NanoFinder 5.5 : Nanoparticle Analysis Software for *sp*-ICP-TOFMS

Kamyar Mehrabi: Kamyarm@ethz.ch

Contents

1. Introduction

NanoFinder is a program designed for the *sp*-ICP-TOFMS user who wants to analyze transient signals (normally nanoparticles, droplets, or cells) for multi-element and multi-transient events simultaneously. For simplicity, we are going to stick to nanoparticle (NP) signals from here; however, the same principles are also true for other transient signals. For the better understanding of the terms and abbreviations, please refer to our publications listed at the end.[1-3]

For development of RAW ICP-TOF-MS data (e.g. the hdf5 file from TofDaqRec) to a list of detected NPs and their quantified mass and particle-number concentrations, the data is processed in 4 distinct steps. The NanoFinder software allows for automated **detection** of NP events, correction of hetero-particle coincident events(**hpCC**), and **Quant**ification of All nanoparticles and hierarchical **clustering** of multi-

metal NP (mmNP). Depending on your study, you could conduct any of the analysis individually or in series by NanoFinder. Below are the 4 data analysis steps.

1. **Detection**:

   a) Select elements and isotopes of interest, generate time traces of selected elements
   b) Determine critical value (L_c) expressions based on compound-Poisson modelling
   c) Determine background (dissolved signal) count rates (λ_{bkgd}) for all elements
   d) Background subtract all time traces
   e) Correct all time traces for split events[2]
   f) Find NP signals above the single-particle critical value ($L_{c,sp}$)

2. **Hetero-particle Coincidence Correction (hpCC):**
   a) Correct data set for particle-coincidence to remove spurious multi-metal (mm) NP signals caused by concurrent measurement of two or more discrete particles with unique element fingerprints

3. **Calibration (not supported by NanoFinder)**
   a) Find microdroplet/Reference nanoparticle and or dissolve standard signals
      Determine mass sensitivities for each element i ($S_{drop,i}$) and sample flow rate to instrument (q_{plasma})

4. **Clustering/Quantification:**
   a) Quantify elemental masses from detectable individual NP signals
   b) Perform hierarchical clustering analysis of mmNP signals; discovery of conserved and non-conserved mmNP types.
   c) Quantify detectable particle-number concentrations (PNCs) of both smNPs and mmNPs
   d) Report smNP and mmNP data

Except calibration step, which requires the user to enter values from their measured standards and determined plasma uptake efficiency, all the rest of the calculation can be done centrally in this program. If calibration data are known (i.e. absolute detection efficiencies of all elements and q_{plasma} for the measured samples), then NanoFinder will provide a complete data analysis from NP identification to quantification of element mass in found NPs and PNCs of found smNP and mmNP classes.

2. Installation

Current version of NanoFinder is developed on Windows operating system. You need to run the **Setup.exe** file form the package to installed "**NanoFinder.exe**" and its host software "Matlab_Runtime". The setup file automatically downloads the free version of "Matlab_Runtime" for its installation. You do not need to have Matlab license for using this program. The only requirement is Microsoft EXCEL.

3. Workflow

We represent a simple example of two data files for ease of explanation. One sample has only gold particles and the other one has gold and silver core-shell NPs. The data files are as Comma separated value (CSV). Generally, Tofwerk icpTOF HDF5 (h5) is also supported as input. A summary of required and generated files are summarized in Table 1 and 2.

Table 1: Files required for each analysis are marked as green in this graph. The orange color indicates cases where either of file would work.

File needed \ Analysis	Detection with SIS	Detection	Summary	Concurrency	Clustering	CONTENTS
MDEF.xlsx	🟩	🟩	🟩	🟩	🟩	
"File name run#".h5	🟧	🟧				icpTOF measurment data file
"File name run#".csv	🟧	🟧				TOF time trace for different element
"File name run#".raw.csv						TOF time trace for used element
"Sample name".xlsx			🟩			TOF time trace for evaluated element
"Sample name".NP time trace.csv				🟩	🟧	TOF time trace for Nanoparticle envents
"Sample name".hpCC.csv					🟧	Concurrency corrected nanoparticle events
"Sample name".Classes.csv						Multi metal nanoparticle clustered events

Table 2: Files generated or modified in each analysis.

