AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2017JD027013

Key Points:
- PCA approach to fast RT is extended to include varying scattering properties
- Retrieval errors due to PCA-based approach are estimated via an ensemble of simulations
- For XCO_2, the errors are generally below ±0.06 ppm for a large range of geophysical conditions

Correspondence to:
P. Somkuti,
ps345@le.ac.uk

Citation:
Somkuti, P., Boesch, H., Vijay, N., & Kopparla, P. (2017). Application of a PCA-based fast radiative transfer model to XCO_2 retrievals in the shortwave infrared. *Journal of Geophysical Research: Atmospheres*, *122*, 10,477–10,496. https://doi.org/10.1002/2017JD027013

Received 21 APR 2017
Accepted 1 SEP 2017
Accepted article online 11 SEP 2017
Published online 12 OCT 2017

Application of a PCA-Based Fast Radiative Transfer Model to XCO_2 Retrievals in the Shortwave Infrared

P. Somkuti[1,2], H. Boesch[1,2], V. Natraj[3], and P. Kopparla[4]

[1]Earth Observation Science Group, Department of Physics and Astronomy, University of Leicester, Leicester, UK, [2]National Centre for Earth Observation, University of Leicester, Leicester, UK, [3]Jet Propulsion Laboratory (NASA-JPL), Pasadena, CA, USA, [4]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Abstract In this work, we extend the principal component analysis (PCA)-based approach to accelerate radiative transfer (RT) calculations by accounting for the spectral variation of aerosol properties. Using linear error analysis, the errors induced by this fast RT method are quantified for a large number of simulated Greenhouse Gases Observing Satellite (GOSAT) measurements ($N \approx 30{,}000$). The computational speedup of the approach is typically 2 orders of magnitude compared to a line-by-line discrete ordinates calculation with 16 streams, while the radiance residuals do not exceed 0.01% for the most part compared to the same baseline calculations. We find that the errors due to the PCA-based approach tend to be less than ±0.06 ppm for both land and ocean scenes when two or more empirical orthogonal functions are used. One advantage of this method is that it maintains the high accuracy over a large range of aerosol optical depths. This technique shows great potential to be used in operational retrievals for GOSAT and other remote sensing missions.

1. Introduction

Carbon dioxide (CO_2) is the most important anthropogenic greenhouse gas, and its increase from the beginning of the industrial era (middle to late eighteenth century) has contributed 2.2 ± 1.1 W/m^2 to the global radiative forcing (Myhre et al., 2013). Space-based observations of CO_2 have the advantage of near-global coverage (dependent on the season), with satellite revisit times on the order of days. Several dedicated satellite missions have been launched recently that focus on observing the spatiotemporal distribution of atmospheric CO_2 with sufficient coverage and precision to estimate regional fluxes between the surface and atmosphere.

The primary space-based measurement concept for CO_2 is based on spectroscopic measurements of CO_2 absorption bands in the shortwave infrared spectral range (SWIR) of reflected sunlight. Compared to CO_2 absorption bands in the thermal infrared (e.g., 15 μm), bands in the SWIR offer the advantage that measurements are sensitive down to the lowermost part of the atmosphere, hence meeting the requirements needed to constrain carbon fluxes from and to the surface. The Greenhouse Gases Observing Satellite (GOSAT) (Kuze et al., 2016), the first dedicated CO_2 mission, for example, measures in three bands within the SWIR and the near infrared (NIR) with high resolution; two of those are used to measure CO_2 at 1.61 μm (so-called weak CO_2 band) and 2.06 μm (so-called strong CO_2 band), whereas the other one at 0.76 μm (O_2 A band) measures atmospheric molecular oxygen. Since the amount of oxygen in the atmosphere is known, any changes in the expected absorption features in the O_2 A band are primarily a result of aerosols and surface pressure. The presence of aerosols and variations in surface pressure will change the path length of the light and thus will induce errors in the retrieved CO_2 columns, if not corrected. Using the information in the O_2 A band, the surface pressure can be constrained better in order to increase the accuracy of the obtained CO_2 columns. Further, there is additional spectral information in the O_2 A band and the strong CO_2 band that allows for a better determination of the aerosol characteristics present in the scene.

Retrieving the columns of atmospheric CO_2 from the aforementioned measurements is primarily achieved through an inversion scheme coupled with a so-called forward model. The forward model describes the physics involved in the propagation of light through the atmosphere, its interactions with the constituents of the atmosphere as well as the surface, and includes the response of the instrument. The inversion scheme used in this work relies on Bayesian optimal estimation and calls the forward model iteratively,

until the forward model fits the measured radiance in terms of a maximum likelihood solution (Rodgers, 2000). With each iteration, a set of parameters (the state vector) is adjusted to minimize the difference between modeled and measured spectral radiances. Once the forward model matches the measurements, the adjusted parameters, such as the concentration of CO_2, are then said to be retrieved. In such a full physics approach (Cogan et al., 2012), the scattering properties of the atmosphere due to aerosols and thin clouds are not only considered in the forward model but are also simultaneously retrieved.

In order to account for scattering in the forward model, a sophisticated multiple-scattering (MS) radiative transfer (RT) model (Natraj & Spurr, 2007; Spurr, 2008; Spurr & Natraj, 2011) has to be utilized to treat the propagation of sunlight through the atmosphere. Compared to a nonscattering model, these calculations can be orders of magnitude more computationally expensive. Ignoring scattering altogether, on the other hand, can lead to significant errors in the top-of-the-atmosphere (TOA) radiances, which then leads to errors in the retrieved quantities (Aben et al., 2007; Natraj et al., 2007). Nelson et al. (2016) have shown that even after filtering for clear-sky conditions, neglecting scattering and absorption by clouds and aerosols can lead to increased errors unless restrictive filtering of the scenes is performed.

Satellites are generating a quickly growing volume of data (from ≈8,000 clear-sky scenes per day for the SCIAMACHY instrument, up to ≈200,000 for NASA's OCO-2, Taylor et al., 2016, or a CarbonSat-like mission, Buchwitz et al., 2013) which necessitates acceleration methods that reduce the computational effort for RT calculations while at the same time retaining a high level of accuracy of the forward model. As pointed out by Chevallier et al. (2007), regional biases in the low sub-ppm (parts per million, 0.1 ppm ≈ 0.03%) range can significantly alter the surface fluxes obtained from flux inversions. Consequently, the adopted RT schemes utilized by retrievals must not result in systematic errors of this order of magnitude that are driven by geophysical parameters like aerosol loadings.

Several fast RT methods have been formulated specifically for greenhouse gas retrievals. The low-streams interpolation (LSI) (O'Dell, 2010) and linear-*k* (Hasekamp & Butz, 2008) methods both belong to the class of spectral binning methods; these methods replace the many (roughly 30,000 per band) line-by-line calculations by a few representative calculations, which they use to derive the result for all spectral points. While linear-*k* requires only a small number of MS calculations, LSI additionally requires low-accuracy line-by-line MS calculations. Accuracy, in the case of RT models utilizing the discrete ordinates method, refers to the number of quadrature angles (see, e.g., Stamnes et al., 1988 for details). A higher number of quadrature angles will increase the numerical accuracy of the solution at the cost of a substantial increase in computational effort (computational time typically increases as the cube of the number of quadrature angles). For both LSI and linear-*k*, the only tuneable parameter, apart from the number of quadrature streams employed for the high-accuracy calculations (which is a parameter of the employed RT models, rather than the acceleration scheme), is the set of bin boundaries (usually approximately logarithmically spaced in the gas optical depth dimension) which has to be chosen by the user for the specific spectral ranges. While this is true for the PCA-based approach as well, it offers the opportunity for the user to choose the number of empirical orthogonal functions (EOFs) used in the reconstruction of the radiances. This provides a second parameter to further tune the reconstruction accuracy according to the requirements of the application. The error introduced by the existing fast RT methods mentioned above shows a dependence on the viewing geometry (see O'Dell, 2010) as well as the total aerosol optical depth (see Hasekamp & Butz, 2008); the PCA-based approach shows little sensitivity to aerosol loading as well as the viewing geometry (as discussed in sections 4.2 and 4.3), which makes it suitable for a much larger range of geophysical parameters.

In this paper, we apply a PCA-based spectral binning method to CO_2 retrievals from high-resolution GOSAT SWIR spectra to characterize the CO_2 retrieval error and its sensitivities. Using linear error analysis, we show that the high accuracy is retained even for larger aerosol loadings.

In section 2, the PCA-based approach is introduced and summarized (a more detailed treatment is provided in Appendix A1). We then describe the setup of the simulations as well as the linear error analysis method in section 3. The results are presented in section 4, where we define a metric to evaluate residual errors and analyze the entire ensemble against this metric. The residual errors then are translated into XCO_2 errors, and their distribution and characteristics are discussed. We conclude section 4 using a subset of simulations to show how different instrument specifications influence the accuracy of the PCA-based approach.

