ADDRESSING BACKGROUND FLUORESCENCE AND UNCERTAINTY ESTIMATION IN RAMAN SPECTRA OF INSOLUBLE ORGANIC CARBON. Gasda, P. J.; Ogliore, R. C.; Taylor, G. J. Hawai'i Institute of Geophysics and Planetology, University of Hawai'i, Mānoa, 1680 East West Road, Honolulu, Hawai'i 96822; gasda@hawaii.edu

Introduction: The current understanding of the chemical structure and the nature of the insoluble organic carbon (IOM) in carbonaceous chondrites comes primarily from studies of its Raman spectra (e.g. [1]). The chemical structure of the IOM has been linked to the metamorphic history of chondrites (e.g. [2]). The majority of these studies use a model (e.g. [3]), to fit their Raman spectra. Fitting a Raman spectrum with a model gives us important spectral parameters, e.g. full width at half maximum, peak position, and peak intensity, which have been correlated with chemical structure information, e.g. maturation grade, disorder, or defect density [4].

It is important to determine if the fitting procedures used are producing good fits and reasonable uncertainty values of the modeled spectrum parameters. Reasonable uncertainties are important for understanding if the IOM is chemically homogenous for a given meteorite. If uncertainty values are overestimated, then information about the IOM is lost. If uncertainties are underestimated we can be tricked into thinking that the sample is heterogeneous. Therefore, it is important to determine the correct uncertainty in every measurement we perform.

The Raman spectrum fitting function used in our work [5] is based on the fitting procedures in use extensively in the literature to describe graphite-like materials [3, 6]. Essentially, the D-band is fit using a Lorentz line shape (Eqn 1), and the G-band is fitted using an asymmetric Lorentz line shape called the Breit-Wigner-Fano (Eqn 2) function.

$$F_L = I_1 \frac{\delta/2\pi}{(x-\omega_D)^2+(\delta/2)^2} \quad (1)$$

$$F_{BWF} = I_2 \frac{[1+2(x-\omega_G)/Q\gamma]^2}{1+(2(x-\omega_G)/\gamma)^2} \quad (2)$$

The variables of F_L and F_{BWF} correspond to the spectral parameters of the D- and G-bands. The variables I_1 and I_2 are the intensity, ω_D and ω_G are the peak positions, and δ and γ are the peak widths of the D- & G-bands, respectively. Q is asymmetry factor of the G-band.

Generally, spectra seen in studies of IOM of the aqueously altered meteorites have interference from high fluorescence background. The source is an unknown organic chemical compound or mix of chemical compounds that varies from spectrum to spectrum. This suggests that the background is a superposition of multiple peaks. A full understanding of the fluorescence phenomena in meteorite samples is likely impossible because the nature of the source is unknown and variable. Typically, the background for a small spectral window containing the relevant peaks is subtracted using a line. We show that this not an accurate representation of a general fluorescence background. A more robust technique that allows for the subtraction of any continuous function is required to accurately model the Raman parameters of the sample.

Second Derivative Fit to Remove Background Fluorescence: Rather than applying a linear background subtraction, we propose Savitzky-Golay Second Derivative (SGSD) fitting. SGSD removes the effect of any smoothly varying background fluorescence component from the Raman spectrum fits. It is necessary to smooth the noisy Raman spectrum before calculating the second derivative. We found that Savitzky-Golay smoothing [7] with a 2nd order smoothing function and a window size of ~27 for our data [8] retains the important Raman spectral structure while greatly reducing the noise.

To assess the accuracy of this method compared to the traditional linear-background-subtraction method, we generated a simulated Raman spectrum around the D and G band with a smoothly varying background of different functional forms (e.g. linear, logarithmic, exponential) and added ~1% noise. Selected variables from the results of our simulations are given in Table 1. Figure 1 shows an example of simulated data with a

Table 1

Linear Background					
Variable	**Actual**	**SGSD**	**Std. Dev.**	**Linear Subtract**	**Std. Dev.**
δ	240	259.69	16.8	240.63	8.19
γ	230	252.37	7.12	230.23	5.38
$I_1/(I_2\ \gamma)$	1.087	1.069	0.151	1.093	0.104
Linear-Logarithmic Background					
δ	240	261.46	13.6	397.27	312.01
γ	230	253.71	6.46	323.80	89.11
$I_1/(I_2\ \gamma)$	1.087	1.079	0.136	1.347	1.136
Inverse Hyperbolic Cosine Background					
δ	240	263.31	7.01	275.53	4.13
γ	230	254.18	2.66	233.43	2.60
$I_1/(I_2\ \gamma)$	1.087	1.078	0.063	1.360	0.064