File generated or modified\ Analysis	Detection with SIS	Detection	Summary	Concurrency	Clustering	CONTENTS
MDEF.xlsx	🟩	🟩	🟩	🟩	🟩	
"File name run#".h5						icpTOF measurment data file
"File name run#".csv						TOF time trace for different element
"File name run#".raw.csv	🟩	🟩				TOF time trace for used element
"Sample name".xlsx	🟩	🟩		🟩		TOF time trace for evaluated element
"Sample name".NP time trace.csv	🟩	🟩				TOF time trace for Nanoparticle envents
"Sample name".hpCC.csv				🟩		Concurrency corrected nanoparticle events
"Sample name".Classes.csv					🟩	Multi metal nanoparticle cluster events
Dendrogram.pdf/.fig					🟩	Dendrogram of second hierarchical clustering
"Progress report charts" .pdf	🟩	🟩	🟩	🟩	🟩	

To run this program:

a) Place a meta_data_excel_file (MDEF) to the same directory as your raw data files.

b) Fill the MDEF as explained in the following section.

c) Open the NanoFinder.exe program. In the user interface of the program, select the required analysis.

d) When you are ready, press "Run." Then there will be a pop-up file browser window. In this window, you need to select the MDEF file.

e) Depending on the type of the process and number of samples, the program starts to run through your data and show you the progress in the user interface and in the pop-up figures.

f) When the processed lamp next to Run bottom in the user interface turns green, it means that the program run is completed.

g) Result are save in the same directory with different file extensions, such .xlsx, .csv, .pdf, .fig, etc., depending on your analysis.

3.1. Raw data file

For full analysis, you need to have the raw time trace (transient signals and background superimposed) of the elements of interest. This program tested for **low time resolution up to 500 µs**. Hence data acquired with much lower than 500 µs resolution (for example 46 µs or 100 µs) is recommended to go through resolution decrease by resizing the data beforehand. For more information, please see the test sample. In the current program, 100'000 is the minimum number of data Points for each element in each run.

3.2. Meta data Excel file (MDEF)

MDEF is an Excel spreadsheet and it contains multiple sheets of necessary information regarding your samples such as data files names, elements of interest, calibration values and so on. The NanoFinder

program uses the MDEF to complete its tasks and eventually writes the summary of the analysis in there. In Table 3, you will find a summary of the required sheets that need to be filled for each analysis.

Table 3: Meta data Excel file required sheets for every analysis indicated in green.

Required MDEF Sheetname \ Analysis	Detection with SIS	Detection	Summary	Concurrency	Clustering
Filename	🟩	🟩	🟩	🟩	🟩
Elements	🟩	🟩	🟩	🟩	🟩
Read me AIO	🟩	🟩			
Line	🟩	🟩			
SIS Hist	🟩				
Read me Den					🟩
Absolute sensitivity(CountsPng)					🟩
q_plasma(mIPs)					🟩

3.2.1. Filename:

After placing the MDEF in the same directory as your data files, shown in the example below. You have two possibilities on how to fill out the **Filename** sheet of MDEF.

Possibility 1 is to fill it manually and put your desire sample name in first column, called "Sample name" and exact name(s) of your sample file name run(s) in the next column(s), called "File name Run 1, 2, …".

