

# The Information Content of Exo-Planetary Spectra

M.R. Line<sup>1</sup>, X. Zhang<sup>1</sup>, G. Vasisht<sup>2</sup>, V. Natraj<sup>2</sup>, P. Chen<sup>2</sup>, H. Knutson<sup>1</sup>, Y.L. Yung<sup>1</sup>

<sup>1</sup>California Institute of Technology, <sup>2</sup>Jet Propulsion Laboratory-California Institute of Technology



## Introduction

Exoplanet spectra are key to understanding their atmospheric compositions. There are currently several retrieval techniques used to determine abundances from these spectra. The most prevalent and simplest is the manual tuning of model abundances and temperatures until a reasonable fit is obtained<sup>1,2</sup>. This technique likely does not provide an optimal solution to the atmospheric state given it is nearly impossible to manually explore all of parameter space by hand. A more systematic approach is the multi-dimensional grid search method or monte-carlo approach<sup>3,4</sup>. These approaches generate millions of model spectra of which the best are determined by a minimum chi-squared criterion. The millions of forward model runs required by the latter techniques can be time consuming and computationally expensive. Furthermore, neither of these techniques provide a robust way to explore the change in information content associated with changes in the quality of the data, such as the signal to noise, additional spectral channels, and resolving powers. In this poster we present the inverse approach to the spectral retrieval problem widely used in the Earth remote sensing community<sup>5</sup>, planetary atmospheres<sup>6,7,8</sup>, and recently exoplanets<sup>9,10</sup>.

## Approach

Using synthetic spectra we determine how well various atmospheric properties can be retrieved by exploring how the information content, degrees of freedom, and retrieval uncertainty improve with different sets of broadband observations such as those from Spitzer and ground based instruments.

## Optimal Estimation

From Bayes Theorem and Gaussian statistics get cost function:

$$J(\mathbf{x}) = (\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_e^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a)$$

Cost function minimized using Levenberg-Marquardt:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + [(1 + \gamma)\mathbf{S}_a^{-1} + \mathbf{K}_k^T \mathbf{S}_e^{-1} \mathbf{K}_k]^{-1} \{\mathbf{K}_k^T \mathbf{S}_e^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}_k)] - \mathbf{S}_a^{-1} [\mathbf{x}_k - \mathbf{x}_a]\}$$

The retrieved atmospheric state has an uncertainty:

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1}$$

A useful metric is the averaging kernel (or retrievability). A parameter who's diagonal element is 1 is constrained entirely by the data, while a value of 0 means the data provided no information in constraining that parameter:

$$\mathbf{A} = \frac{d\hat{\mathbf{x}}}{d\mathbf{x}} = \hat{\mathbf{S}} \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K}$$

The information content and degrees of freedom are, respectively:

$$H = \frac{1}{2} \ln(|\hat{\mathbf{S}}^{-1} \mathbf{S}_a|) \quad d = \text{trace}(\mathbf{A})$$

$H$  describes the reduction in the overall uncertainty in the parameters as a result of the measurements, while  $d$  gives the total number of retrievable parameters allowed by the data in the context of the given forward model.

$\mathbf{x}$  = atmospheric state vector

$\mathbf{y}$  = observation vector

$\mathbf{x}_a$  = *a priori* atmospheric state vector

$\hat{\mathbf{x}}$  = retrieved state-optimal estimate

$\mathbf{F}(\mathbf{x})$  = forward model

$\mathbf{S}_e$  = data uncertainty matrix

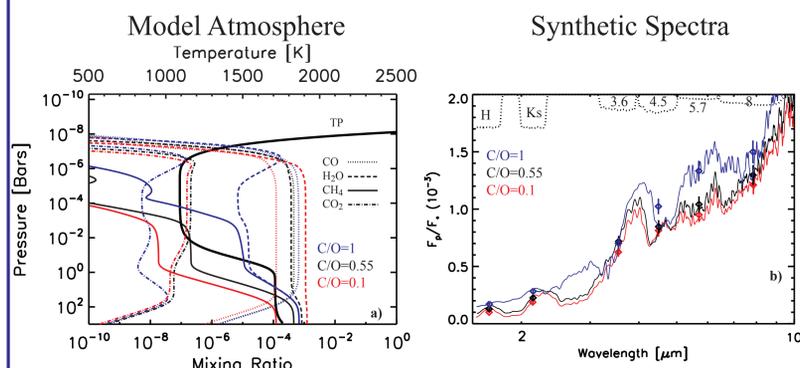
$\mathbf{S}_a$  = *a priori* state uncertainty matrix

