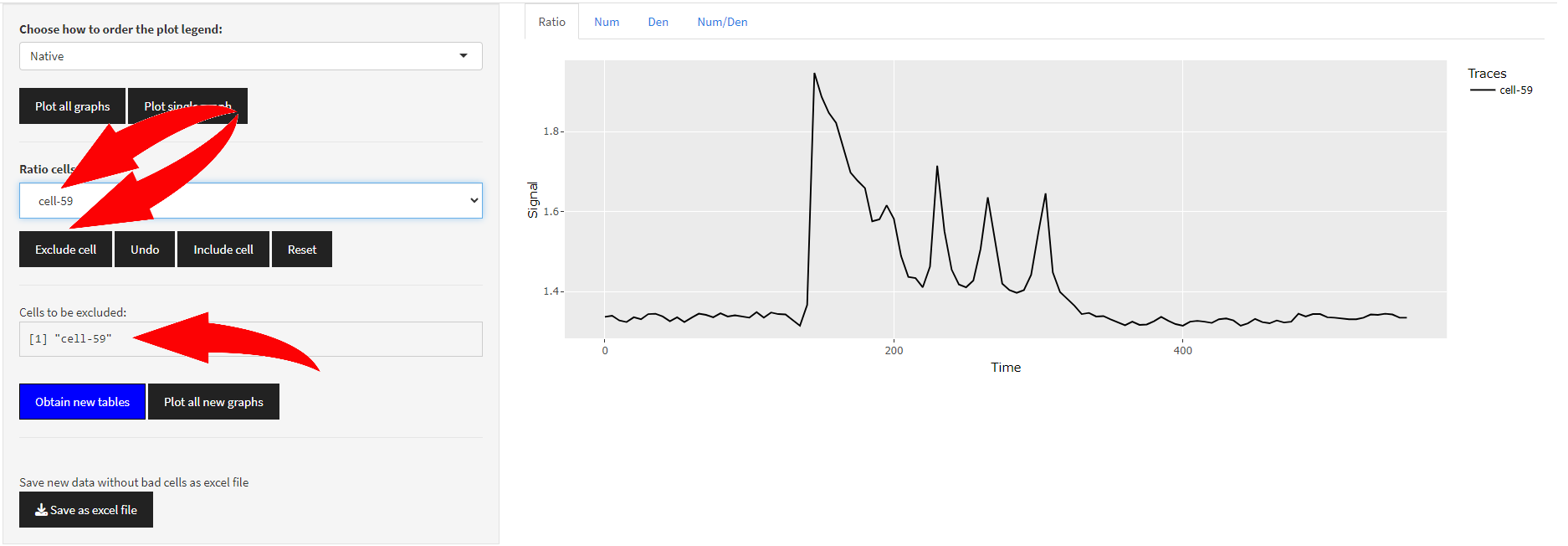


## Excluding traces

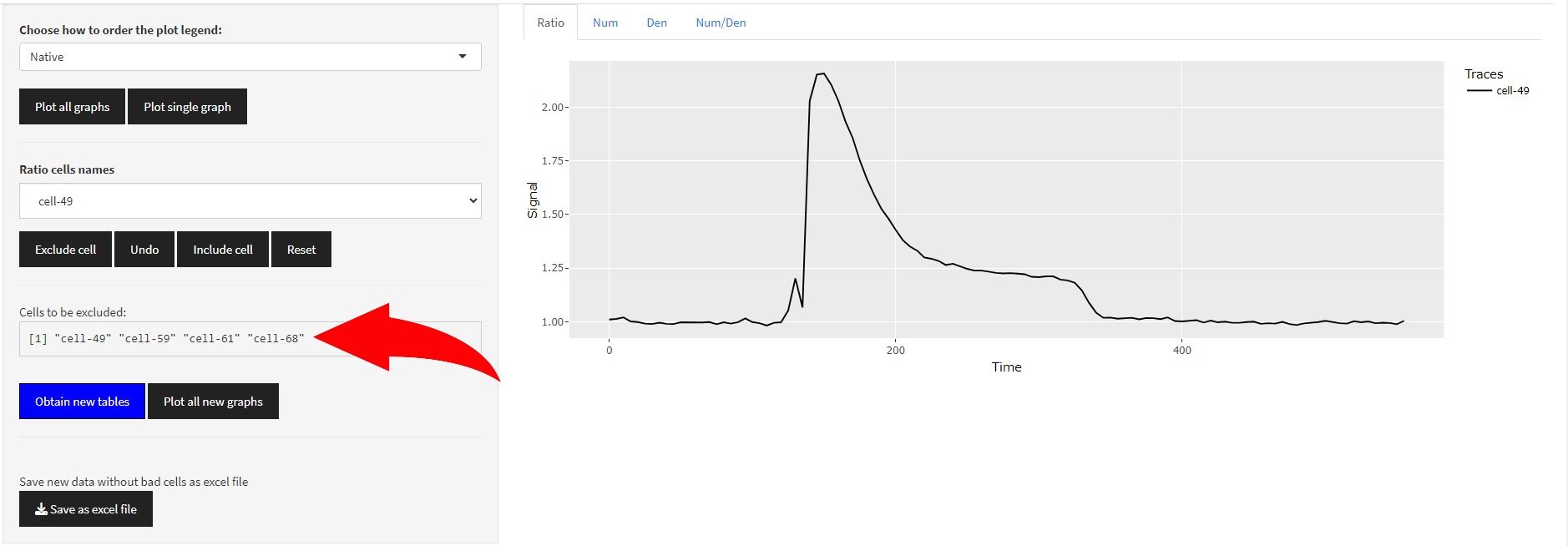
To exclude traces that seem to be outliers there are four buttons that allow user to manipulate data:



To exclude traces one by one user needs to choose the trace in the “Ratio cells names” box and press **“Exclude cell” button**:

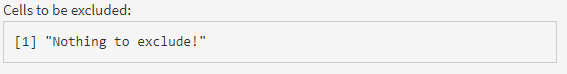


The “Cells to be excluded:” box indicates the name of the excluded trace or a sequence of traces if exclude traces sequentially:

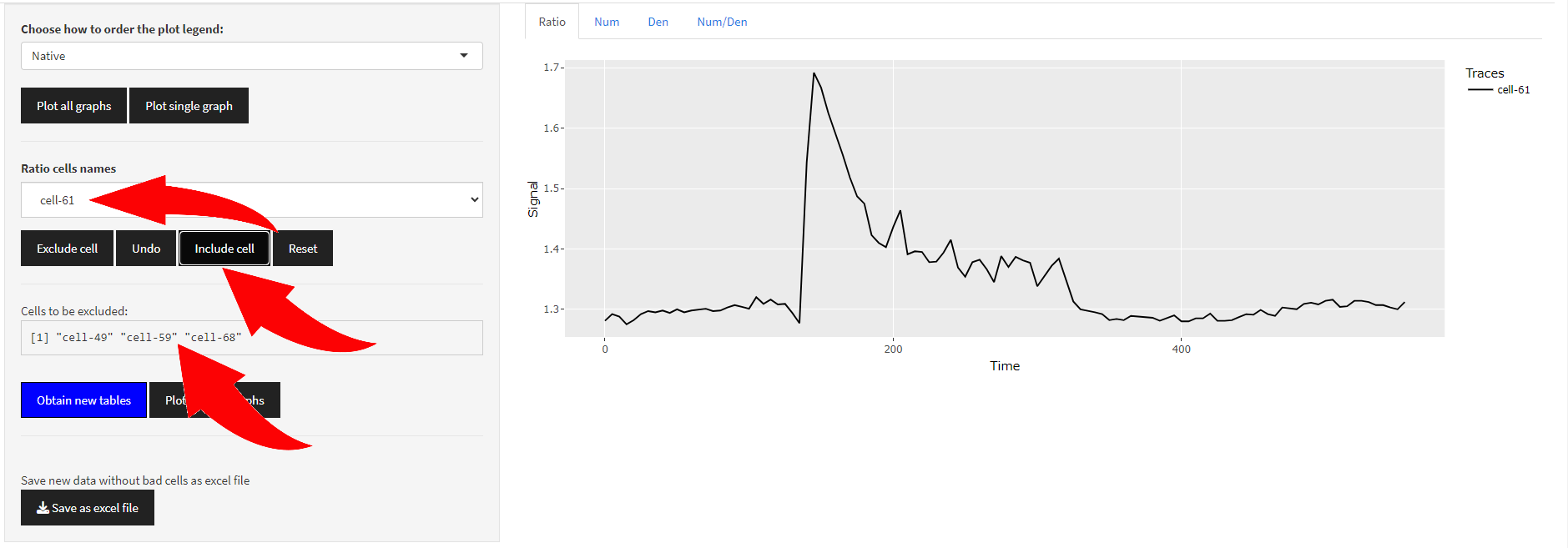


Sequence of trace’s names will be ordered using “Mixed” algorithm, if names were changed to have prefix “cell-” the order will be increasing and for raw names decreasing otherwise.

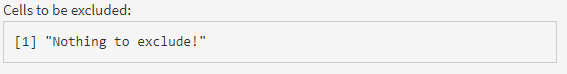
**Button “Undo”** cancel the LAST action. If all actions have been canceled manually user get this message:



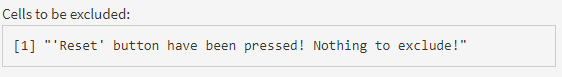
**Button “Include cell”** allows to include in the result table (and exclude from the “Cells to be excluded:” box) any trace of choice. But user is required to pick the name of the trace first in the “Ratio cells names” box.