First column (Sample name) is saved for user-preferred names of each data file, since data file names are usually too crowded to be consider as the given name for that sample. Follow the necessary requirement for a file name: It should not contains special character ({, \, |, /, ?, :, [,…). Give every row a unique name. Feel free to keep the whole sample name column empty; in that case, your file name run 1 will be taken as your sample name.
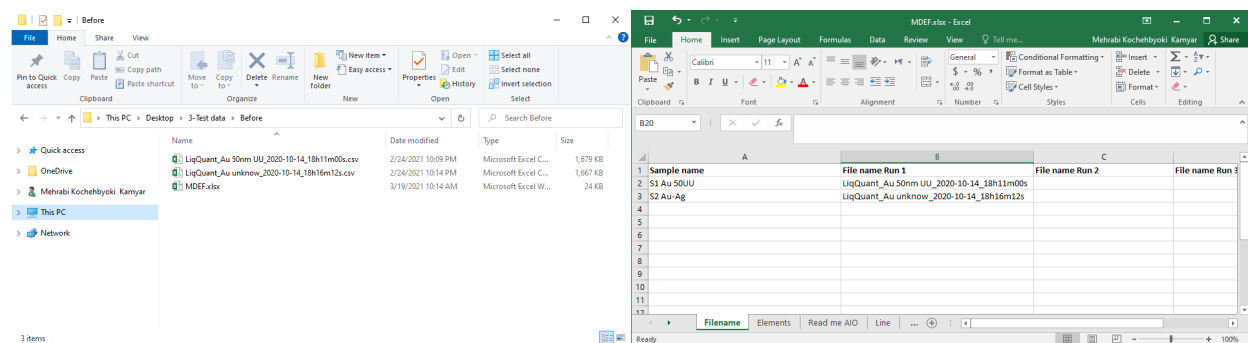

*Figure 1: data directory and example of Filename sheet contents in case of **possibility 1**.*

Possibility 2 is to leave this sheet empty except keeping the headlines. After you executed the NanoFinder program, it will automatically select all of the files in the same directory as selected MDEF and desired format (.csv or .h5) as your data file and fill the Filename sheet. (Notice: in this case, your sample name will be same as your first data file name Run 1.) By specifying the number of replicates in the sheet of

"Read me AIO," you may reshape your file name according to your preference here. (*Important: all of your sample names need to have the same number of runs; otherwise, the program will produce an error.
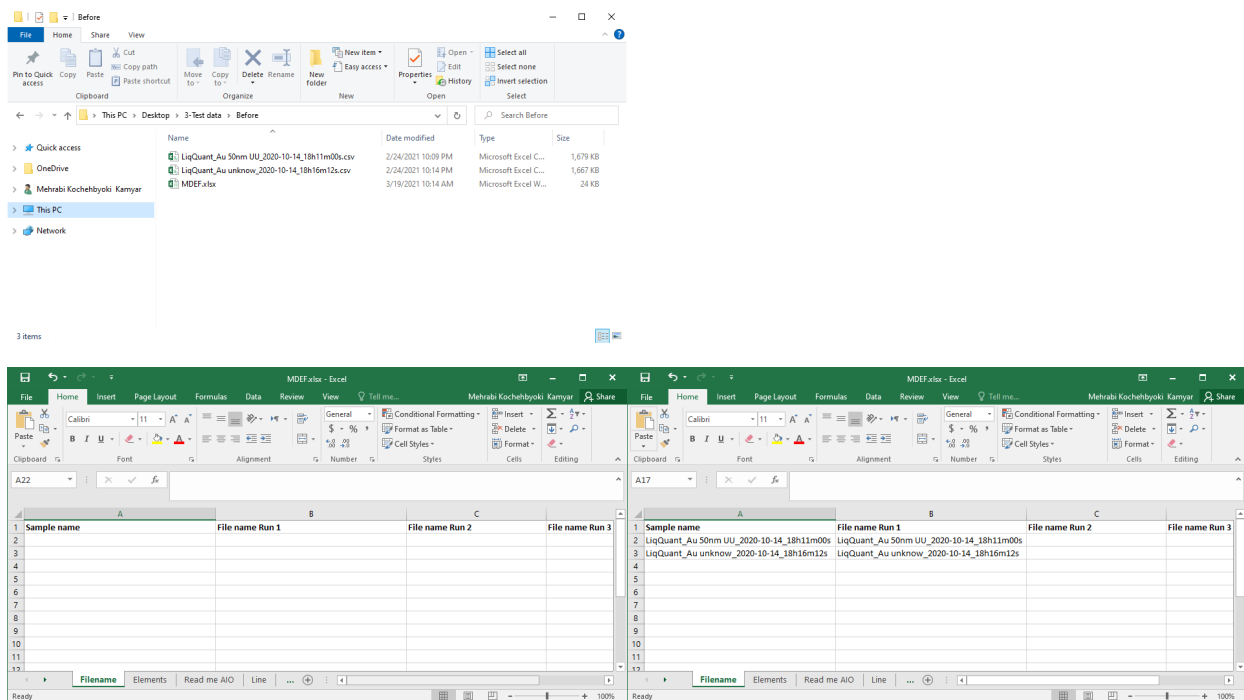

