TOF-ICP-MS Analysis Script Guide

General Instructions:

1. **Running the Script:**

To run the script effectively, you should install a source code editor such as VS Code or PyCharm. Keep in mind that the websites to download these editors may change in the future, so it's a good idea to search for the most recent versions online.

- PyCharm

https://www.jetbrains.com/pycharm/

- VS code

https://code.visualstudio.com/download

Install python, you can install extension form VS code (image) or PyCharm

or install it directly from this site:

https://www.python.org/downloads/

- Install miniconda (Anaconda)

https://docs.anaconda.com/miniconda/miniconda-install/

First, you need to install all the required Python packages using pip. Open your terminal VS code or command prompt and run:

```
pip install streamlit, pandas, plotly, numpy, scikit-learn, statsmodels, mendeleev, scipy, time, io, warning, re
```

- pip install streamlit pandas plotly numpy scikit-learn statsmodels mendeleev scipy time io warning re

Addressing Installation Errors

If you encounter the following error:

Streamlit is not recognized as an internal or external command, operable program or batch file

This might indicate that your environment is not set up correctly. To fix this, you can create and activate a new conda environment and try to install the packages:

conda create --name myenv

conda activate myenv

pip install streamlit ….

Running the Streamlit Application

After ensuring that all packages are installed without errors:

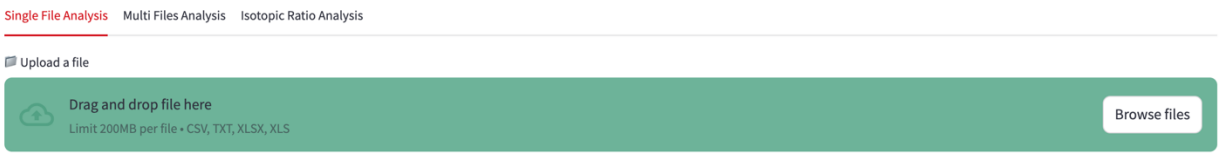
- Open the folder where your script (allapp.py or the script you want to run) is located.
- In your terminal, navigate to this folder using the cd command, or open the folder in your file explorer and use the “Open in Terminal” or “Open in Command Prompt” option.

Run the following command to start the Streamlit app:

streamlit run app.py

Replace app.py with the actual name of your script if it’s different.

Upon running the script, Streamlit will open in your default web browser, displaying the front page of your program. The page should look similar to the image below.



The script has three tabs, each tailored for a specific type of data analysis:

- Single File Analysis: This tab is for analyzing one file at a time with a more detailed approach, you can also combine different files Nu Quant Vitesse, SPCal format files.

- Multi Files Analysis: This tab is designed to handle and analyze multiple files simultaneously (only for Nu Quant Vitesse).
- Isotopic Ratio Analysis: This tab is specifically for isotopic ratio analysis. The file used here should be prepared exclusively for isotopic ratio analysis, containing isotope data from Nu Quant Vitesse.

Make sure that when working with the isotopic ratio analysis tab, you only upload the CSV file with the isotope data, as it's specifically meant for that type of analysis.

Clean data

To ensure that the lines of code related to cleaning the data based on the FWHM (Full Width at Half Maximum) are properly handled, you should follow these steps:

a- Search for the specific lines in the script:

- Use Command + F on a Mac or Ctrl + F on a PC.
- Search for the following lines:

df = clean_data(df)

dd = clean_data(dd)

