MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886

February 18, 2004

Telephone: 978-692-4764
Fax: 781-981-0590

To: Deuterium Array Group

From: Alan E.E. Rogers

Subject: Method for smoothing spectra and removing CW RFI

1] Maximum likelihood smoothing

The maximum likelihood smoothing of a spectrum with portions to be deleted or downweighted can be obtained from a weighted least squares fit of a Fourier series to the original unsmoothed spectrum. A weighted least squares solution to

$$X = As + n$$

where X = vector of original spectrum
A = steering or design matrix
s = vector of model coefficients
n = error vector

which minimizes chi-squared is given by

$$\hat{s} = \left(A^H wA\right)^{-1} A^H wX$$

and $\hat{X} = A\hat{s}$

where w = weight matrix with zeros off diagonal elements
for which $w_{ii} = 1$ for accepted points
and $w_{ii} = 0$ for deleted points

H denotes the hermitian conjugate and -1 the inverse of the matrix. $\hat{X}$ is the smoothed spectral estimate.

Since the spectrum is real it is convenient to make

$$A_{ij} = \cos(\pi x j) \text{ for even values of } j$$
$$= \sin(\pi x (j-1)) \text{ for odd values of } j$$

where $x = (i - fstart)/(fstop - fstart)$ and use real coefficients. *Fstart* and *fstop* are the start and stop indices in the original 1024 point unsmoothed spectrum. With this modeling scheme

$\left(A^H wA\right)$ is real and an efficient real matrix inversion can be used. The number of model coefficients determines the smoothing resolution. For example

if $\quad fstart = 0$
$\quad fstop = 1023$
$\quad j_{max} = 255$

the original 244 Hz resolution is smoothed to 244 × (1024/128) = 1953 Hz. With these parameters the computations, which involves the inversion of a 256 × 256 matrix takes about 1 second on a current generation PC.

2] RFI removal

If the difference is taken between the original unsmoothed spectrum and the smoothed spectrum

$$d = X - \hat{X}$$

the presence of signals which are unresolved by the smoothed spectrum will show up in the difference or residual spectrum. The random noise in the residual spectrum should have a standard deviation of

$$\sigma_d = \left(\sigma_u^2 - \sigma_s^2\right)^{1/2}$$

where

$$\sigma_u = \left(B_u T\right)^{-1/2}$$

$$\sigma_s = \left(B_s T\right)^{-1/2}$$

B_u and B_s are the resolution bandwidth of the unsmoothed and smoothed spectra respectively and T is the integration time. Any points in the residual spectrum which exceed a threshold of σ_T have a probability of approximately

$$\left(2\pi\right)^{-1/2} e^{-\sigma_T^2/2}$$

of being noise. For a value of $\sigma_T = 4$ the residual is very unlikely to be noise and another pass can be made through the smoothing estimation with the weight input for points above threshold set to zero. In practice it may also be useful to set the weight of points adjacent to the points above threshold to zero. Figure 1 shows an example of a spectrum with CW RFI along with the smoothed version and a smoothed version with the points above a threshold of 4 sigma unweighted.

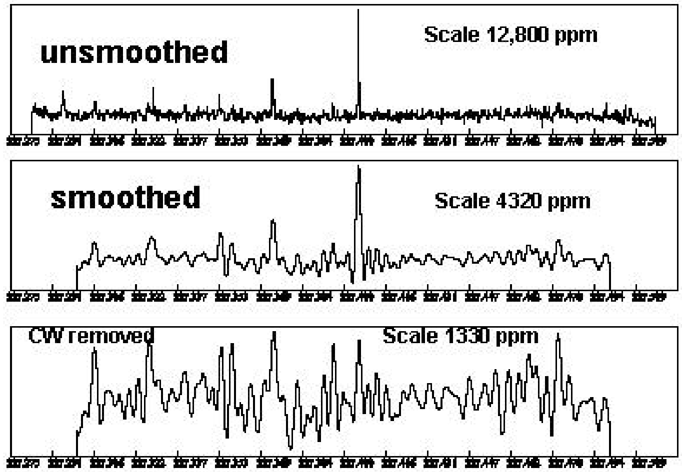
unsmoothed
Scale 12,800 ppm
smoothed
Scale 4320 ppm
CW removed
Scale 1330 ppm

Figure 1.