*Figure 2: data directory and example of Filename sheet contents in case of **possibility 2**, before and After execution of program.*

3.2.2. Elements

Here you write the name of the elements that you are interested in. Element names could be things like [56Fe]+, CeO+ or Au, its only need to be the same as your data file. If you wish to combine signal of isotopes or elements of interest, place them in the same row in cells next to each other. There is no limitation of the number of element that you want to analyzed or combined, as shown in Figure 3.
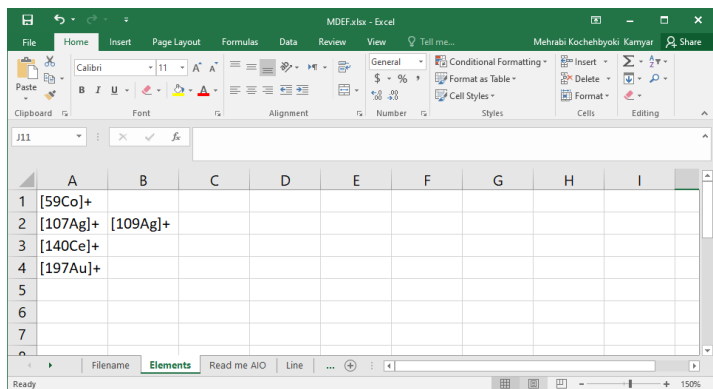

Figure 3: Example of filled Elements sheet of MDEF

3.2.3. Read me AIO:

This sheet is the center of all analysis. It is also containing some of the necessary settings for **Detection** analysis. You only have to fill this sheet if you want to do Detection analysis. For filling the contents of this sheet, you could either directly use the graphical user interface (GUI) of NanoFinder or enter settings directly in MDEF. As it is shown in Figure 4, you must indicate your chose in the GUI.

Conversion factor to counts: Mass spectrometer raw data could be as count/extraction, count/s or etc. To convert raw data to counts, here one need to enter the conversion factor. **If your data file is in HDF5(.h5), put 0 here and the program automatically converts the data into counts from metadata stored in the HDF5 data files(NbrWaveforms).**

Threshold lower boundary: It is the lower threshold of particle detection. Lower than 1 is not recommended.

Nanoparticle to Noise: NanoFinder use dynamic thresholding. It allows the particle thresholding for each element be define with its own level of false detection rate. Basically, this value defines how many times your detected particle events need to be higher than false positive events. For example, 100 here means that if we consider compound Poisson noise, for each 100 detected particles of specific element only 1 false positive is predicted, i.e. the error of particle detection will be 1%.

True to false positive ratio: This is related signal drift compensation. If you do not want to correct for signal drift, then place a "1." Generally, it is better to have it at 50.

Start data point: This indicates where in the dataset the software needs to start the analysis. If you want it to start from beginning of the acquired time trace place a "1" or a "0" in this field. Otherwise specified the row number of the data you interested to start with.

End data point: This indicates where in the dataset the software needs to terminate the analysis (i.e. stop looking for NP events). If you want to process the entire file, type a "0" here. Otherwise specified the row number of the data you interested to end with.