The same message will appear if all traces have been included in the result table manually:

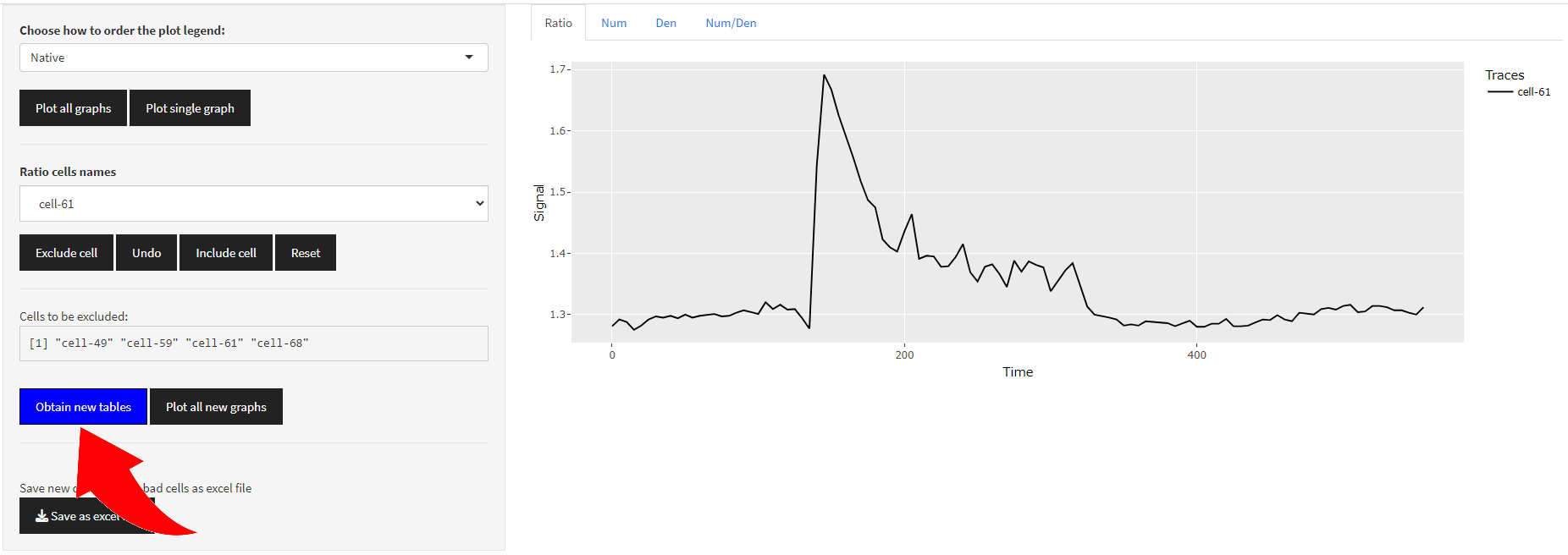


**Button “Reset”** clears the list of excluded traces with the following message:



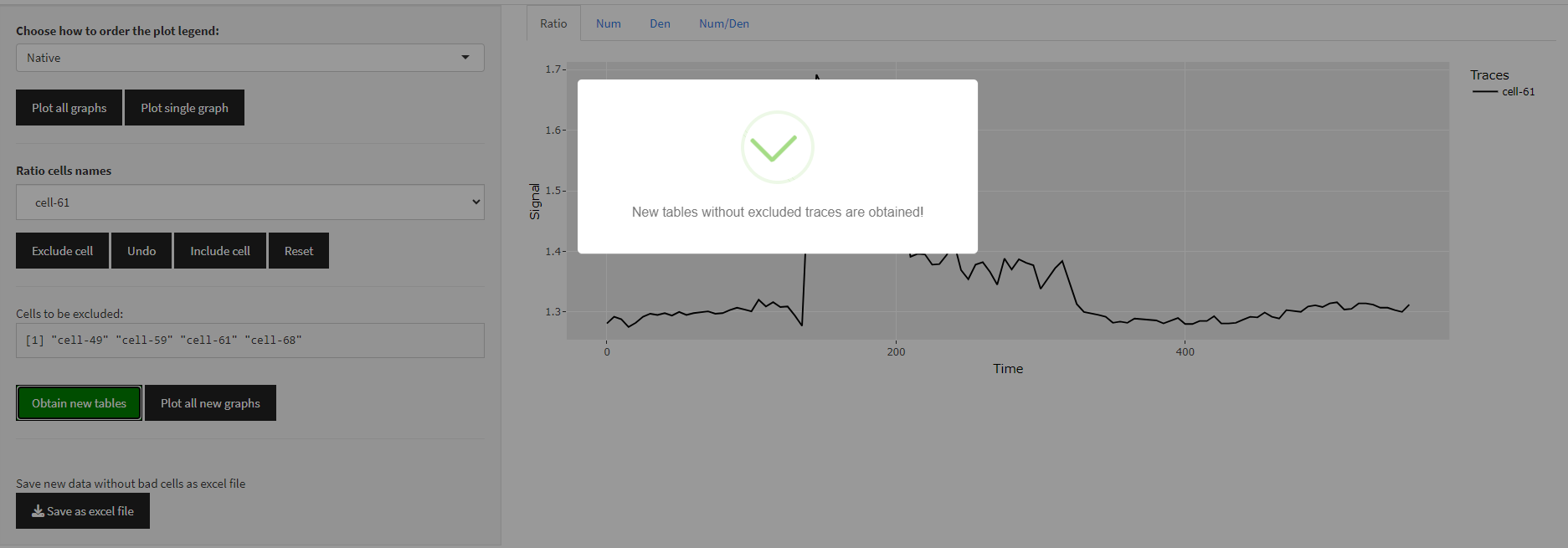
## Discarding unnecessary traces

When all the “bad” traces were chosen in order to save changes the “Obtain new tables” button (in blue, indicating mandatory action) should be pressed.

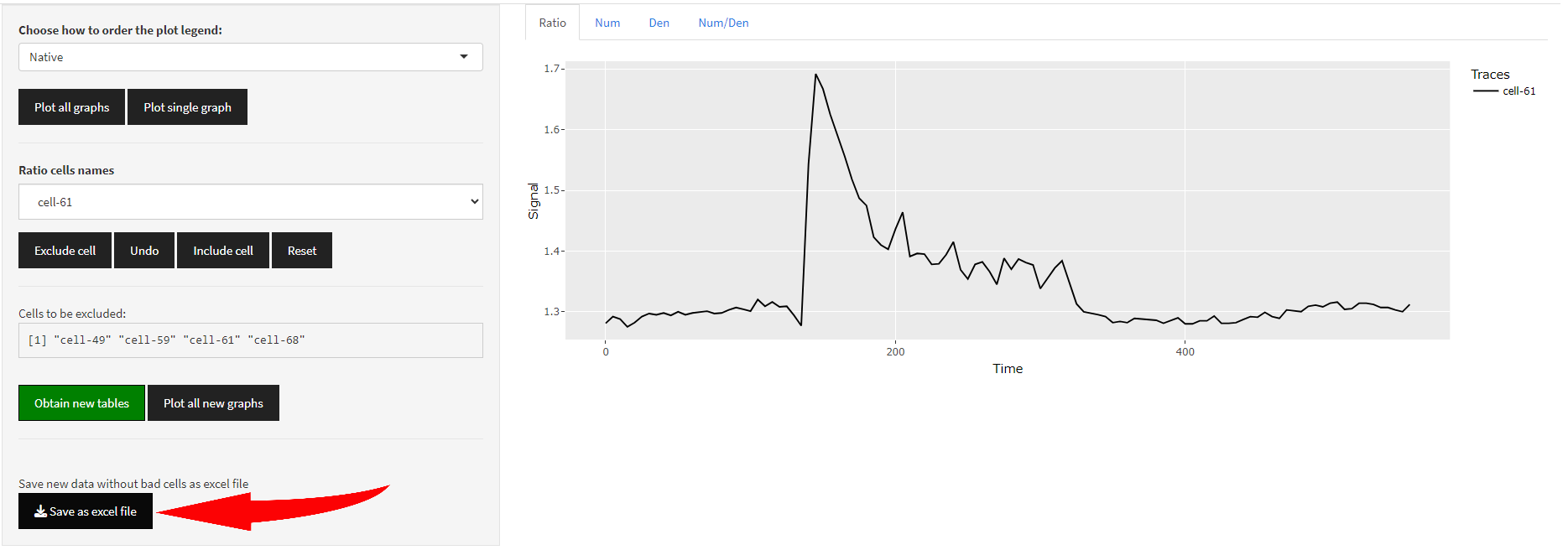


In order to finish the process, user needs to wait for the notification with the green check mark to appear.

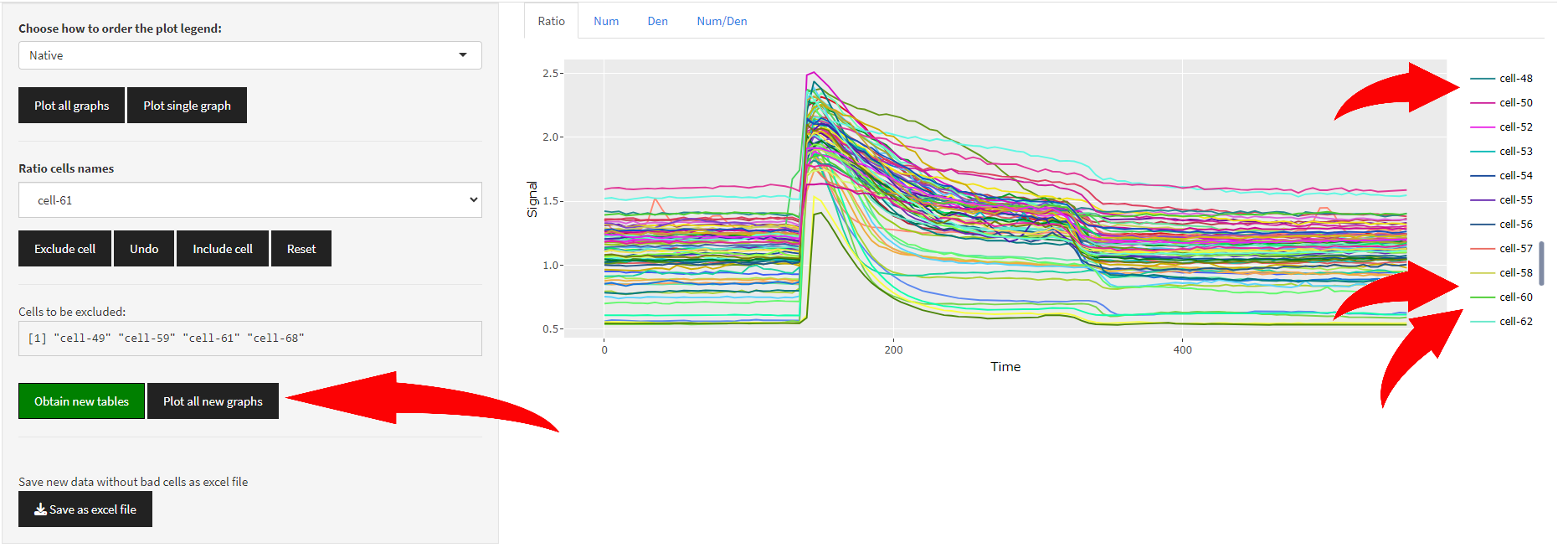
After pressing the “Obtain new tables” button, it turns green indicating that action has been taken.



After that it is possible to save the new table as excel file (**{*initial name of the file}*-CleanTable.xlsx**).



Or plot all new graphs without excluded traces.