$\mathbf{K} = \frac{d\mathbf{F}}{d\mathbf{x}} = \text{Jacobian}$

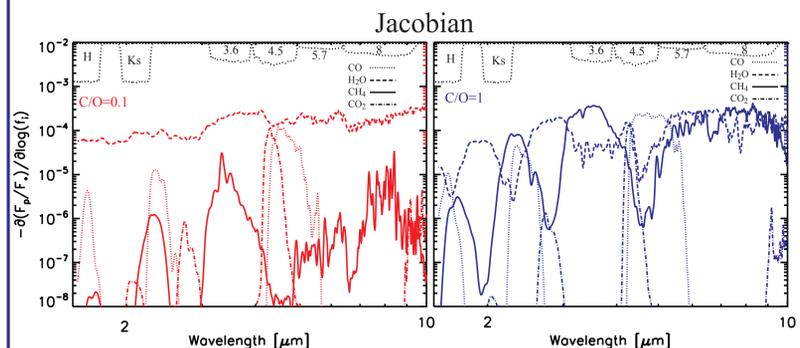
## Forward Model

- Atmospheric state vector characterized by 8 parameters
- Analytical T-P profile<sup>11,12</sup> governed by 3 opacities,  $\kappa_{v1}$ ,  $\kappa_{v2}$ ,  $\kappa_{IR}$ , and a partitioning between the two visible streams,  $\alpha$  (4 total parameters)
- Constant-with-altitude mixing ratios for H<sub>2</sub>O, CH<sub>4</sub>, CO, and CO<sub>2</sub> (4 total parameters)
- HiTemp<sup>13</sup> H<sub>2</sub>O, CO, and CO<sub>2</sub> line lists, STDS<sup>14</sup> CH<sub>4</sub> line list
- H<sub>2</sub>-H<sub>2</sub>/H<sub>2</sub>-He opacities<sup>15</sup>
- RFM (Reference Forward Model) for LBL radiative transfer<sup>16</sup>

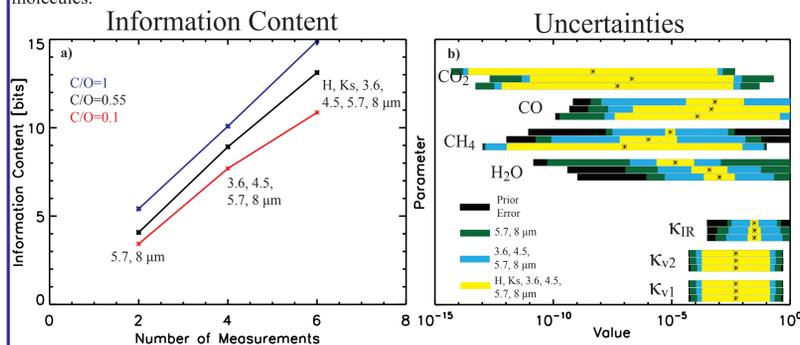
## Synthetic Spectra Experiment



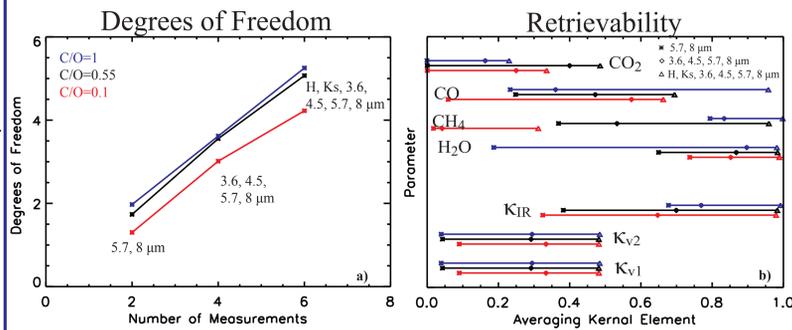
**Figure 1-** (a) Model atmospheres for three different C/O ratios and solar metallicity generated using a chemical kinetics model<sup>17</sup> for a Jupiter-like planet around a sun-like star at 0.05 AU. The solid curve labeled "TP" is the temperature profile. The rapid increase in temperature near the top is an artificially added thermosphere. The red curves are the mixing ratios for C/O=0.1, black C/O=0.55 (solar) and, blue C/O=1. (b) Synthetic spectra generated using the mixing ratios and temperature profile in (a). The diamonds with the error bars (S/N~15) are the fictitious observations of the higher resolution spectra (solid curves) sampled over the instrumental filter functions (dotted curves at top) for the CFHT-WIRcam H and Ks bands and the Spitzer IRAC 3.6, 4.5, 5.7, and 8  $\mu$ m bands.