Number of Runs per sample: If you measured your samples in more than one data file, specify the number of run per sample here. All of your samples must have same number of runs.

Data file format (h5 or csv): Raw data format for Detection analysis need to be specified here. There are two possibilities, HDF5(h5) file (default of the icpTOF) or comma separated value (CSV). For further clarification on the correct format of CSV file, please look at the test file.
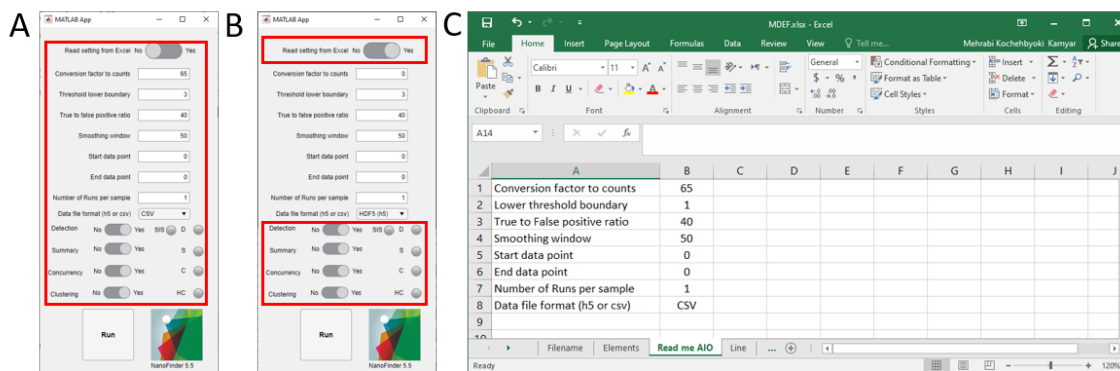
3.2.4. Line

In this page, the line expressions for the critical values are input. We encourage users to read more about fundamental of this expression from Gundlach-Graham, et al paper.[1] If you have the single-ion-signal histogram from your instrument (this should be measured preceding the analysis of any standards or samples) it is possible to calculate the single-ion values with current program. If not, you may use expression from other days. In worst case, use values in the test data here. You may choose a more conservative **True to false positive ratio** (for example 100 or more) if your critical value expression is outdated.

If you have a single-ion-signal histogram and wish to calculate the critical value expression. In the **Line** sheet, delete all columns except the first column (Alpha(rate)) and copy the single-ion-signal histogram to the **SIS Hist** sheet as explained in following section. If you wish to calculate critical value expression for a different set of Alpha rates from the ones presented in Figure 6, simply delete the present values and add new ones.
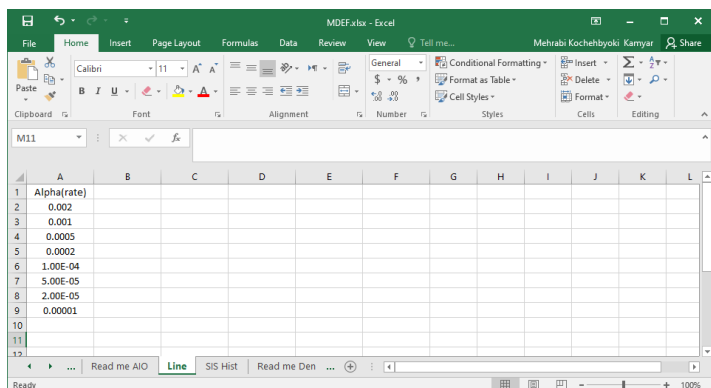