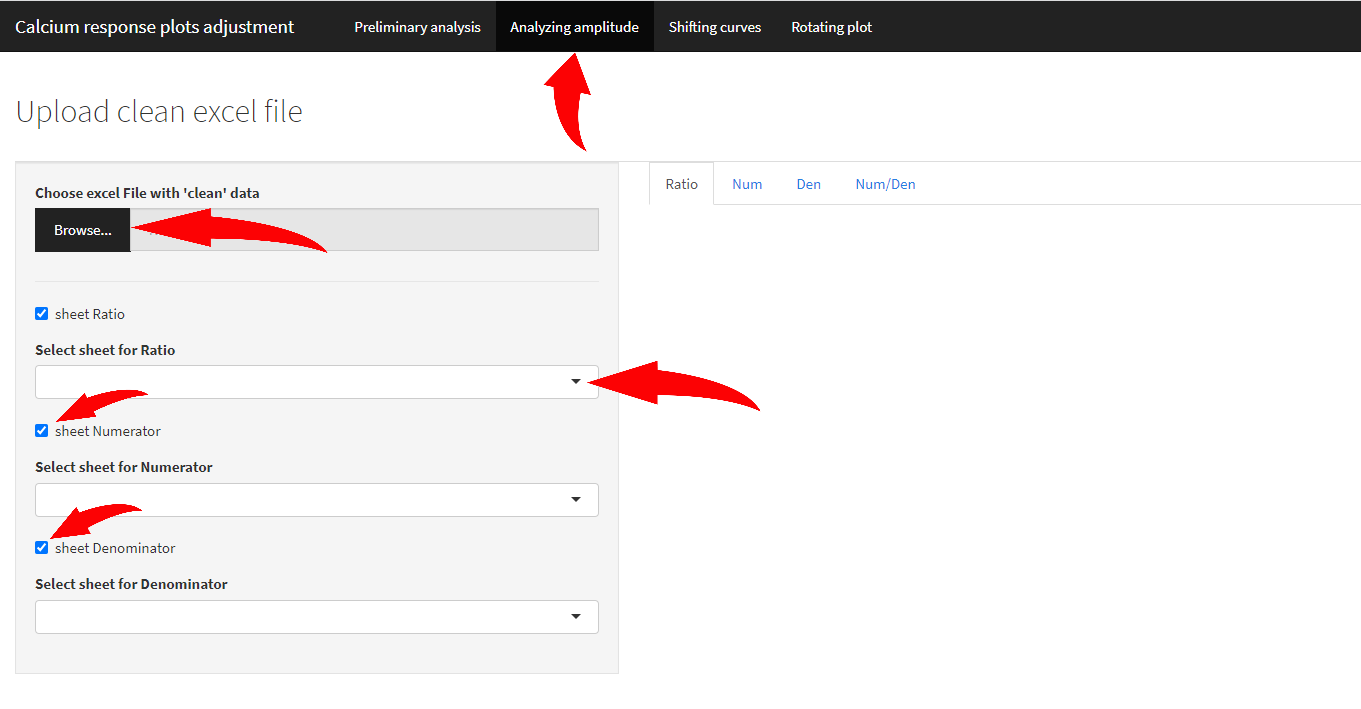


Even after that it is possible to look through the whole set of initial traces using the options above and pressing **“Plot single graph” button**. Excluded traces can be included again in the result table by choosing them in **“Ratio cells names” box** and pressing **“Include cell” button**. **After that it is necessary to press “Obtain new tables” button again to save changes!**

# TAB: Analyzing amplitude

## Uploading file with the “clean data”

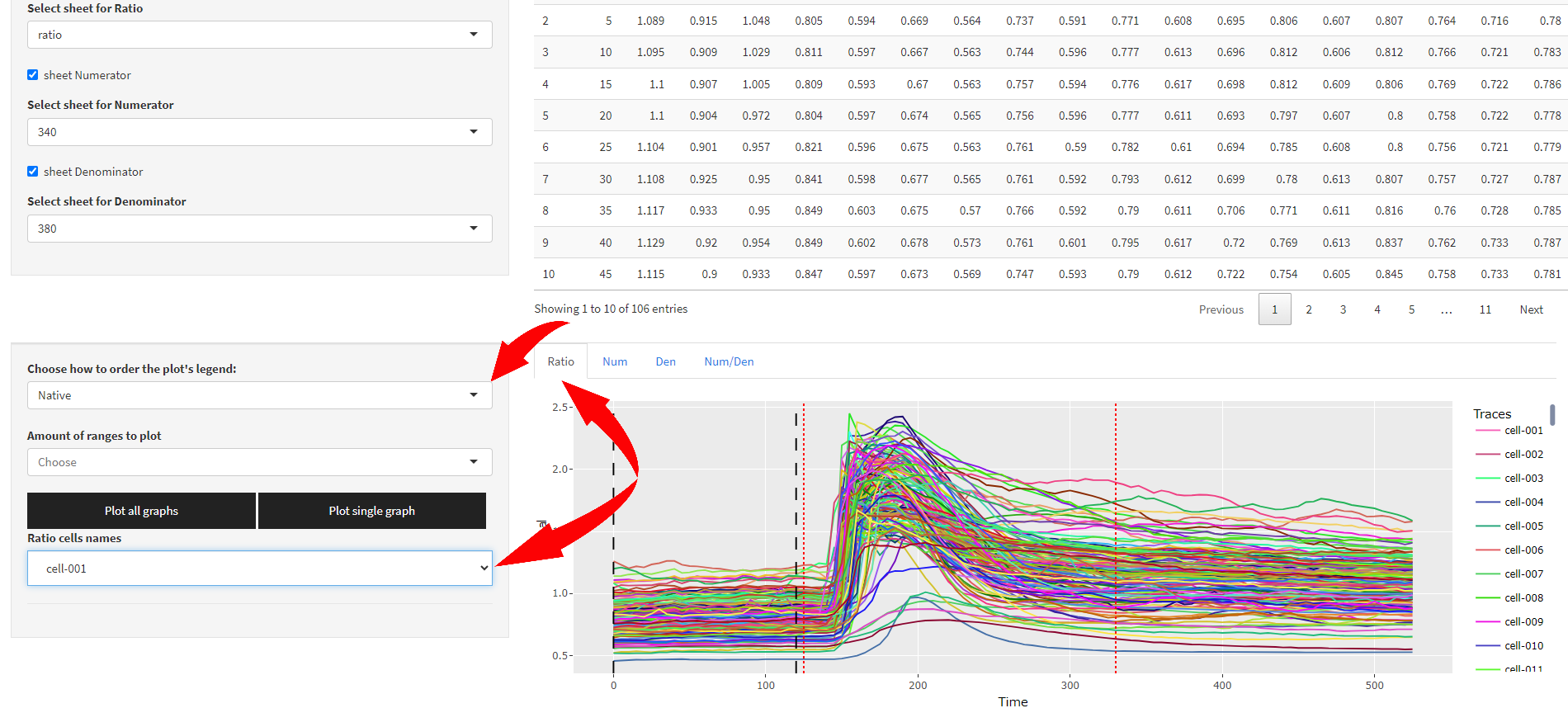
For this step it is preferable to use files processed on the previous stage **(…\*-CleanTable.xlsx)**. However, it should work with ‘raw’ files as well. Uploading process is pretty much the same as on the previous step.



If a file has only one sheet with necessary information all the other checkboxes should be unchecked.

## Plotting data

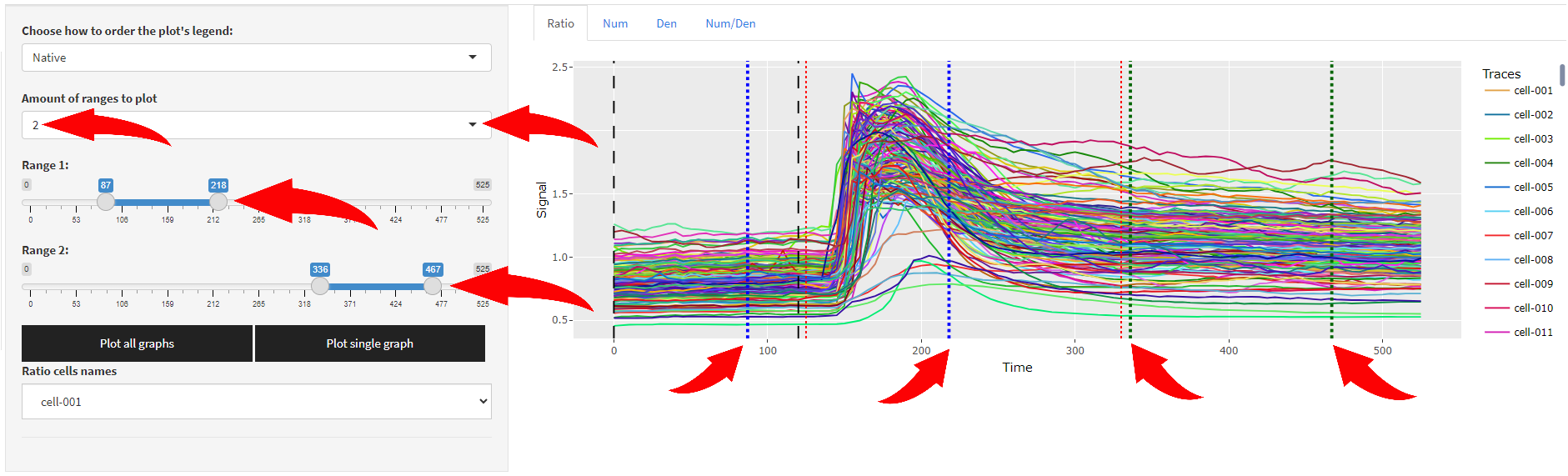
On the next step it is possible to plot all the graphs and sort them (trace’s names). For a single plot there is a menu allowing to choose a trace by its name.



**Each tab (“Ratio”, “Num”, “Den”, “Num/Den”) generates its own menu! But it works for “Plot single graph” button only.**

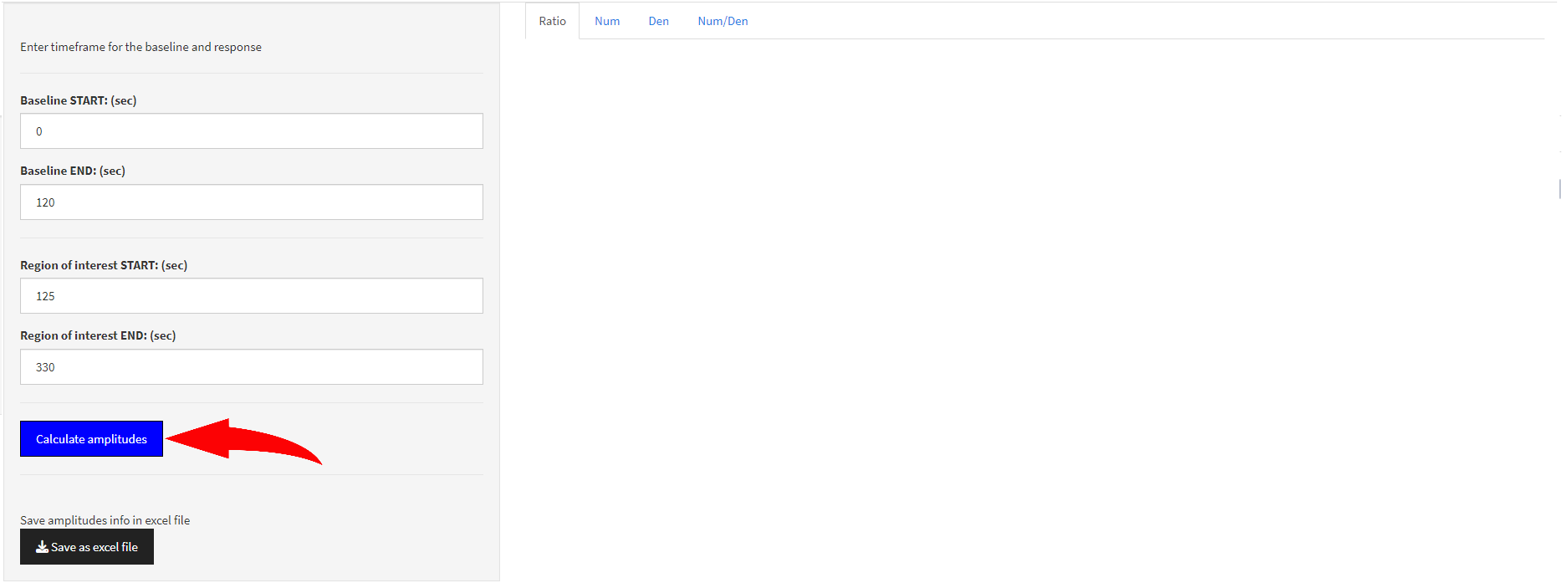


For all graphs as well as for the single graph it is possible to mark several regions of interest on the plot rendered. It is made for convenience so it is possible to find several regions of interest visually. However, only the region of interest marked as red dotted lines will be taken into account. So, all the additional ranges on the plot are exclusively for visualization.