2. PCA-Based Approach

RT calculations require optical properties for a given atmosphere and surface. The atmosphere is typically divided into multiple layers, where the optical parameters are assumed to be uniform within a layer but different between layers. The most commonly used parameters are the optical depth τ, the single-scattering albedo ω, and a representation of the scattering phase function, usually some expansion coefficients β. In addition to the atmospheric parameters, a surface model is employed; in our case we use a purely Lambertian surface model for land scenes that is defined solely by the Lambertian albedo parameter ρ. For ocean scenes, we employ the isotropic Cox-Munk surface model (Cox & Munk, 1954) that is parametrized by the magnitude of the wind speed (but not the direction); the refractive index of water is assumed the same for every scene.

A TOA spectrum (see Figure 1) can be obtained by line-by-line RT calculations for the spectral range of interest. The RT calculations need to be carried out on a fine spectral grid, which in some cases means a few tens of thousands of points. Thus, a set of atmospheric optical properties that fully define the model atmosphere consist of a large number of $\tau_{i,j}$, $\omega_{i,j}$, and $\beta_{i,j,m}$, where i is the index describing the spectral dimension, j is the layer index, and m denotes the expansion coefficient order. Due to the layered structure of the model atmosphere, one can consider the optical properties at a grid point i as profiles; for example, $\boldsymbol{\tau}_i$ is the optical depth profile for a specific wavelength index i. More details on the construction of the optical properties themselves can be found in Appendix A1.

Similar to the LSI and linear-k methods, the PCA-based approach makes use of the large amount of information redundancy within the spectral range of an absorption band that is present in the optical properties. The representative profiles of the optical properties can be inferred from PCA. Within the small spectral range of a typical absorption band, the scattering properties and surface reflectance vary little compared to the dynamic range of gas absorption. This is even more so if the spectral points and the corresponding sets of optical parameters are previously divided into "bins" that are defined in gas optical depth space, where ideally all points within a bin are optically similar. The computationally expensive MS calculations can be performed for representative profiles for a bin (rather than for every point in the bin), and the result for all points in that bin can then be derived from that rather effortlessly.

We utilize the binning scheme described in Kopparla et al. (2016), which proposes 11 bins in total gas optical depth space, where each bin is further subdivided into two bins according to the single-scattering albedo. It is acknowledged that the binning strategy can have an impact on the reconstruction accuracy and therefore on the results derived from the reconstructed radiances. However, as long as the same binning strategy is used for every scene, we expect mainly a shift in the magnitude of the relevant results (e.g., XCO_2 errors), the dependency on geophysical parameters to be similar when compared to a different binning strategy. The issue of binning strategy is explored in detail in Kopparla et al. (2017).

The following paragraph briefly outlines the steps needed to approximate a high-accuracy RT calculation for a band using the PCA-based approach. It is assumed that assignment of spectral points into bins has already taken place, and each step has to be repeated for every bin. A detailed elaboration on all steps is found in the Appendix, as well as in Natraj et al. (2005, 2010) and Spurr et al. (2013). It is assumed that a fast, low-accuracy (two-stream) line-by-line calculation of the TOA radiances ($I_{\mathrm{low},i}$) is performed beforehand for every spectral point in the band—the PCA-based approach then adjusts those calculations to approximate the high-accuracy line-by-line TOA radiances using correction factors J_i:

$$I_{\mathrm{approx},i} = I_{\mathrm{low},i} \cdot \exp(J_i). \tag{1}$$

1. Construct the optical state matrix $\mathbf{O} = \{\boldsymbol{\tau}, \boldsymbol{\tau}^{(\mathrm{Ray})}, \mathbf{q}, \mathbf{c}, \boldsymbol{\rho}\}$ which contains the total optical depth profiles $\boldsymbol{\tau}$, the optical depth profiles due to Rayleigh scattering $\boldsymbol{\tau}^{(\mathrm{Ray})}$, the aerosol scattering coefficients $\mathbf{q}$ (aerosol absorption is already included in the total optical depth $\boldsymbol{\tau}$), the aerosol interpolation coefficients $\mathbf{c}$ (see Appendix A1 for the meaning of $\mathbf{q}$ and $\mathbf{c}$), and the Lambertian surface albedos $\boldsymbol{\rho}$. $\mathbf{O}$ is then transformed by taking the natural logarithm elementwise: $\mathbf{F} = \ln(\mathbf{O})$. $\mathbf{F}$ is subsequently mean subtracted (the mean of every column is subtracted from every element in that column) to form $\overline{\mathbf{F}}$, from which the principal components (PC) are extracted. Note that the inclusion of $\mathbf{q}$ and $\mathbf{c}$ in $\mathbf{O}$, which was not done in previous work, is crucial to fully take the spectral dependence of the scattering properties of aerosols into account.

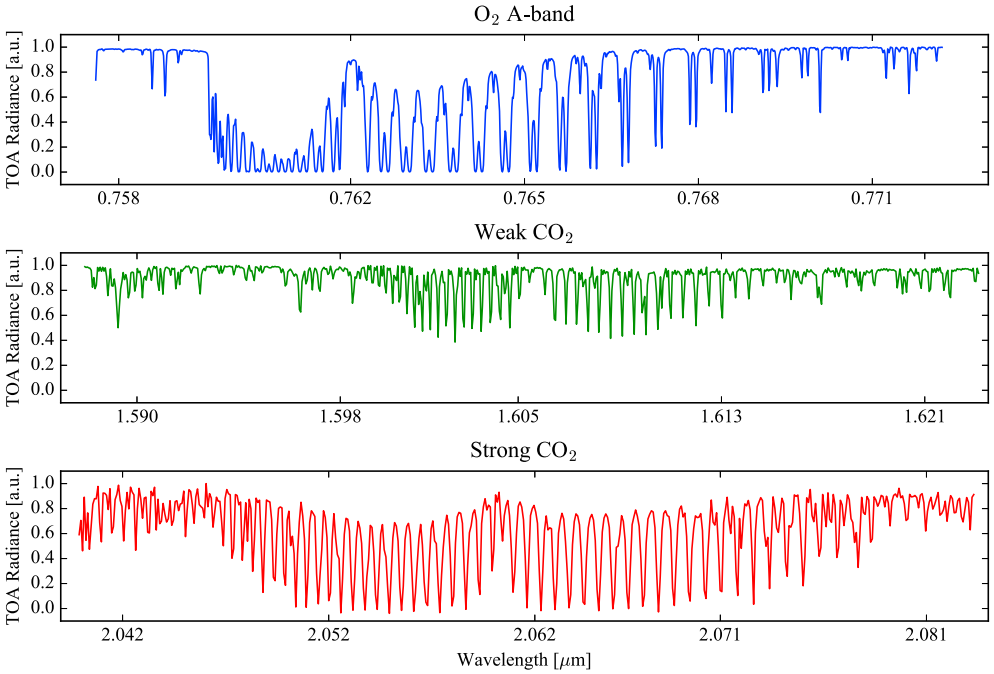

Figure 1. Typical modeled spectra using GOSAT instrument specifications. (top) O_2 A band, (middle) weak CO_2 band, and (bottom) strong CO_2 band. All spectra have been normalized by the highest value within the respective band.

2. PCs are calculated by solving the eigenproblem $\mathbf{C}\mathbf{V}_k = \eta_k \mathbf{V}_k$ with eigenvectors, or EOFs, $\mathbf{V}_k$ and eigenvalues η_k for the covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \frac{1}{N_s - 1}\left(\bar{\mathbf{F}}^T \bar{\mathbf{F}}\right), \tag{2}$$

N_s being the number of spectral points in the given bin. The eigenvectors then are scaled to obtain $\mathbf{W}_k = \sqrt{\eta_k}\,\mathbf{V}_k$, before the original mean-removed transformed optical properties are projected onto the scaled EOFs to yield the PCs $\mathbf{P}_k$:

$$\mathbf{P}_k = \frac{1}{\eta_k}\bar{\mathbf{F}}\,\mathbf{W}_k. \tag{3}$$

3. For every bin, a set of RT calculations with optical parameters representative of that bin have to be performed: one for the so-called mean optical state $\mathbf{O}^{(0)}$ and $2 \times N_{\mathrm{EOF}}$ calculations for perturbed optical states $\mathbf{O}^{(\pm k)}$. N_{EOF} is the total number of EOFs used for the reconstruction and is typically between 1 and 5. Binned mean optical properties $\mathbf{O}^{(0)}$ are constructed simply as the mean of all spectral values in the bin, following a potential back transformation from logarithmic space. For every EOF k used in the reconstruction, two optical states $\mathbf{O}^{(\pm k)}$, each perturbed in EOF space, have to be created as well. These calculations can be seen in detail in Appendix A5.
4. The optical states $\mathbf{O}^{(0)}$ and $\mathbf{O}^{(\pm k)}$, along with other auxiliary parameters (layer altitudes, solar and viewing angles, etc.), contain all information needed to perform RT calculations. For every one of these optical states, we run two RT calculations, one each in low-accuracy mode and one in high-accuracy mode (e.g., low and high number of quadrature streams), in order to calculate the ratios J:

$$J^{(0)} = \ln\left[I_{\mathrm{high}}\left(\mathbf{O}^{(0)}\right)/I_{\mathrm{low}}\left(\mathbf{O}^{(0)}\right)\right], \tag{4}$$

$$J^{(\pm k)} = \ln\left[I_{\mathrm{high}}\left(\mathbf{O}^{(\pm k)}\right)/I_{\mathrm{low}}\left(\mathbf{O}^{(\pm k)}\right)\right]. \tag{5}$$

