Instructions for *IntelligentAcquisition Exclusion* and *IntelligentAcquisition Inclusion* applications

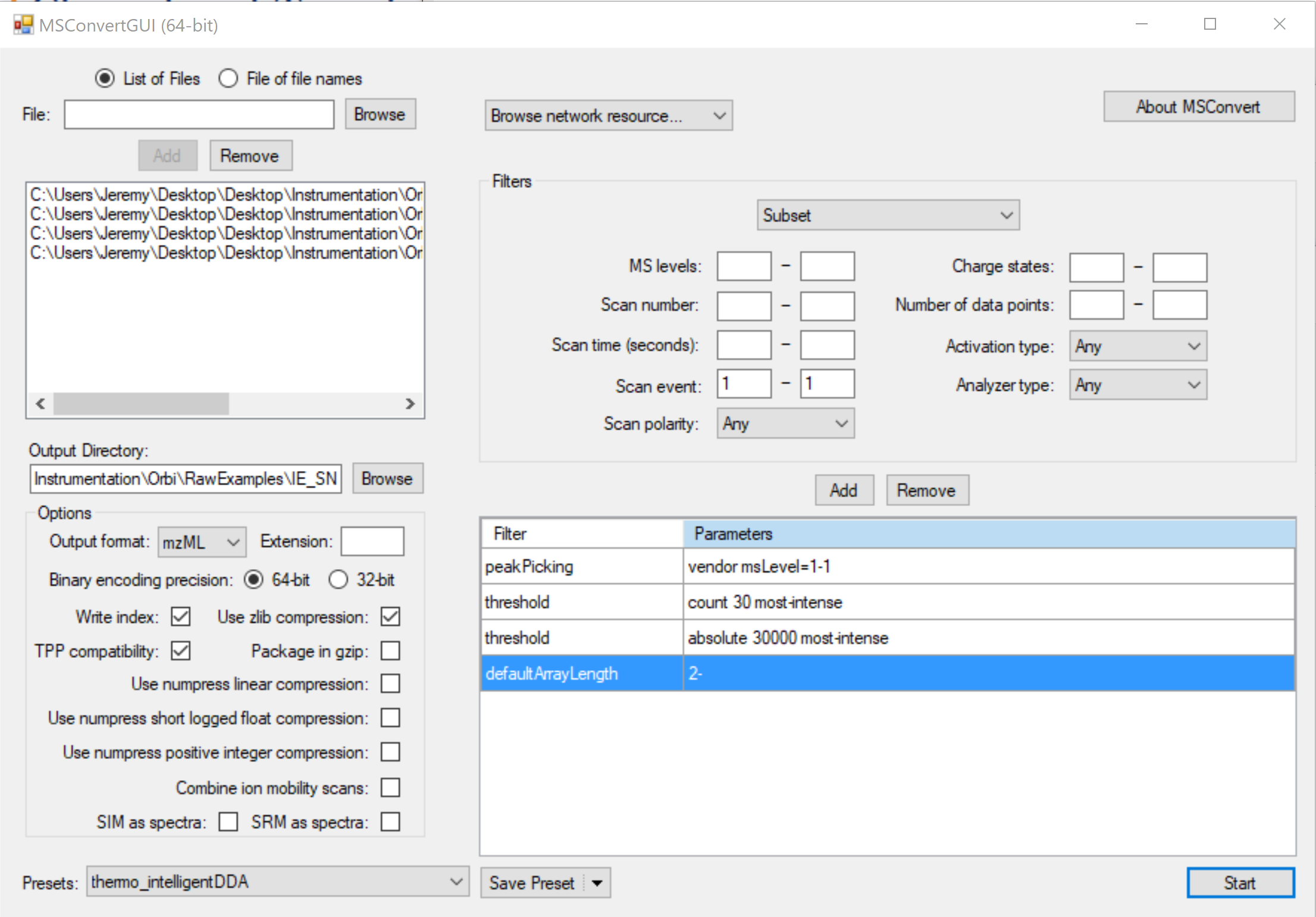
Jeremy P Koelmel 04/22/2019

***IntelligentAcquisition Exclusion***

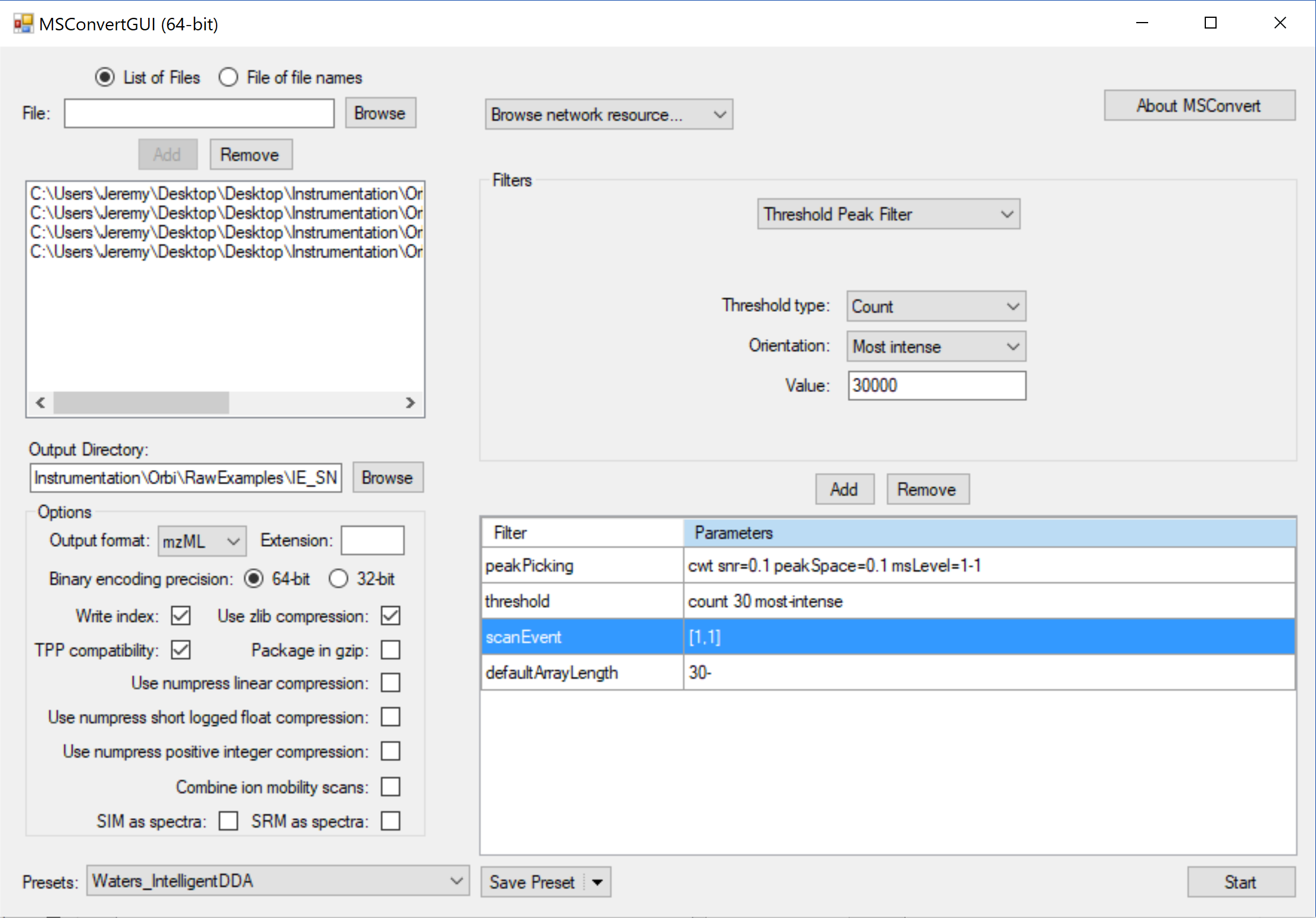
**Part 1. Conversion Parameters (use MSConvert)**

* Export as mzML
* PeakPicking (use CWT for waters, otherwise vendor is fine)
* Threshold: count – most intense – calculate this values based on total number of exclusion list you desire, as well as the topN used. Count = ExclusionList\*topN. This is optional but will drastically reduce file size.
* Absolute intensity: not necessary, but if there is a certain tensity at which even if MS/MS is acquired no useful information will be obtained (eg. noise), than it is a good idea to set a cutoff to reduce the size of the exclusion list.
* Default array 2- (if an ms spectrum contains no peaks or one peak it can cause errors, generally this is not the case, but just encase…)
* In the case of waters scan event 1 should be the only one taken, to filter out the scans for mass accuracy calibration.

Thermo:



Waters:



**Part 2. Running R script**

1) Open the script in R studio or a text editor

2) At the top of the code there is one parameter, save a new version of the code before editing.

The only parameter to manually change is on the first few lines: which instruments are you using? If your instrument is not shown go with thermoQE and then manually change the exclusion list format after you run the code

3) For R studio select all, and click run, or paste the code from the text editor into the R console.

4) MAKE SURE TO CHECK BEHIND THE CONSOLE (pop up boxes will now appear for all inputs)

5) Inputs: read the dialogue boxes for descriptions and fill out inputs.