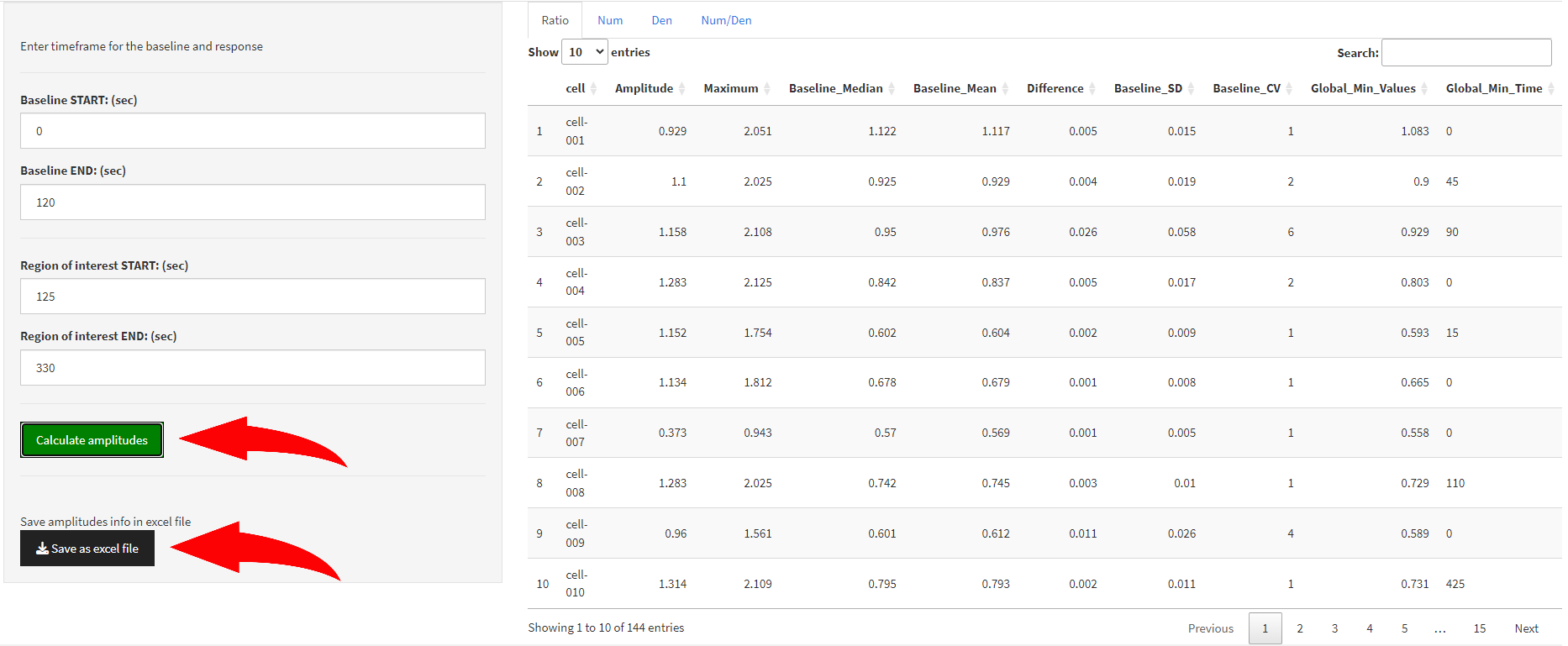


## Calculating amplitudes

The next element on the tab allows user to calculate amplitudes and obtain result table by pressing the “Calculate amplitudes” button.



Which turns green if you press it and the result table will appear. To save it as excel file press the following button below.

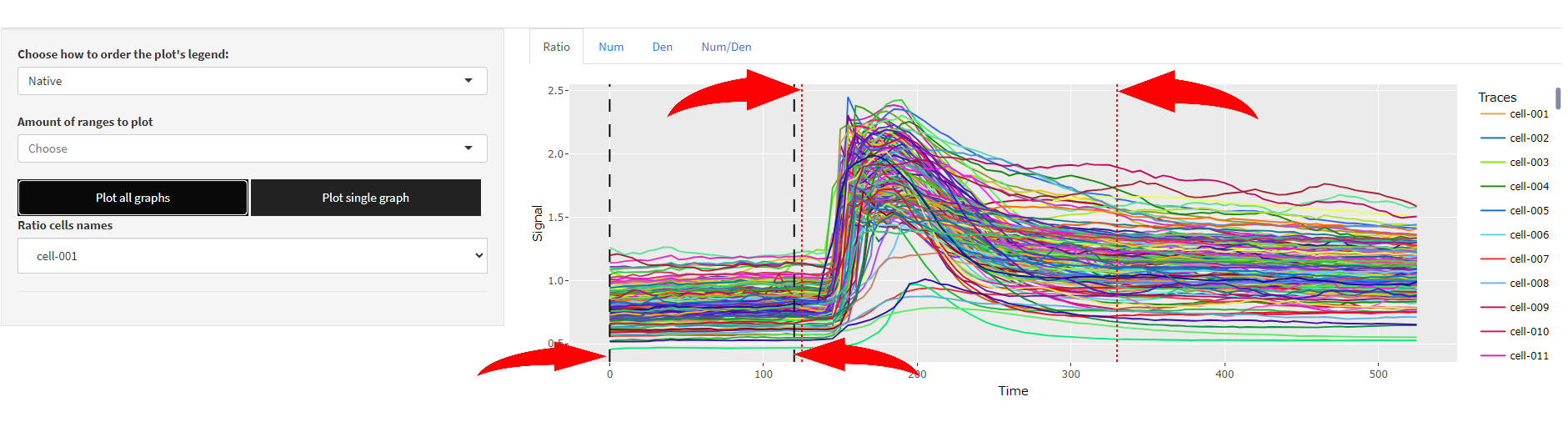


## Amplitudes result table

Altogether there are **two** excel sheets for every data type (Ratio/Num/Den etc.) that are formed when amplitudes are calculated.

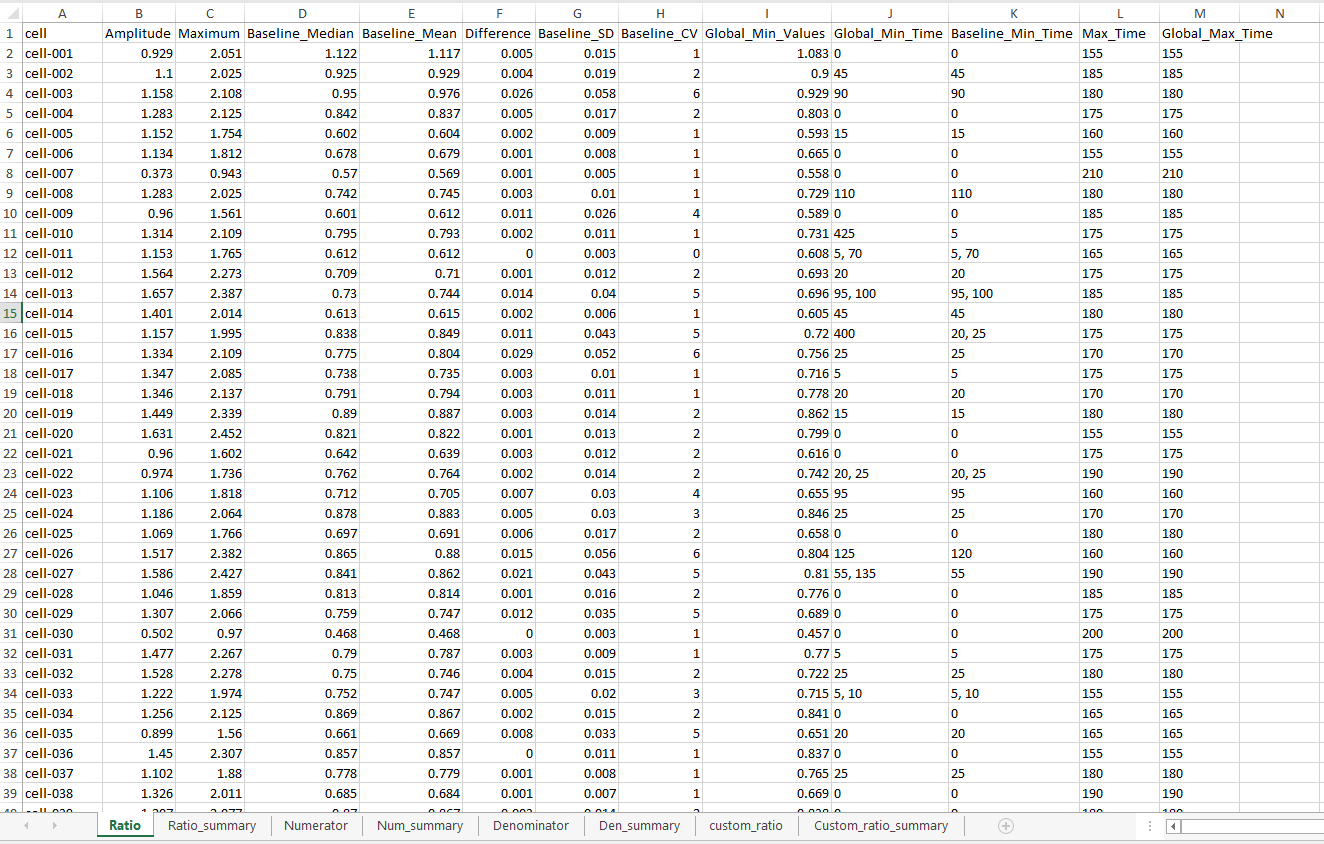
The **first** – statistic for each trace (cell). Column specification:

**cell** — the name (number) of the trace (cell)  
**Amplitude** — difference between **Maximum** and **Baseline\_Median  
Maximum** —maximum value in an interval between the **Region of interest START: (sec)** and **Region of interest END: (sec)**  
**Baseline\_Median** — median value in an interval between **Baseline START: (sec)** and **Baseline END: (sec)**