5. Using the principal components, the ratios are then used to map the correction factors back into spectral (e.g., wavelength or wave number) space, denoted by the index i:

$$J_i = J^{(0)} + \sum_{k=1}^{N_{\mathrm{EOF}}} \frac{J^{(+k)} - J^{(-k)}}{2} P_{k,i} + \sum_{k=1}^{N_{\mathrm{EOF}}} \frac{J^{(+k)} - 2J^{(0)} + J^{(-k)}}{2} P_{k,i}^2. \tag{6}$$

The final step is then to multiply the low-accuracy result $I_{\mathrm{low},i}$ by the correction factors J_i as mentioned in equation (1) to obtain the approximation to the high-accuracy calculation $I_{\mathrm{high},i}$.

3. Setup of Geophysical Simulations

To realistically characterize the CO_2 retrieval errors and the spectral residuals of the PCA-based method, we perform a set of simulations that represent the global distribution of expected scenarios covering a wide range of geophysical parameters. Roughly 30,000 model atmospheres have been used to compute the TOA radiances which correspond to locations of GOSAT soundings from two seasons: May, June, and July 2011 (summer) and November, December, and January 2011/2012 (winter). The GOSAT soundings have been chosen for near-uniform global coverage (cutoff at −60° southern latitude to exclude Antarctica), such that there are at most five locations in every 2.0 by 2.0° grid box; we choose clear-sky scenes only. The cloud screening was carried out prior to the simulations via an O_2 A band fit, where cloudy scenes were identified when the apparent surface pressure deviated more than 30 hPa from the value obtained from European Centre for Medium Range Weather Forecasts (ECMWF) ERA-Interim data.

For each season, simulations for ocean glint scenes have been performed as well, where the wind speed parameter is also taken from ECMWF ERA-Interim. The two seasons are in contrast to each other mostly in terms of solar zenith angles and the associated signal level.

The model atmospheres for each individual sounding are based on Boesch et al. (2013). Temperature, humidity, and gas profiles are extracted from ECMWF ERA-Interim and the model CO_2 from the Copernicus Atmosphere Monitoring Service (CAMS) respectively; the spectrally invariant surface albedos (over land) per band are estimated from the radiances in the measured GOSAT spectra. Aerosol profiles are calculated on a per scene basis from CAMS and consist of five different tropospheric types (sea salt, dust, organic matter, black carbon, and sulfate), which are partly further differentiated by size bins and dependence on relative humidity. Every scene contains a thin cirrus cloud aerosol mixture with a constant total optical depth of 0.005. The Gaussian height profile is parametrized by a latitude-dependent mean height and width according to Eguchi et al. (2007).

Due to the global distribution of the sounding locations (see Figure 2), the simulations cover a wide range of scenes with different surface types, varying aerosol loadings, and solar zenith angles.

3.1. Linear Error Analysis

The forward model error (see section 3.1) is influenced by the measurement noise covariance $\mathbf{S}_\epsilon$, which is assumed to be a diagonal matrix containing the squared noise equivalent radiances. The three bands in the SWIR utilized for full physics XCO_2 (dry-air column-averaged carbon dioxide mole fractions) retrievals exhibit different radiance levels, and therefore, the associated signal-to-noise ratios are different for each band due to the spectral dependence of the reflective properties of the various surfaces. We use the linear sensitivity analysis framework described in Rodgers (2000), in which the retrieval error can be estimated from the forward model error. The error of the state vector $\Delta\mathbf{x}^{\mathrm{EOF}}$ is derived from the gain matrix $\mathbf{G}$, which maps spectral features into state vector space:

$$\mathbf{G} = \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}, \tag{7}$$

$$\Delta\mathbf{x}^{\mathrm{EOF}} = \mathbf{G}\Delta\mathbf{f} = \mathbf{G}\left[\mathbf{f}^{\mathrm{true}}(\mathbf{x},\mathbf{b}) - \mathbf{f}^{\mathrm{approx}}(\mathbf{x},\mathbf{b})\right]. \tag{8}$$

$\mathbf{S}_a$ is the prior covariance matrix (Rodgers, 2003), and the noise covariance $\mathbf{S}_\epsilon$ is calculated from the standard deviation of the GOSAT out-of-band signal (Cogan et al., 2012). The weighting functions $\mathbf{K} = \partial\mathbf{f}^{\mathrm{true}}/\partial\mathbf{x}$ are obtained from the high-accuracy simulation $\mathbf{f}^{\mathrm{true}}$, which is a line-by-line simulation using 16 quadrature streams (8 in each half-space). $\mathbf{f}^{\mathrm{approx}}$ is the simulation using the PCA-based approach. Both simulations $\mathbf{f}^{\mathrm{true}}$ and $\mathbf{f}^{\mathrm{approx}}$ are evaluated using the same state vector $\mathbf{x}$ and auxiliary parameters $\mathbf{b}$ so that the only difference between the two runs is the RT portion of the forward model. $\mathbf{b}$ contains non–state vector quantities that are still needed for forward model computations, such as the volume mixing ratios of nonretrieved gases (O_2, CH_4, and H_2O) and various instrument parameters.

The parts of the state vector error that correspond to the CO_2 profile ($\Delta\mathbf{x}^{\mathrm{EOF}\ (CO_2)}$) are converted to a column-averaged value using the pressure weighting function $\mathbf{h}$ (O'Dell et al., 2012):

$$\Delta XCO_2 = \mathbf{h}^T\Delta\mathbf{x}^{\mathrm{EOF}\ (CO_2)}. \tag{9}$$

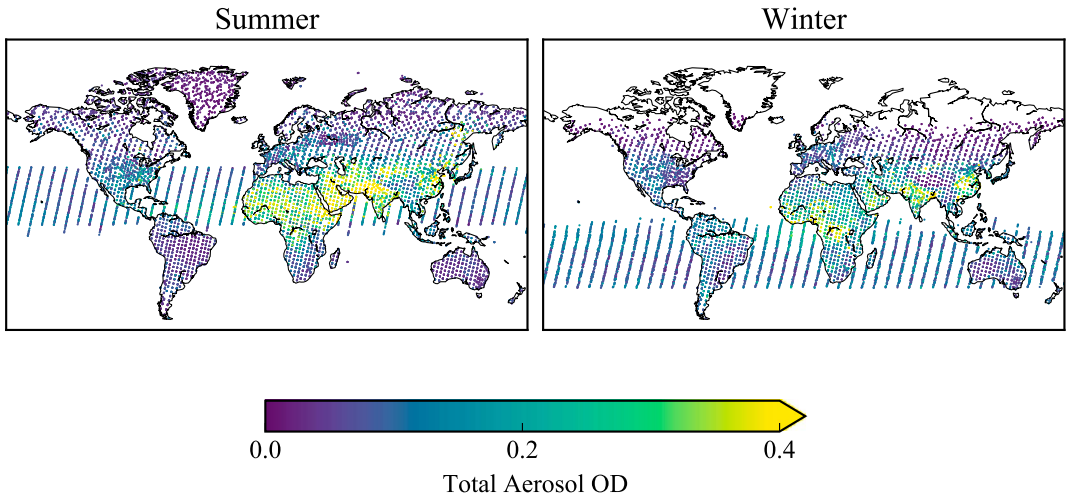

Figure 2. Locations for which forward model simulations are performed and corresponding total aerosol optical depths. For both seasons, the coverage is almost global, where the shift toward high southern latitudes is due to the changing solar zenith angle for the winter season. Note the two different latitudinal bands covered by the glint observations over the oceans in the two seasons. Apart from Boreal forests and Greenland, most landmasses are covered in both seasons.