6) The results will be generated in the same folder as the imported mzML file.

***IntelligentAcquisition Inclusion***

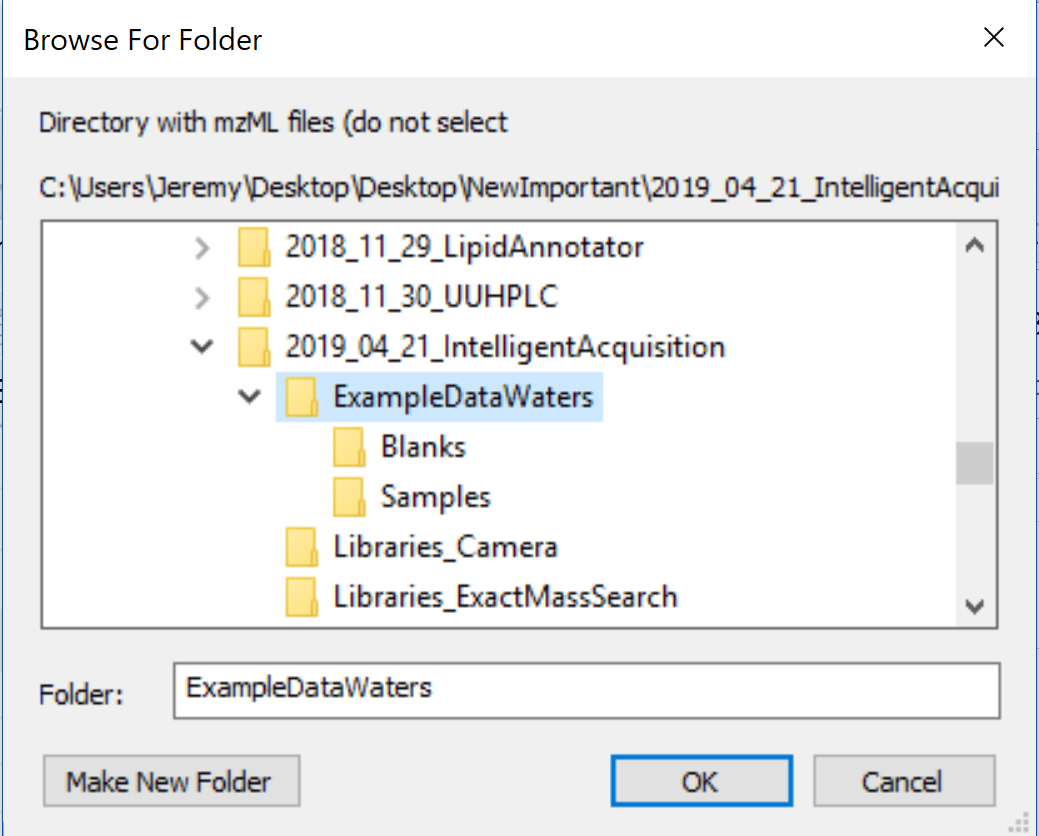
**Part 1. Files to run**

**Name files with "POS" or "NEG" in the filename. For blanks include "blank" somewhere in the filename (as well as "POS" or "NEG")s. This is case sensitive!**