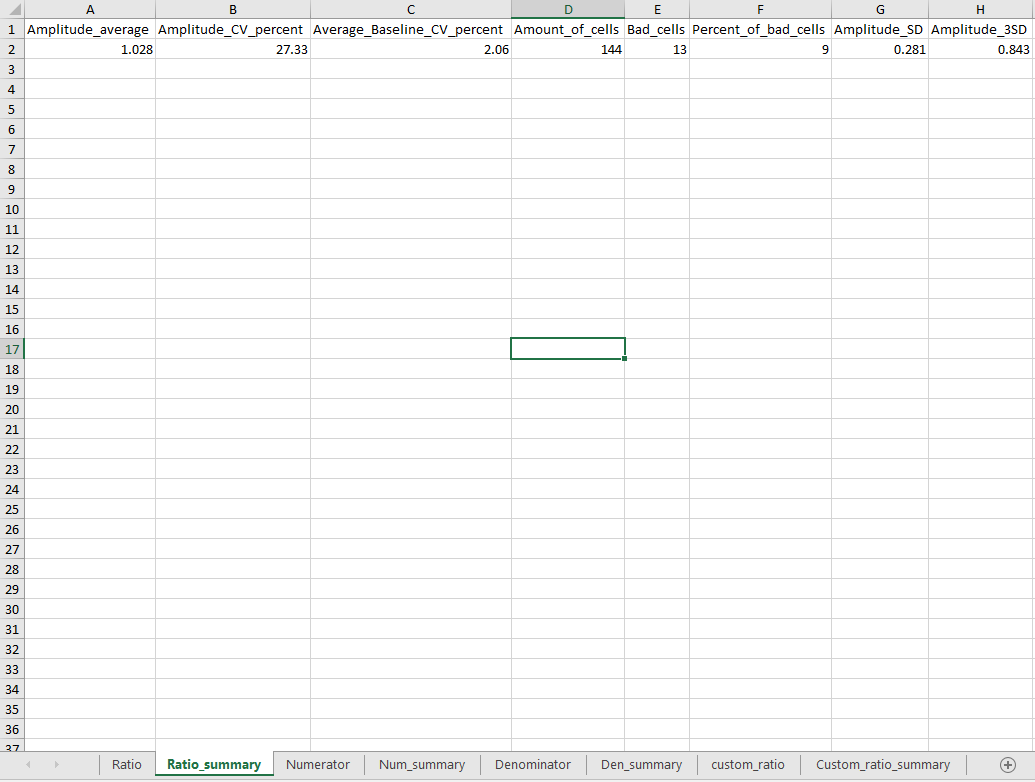
\*Region of interest marked on a plot for all traces as red dotted line, baseline – as black long dash line. 

\*This values can be adjusted in a box below.

**Baseline\_Mean** — mean value between Baseline START: (sec) and Baseline END: (sec)  
**Difference** — difference between **Baseline\_Median** and **Baseline\_Mean   
Baseline\_SD** — standard deviation for **Baseline\_Mean  
Baseline\_CV** — coefficient of variation for **Baseline\_Mean  
Global\_Min\_Values** —a global minimum value for each cell **Global\_Min\_Time** —atime value for theglobal minimum value for each cell  
**Baseline\_Min\_Time** —a minimum value for the baseline region  
**Max\_Time** —a time value for the **Maximum** value for each cell  
**Global\_Max\_Time** —a time value for the global maximum value (if other than the **Maximum** value) for each cell



The second – all traces statistics.



Column specification:

**Amplitude\_average** — mean for all the amplitudes  
**Amplitude\_CV\_percent** — coefficient of variation for Amplitude\_average (in percents)  
**Average\_Baseline\_CV\_percent** — coefficient of variation for Average\_Baseline\_CV\_percent (in percents)  
**Amount\_of\_cells** — an amount of cells were taken into account  
**Bad\_cells** — an amount of cells that were excluded  
**Percent\_of\_bad\_cells** — 100\***Bad\_cells**/**Amount\_of\_cells** (in percents)  
**Amplitude\_SD** — standard deviation for **Amplitude\_average**  
**Amplitude\_3SD** — 3\*standard deviation for **Amplitude\_average**

# TAB: Shifting curves

## Explaining algorithms

This tab allows user to adjust curves for every trace in order to dismiss possible delay between cells in the response to any compound introduced during the experiment. Altogether, there are two possible algorithms to adjust curves:

* **CCF (cross correlation function)**
* **Alignment of maximums**

**CCF (cross correlation function)** — in this case for each pair of traces a cross-correlation matrix is found. In order to find such matrix each trace is shifted from another one to the right and to the left (the value for that is called “maximum lag for CCF”). If, for example,

“maximum lag for CCF” = 2,

two columns of sample data (two sequences): (2, 4, 6, 5, 8, 9, 10) and (21, 43, 63, 58, 82, 92, 102) will be shifted to the right and to the left for 2 steps:

(2, 4, 6, 5, 8, 9, 10)  
(21, 43, 63, 58, 82, 92, 102)

(2, 4, 6, 5, 8, 9, 10)  
(21, 43, 63, 58, 82, 92, 102)

(2, 4, 6, 5, 8, 9, 10)  
(21, 43, 63, 58, 82, 92, 102)

(2, 4, 6, 5, 8, 9, 10)  
(21, 43, 63, 58, 82, 92, 102)

(2, 4, 6, 5, 8, 9, 10)  
 (21, 43, 63, 58, 82, 92, 102)

For every intersection of values, like in the last example:

X = (6, 5, 8, 9, 10)  
 Y = (21, 43, 63, 58, 82)

the following parameter is found:

= ,

Where and are mean values of each sequence, *N* – amount of values in each intersection of sequences (*X* and *Y* in the example).

Maximum acf value represents the higher correlation between two curves. This value corresponds to the lag value to apply to the second curve in order to get aligned curves. These lag values are put in the correlation matrix. Having this matrix allows to find a trace with the earliest response and shift every other trace to the left in order to get maximum correlation between curves.

**This algorithm is very slow but much more precise.** It is better to use it only if the amount of traces are lower than 300 and “maximum lag for CCF” lower than 60. Otherwise, it can take a long time to calculate.

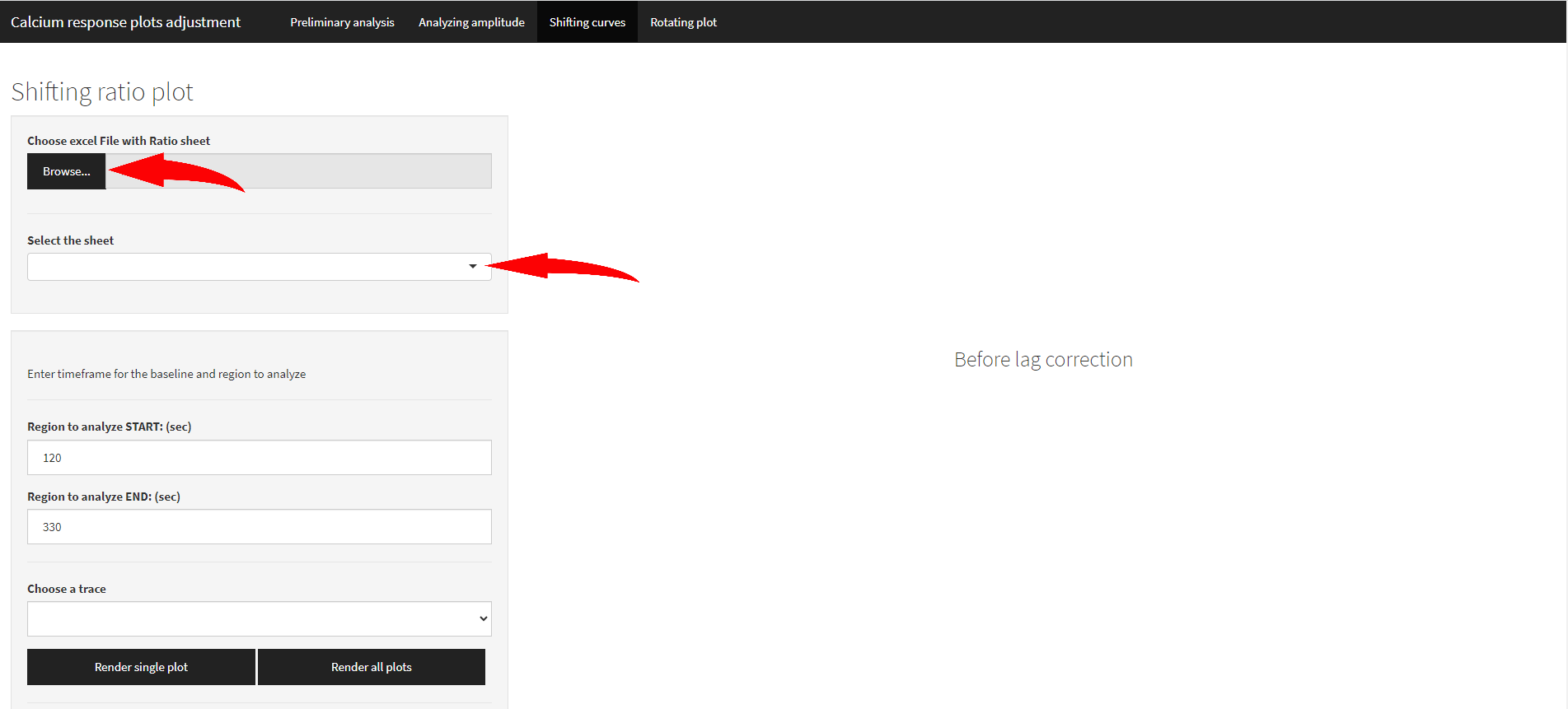
**Alignment of maximums** — finds local maximums that are more likely related to a response to any temporary compound introduction. It means that a **response on the plot should be represented only as a local maximum**, i.e. maximum values before and after the region of interest won’t be taken into account.

