**STOCSYE Users Guide**

Sands *et al.* Statistical Total Correlation Spectroscopy Editing of 1H NMR Spectra of Biofluids: Application to Drug Metabolite Profile Identification and Enhanced Information Recovery *Anal. Chem.* 2009, *81,* 6458–6466.

**MATLAB code required:**

Code required:

STOCSYE.m

stocsyCS.m

A detailed description of input (required and optional) and output arguments is provided in the STOCSYE help:

>> help STOCSYE

**Basic run:**

To run STOCSYE with default settings:

>> dataN = STOCSYE(X, ppm, driver);

Where dataN is the set of STOCSY-edited spectra returned from scaling the original data set X based on correlations to peak indices included in driver. In terms of the required input parameters X is the original spectral data, and ppm the corresponding ppm scale, driver is the ppm value of the peak from which to drive the initial STOCSY from. Driver can also be a vector of ppm values in which case STOCSY-editing will be driven from each, and the results combined to produce the final edited spectra.

For example:

>> dataN = STOCSYE(X, ppm, [3.713, 2.601, 4.272]);

**Additional input arguments:**

To run STOCSYE with some user specified settings:

>> dataN = STOCSYE(X, ppm, driver, cutoff, all, noise, extra, mode)

These arguments (and default settings) are described below.

**cutoff** = correlation threshold (default = 0.9). Indices correlating to driver with a correlation coefficient (r) r2 > cutoff will be scaled and background corrected if signal intensity (I) I<LOD (limit of detection). If 0.9 is unsuitable (owing, for example, to reduction in correlation values caused by peak overlap) different correlation cutoff values may be more suitable. Plotting correlation to each peak prior to running STOCSYE may be useful in finding a suitable cutoff:

>> cor = stocsyCS(X, driver);

>> figure;

>> plot(ppm, cor.^2)

If running STOCSYE from multiple peaks, different cutoffs may be entered for each peak (e.g., for three driver peaks, cutoff = [0.9, 0.8, 0.9])

**all** = describes which correlations to use. All = ‘pos’ uses positive correlations to peaks only (i.e., use for scaling structural correlations) while All = ‘all’ includes negative correlations (i.e., will scale any correlation with r2 > cutoff) (default = ‘pos’).

**noise** = start and stop ppm values of spectral region including only noise (for background correction) (default = between 9.5 and 10 ppm).

**extra** = defines the number of indices each side of each region which are used to identify where background should be replaced and generate the replacement. Basically it is important that extra is sufficiently large that the region driver+/- extra includes a section in-between peaks for local background estimation. The default value (0.02 ppm either side of each region to be scaled) should be sufficient for most data sets.

**mode** = defines how the sections to scale and replace are defined. ‘bysam’ (default) specifies that data scaling and background correction continues either side of each identified drug peak region until the signal intensity for each sample reaches a local minimum. However in some cases (for example in regions of peak shifting) this may result in peaks not being scaled for some samples (if their peak apex falls outside the original r2 > cutoff region). In this case ‘bymean’ may produce better results, in this mode, data scaling and background correction for each sample continues either side of each identified drug peak region until the signal intensity of the mean spectrum reaches a local minimum.

**Output:**

>> [dataN, cor, out] = STOCSYE(X, ppm, driver);

**cor** = matrix of r2 values. Each row corresponds to the correlation across the spectrum for each input driver, and the final row is the maximum correlation across all peaks at each point.

**out** = structure containing values of the running parameters (driver, cutoff, all, noise, extra, mode).

**Any queries?**

caroline.sands01@imperial.ac.uk