4. Results

4.1. Spectral Residuals

Before focusing on the entire ensemble of model simulations, we pick out a single case and take a closer look at the spectral residuals for this scene. We use a quantity that gives the relative error of the radiances on a per wavelength basis with respect to the full line-by-line calculation $I_{\mathrm{lbl},i}$ for each wavelength index i as

$$\Delta I_{\mathrm{rel},i} = \frac{I_{\mathrm{approx},i} - I_{\mathrm{lbl},i}}{I_{\mathrm{lbl},i}}. \tag{10}$$

This quantity is shown for all three bands using 1, 2, 3, and 5 EOFs per bin (N_{EOFs}) in Figure 3. The particular scene chosen here corresponds to a GOSAT measurement near Beijing (land). For this model atmosphere, the total aerosol optical depth (AOD) is high with a value of 0.82; the viewing zenith is off nadir at ≈14°, and the solar zenith angle is around 25°. This scene, whose reconstruction accuracy is shown in Figure 3, is challenging because of the high aerosol loading, which increases the MS contribution relative to the single-scattered component, thereby amplifying the difference between the low-accuracy (two stream) and high-accuracy calculations; this is therefore a harder problem for the PCA model compared to a scenario with low AOD.

Since the relative error (equation (10) is calculated for each pixel individually, spectral points in the center of deep absorption lines with very small TOA radiances show larger relative errors, mostly in the strong CO_2 band, which also increases the residual root-mean-square (RMS). The interquartile range (IQR) of the residual absolute values is hence also stated in Figure 3, which is less sensitive to single outliers, and serves as a robust metric to evaluate the magnitude of the residuals for the entire spectral band.

The residuals in Figure 3 show that the spectrally varying scattering properties are well compensated for, thanks to the extension of the formalism to include necessary quantities in the decomposition and reconstruction. There is no significant overall or piecewise slopes unlike in earlier work (see, e.g., Figure 1 in Natraj et al., 2010). For the sounding in Figure 3, the signal-to-noise ratios for the three bands are 182, 298, and 286, respectively. Taking the inverse of these ratios and stating them as percentages reveals that the noise levels as a fraction of the continuum level radiance for this particular sounding are approximately 0.55%, 0.34%, and 0.35%. Thus, the spectral residuals are roughly 1 order of magnitude smaller than the noise levels when using 1 EOF.

Despite the presence of a few points with large relative errors in the line centers, the expected result is obtained with both RMS and IQR decreasing with N_{EOFs} increasing. The biggest change is seen going from 1 to 2 EOFs, with a further decrease for 3 EOFs. The difference between 3 and 5 EOFs is largest in the weak CO_2 band for this case; however, the improvement is not the same in the other two bands.

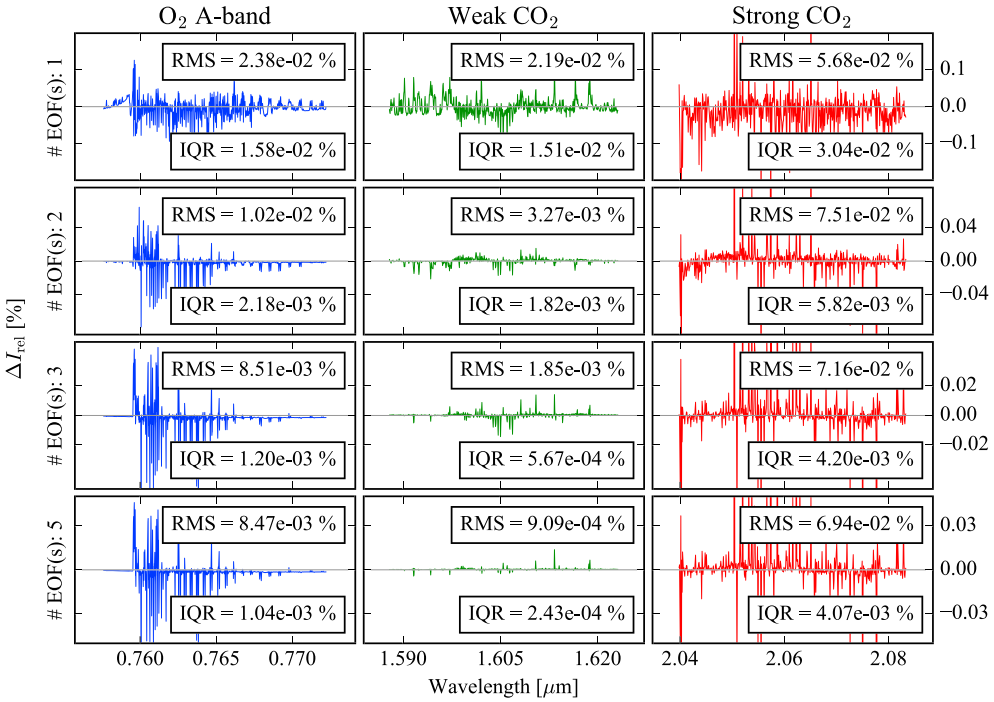

Figure 3. Sample residuals using different numbers of EOFs to reconstruct the TOA radiances (see full modeled spectra in Figure 1). Note the spectral flatness of the residuals as well as the changing scales with increasing number of EOFs. RMS is the root-mean-square, and IQR is the interquartile range.

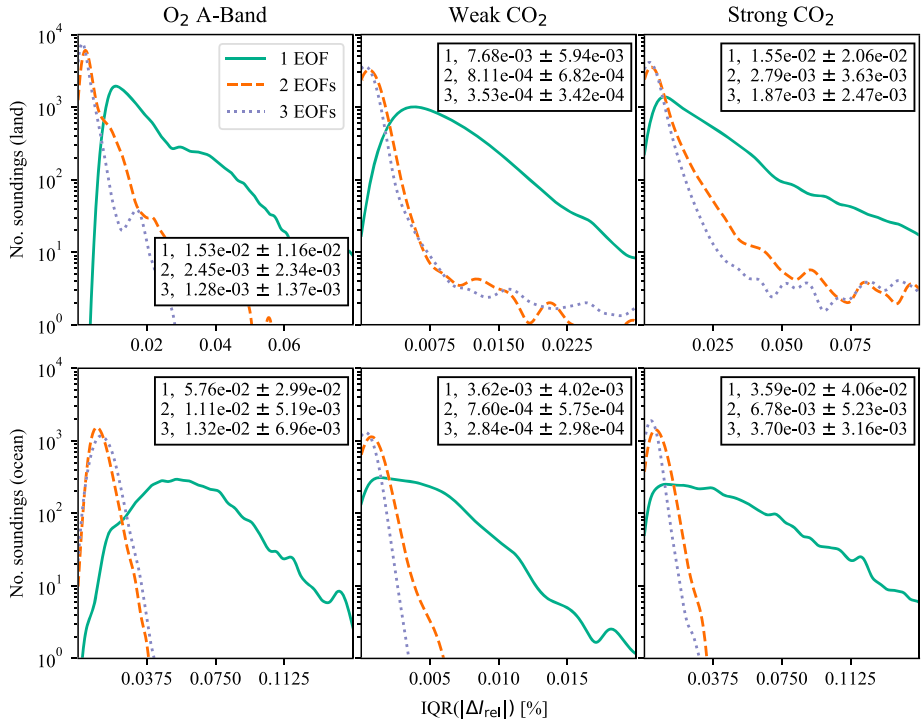

Figure 4. Histograms of the spectral residuals. The abscissae represent the interquartile ranges of the absolute values of the relative errors $|I_{rel,i}|$ (see equation (10). The inset text boxes in each subplot show the median and the interquartile ratios for each distribution (number of EOFs).