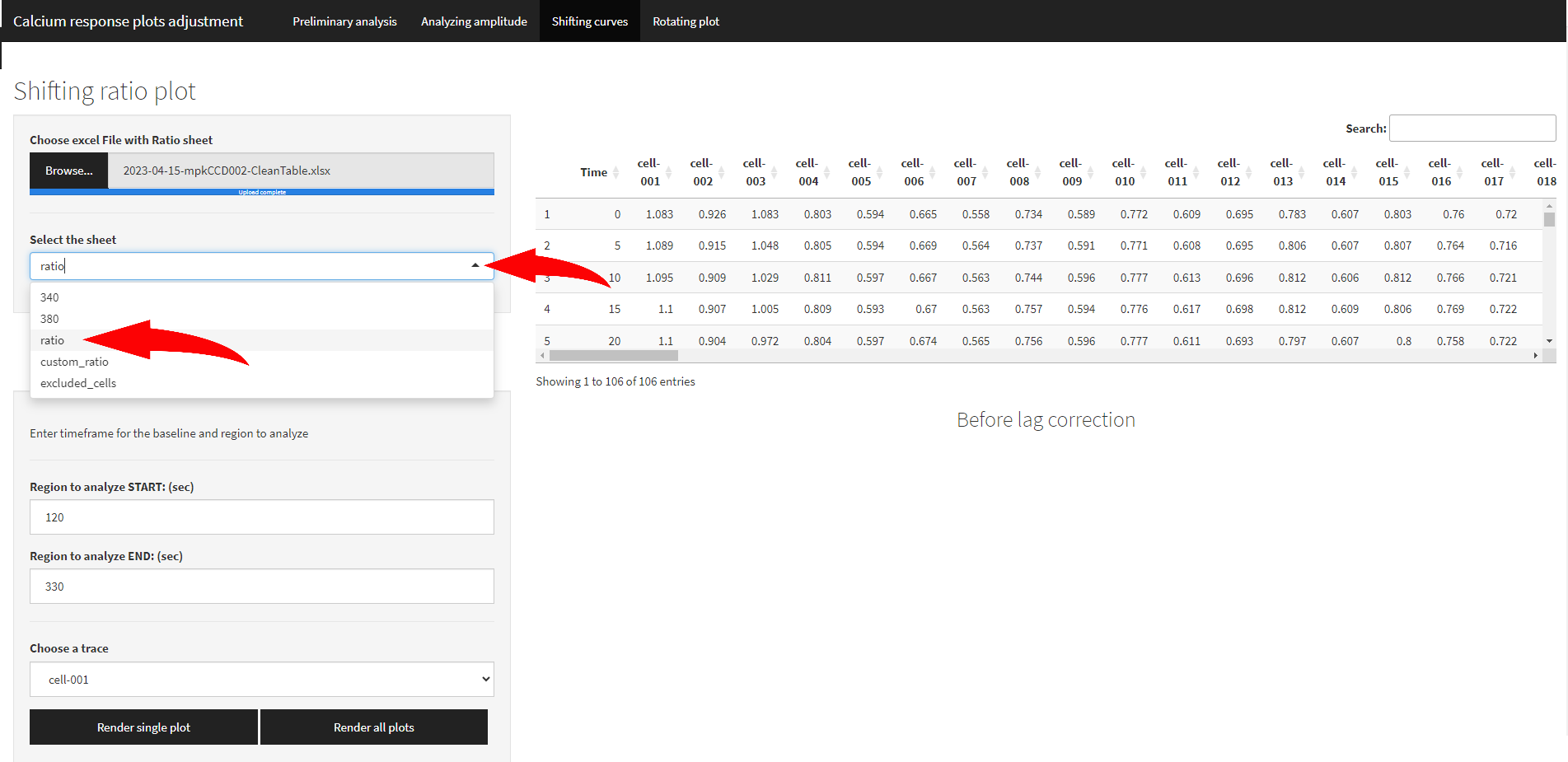
Each index and related time point of every trace’s local maximum are stored in a list where the trace with earliest local maximum can be found. Using the list’s information algorithm finds lag values for each trace to shift them to the left to match the earliest local maximum.

**This algorithm is very fast but less precise especially if a local maximum of any trace is represented by two close peaks that are equal. What’s more important, using this algorithm before launching CCF algorithm won’t increase the CCF’s performance.** The longest stage for CCF algorithm is calculating the correlation matrix with recommended lag values and it will be calculated in any case.

## Uploading files

For this step it is preferable to use files processed on the first stage **(…\*-CleanTable.xlsx)**. However, it should work with ‘raw’ files as well. Uploading process is pretty much the same as on the previous step. **It should be noted that on this tab only one excel sheet can be processed at a time! That is why user needs to specify it after loading the file.**

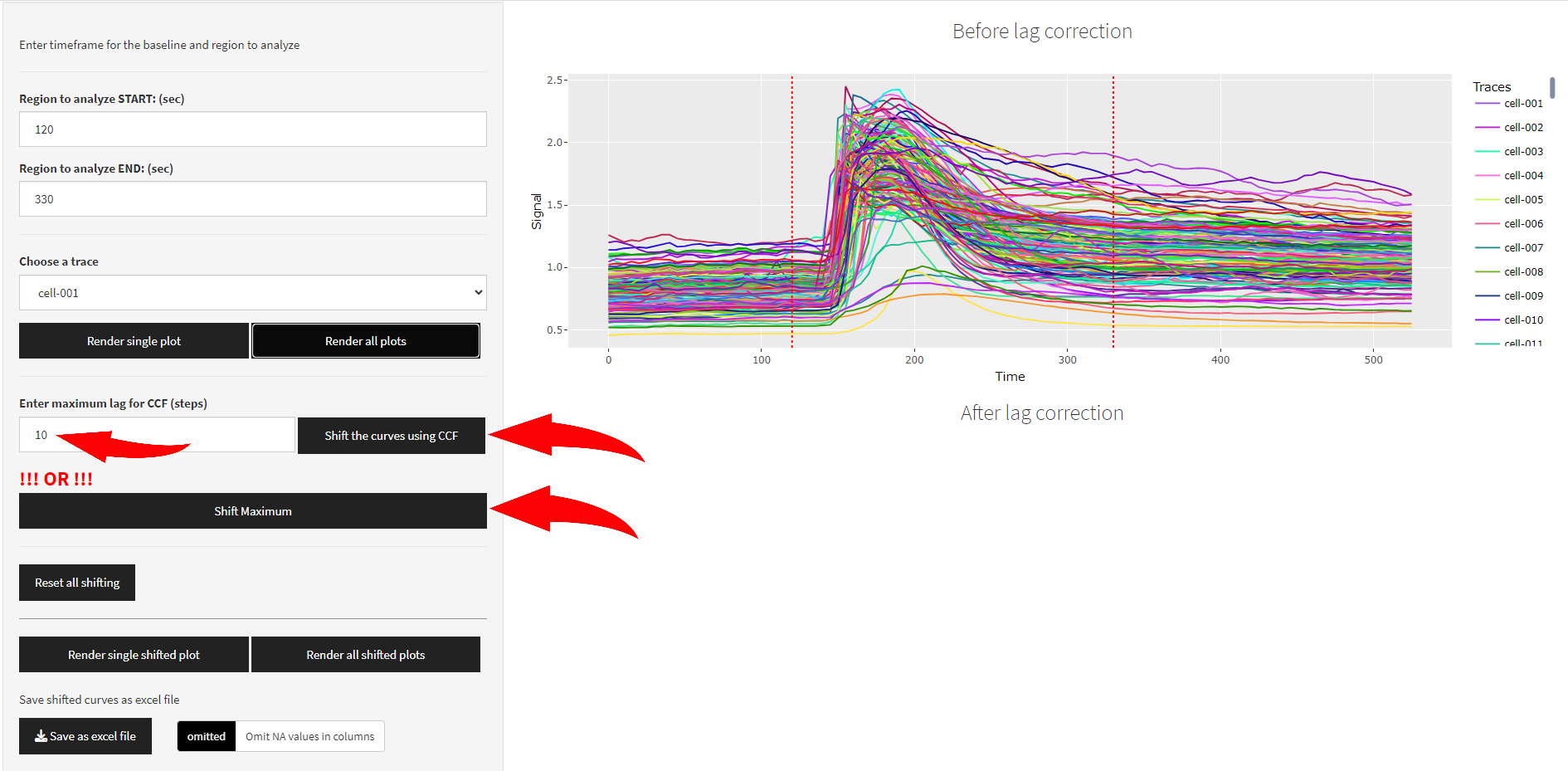




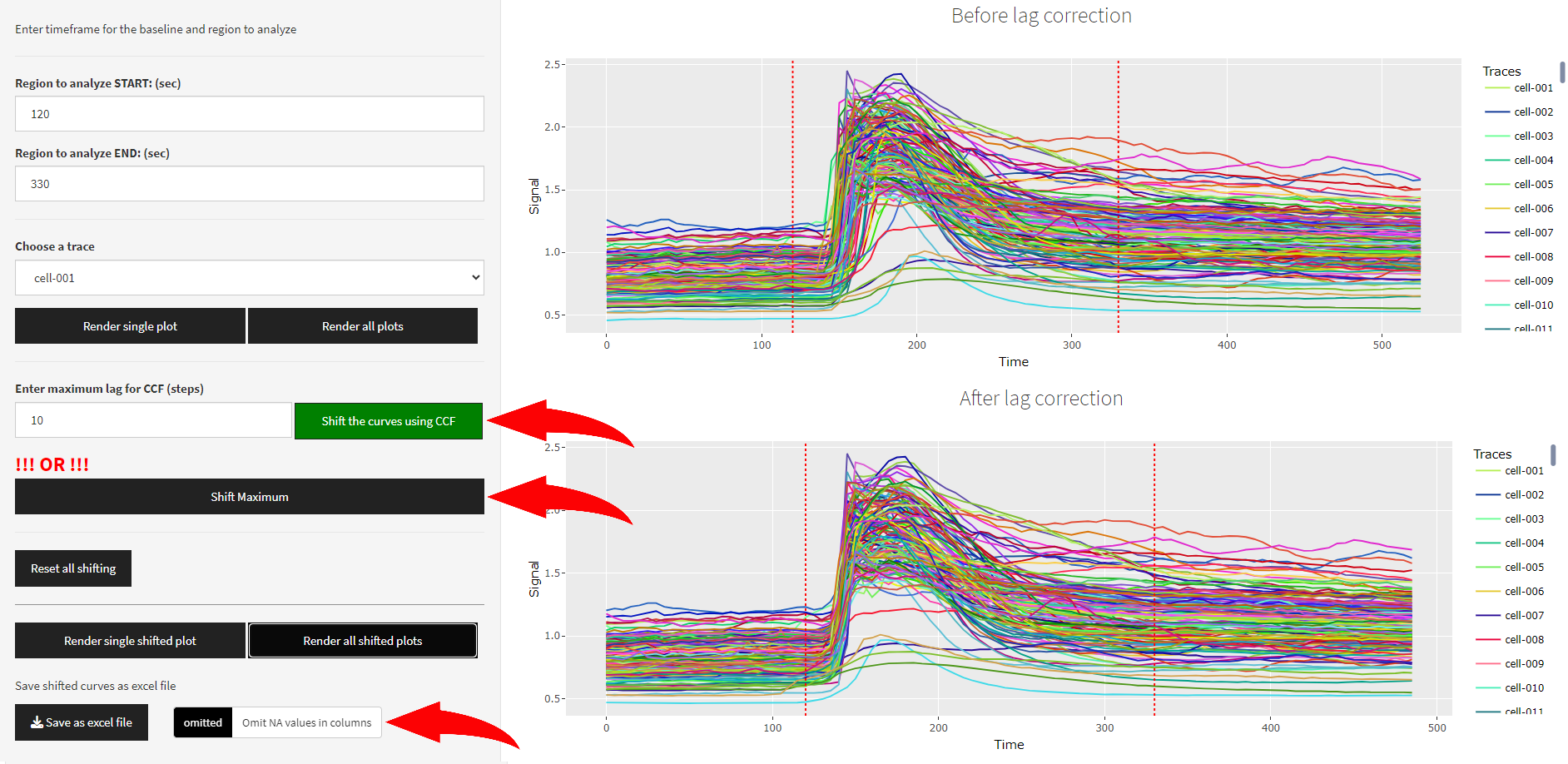
Next come buttons to render single plot with “Choose a trace” option and all the plots with region to align (to consider as a region with maximum). If all the curves are similar it is better to increase the range to make **CCF** algorithm more precise. On the contrary, using **Alignment of maximums** algorithm it is better to narrow the range to avoid including unspecific maximums, if they are greater than the response.