Figure 5: Example of emptied Line sheet for calculation of critical value expression.
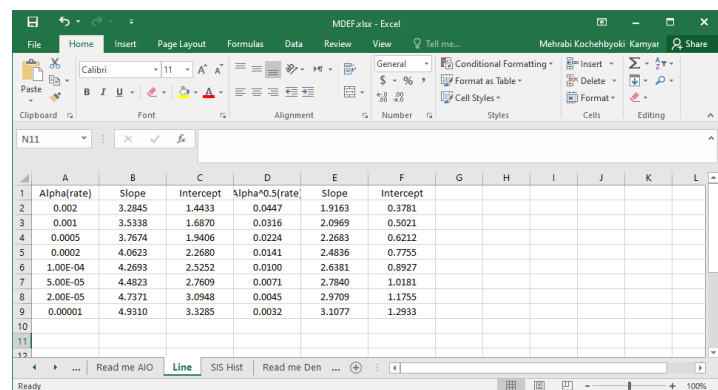

Figure 6: Example of filled Line sheet with critical value expressions.

3.2.5. SIS Hist

This sheet only required if you wish to calculate critical value expressions of your instrument for **Detection with SIS** analysis. If you already have the critical line expressions, this sheet will not be used. Single-Ion-Signal in first column and its repetition in the second column as shown in Figure 7.
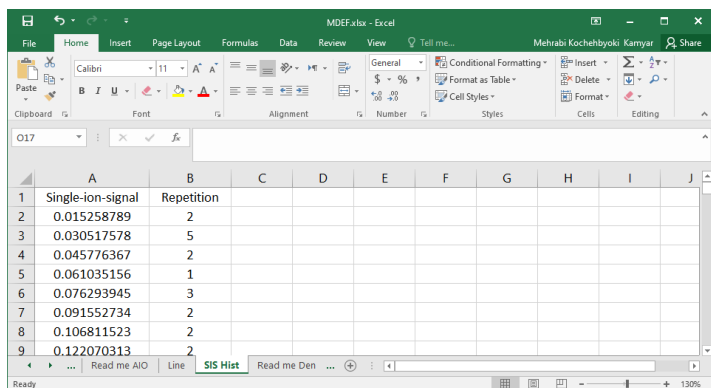
Figure 7: Single-ion-signal histogram placed in the SIS Hist sheet of MDEF file

3.2.6. Read me Dem:

This sheet only required for the **Clustering** analysis. In this sheet, you need to enter information as summarized below.

Cutoff1: A number between 0 and 2, for the first Clustering analysis. This cutoff is used for intra-sample clustering analysis.

*The distance used for clustering is correlation distance, which is defined as one minus the sample correlation between points. A 0 means there is no distance, i.e. high correlation, and a 2 means the highest distance and anti-correlation.

Occurrence in cluster rate: The minimum ratio of the elements that need to happen in each cluster to be taken to account for cluster proxy of second hierarchical clustering.

Max number of clusters: Max number of cluster in each data file if this number is exceeded, an error will arise and Cutoff 1 needs to be increased. This error normally happens if cutoff1 is too low.

Cutoff2: A number between 0 and 2, for the second clustering analysis. This cutoff is use for inter-sample clustering analysis and it is normally smaller than Cutoff1.

File name addition text: Add what your csv file has in addition to first column of filename here. For example, for **coincidence-corrected** data use (**.hpCC.csv**), for a non-coincidence-corrected file from **Detection** use (**.NP time trace.csv**) and for other files, just make sure it is in comma separated format and add (**.csv**) here.

3.2.7. Absolute sensitivity (CountsPng)

This sheet is only required if you wish to quantify and cluster your data (**Clustering** step). As the name suggests, you need the absolute sensitivity of each element and sample that you have in the analysis in the unit of counts per nanogram (counts/ng). The absolute sensitivity is corrected for transport efficiency and it is used for conversion of nanoparticle counts to mass (in ng).

*If you don't have these calibration values, just make a table full of ones (1's) instead. This will allow for the clustering analysis to proceed; however, please remember that you will have arbitrary units and the reported data masses are not correct. This allows you to still see how your mmNPs are clustered; though, the clustering may be slightly different if you use actual mass of elements in NPs rather than just isotope intensities.