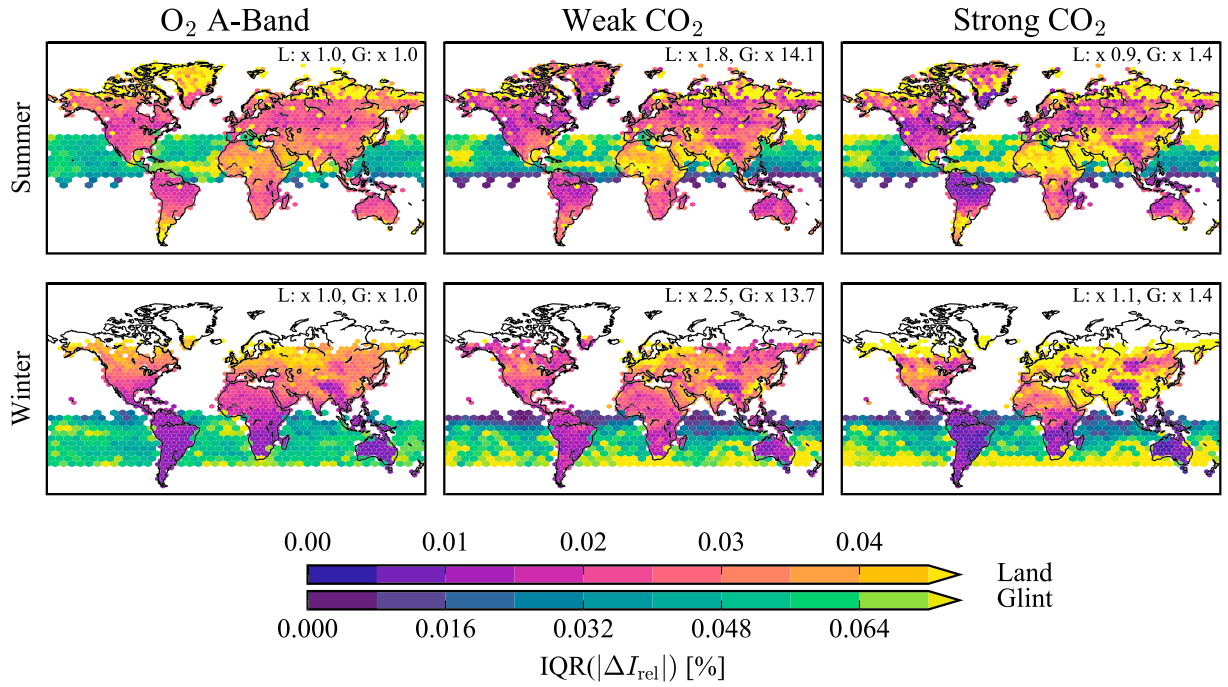

Figure 5. Global map of residual magnitudes when using 1 EOF. Bins for ocean simulations have been chosen to be large enough to allow for continuous patches despite the spacing of orbits (see Figure 2), which makes a visual assessment easier. The factors shown in the top right corner of each map indicate the factors, with which the respective ensembles have been multiplied (L for land and G for glint) such that the values can be plotted using the same color map. Regions over land with larger residuals are easily identified, being dusty deserts (Sahara and Arabian peninsula), regions with large amounts of sea salt (high southern latitudes) and aerosol outflow regions (mid-Atlantic, western North Pacific) as well as high latitudes with higher solar zenith angles.

We investigate the residuals for the entire ensemble of simulations by assessing the histograms of the IQRs in Figure 4. The histograms are generally consistent with the trend shown in Figure 3; increasing the number of EOFs typically reduces the spectral residuals. For simulations that employ the Lambertian surface model (land), both the O_2 A band and the strong CO_2 band show an improvement when increasing the number of EOFs from 2 to 3, whereas the improvements for the weak CO_2 band are only marginal when considering the entire ensemble. For the Cox-Munk surface model, on the other hand, the major improvement seems to be when going from 1 to 2 EOFs.

Ocean glint scenes exhibit systematically higher residuals for the O_2 A band and the strong CO_2 band, and lower residuals for the weak CO_2 band, when compared to land scenes. This result is counterintuitive, since glint scenes are more dominated by the SS contributions and thus have smaller MS contributions compared to the SS magnitude. With a small MS contribution, the reconstruction accuracy will depend linearly on the the ratio of MS to SS radiances. However, when the MS contributions are higher (in our case up to ≈120% for land scenes and ≈50% for ocean scenes), we are no longer in the linear regime. Here the residuals will depend on the observation geometry, surface model, and aerosol scattering properties in a nontrivial fashion.

To explore the spatial structure of the error distribution in Figure 4, IQR ($|I_{\mathrm{rel},i}|$) is displayed on global maps in Figure 5. Due to the varying scales of the residuals for each map, the respective data are scaled by a factor that is indicated in every subplot. The maps reveal apparent drivers for the residuals. For glint simulations, the largest residuals are seen in regions associated with heavy aerosol loadings, such as the mid-Atlantic Ocean with its desert dust outflow and larger viewing and solar zenith angles. Over land, the residuals are larger with increasing solar zenith angles as well as increasing aerosol optical depths.

4.2. XCO_2 Errors

Section 4.1 highlights certain regions and scenarios for which the PCA-based approach produces larger residuals. These regions, however, do not directly translate into the regions with the largest XCO_2 errors, as the signal-to-noise ratio and the gain matrix affect those errors according to equation (7).

Figure 6 shows the calculated errors for the two seasons distributed on global maps as well as separate histograms for land and ocean simulations. There is a notable difference, between the summer and winter

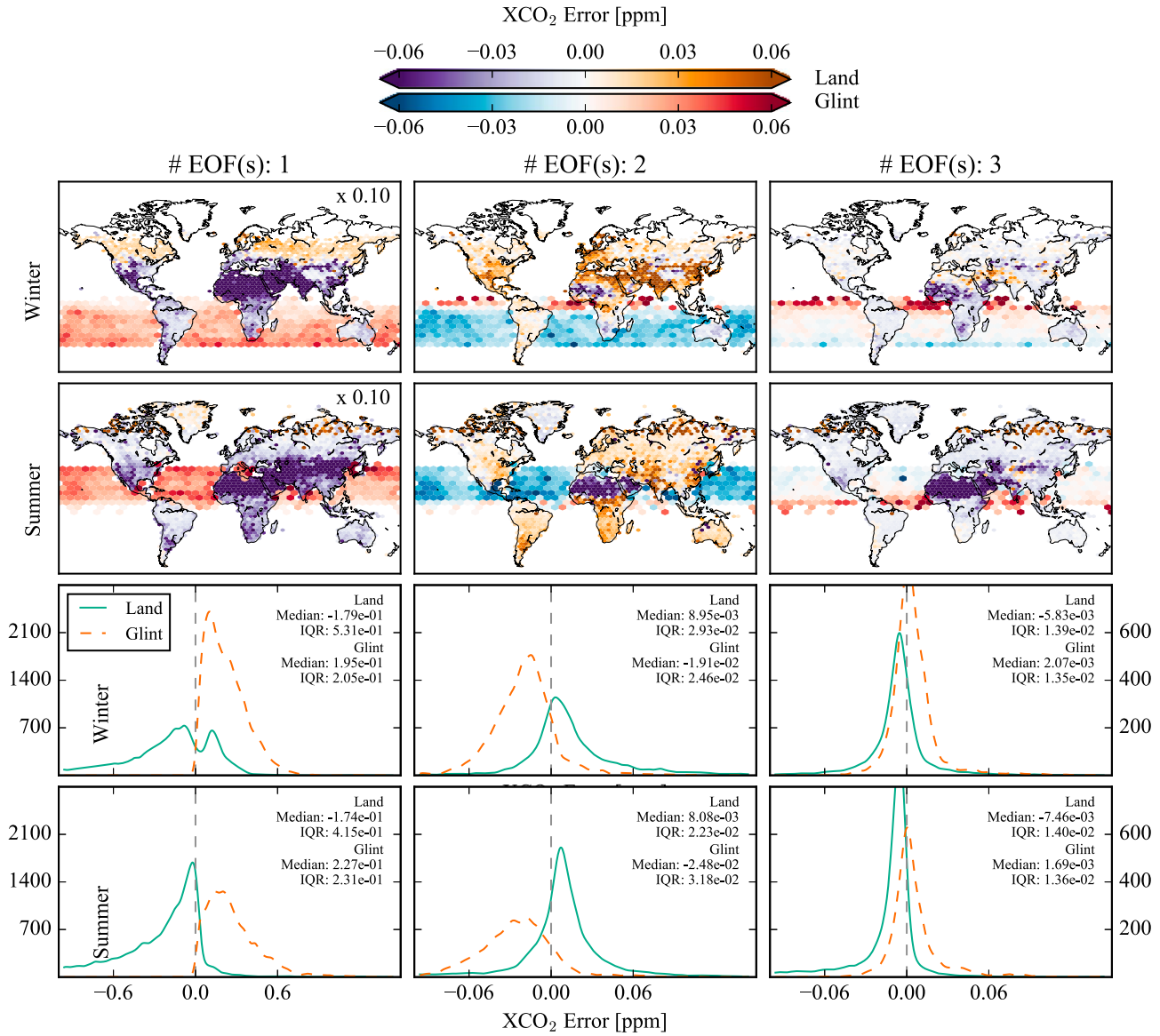

Figure 6. Estimated XCO_2 errors displayed as both maps and histograms. For the case of using only 1 EOF, the errors on the map were scaled by a factor of 0.1 such that the same color bar could be used. Regions with higher surface reflectance and thus higher SNR clearly result in higher XCO_2 errors (deserts) over land. Over ocean, errors are increased in the mid-Atlantic due to larger AODs (compare Figure 2) and the larger viewing angles at the edges of the glint observation bands.

seasons, in XCO_2 error behavior for scenes over the Sahara, the Arabian peninsula, and over parts of India; these are all regions with high signal-to-noise ratio measurements (or equivalently, high albedo in the strong CO_2 band). For the summer season, the XCO_2 error is larger over these areas compared to others when 1, 2, or 3 EOFs are used. In contrast, for the winter season, this is true only when 1 EOF is used.