## Shifting curves

At first it is better to “Render all plots” and specify the region to analyze marked as red dotted line. Then it is possible to launch **ONE** of two algorithms by pressing the corresponding button. For **CCF** algorithm a maximum lag value should be entered. **The number represents amount of time points (values) that can be skipped/shifted to the left, not seconds.** It means that, for example, 10 stands for 10 time points (5 seconds each, or another value if used in experiment) which gives 10\*5=50 seconds maximum lag.



After pressing a button it turns green. After that if the result in unsatisfying user can press “Reset all shifting” and proceed with another algorithm. **Sequential launch of two algorithms can lead to unexpected behavior! In this case both buttons turn red.**

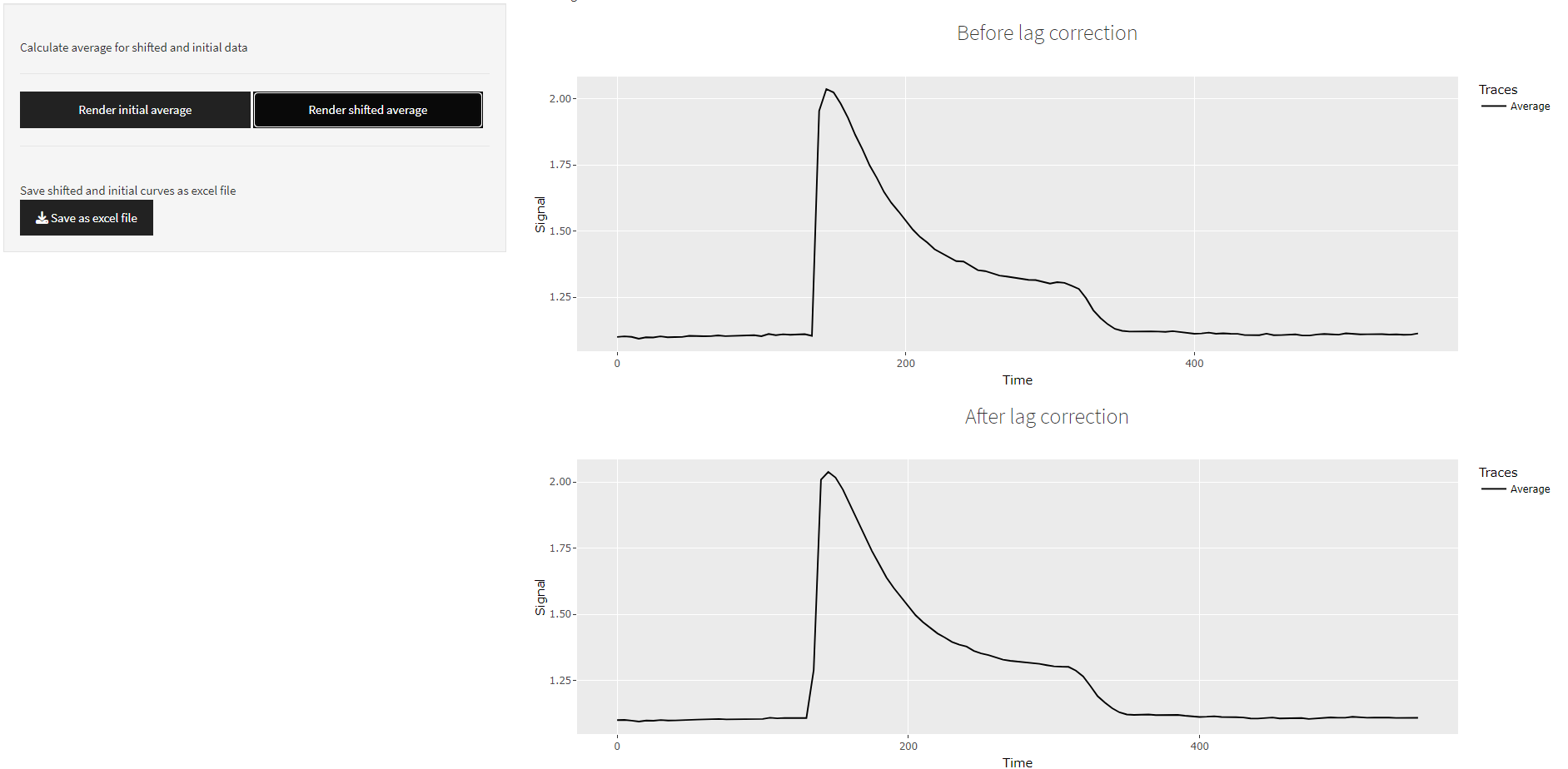


After launching the shifting process a status bar appears in the right lower corner. When it’s done a notification windows appears that tells the maximum lag achieved by the **CCF** algorithm. It means that the result plot for all the curves contain empty values (NA) and their amount equal this number. By default “Omit NA values in columns” switcher is on, so they are discarded. To keep them switch it to off. **Such data with NA values cannot be used in any further calculations using this program. To use the result file further “Omit NA values in columns” switcher should be in ON position.**

Shifted curves and their values can be saved by pressing “Save as excel file” button (**…\*-Shifted.xlsx**).

## Average curves

To verify the result of shifting curves an average curve before and after shifting can be plotted using the last UI element on the page.

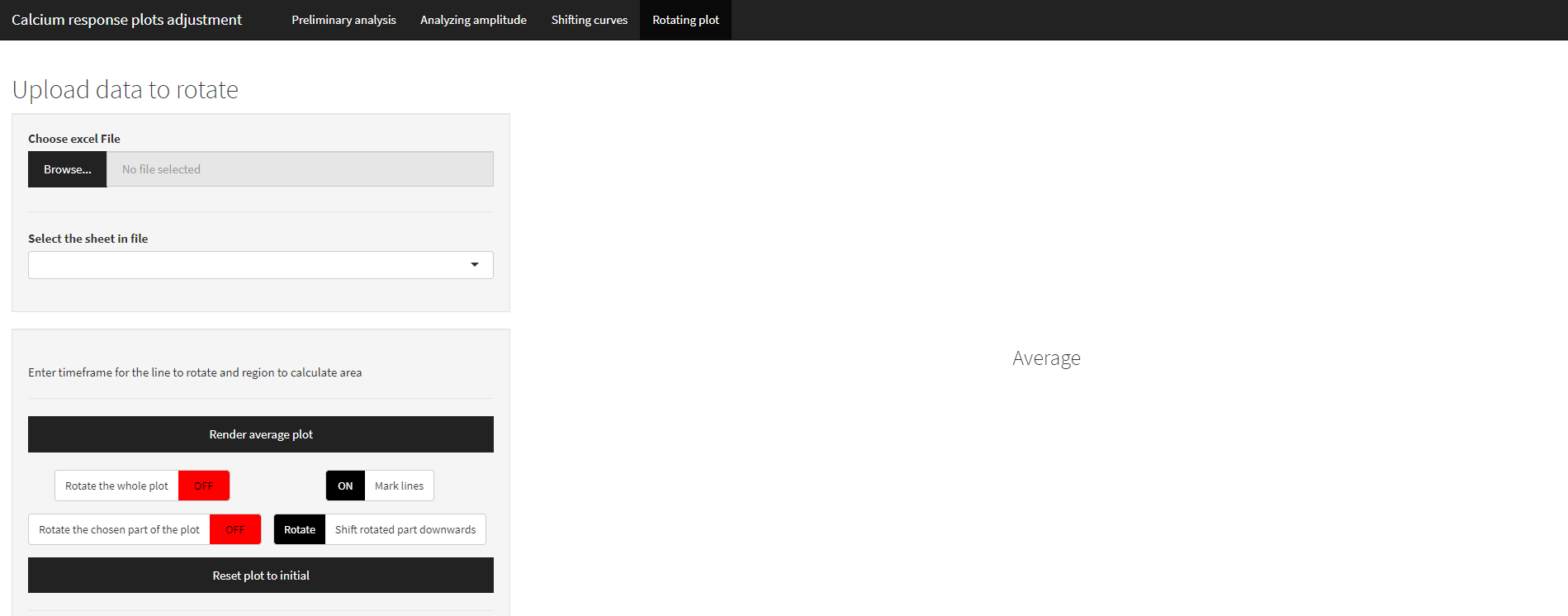


Plots appear after pressing the corresponding buttons. Data table before shifting and after **!!!with an additional column that represents mean values for all the traces!!!** can be saved as excel file as well   
(**…\*-Shifted\_with\_Average.xlsx**).

# TAB: Rotating plot

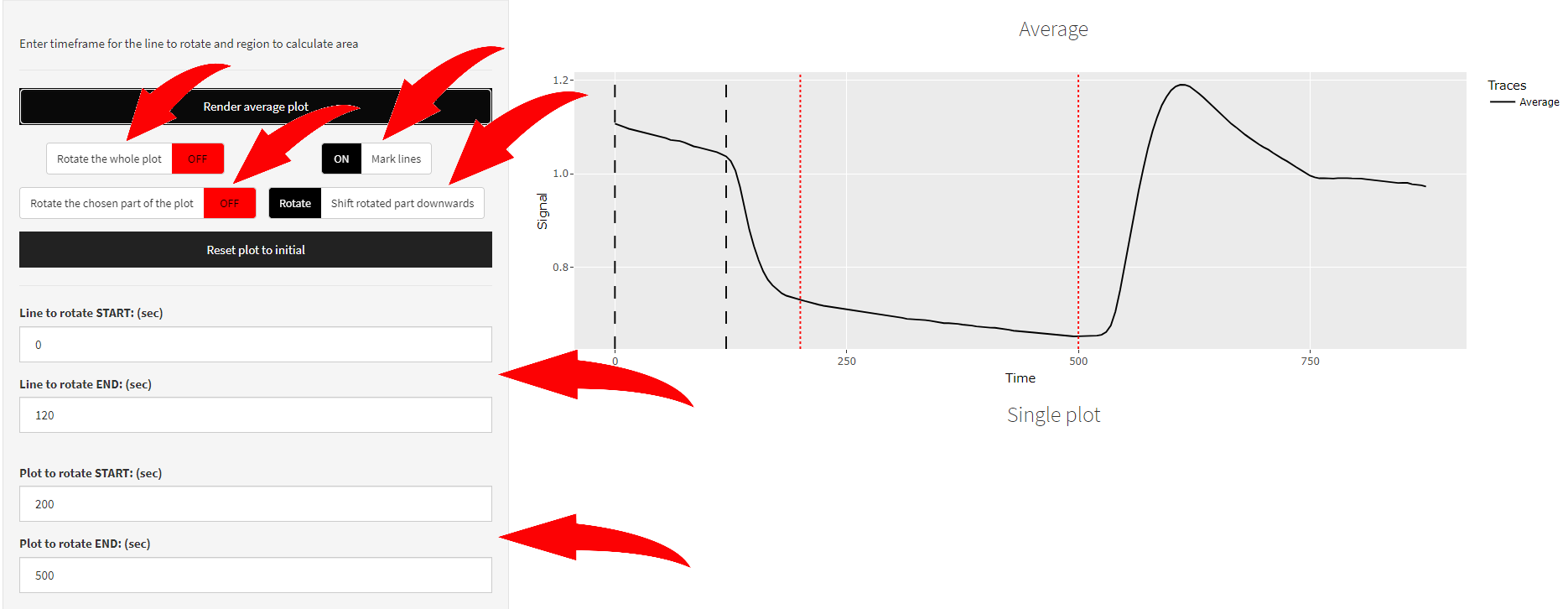
## Uploading files

For this step it is preferable to use files processed on the previous stage (**…\*-Shifted.xlsx**). However, it should work with ‘raw’ files as well. Uploading process is pretty much the same as on the previous step. It should be noted that on this tab only one excel sheet can be processed at a time! That is why user needs to specify it after loading the file.



## Rotating average curves

The next UI element allows user to render an average plot. “Line to rotate” region (black long dash lines) stands for the baseline that can be adjusted (aligned with X axis) independently. For this “Rotate the chosen part of the plot” switcher is used **but at first it is necessary to proceed with the rotation of the whole average curve**. To do that “Plot to rotate” region exists (red dotted lines). By pressing the switcher “Rotate the whole plot” for the region of interest a linear regression model will be constructed and according to it every value will be recalculated. Then “Rotate the chosen part of the plot” switcher can be used to adjust baseline only. “Shift rotated part downwards” switcher allows to shift base line up (the highest initial baseline value) or down (to the lowest initial baseline value). “Mark lines” switcher just removes/renders dotted and long dash lines.

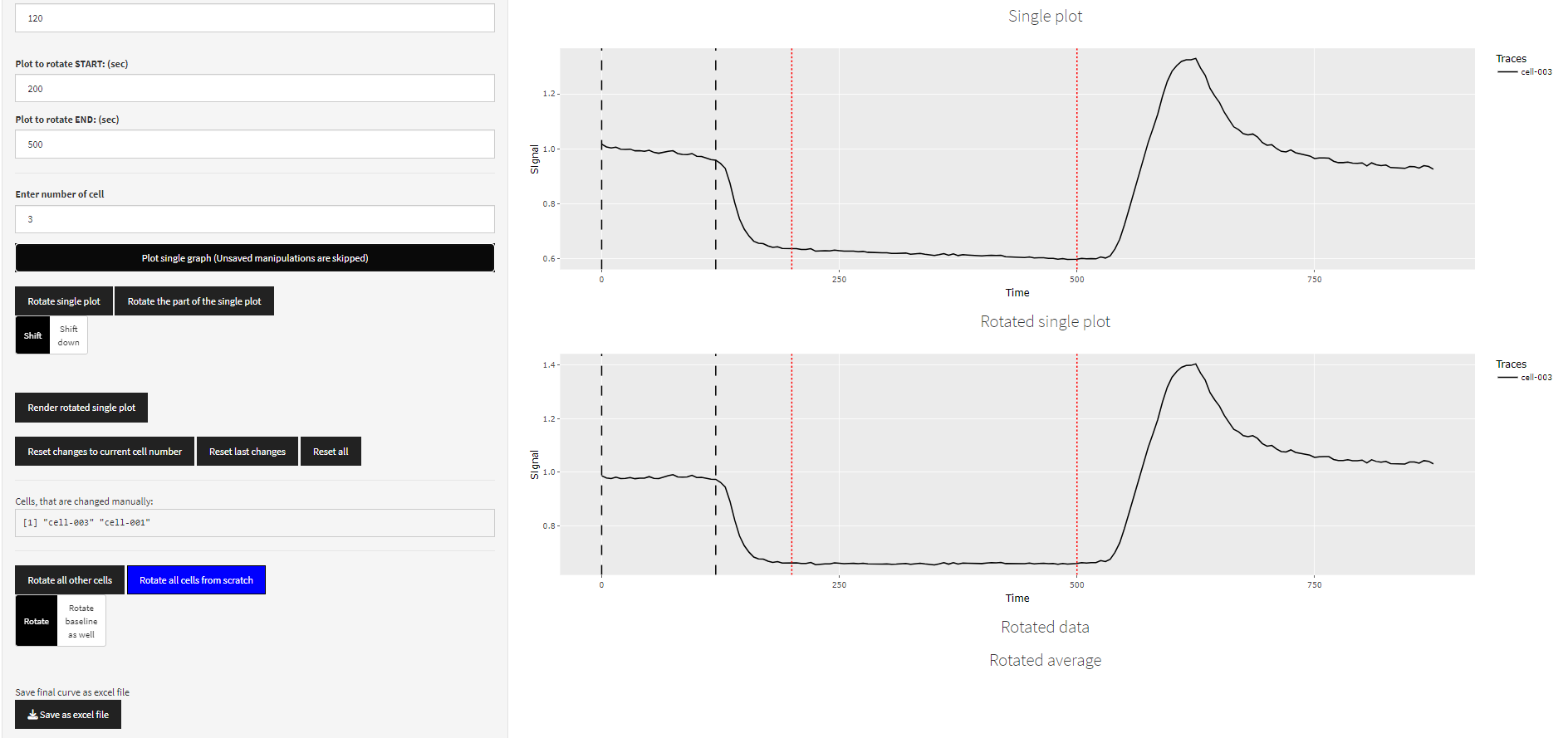


## Rotating single curve

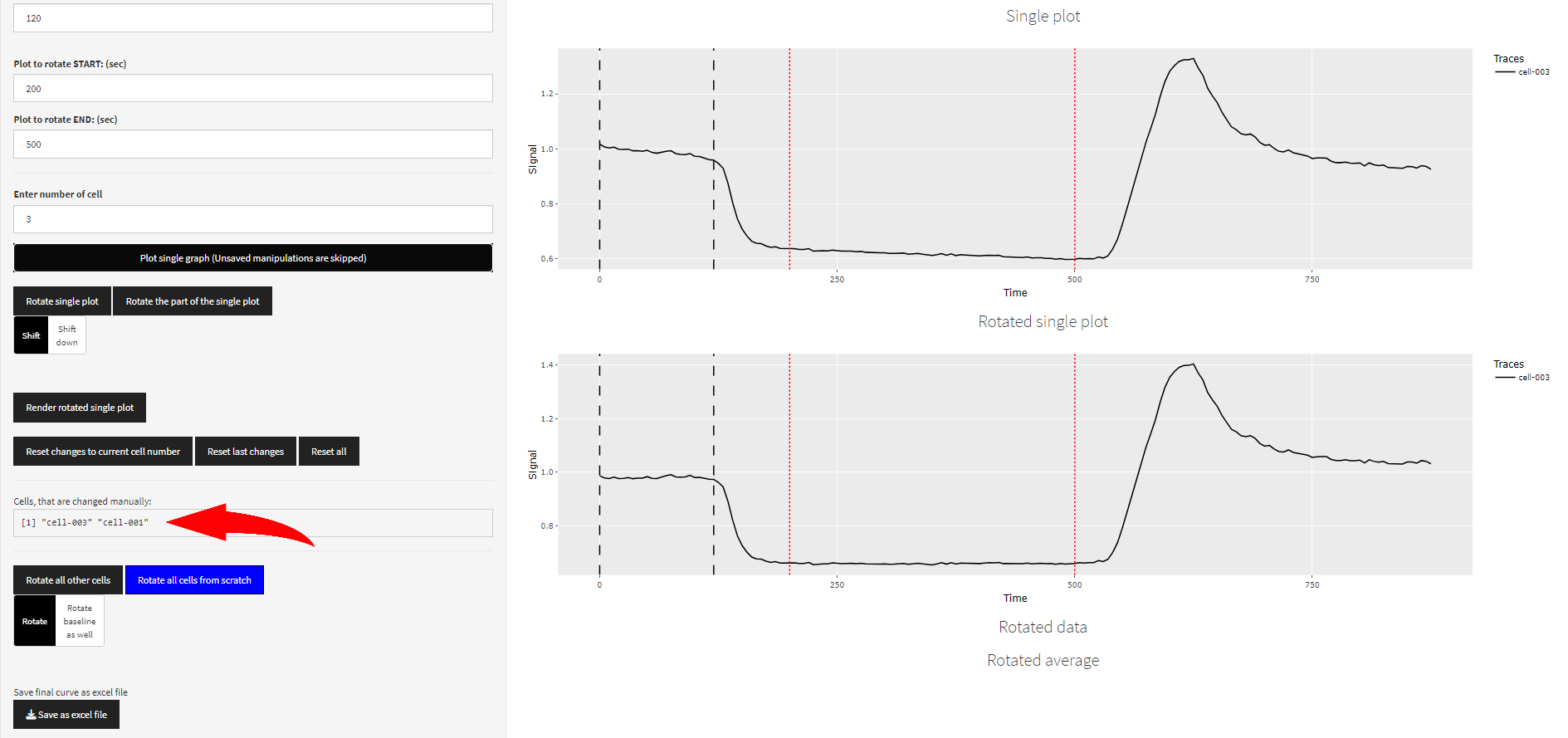
In the same manner it is possible to rotate each trace one by one, save this traces and automatically rotate all the other traces.



“Shift down” switcher in this case works the same as “Shift rotated part downwards” switcher and “Rotate baseline as well” in case of automatic rotation of all the other cells.



Cells, that are changed manually listed in a box. By pressing “Rotate all other cells” only traces not altered will be changed automatically and a table with rotated values appears. “Rotate all cells from scratch” button rotate all curves automatically and discard all manual changes.



Results can be saved as excel file on that stage. Rotating average curve and rotating all cells from scratch is the same process if all the ranges are the same. However, rotation of average curves gives only image that can be saved by copying it. In order to save it as a file “Rotate all cells from scratch” (in blue, which means essential) button should be pressed. Then data can be saved. Rotated average in this case would be identical as in the UI element above with rotated average plot.

## Calculating area

Using the last UI element it is possible to calculate area for every single trace manually or for just a portion and then calculate all the rest like on the previous step.



To see the current state a “Mark area” button should be pressed. Baseline linear model is represented as a blue line, and area as a green one.



In order to do that one needs to specify baseline range (to this range linear regression model will be applied to create a cut off for area measurment). This model is calculated for each trace independently and can be changed for every trace.

Next step is to specify the region of interest “Calculate area START/END”. **It is essential to press “Calculate area for all curves” button to calculate area automatically and only after that change it for every single trace of interest. User needs to obtain the table with calculated areas to modify row corresponding to a trace one by one!**

Both regions can be changed for a chosen trace. The current area after changing this values is showed for the chosen curve and user can see the value in “The calculated area is:” box. To save changes “Save manually calculated area for the current curve” button should be pressed so it will replace the row (for this trace) in the result table respectively.

“Calculate statistics” button can be pressed when all changes have been made to obtain summary over all traces. Both these tables will be present in the result excel file when pressing “Save statistics as excel file”.