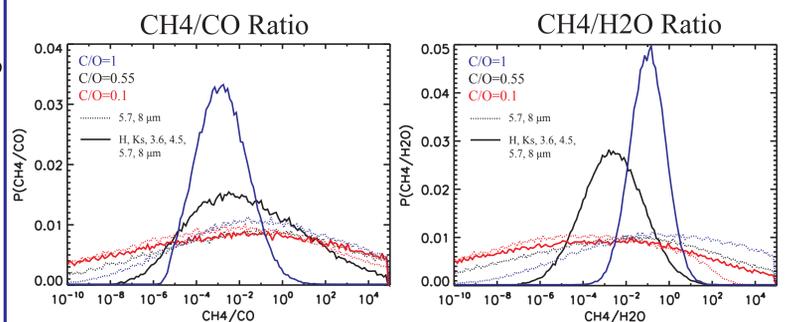
**Figure 2-** The Jacobian describes the sensitivity of flux in each wavelength to each of the parameters we wish to retrieve. Note how all the channels over this wavelength region are sensitive to water in the low C/O case. Because of the overwhelming sensitivity of the flux to H<sub>2</sub>O in all channels, it is difficult to retrieve the abundances of the carbon species with any useful precision. The ground-based H and Ks bands are generally only sensitive to water with little interference from other gases. As the C/O ratio increases, the observations become more sensitive to the carbon bearing species due to their increased abundances. The IRAC 4.5  $\mu$ m bandpass becomes sensitive to both CO and CO<sub>2</sub>, while the IRAC 3.6  $\mu$ m bandpass becomes sensitive to CH<sub>4</sub>. The 5.7  $\mu$ m channel also has some sensitivity to CO. The H and Ks bands become polluted with CH<sub>4</sub> and CO making it more difficult to retrieve water. An ideal measurement for aiding in the further constraint of water would fall between the H and Ks or between the Ks and 3.6  $\mu$ m bands as these locations are generally devoid of absorption from most other molecules.



**Figure 3-** Information content and retrieval uncertainties as a result of the different broadband observations. The retrieval uncertainties and information content are initially determined for the synthetic spectra sampled with only 2 broadband observations, the IRAC 5.7 and 8  $\mu$ m bands. The uncertainties and information content are then recalculated with 4 broadband observations, the IRAC 3.6, 4.5, 5.7, and 8  $\mu$ m bands. Finally the process is repeated with all 6 broadband observations, the CFHT-WIRcam H and Ks bands and the Spitzer IRAC 3.6, 4.5, 5.7, and 8  $\mu$ m bands. (a) Information content resulting from the different sets of broadband measurements for the three different C/O ratios (red C/O=0.1, black C/O=0.55, and blue C/O=1). This plot tells us that the addition of the extra data points improves the overall uncertainty relative to our prior uncertainty in the model parameters. (b) Retrieval uncertainties in each parameter for each set of 2 (green bars), 4 (light blue bars), and 6 (yellow bars) point observations and the three different C/O ratios. The black "error" bar is our assumed prior error. For each parameter there are 3 sets of error bars. The top most in each set is for C/O=1, the middle for C/O=0.55, and bottom for C/O=0.1. Note how the uncertainties for CH<sub>4</sub> and CO improve with increasing abundance. The x-axis has units of volume mixing ratio for the gases and units of (cm<sup>2</sup>/g) for the opacities.



**Figure 4-** Total degrees of freedom and the retrievability for each parameter. As in Figure 2, these quantities are calculated for the three different C/O ratios and for each set of the three sets of observations (2 points, 4 points, and 6 points). (a) Total degrees of freedom as a function of data set. The total degrees of freedom represent the total number of independent quantities that can be retrieved. Like the information content, increasing the number of data points increases the total available degrees of freedom. (b) The retrievability (averaging kernel) of each parameter. A value of 1 means that the parameter can be constrained solely by the data. A value of 0 means that the data provides no useful constraint, and the parameter is solely determined by the prior assumptions. The \* is the retrievability for each parameter using only 2 data points, the diamond with 4 data points, and the triangle with 6 data points. The lines connecting the points are just for aesthetic purposes. Blue lines are C/O = 1, black C/O=0.55, and red C/O=0.1.