**At least 2 files need to be run**, for example a file and a blank, or two full scan files.

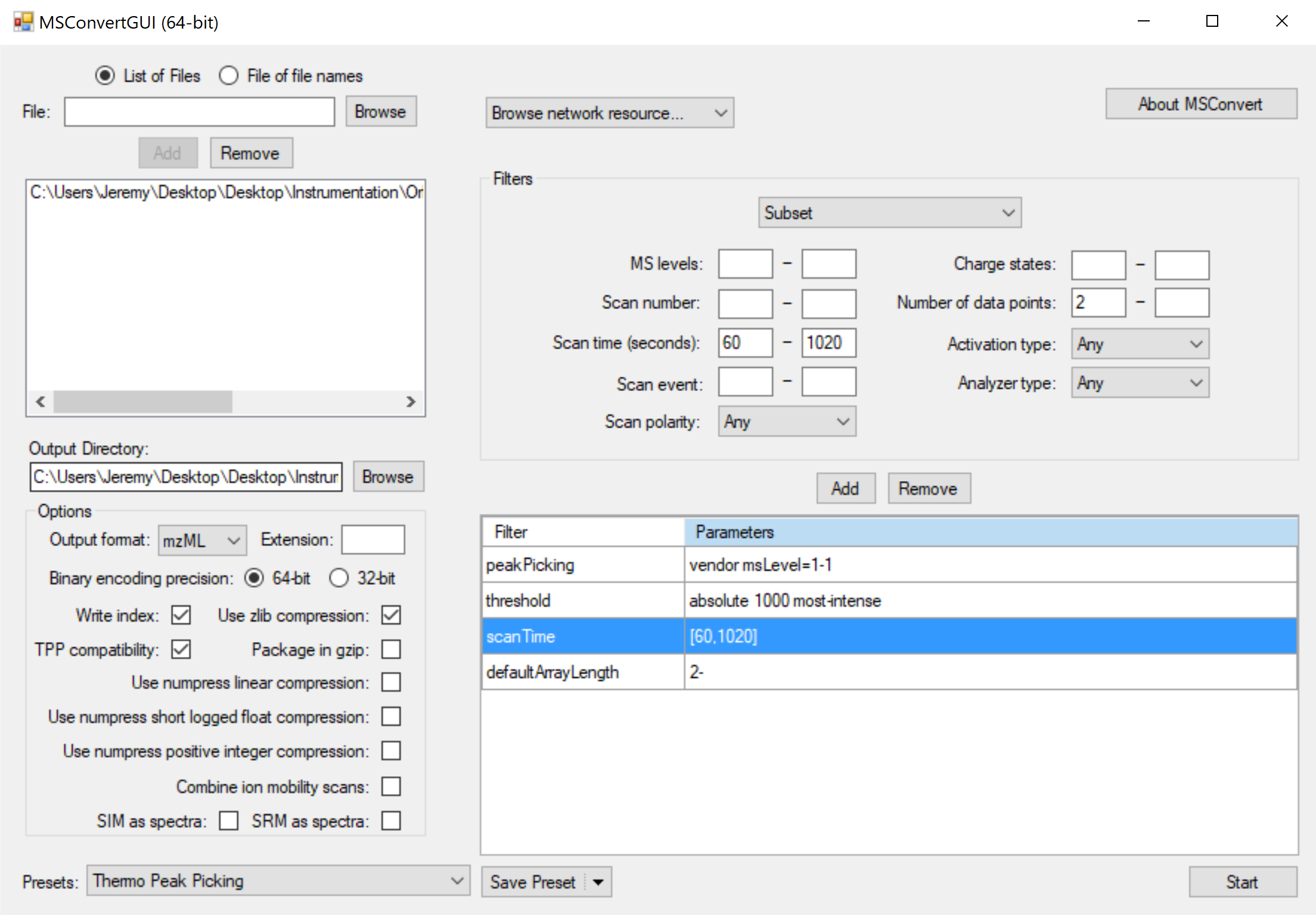
A blank is not necessary, but the blank filtration step is VERY helpful, and hence including at least one blank (ideally three) is recommended. Extraction blanks are better than solvent blanks, as noise introduced during sample preparation, etc. will be removed.

Note for MinFrac (which is the minimum fraction in which a peak is detected across all samples) the samples and blanks should be put in separate folders. Than the parent directory with the 2 folders (blanks and samples folder) can be chosen. See the example data folder and figure below.



**Part 2. Conversion Parameters (use MSConvert)**

Use parameters you would for peak picking. Do not include threshold:count as above, see example below. Change absolute threshold based on the instrument type (what signal can be considered noise?). For Waters include the scan event filter [1-1] and cwt for peak picking. Order does matter.



**Part 2. Running the code**

Use parameters you would for peak picking. Do not include threshold:count as above, see example below. Change absolute threshold based on the instrument type (what signal can be considered noise?). For Waters include the scan event filter [1-1] and cwt for peak picking. Order does matter.

1) Input parameters: At the top of the code there are a number of parameters, save a new version of the code before editing. The most important parameters are: **instrument type** – waters, thermoQE, or thermoLumos…

If none of those are your instruments select thermoQE and you can edit the output file after running the code to fit your format. Also **searching by exact mass (ExactMass\_LibrarySearch)**, if you do not have an exact mass library to reduce your inclusion list by (see format in the example metDNA precursor library) than set to FALSE. Even if set to TRUE both an inclusion with reduction by those only contained in the library and without reduction to those contained in the library will be generated,

Split by abundance OR retention time is another parameter which may be of interest. If the inclusion list is over the designated size it will either be:

a) split so that the first inclusion list has the most abundant ions, the second the second most, etc.

or b) split so that each inclusion list has precursor ions distributed as far apart in retention time as possible.

b) is the default and recommended in the case where all inclusion lists will be implemented.

Some parameters will be edited when your run the code, these include:

Threshold and blank parameters are given. Set the threshold high, so that only ions are included which would give good MS/MS spectra.

Note that there is a polarity parameter, this must be edited to the correct polarity (negative or positive, spelled exactly as such)

2) the outputs will be generated in the same folder as the inputs.