Radiative transfer speed-up combining optimal spectral sampling with a machine learning approach

Steffen Mauceri[1*], Chris O'Dell[2], Vijay Natraj[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA
*Steffen.Mauceri@jpl.nasa.gov

Abstract: The Orbiting Carbon Observatories-2 and 3 make space-based measurements in the oxygen A-band and the weak and strong carbon dioxide (CO2) bands. A Bayesian optimal estimation approach is employed to retrieve the column averaged CO2 dry air mole fraction from these measurements. This retrieval requires a large number of polarized, multiple-scattering radiative transfer (RT) calculations for each iteration. These RT calculations take up the majority of the processing time for each retrieval, and slow down the algorithm to the point that reprocessing data from the mission over multiple years becomes very expensive. To accelerate the RT calculations and thereby ease this bottleneck, we have developed a novel approach that enables reproduction of the spectra for the three OCO-2/3 instrument bands from radiances calculated at a small subset of monochromatic wavelengths. This allows reduction of the number of monochromatic RT calculations by a factor of 10 - 20 and can be achieved with radiance errors of less than 0.1% with respect to the existing algorithm. The technique is applicable to similar retrieval algorithms for other greenhouse gas sensors with large data volumes, such as GeoCarb, GOSAT-3, and CO2M.

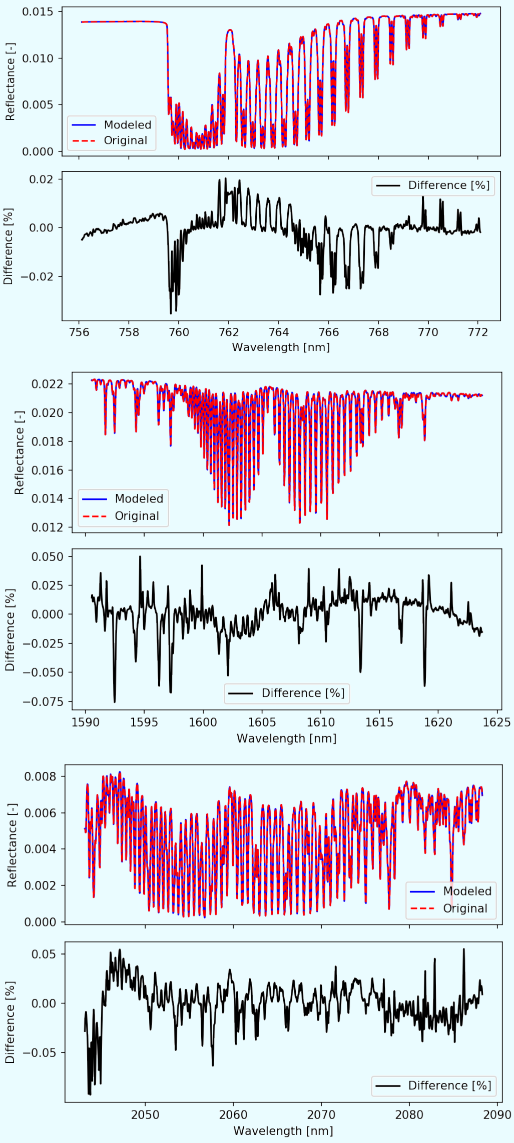

Figure 1: Reconstructed and convolved spectrum and error for one representative spectrum for O2A-band (top), WCO2-band (middle), and SCO2-band (bottom).

Methodology:

For the forward radiative transfer calculations we model the full high-resolution spectrum (27,000 wavelength for O2A-band, 11,000 wavelengths for weak and strong CO2-band) from the calculated radiances at a subset of 800 monochromatic wavelengths (number of wavelengths can be adjusted for a different speed-accuracy trade-off). This is possible due to the high correlation of individual wavelenths, evident in the few number of principal components that can discribe the OCO-2/3 bands (see Figure 3).

To extrapolate from the subset of calculated radiances, x, to the full spectrum, Y, we first project the radiances into log space to account for Beer's law. Next, we calculate the radiance at each wavelength of the full spectrum from a linear combination of log-transformed radiances at a subset of wavelengths, using matrix, A. Finally, we take the exponent of this product to get back to the original radiances.

$$Y = \exp(\log(x) * A)$$

The matrix, A, is found by minimizing the root mean square error between the reconstructed spectrum, Y, and the true spectrum, y, for a set of training spectra.

To determine which wavelengths we use to extrapolate from we utilize an autoencoder. While training the autoencoder we incrementally remove wavelengths that do not affect the reconstruction error, until there are no wavelengths left. The order in which the wavelenghts were removed is representative of how much information they contain to reconstruct the full spectrum.

Results:

The ordering of the wavelengths for the O2A-band as determined by the autoencoder from least to most informative is shown in Figure 2. The most informative wavelengths are in the continuum at both ends of the band as well as deep in the absorption bands. This suggests that the degrees of freedom of the spectra in the O2A-band are mostly constrained by the reflected radiance at wavelengths where there is little to no absorption as well as wavelengths with high absorption. The wavelengths in the continuum contain information about the solar and viewing geometry as well as aerosols and surface reflectance. The wavelengths in the absorption bands contain information about the abundance of various absorbing gases in the atmosphere.

The reconstructed and convolved spectrum as well as the difference to the original spectrum for the three OCO-2/3 bands is shown in Figure 1. The wavelength dependent difference is shown for one representative spectrum for each band.

Over a testing set of spectra that were not used to fit matrix, A, the relative error to continuum for the O2A-band, weak and strong CO2-band is 0.006%, 0.012%, and 0.014%, respectively.

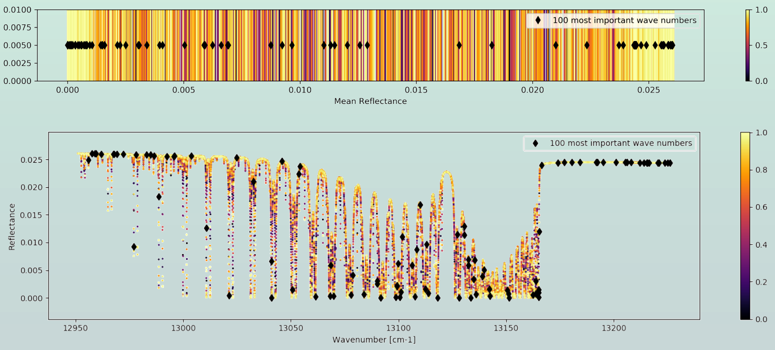

Figure 2: Informativity of each wavelength as determined by an autoencoder for the O2A-band. The 100 most important wavelengths are highlighted with black diamonds. The most informative wavelengths are shown in yellow, the least informative wavelengths in black.

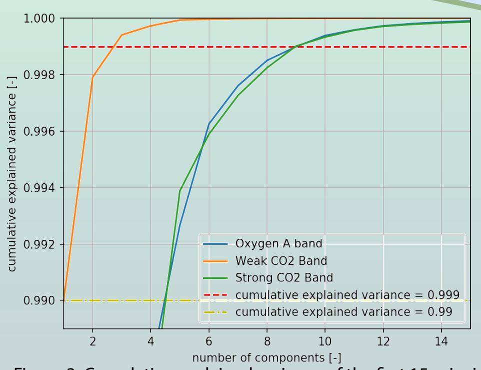

Figure 3: Cumulative explained variance of the first 15 principal components for the three OCO-2/3 instrument bands. O2A-band is shown in blue, WCO2-band in orange, and SCO2-band in green.