linear-logarithmic fluorescence background. The simulated data is fitted using the linear-background subtraction model (red). Figure 2 is the 2nd-derivative of the same data with the SGSD fit (blue). To first order, it appears that the fit in Figure 1 is adequate, but we see in Table 1 that it is not accurate and tends to have poor precision. The fit of the second derivative peaks in Figure 2 also looks adequate, and Table 1 confirms that it is accurate. Table 1 shows that the linear-subtraction model works as well or better than the SGSD method when the background is linear. However, when the background type deviates from linear, SGSD tends to retain the actual simulated values much better with good precision.

Monte Carlo Uncertainty Estimation: We seek uncertainties of the spectral parameters (e.g. peak position, peak width) given in Eqns 1 and 2. Traditionally, the fit parameters and their uncertainties have been calculated using a curve-fitting computer algorithm, such as the IDL procedure CURVEFIT. CURVEFIT uses a chi-squared (χ^2) minimization technique that requires the user to provide an array of weights (typically equal to the squared inverse of the 1σ measurement uncertainty) and a fitting model. A common mistake is to perform an "unweighted fit" by setting all these weights equal to one. This is equivalent to assigning an uncertainty of one to each measured data point (most Raman spectra are recorded in units of CCD counts). However, the actual measurement uncertainty of each point is much larger than one count because of signal noise. If CURVEFIT calculates the best-fit model parameters based on the unweighted fit, the resulting χ^2 value is far too large. The uncertainty values on the parameters returned by CURVEFIT are computed when χ^2 increases by one. Since χ^2 is far too large, these uncertainties are greatly underestimated.

To obtain a better estimate of the model parameter uncertainties, we propose the following technique based on a Monte Carlo bootstrap method [9]. We generate an array of empirical uncertainties (σ_i) by computing the distance between each measured data point and the Savitzky-Golay smoothed spectrum. From this array, we randomly draw values (allowing repeats) that we add to the original smoothed spectrum. Using this new simulated data spectrum, we recalculate our 2nd derivative fit parameters as described above. Repeating this process 10^3 times yields 10^3 values for each fit parameter. The 2σ uncertainty of our fit parameters is simply twice the sample standard deviation of these 10^3 values for each parameter.

Conclusions: We proposed that the SGSD method can be applied to Raman spectra of the insoluble organic matter in carbonaceous chondrites. SGSD produces fits with greater accuracy than the chi-squared minimization techniques when fitting non-linear backgrounds. Precision on the modeled spectral parameters are estimated using the Monte Carlo bootstrap method.

References: [1] Busemann, H. et al. (2007) MAPS, *42*, 1387-416. [2] Bonal, L. et al. (2007) *GCA*, *71*, 1605-23. [3] Ferrari A. C. and Robertson J. (2000) *Phys. Rev. B, 61*, 14095-107. [4] Lucchese M. M. et al. (2010) *Carbon, 48*, 1592-7. [5] Gasda, P. J. and Taylor, G. J. (2013) *LPSCXLIII*, Abstract#1029 [6] Bonal, L et al. (2006) *GCA*, *70*, 1849-63. [7] Zhang, D. and Dor, B.-A. (2000) *Appl. Spec.*, *54*, 1379-1383. [8] Holler, D. et al. (1989) *Appl. Spec.*, *43*, 877-882. [9] Chernick, M. R. 1999, Bootstrap Methods: A Practitioner's Guide, Wiley Series in Probability and Statistics, (New York: Wiley)

Figure 1

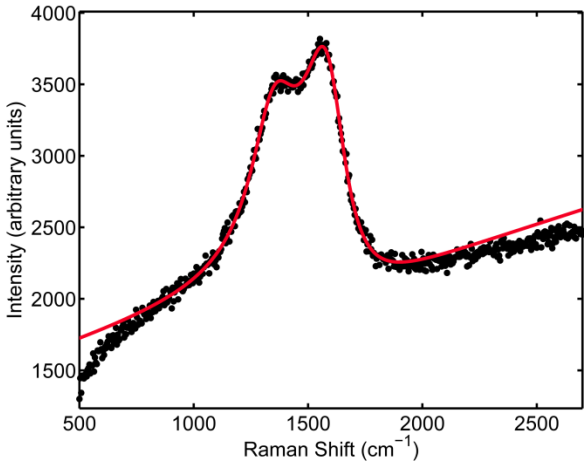

Figure 2