Table 1
Percentage of Scenes With XCO_2 Errors Larger Than ±0.1 ppm

# EOFs	Land (summer)	Glint (summer)	Land (winter)	Glint (winter)
1	67.09%	82.10%	78.85%	76.84%
2	5.59%	1.36%	5.76%	0.29%
3	5.63%	0.75%	2.18%	0.47%
5	0.39%	0.78%	0.21%	0.40%

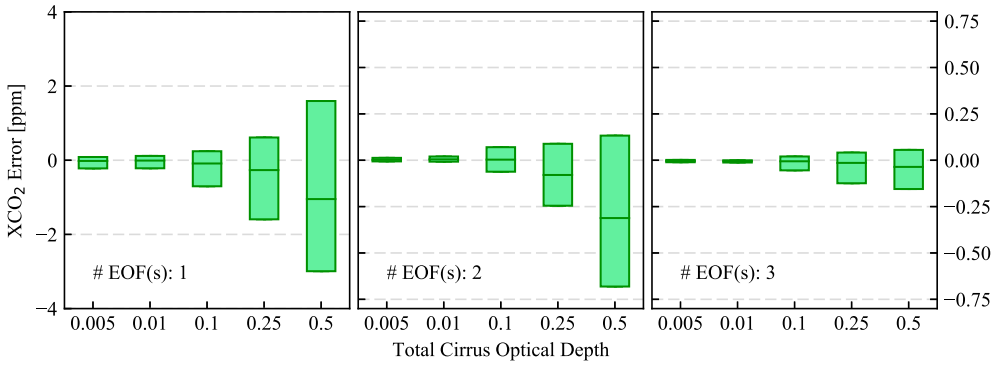

Figure 7. XCO_2 errors at various cirrus cloud optical depths for a small subset of the winter season scenes. There is a dramatic reduction of the error magnitude going from 1 EOF to 2 EOFs, especially for cases with cirrus optical depth larger than 0.1. The tick labels for the 1 EOF case are seen on the left-hand side of the three plots, whereas the tick labels for the cases of 2 and 3 EOFs are seen on the right-hand side. Whiskers in this plot have been suppressed.

This difference can be seen in the spectral residuals (Figure 5) and is driven mainly by the changing solar zenith angle, which is roughly between 10° and 20° in the summer and between 30° and 60° in the winter season.

The effect of changing the number of EOFs on the dependence of the residual errors with respect to optical properties is also seen in Figure 6. For summer scenes over the Sahara, the estimated XCO_2 error is largely independent of the total AOD using only 1 EOF but shows a rather clear dependence when using 2 or 3 EOFs. The exact opposite behavior is seen for the winter season. SNR dependence of the estimated error is similar to the AOD dependence: for summer scenes, using more than 1 EOF leads to a clear dependence, whereas for winter scenes, this dependence is much weaker.

The generally larger spectral error for glint observations (see Figure 4) translates into a larger mean XCO_2 error. Land scenes in the summer season show a higher error over the Sahara and Arabian peninsula, coinciding with those scenes with a high signal-to-noise ratio. There is a change in sign for the overall error for land observations, but the magnitude of the errors is already small with an IQR of the distributions of around 0.06 ppm and lower. Glint-type soundings using the Cox-Munk surface model show higher scatter for each season, when only 1 EOF is used, and are comparable to land scenes when using 2 or 3 EOFs. To summarize the dependence of the XCO_2 error on the number of EOFs, the percentage of scenes with errors larger than 0.1 ppm is stated in Table 1.

4.3. Influence of Cirrus Optical Depth

To assess the errors induced by different cirrus optical depths, the XCO_2 errors of a small subset of $N \approx 500$, globally distributed scenes from the winter/land season set, were additionally simulated for four different total optical depths of cirrus clouds: 0.01, 0.1, 0.25, and 0.5. Boxplots illustrating the results are shown in Figure 7.

Since XCO_2 retrievals tend to be filtered very restrictively using the total retrieved cirrus OD (e.g., a filter threshold of <0.05 as stated in Cogan et al., 2012), the cases of 0.01 and 0.1 are most relevant. For the latter case, the standard deviation (as well as the IQR) of the errors in this subset increased roughly by a factor of 2 to 3 compared to the original scenario of 0.005. These IQRs do not exceed 0.7 ppm using 1 EOF and decrease to 0.05 ppm using 3 EOFs. Comparing these numbers to the those in Figure 6, one can see that the IQRs of the respective distributions roughly doubled/tripled when the total cirrus OD was increased to 0.1.

Table 2
Land Cover Class Indices Associated With the Four Surface Types, European Space Agency (2016)

Surface type	Land cover class indices
Tree cover	50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90
Bare area	200, 201, 202
Sparse vegetation	120, 121, 122, 130, 140, 150
Water	220

The conclusion here is that high-altitude aerosols, such as cirrus clouds, have a similar effect on the reconstruction accuracy, and ultimately the retrieval error, as aerosols in the lower troposphere in the glint observation mode. The PCA-based approach performs well for cirrus optical depths as large as 0.5 when 3 EOFs are used.

4.4. Varying Instrument Configuration

The XCO_2 error from the PCA-based method will change for different instruments. To evaluate how the PCA-based approach performs for other instruments, we perform the linear error analysis for a subset ($N \approx 5500$) of scenes using different instrument models. The assumed

Table 3
Instrument Model Characteristics; the FWHMs of the Gaussian Instrument Line Shape Functions Are Stated for All Three Bands (O_2 A Band, Weak CO_2, and Strong CO_2)

Instrument	ILS FWHM	Sampling ratio
GOSAT	0.35, 0.25, 0.24 (cm^{-1})	1.4
A (OCO-2 like)	0.042, 0.076, 0.097 (nm)	2.5
B (CarbonSat like)	0.1, 0.3, 0.55 (nm)	3
C (Sentinel 5 like)	0.4, 0.25, 0.125 (nm)	3

instrument models are described by their spectral resolution, that is, the full width at half maximum (FWHM) of the Gaussian instrument line shape (ILS) functions, the noise model, and the dispersion relation as derived from the sampling per FWHM. A potentially different field of view or footprint size is not taken into account. The spectral windows and the viewing geometries also remain unchanged to allow for a direct comparison between the simulations with the only difference being the instrument models. The scenes themselves still correspond to specific GOSAT sounding locations found in both the summer and winter season sets.

The subset of scenes was chosen to contain predominantly four different surface types: tree cover, bare areas, sparse vegetation, and water. A scene falls into one of these categories if at least 95% of the land cover class grid boxes enclosed by the GOSAT footprint belong to one of the listed indices in Table 2. To that end, we utilize the 2008–2012 epoch of the ESA CCI Land Cover map project (European Space Agency, 2016) at 300 m resolution.

We consider three additional instruments, all of which are grating-spectrometer types, as opposed to GOSAT being a Fourier transform spectrometer. The first one is an OCO-2-like (Crisp et al., 2017) instrument that features high SNR and high resolution across all three bands. The second instrument is to resemble the Earth Explorer 8 candidate mission CarbonSat (Buchwitz et al., 2010) which has lower resolution than the first instrument, especially in bands 2 and 3. Finally, the last instrument resembles Sentinel 5-like specifications (Ingmann et al., 2012), with a much lower resolution in the O_2 A band but higher-resolving power in the strong CO_2 band at 2.0 μm. The 2.0 μm band is not present in the design specifications of Sentinel 5; we have assumed the instrument characteristics for that band from earlier studies (European Space Agency, 2011, 2012). Table 3 lists the spectral characteristics for the various instruments; the noise models are found in Appendix A7. The SNR of instrument A is generally the highest, followed by Instrument C, GOSAT, and Instrument B.

The comparison of the estimated XCO_2 errors for the four instrument types is visualized in Figure 8, where the scenes have been aggregated according to AOD bins. For these simulations we have used 3 EOFs for the radiance reconstruction.

Instrument C exhibits systematically larger errors than instrument B for all surface types apart from ocean, indicating that the spectral resolution in the strong CO_2 band is of higher relevance to the retrieval error than

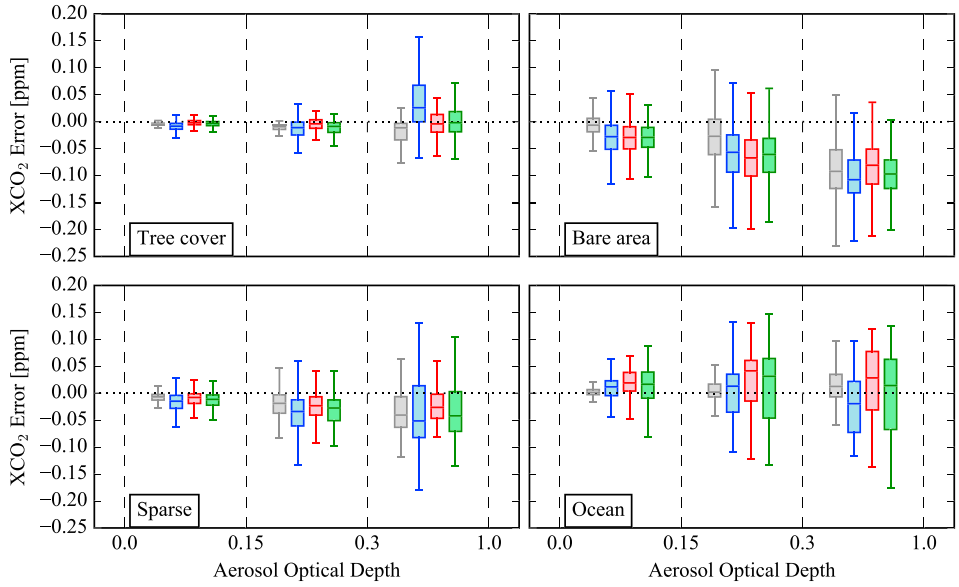

Figure 8. XCO_2 errors (using 3 EOFs) of scenes grouped into aerosol optical depth bins (the grey dashed lines indicate the bin boundaries) for the four instruments listed in Table 3 (left to right, GOSAT: grey, A: blue, B: red, C: green) for different surfaces.

the noise levels. When only 1 EOF is used (not shown), the dependence on the AOD for all surface types and instrument models is much higher.

From this analysis, we conclude that if 3 EOFs are used, the PCA-based approach is able to reconstruct the radiances to an accuracy at which the forward model errors result in less than ±0.2 ppm error in the retrieved XCO_2 for a variety of instrument models. Using only 1 EOF will inadvertently cause biases in high-AOD scenes on the order of 1 ppm.

4.5. Considerations Regarding the Computational Performance

A realistic assessment of the computational savings or the performance of the PCA-based fast RT method is dependent on the employed RT model(s) but stated to be roughly between 1 and 2 orders of magnitude (Kopparla et al., 2016).

We can make an assessment of the expected speedup by breaking down the full TOA radiance calculation into its constituent parts. Ignoring the overhead from the PCA computations, which for this type of application tends to be on the order of a few seconds, as well as from preparing the optical inputs for the binned calculations, the upper limit for the speedup S for a band can be estimated by taking the ratio between the computational times for the line-by-line and PCA RT calculations. The line-by-line durations are simply the number of spectral points multiplied by the average time of one full calculation for a single point t_{high}. The duration for the PCA approach is the sum of the line-by-line low-accuracy runtime $N_{\text{tot}} \cdot t_{\text{low}}$ and the runtime of both high- and low-accuracy binned calculations $(t_{\text{high}} + t_{\text{low}}) \cdot N_{\text{bins}} \cdot (2N_{\text{EOFs}} + 1)$:

$$S = \frac{t(\text{line-by-line})}{t(\text{PCA method})} = \frac{N_{\text{tot}} \cdot t_{\text{high}}}{N_{\text{tot}} \cdot t_{\text{low}} + (t_{\text{high}} + t_{\text{low}}) \cdot N_{\text{bins}} \cdot (2N_{\text{EOFs}} + 1)}, \tag{11}$$

$$= \frac{t_{\text{high}}}{t_{\text{low}} + (t_{\text{high}} + t_{\text{low}}) \cdot N_{\text{bins}} \cdot (2N_{\text{EOFs}} + 1)/N_{\text{tot}}}, \tag{12}$$

with

$$t_{\text{high}} = t_{\text{SS, high}} + t_{\text{MS, high}}, \tag{13}$$

$$t_{\text{low}} = t_{\text{SS, low}} + t_{\text{MS, low}}. \tag{14}$$

N_{tot} is the number of spectral points in the band, N_{bins} is the number of bins in the band, and N_{EOFs} is the number of EOFs used for each bin. The t_* represent the time it takes to compute the full (SS + MS) TOA radiance for a single wavelength point, running in either low- or high-accuracy mode. One can further simplify the expression above by assuming that a low-accuracy calculation of the full radiance field for any wavelength takes a certain fraction $\xi = t_{\text{low}}/t_{\text{high}}$ of the time of a high-accuracy calculation:

$$S \approx \frac{1}{\xi + (1 + \xi) N_{\text{bins}} \cdot (2N_{\text{EOFs}} + 1)/N_{\text{tot}}}. \tag{15}$$

In the case of discrete ordinates solvers, the ratio ξ depends mostly on the number of chosen computational quadrature angles, the total number of layers, and the number of analytical weighting functions to be calculated. In our setup, we use 19 layers and 5 profile weighting functions with 16 streams for the high-accuracy mode; $\xi \approx 0.005$ for both Lambertian and Cox-Munk surface models. The large reduction in computational time is due to the use of a dedicated two-stream model (Spurr & Natraj, 2011) rather than just running a general MS model with two quadrature streams. Note that ξ can change dramatically depending on the utilized RT model and is about an order of magnitude smaller if polarization is taken into account (O'Dell, 2010).

An example of the speedup as a function of N_{EOFs} and N_{bins} is shown in Figure 9, along with indicators of the choices for the simulations presented earlier. The speedup for these choices ranges from about 50 up to about 150, which falls in line with the numbers reported in Kopparla et al. (2016) and Natraj et al. (2010). The considerations here have been made for a scenario in which the number of EOFs is the same for every bin, and further optimization could reduce the number of binned calculations by reducing the number of EOFs for certain bins.

Regardless of the choice of N_{bins} and N_{EOFs}, equation 15 is bounded by ξ, in the limit of $N_{\text{bins}} \to 0$, $S \approx 1/\xi$. This upper bound represents a limit for RT acceleration methods for which full line-by-line calculations using a low-accuracy model or mode have to be performed.

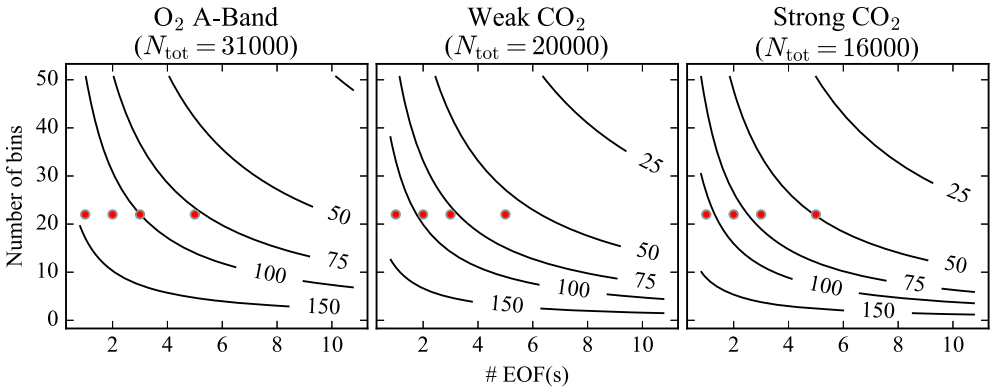

Figure 9. Speedup S for $\xi = 0.005$ and three different values of N_{tot}, each corresponding to the number of spectral points for the three bands. The location of the circles signifies the number of EOFs and number of total bins (22) used for the simulations. Black lines represent contour lines of constant S.

5. Summary

We have extended the PCA-based fast RT method introduced in Natraj et al. (2005) to account for spectrally dependent aerosol scattering properties and applied it to GOSAT-like measurements of CO_2. In order to assess the reconstruction accuracy of TOA radiances in SWIR bands, a large set of simulations was performed, and the approximated TOA radiances were compared to high-accuracy line-by-line calculations.

From these direct comparisons, the effect on retrieved quantities can be inferred via linear sensitivity analysis that also takes into account the instrument noise levels. Using 2 EOFs rather than only 1 for the reconstruction increases the accuracy by roughly an order of magnitude (as indicated by Figure 4), while using 3 EOFs does not increase the accuracy for the entire ensemble significantly but changes the spatial distribution. The radiance residuals are below that of those reported in alternative methods when more than 1 EOF is used, however, at the cost of more binned calculations when compared to LSI or linear-k.

To understand the estimated errors in terms of their magnitude, the XCO_2 errors due to the PCA approach can be compared to other errors that make up the total error budget of a retrieval. We have shown that using 3 EOFs in the PCA approach results in XCO_2 errors mostly smaller than $\pm$0.06 ppm for GOSAT measurements. Simulations using an OCO-2-like instrument model showed that the errors are mostly smaller than $\pm$0.2 ppm. Connor et al. (2016) report in detail the errors for OCO-2-based retrievals that stem from various sources of uncertainties, and we can therefore state that the errors due to the PCA approach are about $\leq$10% ($\leq$5%) of the reported total (variable) error. They also report the error induced by a less-informed aerosol prior. These errors are about 0.47 $\pm$ 0.63 ppm (June/land nadir) and about 0.37 $\pm$ 0.72 ppm (June/ocean glint) and are thus in the same order of magnitude as the errors due to the PCA-based approach (1 EOF). Using only 1 EOF for XCO_2 retrievals from OCO-2 measurements can significantly add to the total retrieval error.