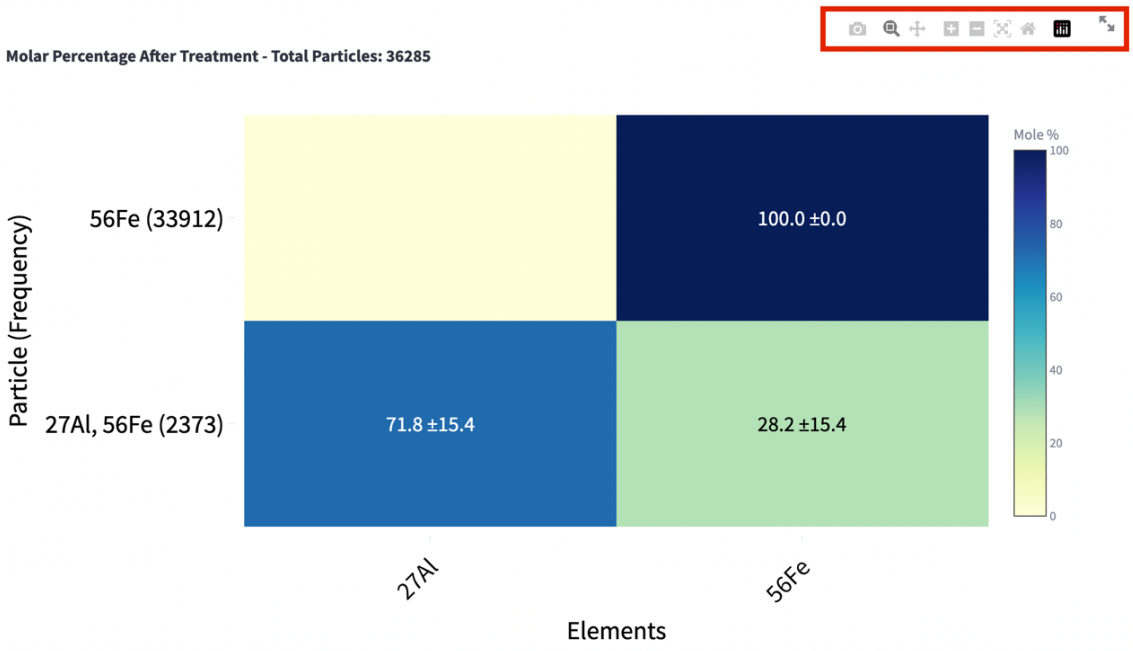
b- Comment out the lines:

If you decide that you don't want to clean the data based on the FWHM, you should comment out these lines by adding a # at the beginning of the line. This will prevent the data cleaning function from being executed.

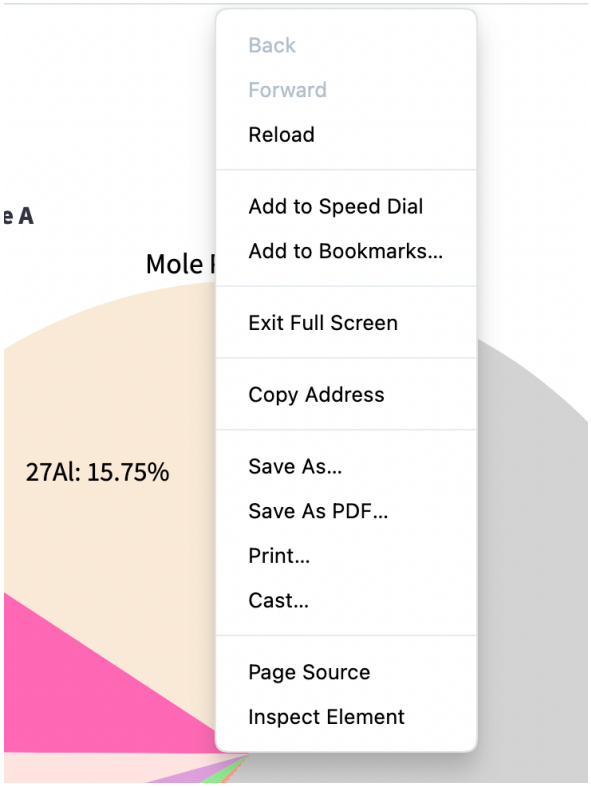
```
if df is not None:
                #st.write("Before Cleaning:")
                #st.dataframe(df)

                df = clean_data(df)

                if df is not None:
                    #st.write("After Cleaning:")
                    #st.dataframe(df)
```

Save the script add re-run the script.

2. **Uploading Files:**
    - To begin, use the 'Upload files' section to add your data files. The tool accepts file format csv form Nu Quant Vitesse. You can either drag and drop your files into the designated area or click 'Browse files' to select files from your directory.
3. **File Management:**
    - Once your files are uploaded, they will be listed along with their respective sizes.
    - Assign letters to each file for easy identification. You also need to enter the dilution factor and the acquisition time for each file to convert number of particles (event number) to Particles/mL and ensure the filename and assigned letter remain unchanged.
    - Besides, a table will be included to present information about each file, including the assigned letter, total particles, and calibration transport rate.
4. **Download CSV:**
    - Under each chart/table generated, there is an option to download the data as CSV by clicking on the buttons.
5. **Tool Bar:**
    - For each graph generated, you can find a toolbar in the top right corner by moving your cursor there. This toolbar offers various interactive options such as zooming in, zooming out, and downloading the plot as a PNG.



- If you use the Opera browser, you can right-click the image and save as PDF. See below as illustration.



Section Descriptions and Usage:

- Multi files Analysis

1. **Elemental Distribution:**
   - By selecting the "Show Mass and Mole" checkbox in the sidebar, you can generate pie charts that shows the distribution of elements based on mass and mole percentages. You can select any combinations to display by clicking on the checkbox.
   - You have the flexibility to customize your analysis by adjusting the "Threshold for Others" percentage. This feature allows you to set a threshold value, above which minor elements are grouped into an "Others" category.
2. **Mass Distribution Analysis:**

- By selecting the "Perform Mass Distribution Analysis" checkbox, you can generate histograms that display the distribution of particles per milliliter against their mass.
- The "Select max value for x-axis (Mass fg)" slider sets the maximum value on the x-axis, allowing you to focus on a specific range of masses.

3. **Distribution Analysis:**
    - The Distribution Analysis section provides a visualization of the statistical distribution of elements within your sample. By selecting the "Perform Distribution Analysis" checkbox, you can generate plots that represent the probability distributions of selected elements.
    - The dropdown menu for selecting the distribution type offers several options, each suitable for different kinds of data and analysis needs. Below is a table that discusses each distribution type, simple description, and the meaning of each test:

Table 1: Distribution Type information

Distribution Type	Description
Gaussian	A bell-shaped curve representing a normal distribution.
Poisson	A distribution representing the probability of a given number of events occurring in a fixed interval.
Exponential	Describes the time between events in a Poisson process.
Binomial	Represents the number of successes in a fixed number of independent experiments.
Log-normal	A distribution of a variable whose logarithm is normally distributed.
Gamma	A two-parameter family of continuous probability distributions.

Weibull	A versatile distribution used for modeling life data, reliability, and failure times.
Cubic Spline	A piecewise polynomial function that can model complex relationships.

Ternary Diagrams:

- By selecting the “Perform Ternary Analysis” checkbox, you can visualize the relationship between three different elements in your sample. You will need to select three elements for the ternary diagram by clicking on the dropdown menu and choosing the desired elements. The selected elements will be represented on the three corners of the ternary diagram.
- It's important to note that if the selected elements have a very low or high concentration, the plot might appear cluttered or skewed. In such cases, consider adjusting the element selection or normalizing the data before plotting. Additionally, ensure that the elements you select are relevant to the compositional analysis you intend to perform, as irrelevant selections might not provide meaningful insights.
- Also, the elements selected for analysis will need to be present in all the CSV files analyzed.

4. **Mole Ratio Analysis:**
    - By selecting the "Perform Mole Ratio Analysis?" checkbox, you can analyze the ratio between two different elements in your sample. You will need to select the elements for the mole ratio analysis by clicking on the dropdown menus and choosing the desired elements. The mole ratio will be plotted on the X-axis, and the frequency of each ratio will be displayed on the Y-axis. If the selected element does not contain in any of the files uploaded, there would be a warning message saying “No valid data available for plotting”.

5. **Heatmap Analysis：**

- By selecting the "Perform Heatmap Analysis?" checkbox, you can visualize the relationship between different elements in your sample using a heatmap. You will need to select the file for the heatmap analysis and adjust the parameters such as the color scale, start, and end combinations. The heatmap will display the frequency of particle detection for each element combination, with colors representing the intensity of mole percentage of each element in the particle.

Single file analysis

1. **Elemental Distribution:**
    - By selecting the "Show Mass and Mole" checkbox in the sidebar, you can generate pie charts that shows the distribution of elements based on mass, and mole percentages.
    - You have the flexibility to customize your analysis by adjusting the "Threshold for Others" percentage. This feature allows you to set a threshold value, above which minor elements are grouped into an "Others" category.
2. **Heatmap Analysis**:
    - By selecting the "Perform Heatmap Test?" checkbox, you can visualize the relationship between different elements in your sample using a heatmap. This feature allows you to select the data type (mass, mole, and mole%) for the heatmap and customize various parameters such as the color scale, font size for numbers, and the range of combinations to be analyzed.
    - The heatmap displays the frequency of particle detection for each element combination, with colors representing the intensity of mole percentage of each element in the particles.

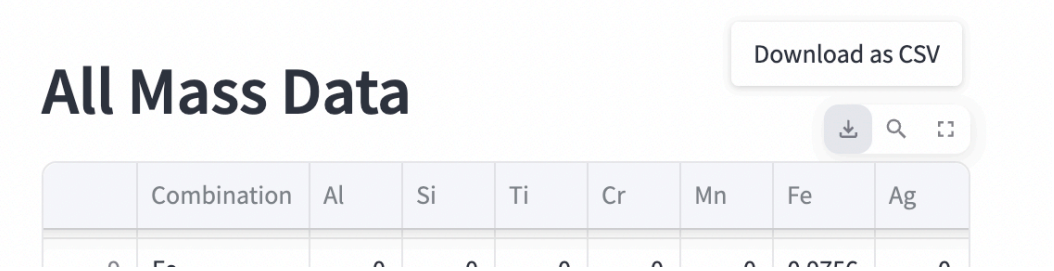
3. **Mass Distribution Analysis**:
    - By selecting the "Perform Mass Distribution Analysis?" checkbox, you can view the mass distribution of particles in your sample. You can select up to three elements to view their histograms, adjust the bin size, and set the maximum value for the x-axis.

- Additionally, you can customize the colors for different data categories such as all data, single detections, and multiple detections. This analysis provides a detailed histogram showing the frequency distribution of particle masses for the selected elements.

4. **Single and Multiple Element Analysis**:
    - By selecting the "Single and Multiple Element Analysis?" checkbox, you can analyze the distribution of single and multiple elements within your sample. This feature allows you to set thresholds for multiple and single elements, and it generates pie charts representing the proportion of different elements detected as single or in combination.

5. **Element Distribution Analysis**:
    - By selecting the "Perform Element Distribution" checkbox, you can analyze the distribution of specific elements within your sample. You can select elements to analyze and exclude from combinations, as well as set a count threshold for display.
    - This analysis generates a pie chart showing the distribution of particles containing the selected elements, providing insights into the elemental composition of the sample.

6. **Raw Data after Combination**:
    - By selecting the "Export Raw Data?" checkbox, you can choose to export the raw data after combination for further analysis. This feature allows you to select the data type to aggregate and display, providing the option to download the combined raw data as a CSV file.



7. **Clustering Analysis**:
    - By selecting the "Perform Clustering Analysis?" checkbox, you can perform hierarchical clustering on your sample data. This feature allows you to select the clustering method and the number of clusters.
    - Below is a summary table for each clustering method.

Table 1: Clustering Method Information

Name	Description	Meaning/Explanation
Hierarchical	Builds a hierarchy of clusters either through agglomerative (bottom-up) or divisive (top-down) methods.	Useful for understanding the data structure at various levels of granularity.
Spectral	Uses the eigenvalues of a similarity matrix to perform dimensionality reduction before clustering in fewer dimensions.	Effective for clustering data that is not linearly separable in its original space. Often used in image segmentation and community detection in networks.
Gaussian	Assumes data is generated from a mixture of several Gaussian distributions and uses the Expectation-Maximization algorithm to estimate the parameters.	Suitable for data that can be modeled as a combination of multiple Gaussian distributions. Provides probabilistic cluster assignments.
K-Means	Partitions data into K clusters by minimizing the sum of squared distances between data points and the centroids of the clusters.	One of the most commonly used clustering methods. Works well for spherical clusters of similar size.
Mini-Batch K-Means	A faster, approximate version of K-Means that uses mini-batches of data to update the cluster centroids.	Suitable for large datasets where standard K-Means would be too slow. Provides a balance between speed and accuracy.
Mean Shift	A non-parametric method that iteratively shifts data points towards the mode of the distribution until convergence.	Works well for discovering clusters of arbitrary shape. Computationally intensive for large datasets.

Isotopic ratio analysis

1. **Mole Ratio Analysis:**

- Enable the "Perform Mole Ratio Analysis?" checkbox to analyze the mole ratios of selected elements.
- You need to choose two elements for the ratio and set parameters such as bin size and maximum x-axis value.
- The resulting histogram shows the frequency distribution of the mole ratios.
- If the chosen elements are not present in the file, a warning message will be displayed.

2. **Isotopic Ratio Analysis:**

   - To perform an isotopic ratio analysis, select the "Perform Isotopic Ratio Analysis?" checkbox.
   - You can choose two elements for the isotopic ratio and set parameters like the element for the x-axis and maximum x-axis value.
   - The resulting plot shows the isotopic ratio distribution with confidence intervals.
   - If selected elements are missing in the data, a warning message will be shown.

3. **Ratio Counts Analysis:**

   - Select the "Perform Counts Ratio Analysis Bars?" checkbox to analyze the counts ratio of selected elements.
   - You can set the elements for the ratio, bin size, and maximum x-axis value.
   - The histogram shows the frequency of the counts ratio.
   - If the elements chosen are not available in the data file, a warning message will alert the user.