Figure 9: Example filled Absolute sensitivity(CountsPng) sheet

3.2.8. q_plasma(mlPs)

This sheet only required if you wish to quantify and cluster your data (**Clustering** step). In this MDEF sheet, you have to fill two columns and one row per each sample in the similar order as in **Filename**. The First column contains the q-plasma (uptake of sample to plasma), which is the flow rate of sample into nebulizer corrected for transport efficiency. For example, if you have 100 **µL/min** nebulizer flow and 10% efficiency then this value is 0.000167 **mL/s. (Be sure to be careful with the units.)**

The second column contains the measurement time divided by dilution factor of the sample for each sample in filename. For example, if your measurement time was 120 sec and the dilution was 100 times, then this number is 120 s / 100 = 1.2. Make sure to also add the time of all of your runs per sample, if you are merging files.

*If you don't have these values just make a table full of ones (1's). This will still work; however, please remember you will have an arbitrary particle number concentration (PNC) and the reported data PNC are not correct.

Figure 10: Example filled q_plasma(mlPs) sheet

3.3. Analysis result

As it was shown in table 2 above, every process generates some result file. in this section we tried to elaborate on the detail result that each analysis provides.

3.3.1. Detection result

Detection step as input need raw data file and specified sheets in MDEF, and as output it provide nanoparticle time trace **("sample name".NP time trace.csv)** signal of multi-element. Beside NP time trace, it also provides you with the raw data with same element list as specified in element sheet of MDEF and in the unit of counts for all runs of each samples (**"File name run#".raw.csv**). Finally, in an Excel file called, **"Sample name".xlsx**, it stores all the useful information (e.g. average particle signal and number of detected particle) about your nanoparticle detection of that sample.

3.3.2. Summary

Summary step will combine all the useful analyzed data from every **"Sample name".xlsx** files and make a summary of them in multiple sheets in **MDEF**.

3.3.3. hpCC result

Hetero-particle coincidence correction (hpCC) uses **"sample name".NP time trace.csv**. After process is done it generate **"sample name".hpCC.csv** which contains all nanoparticle signal as in NP time trace, but this time corrected for hpCC. Moreover, an overview of the hpCC could be find in the **Sample name".xlsx** where there is one sheet with All the unique particle type and their repetition before (**Before hpCC1**) and after (**After hpCC1**) coincidence correction.

3.3.4. Quant. & Clustering

As input, this step use Nanoparticle signal and designated MDEF sheets as explained above. By default, it set to read **".hpCC.csv"** file (setting in **"Read me Dem"** sheet of **MDEF)** as input but you may change it to

".NP time trace.csv" or your own data file as just **".csv"**. The result is including quantification of single metan nanoparticle (smNP), and clustering and quantification of multi metal nanoparticle (mmNP). By quantification we mean both in terms of mass concentration, particle concentration and mass of individual particles. By clustering mean, group mmNP in a logical format that particles with similar fingerprint group in a same class. All individual events in unit of mass, including smNP and different classes of mmNP for each sample will saved in **"Sample name".Classes.csv**. Other quantification such as concentration, representative proxy particle mass and others will save in multiple sheets in MDEF.

4. Conclusion

5. References:

1. A. Gundlach-Graham, L. Hendriks, K. Mehrabi and D. Gunther, Monte Carlo Simulation of Low-Count Signals in Time-of-Flight Mass Spectrometry and Its Application to Single-Particle Detection, *Anal. Chem.*, 2018, **90**, 11847-11855.
2. A. Gundlach-Graham and K. Mehrabi, Monodisperse microdroplets: a tool that advances single-particle ICP-MS measurements, *J. Anal. At. Spectrom.*, 2020, **35**, 1727-1739.
3. K. Mehrabi, D. Gunther and A. Gundlach-Graham, Single-particle ICP-TOFMS with online microdroplet calibration for the simultaneous quantification of diverse nanoparticles in complex matrices, *Environmental Science-Nano*, 2019, **6**, 3349-3358.