**Figure 5-** Probability distribution functions of ratios of carbon species to some oxygen species via Monte Carlo runs based on uncertainties in the mixing ratios from figure 3. These ratios can be used to diagnose different chemistries. For instance, the CH<sub>4</sub>/CO ratio can be used as an indicator of disequilibrium chemistry and the CH<sub>4</sub>/H<sub>2</sub>O ratio can be used as a proxy for the C/O ratio. The ratio distributions for the 2 band (dotted curves) observations are compared with those of the 6 band (solid curves) observations for the three different C/O ratios (red C/O=0.1, black C/O=0.55, blue C/O=1.0). The 2 band observations offer little hope in distinguishing the two different C/O ratios using the CH<sub>4</sub>/CO and CH<sub>4</sub>/H<sub>2</sub>O ratios as proxies. The 6 band observations are enough to break some of the degeneracies between overlapping spectral features and allow some constraint on the CH<sub>4</sub>/H<sub>2</sub>O ratio, though the low C/O ratio case is plagued with uncertainties in the carbon species.

## Conclusions

The optimal estimation technique offers a powerful formalism to study the information content and retrieval uncertainties. Broadband photometry points can offer good constraints on the water abundance as it is present everywhere in the spectra. The carbon species, especially CO<sub>2</sub>, are poorly constrained because they are not as prevalent in the spectra, unless their sensitivities surpass that of H<sub>2</sub>O's in the observed channels. Typically, the greater the abundance of a gas, the greater its retrievability and the lower its uncertainty. The large carbon species uncertainties propagate into the CH<sub>4</sub>/CO and CH<sub>4</sub>/H<sub>2</sub>O ratios making it difficult to distinguish the different C/O ratios with just 2 broadband points, but the constraint improves when using all 6. Future work will identify bandpasses that will offer the greatest increase in the information content, retrievability, reduction in the uncertainties.

## References

- Swain, M.R., Vasisht, G., Tinetti, G., et al. 2009a, ApJL, 690, L114 2. Tinetti, G., Deroo, P., Swain, M.R., et al. 2010, ApJL, 712, L139 3. Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24 4. Madhusudhan, N., Harrington, J., Stevenson, K.B., et al. 2011, Nat., 469, 64 5. Rodgers, C.D., Inverse methods for atmospheric sounding, Theory and Practice, 2000 6. Nixon, C.A., Achterberg, R.K., Conrath, B.J., et al. 2007, Icarus, 188, 47 7. Fletcher, L.N., Drossart, P., Burgdorf, M., Orton, G.S., & Encrenaz, T. 2010, A&A, 514, A17 8. Irwin, P.G.J., Teanby, N.A., de Kok, R., et al. 2008, JQSRT, 109, 1136 9. Lee, J.-M., Fletcher, L.N., & Irwin, P.G.J. 2011, MNRAS, 1983 10. Line, M.R., Zhang, X., Vasisht, G., et al. 2011, arXiv:1111.2612 11. Guillot, T. 2010, A&A, 520, A27 12. Parmentier, V., & Guillot, T. 2011, EPSC-DPS Joint Meeting 2011, held 2-7 October 2011 in Nantes, France 13. Rothman, L.S., Gordon, I.E., Barbe, A., et al. 2009, JQRT, 110, 533 14. Borysov, A., Champion, J.P., Jorgensen, U.G., & Wenger, C. 2002, Molecular Physics, 100, 3585 15. Borysov, A. 2002, A&A, 390, 779 16. <http://www.atm.ox.ac.uk/RFM/> 17. Line, M.R., Vasisht, G., Chen, P., Angerhausen, D., & Yung, Y.L. 2011a, ApJ, 738, 32

## Acknowledgements

We thank Aaron Wolf, Alejandro Soto, and Cheng Li for useful statistics discussions and Leigh Fletcher for providing us with the high temperature methane line list. M. Line is supported by the JPL Graduate Fellowship funded by the JPL Research and Technology Development Program. XZ and YLY are supported by a grant from the PATM program of NASA to the California Institute of Technology. P. Chen & G. Vasisht are supported by the JPL Research & Technology Development Program, and contributions herein were supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration