MetAssimulo:Simulation of Realistic NMR Metabolic Profiles

Harriet Muncey[1], Rebecca Jones[1], Maria De Iorio[1]and Timothy M D Ebbels*[2]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK. [2]Biomolecular Medicine, Department of Surgery and Cancer, Imperial College London, UK.

Email: Timothy M D Ebbels*- t.ebbels@imperial.ac.uk;

*Corresponding author

Default Parameters

'Display Output' '1' : *Display simulated spectrum or not.*

'Same concentration for both mixtures' '0' : *Use mixture 1 concentration file for both mixtures.*

'SNR '1000' : *Signal to noise ratio.* '# of replicates(control)' '1' : *Number of replicates to be simulated for mixture 1 (i.e. control group).*

'# of replicates(case)' '1' : *Number of replicates to be simulated for mixture 2 (i.e. case group).*

'Mix1 Conc Filename' 'Input/normal_urine.txt' : *Filename for output of mixture 1 simulated concentrations.*

'Mix2 Conc Filename' 'Input/paraquatpoisoning.txt': *Filename for output of mixture 2 simulated concentrations.*

'x-Axis Start' '0' : *Minimum x-axis value (ppm).*

'x-Axis End' '10' : *Maximum x-axis value (ppm).*

'# of data points' '10000' : *Number of data points to be used (x-values).*

'NSSD Location' 'Local_Database/' : *Location of NSSD.*

'# of bins' '32' : *Number of bins to be used for* σ_n

'Boundary of Standard Peak' '0.2' : *Upper boundary of exclusion region for TSP peak.* 0.(1)

'Excl. Region Lower Bound' '4.5' : *Lower boundary of exclusion region for water and urea peaks.*(1)

'Excl. Region Upper Bound' '6.0' : *Upper boundary of exclusion region for water and urea peaks.*(1)

'Experiments File' 'Input/experiments.txt' : *Filepath for experiment data.*

'Protons File' 'Input/protons.txt' : *Filepath for metabolite proton numbers input (values can be estimated from chemical formula).*

'# of bins (baseline correction)' '32' : *Number of bins to be used in baseline correction.*

'Local Names File' 'Input/local_names.txt' : *Filepath for list of NSSD metabolite names.*

'Synonym Converter' 'Input/synonym_converter.txt' : *Filepath for synonym conversion input data.*

'Bandwidth Indiv' '7' : *Bandwidth, as a number of data points, for kernel smoothing of individual spectra.*

'Bandwidth Mixture' '7' : *Bandwidth, as a number of data points, for kernel smoothing of composite spectrum.*

'Threshold Indiv' '0.8' : *Threshold, as a percentage of the maximum intensity, for kernel smoothing of individual spectra.*

'Threshold Mixture' '0.8' : *Threshold, as a percentage of the maximum intensity, for kernel smoothing of composite spectra.*

'Is Mix2 given as fold changes of mix 1?' '0' : *A value of '1' indicates that the mean and standard deviation of concentrations for the second mixture are ratios of the first. '0' indicates that the second mixture concentrations are not a ratio of the first.*

'Peak Detection' '1' : *A value of '1' indicates peak detection is turned on, whilst '0' is off.*

'Threshold for Peak Shift' '10' : *Threshold for determining whether to shift a peak or not.*

'Broadness' '3' : *Peak Broadness*

'pKa Mean' '4.92' : *Mean pKa value calculated from available pKa data.*

'pKa SD' '0.1' : *Standard deviation for pKa values. 0.1 was chosen in order to ensure the variance was small.*

'Max distance between peaks (ppm)' '0.05'

'Maximum shift (ppm)' '0.2' : *Maximum peak shift (ppm) and also 3xsigma of normal distribution used to calculate [acid limit – base limit].*

'Peak Shift On' '0' : *A value of '0' will generate spectra without including peak shifts. '1' turns on the peak shift function.*

'Update peak shift input files' '0' : *A value of '1' will output the peak shift input data replacing metabolite names with the NSSD synonyms. '0' turns this function off.*

'Same pH for all mix1 reps' '1' : *A value of '1' means that each replicate for mixture 1 will be computed using the same pH value. '0' means that pH values for each replicate will be simulated individually.*

'Same pH for all mix2 reps' '1' : *As above, for mixture 2.*

'Mix1 pH Mean' '7.4' : *Mean pH value for mixture 1.*(2)

'Mix2 pH Mean' '7.4' : *Mean pH value for mixture 2.*(2)

'Mix1 pH SD' '0.5' : *Standard deviation of pH value for mixture 1. 0.5 was chosen in order to ensure the variance was small*

'Mix2 pH SD' '0.5' : *Standard deviation of pH value for mixture 2.*

'Original pKa file' 'Input/peak_shift/pKa_list.txt' : *Filepath for available pKa input data (values from (3)).*

'Original multiplet boundaries' 'Input/peak_shift/just_multiplets.txt' : *Filepath for multiplet boundary data.*

'Original multiplet peak info' 'Input/peak_shift/multiplets.txt' : *Filepath for individual multiplet peak data.*

'Original acid base limit' 'Input/peak_shift/acid_base_list.txt' : *Filepath for acid base limits data.*

'Converted pKa file' 'Input/peak_shift/converted_pKa_list.txt' : *Filepath for pKa data converted to NSSD names for input.*

'Converted multiplet boundaries' 'Input/peak_shift/converted_just_multiplets.txt' : *Filepath for multiplet boundary data converted to NSSD names for input.*

'Converted multiplet peak info' 'Input/peak_shift/converted_multiplets.txt' : *Filepath for individual multiplet peak data converted to NSSD names for input.*

'Converted acid base limits' 'Input/peak_shift/converted_acid_base_list.txt' : *Filepath for acid base limit data converted to NSSD names for input.*

'ppm values' 'ppm_Values_X.txt' : *Filename for x values (chemical shift data points) output.*

'Output Mix 1' 'Spectra_Mixt1.txt' : *Filename for mixture 1 intensities output.*

'Output Mix 2' 'Spectra_Mix2.txt' : *Filename for mixture 2 intensities output.*

'Output Metabolites not included' 'Metabolites_NOT_Included.txt' : *Filename for list of metabolites MetAssimulo was unable to process.*

'Output Simulated Concs Mix1' 'Concentrations_Mix1.txt' : *Filename for simulated concentration output for mixture 1.*

'Output Simulated Concs Mix2' 'Concentrations_Mix2.txt' : *as above, for mixture 2.*

'Output Converted pKa file' 'Input/peak_shift/converted_pKa_list.txt' : *Output filepath for pKa data converted to NSSD names.*

'Output Converted multiplet boundaries' 'Input/peak_shift/converted_just_multiplets.txt' : *Output filepath for multiplet boundary data converted to NSSD names.*

'Output Converted multiplet peak info' 'Input/peak_shift/converted_multiplets.txt' : *Output filepath for individual multiplet peak data converted to NSSD names.*

'Output Converted acid base limits' 'Input/peak_shift/converted_acid_base_list.txt' : *Output filepath for acid base limit data converted to NSSD names.*

'Output pKa list' 'pKa_list.txt' : *Output filename for pKa values used.*

'Output acid base limits' 'acid_base_list.txt' : *Output filename for acid base limits used.*

'Output Mix 1 pH list' 'pH_Mix1.txt' : *Output filename for pH values used in simulating mixture 1.*

'Output Mix 2 pH list' 'pH_Mix2.txt' : *Output filename for pH values used in simulating mixture 2.*

References

1. Ross A, et al.: *NMR Spectroscopy Techniques for Application to Metabonomics. In: Lindon, J(ed) et al. The Handbook of Metabonomics and Metabolomics.* Elsevier, Oxford 2007.

2. Lauridsen M, Hansen SH, Jaroszewski JW, Cornett C: **Human urine as test material in H-1 NMR-based metabonomics: Recommendations for sample preparation and storage**. *ANALYTICAL CHEMISTRY* 2007, **79**(3):1181–1186.

3. Williams R: **pKa data compiled by R. Williams: research.chem.psu.edu/brpgroup/pKa_compilation.pdf.**