Running STOCSYE in MATLAB

Code required:
`STOCSYE.m`
`stocsyCS.m`

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

Basic run:

```
dataN=STOCSYE(dataO,ppm,peak);
```

Where 'dataN' is the set of STOCSY-edited spectra returned from scaling the original data set 'dataO' based on correlations to peak indices included in 'peak'.

If no other input arguments are included STOCSY editing is run with the following default parameters
cutoff = 0.9
all = 'pos'
noise = between 9.5 and 10ppm
extra = 0.02ppm either side of each region to be scaled
mode = 'bysam'

Additional input arguments:

```
dataN=STOCSYE(dataO,ppm,peak,cutoff,all,noise,extra,mode)
```

A description of the optional additional input arguments is given in STOCSYE type:
`help STOCSYE`

cutoff = correlation threshold. Indices correlating to peak with a correlation coefficient (r) r^2 > 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(dataO,peak);
figure;plot(ppm,cor.^2)
```

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

all = ‘pos’ uses positive correlations to peaks in ‘peak’ only (i.e. use for scaling structural correlations) while ‘all’ includes negative correlations (i.e. will scale any correlation with r^2 > cutoff).

noise = start and stop ppm values of spectral region including only noise (for background correction)

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 peak+/- extra includes a section in-between peaks for local background estimation. The default value 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 r^2 > 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.

Additional output arguments:

```
[dataN cor out]=STOCSYE(dataO,ppm,peak);
```

cor = matrix of r^2 values. Each row corresponds to the correlation across the spectrum for each input peak, and the final row is the maximum correlation across all peaks at each point.

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