We therefore conclude that the PCA-based method allows accurate reconstruction of TOA radiances for a large range of parameters. The accuracy can be tuned to the application by choosing different bins and changing the number of EOFs.

Comparing the errors to those reported for the linear-k method in Hasekamp and Butz (2008), we find that while the errors from the PCA method generally increase with AOD, the overall scale of the errors is substantially lower than for linear-k—especially at aerosol optical depths larger than 0.5. The linear-k method, however, does not require line-by-line low-accuracy calculations using an MS model and is thus faster than the PCA approach.

Appendix A

A1. Atmospheric Optical Properties

We use a setup similar to earlier work (Kopparla et al., 2016; Natraj et al., 2005, 2010) and consider a one-dimensional model atmosphere that is subdivided into N_l optically homogeneous and plane-parallel layers. For every spectral point i, each layer j is optically defined by the total optical depth $\tau_{i,j}$, the single-scattering albedo $\omega_{i,j}$, and the composite phase function expansion coefficients $\beta_{i,j,m}$ where the index

m runs from 0 to the number of expansion coefficients N_{mom}. A set of these three quantities τ, ω, and β fully define the optical properties of the model atmosphere in terms of the RT computations. Additionally, layer boundary altitudes are defined for the application of the pseudospherical approximation in which the incoming solar beam attenuation is calculated for a curved atmosphere.The total optical depth per layer can be written as the sum of its various contributions, which in our case are due to gas and aerosol extinction, as well as extinction due to molecular (Rayleigh) scattering:

$$\tau_{i,j} = \tau_{i,j}^{(\mathrm{gas})} + \tau_{i,j}^{(\mathrm{aer\text{-}ext})} + \tau_{i,j}^{(\mathrm{Ray})}. \tag{A1}$$

Optical depths due to gas extinction are often derived from spectroscopic line lists and are explicitly calculated from a precomputed lookup table (Benner et al., 2015; Devi et al., 2016; Drouin et al., 2016) for each spectral point i. Similarly, the extinction due to Rayleigh scattering can be computed from an analytic expression that contains the wavelength explicitly (Bucholtz, 1995).

The optical depths due to aerosol extinction, however, are not calculated explicitly for every spectral point. We make use of the methodology in Vogel et al. (2017, private communication), for which the parts relevant to the EOF method are quickly summarized here. Layer-resolved aerosol optical depths for mixture a at a given reference wavelength (in our case at 755 nm) are derived from the global CAMS aerosol product (http://www.gmes-atmosphere.eu/d/services/gac/nrt/nrt_opticaldepth/). The extinction and scattering coefficients $q^{(\mathrm{ext})}$ and $q^{(\mathrm{sca})}$ are calculated for each individual aerosol type using either Mie or T-Matrix calculations. For each observation, extinction and scattering coefficients are constructed from these base properties and calculated for two wavelengths $\lambda^{(\mathrm{beg})}$ and $\lambda^{(\mathrm{end})}$ at the edges of each spectral band; for details see Vogel et al. (2017, private communication). Since the GOSAT spectral bands are relatively narrow, the extinction, and thus the optical depth profiles, can be reconstructed via the Ångström exponent α, which can be calculated as

$$\alpha = -\frac{\ln\left(q^{(\mathrm{beg})}/q^{(\mathrm{end})}\right)}{\ln\left(\lambda^{(\mathrm{beg})}/\lambda^{(\mathrm{end})}\right)}, \tag{A2}$$

where q is either $q^{(\mathrm{ext})}$ or $q^{(\mathrm{sca})}$ at the wavelengths $\lambda^{(\mathrm{beg})}$ and $\lambda^{(\mathrm{end})}$. The extinction and scattering ratios for any wavelength in that specific band can then be obtained:

$$q_{i,a} = q^{(\mathrm{beg})}\left(\frac{\lambda_i}{\lambda^{(\mathrm{beg})}}\right)^{\alpha} = q^{(\mathrm{end})}\left(\frac{\lambda_i}{\lambda^{(\mathrm{end})}}\right)^{\alpha}. \tag{A3}$$

The same calculations also yield the phase function expansion coefficients $\beta_{a,m}^{(\mathrm{aer,\,beg})}$ and $\beta_{a,m}^{(\mathrm{aer,\,end})}$. The expansion coefficients as well as the extinction and scattering ratios are calculated for the band edges only due to the computational effort involved. To obtain the expansion coefficients at any given wavelength, a simple linear interpolation is used:

$$c_i = \frac{\lambda_i - \lambda^{(\mathrm{beg})}}{\lambda^{(\mathrm{end})} - \lambda^{(\mathrm{beg})}}. \tag{A4}$$

Using equation A4, the phase function expansion coefficients for an aerosol mixture a at spectral index i are calculated as

$$\beta_{i,m,a}^{(\mathrm{aer})} = \left(1 - c_i\right)\cdot \beta_{m,a}^{(\mathrm{aer,\,beg})} + c_i \cdot \beta_{m,a}^{(\mathrm{aer,\,end})}. \tag{A5}$$

To obtain the extinction and scattering optical depth at any given wavelength, the reference profile is multiplied by the extinction and scattering coefficients, respectively:

$$\tau_{i,j,a}^{(\mathrm{aer\text{-}ext})} = q_{i,a}^{(\mathrm{ext})} \cdot \tau_{j,a}^{(\mathrm{aer\,,ref})}, \tag{A6}$$

$$\tau_{i,j,a}^{(\mathrm{aer\text{-}sca})} = q_{i,a}^{(\mathrm{sca})} \cdot \tau_{j,a}^{(\mathrm{aer\,,ref})}, \tag{A7}$$

where the partial contributions due to each aerosol mixture a are summed up to obtain the total aerosol optical depth due to extinction and scattering in any given layer:

$$\tau_{i,j}^{(\mathrm{aer\text{-}ext})} = \sum_{a=1}^{N_a} \tau_{i,j,a}^{(\mathrm{aer\text{-}ext})}, \tag{A8}$$

$$\tau_{i,j}^{(\mathrm{aer\text{-}sca})} = \sum_{a=1}^{N_a} \tau_{i,j,a}^{(\mathrm{aer\text{-}sca})}. \tag{A9}$$

The total composite phase function expansion coefficients are composed according to the fraction of scattering that each aerosol mixture contributes compared to the total scattering including molecular Rayleigh scattering:

$$\beta_{i,j,m} = \frac{\tau_{i,j}^{(\mathrm{Ray})} \cdot \beta_{i,j,m}^{(\mathrm{Ray})} + \sum_{a=1}^{N_a} \tau_{a,i,j}^{(\mathrm{aer\text{-}sca})} \cdot \beta_{i,m,a}^{(\mathrm{aer})}}{\tau_{i,j}^{(\mathrm{Ray})} + \tau_{i,j}^{(\mathrm{aer\text{-}sca})}} \tag{A10}$$

The single-scattering albedos $\omega_{i,j}$ are calculated using the aerosol scattering coefficient $q_{i,a}^{(\mathrm{sca})}$:

$$\omega_{i,j} = \frac{\tau_{i,j}^{(\mathrm{Ray})} + \sum_{a=1}^{N_a} q_{i,a}^{(\mathrm{sca})} \cdot \tau_{j,a}^{(\mathrm{aer,ref})}}{\tau_{i,j}}. \tag{A11}$$

Despite the used RT models being blind to the partial contributions, these contributions due to each mixture need to be retained for the reconstruction of the binned optical properties.

A2. Preparing the Optical Properties

For every bin containing N_s spectral points, an $N_s \times N_o$ matrix $\mathbf{O}$ is constructed which contains a total of N_o layer-resolved, bulk or auxiliary optical properties. We consider auxiliary optical properties to be wavelength-dependent quantities which, while not directly used in RT computations, are used to derive such quantities. An example would be the aerosol interpolation fraction c_i (equation (A4)). All properties may be transformed using a suitable function beforehand, which we denote as f^F, and $f^F(\mathbf{O}) = \mathbf{F}$. This function f^F should have an inverse $f^B = (f^F)^{-1}$ that transforms the quantity back into its original space, such that $f^B(\mathbf{F}) = \mathbf{O}$. Previous work (Kopparla et al., 2016; Natraj et al., 2005, 2010; Spurr et al., 2013) chose f^F to be the natural logarithm in order to compress the range of gas optical depths, which results in less PCs needed to capture the same amount of variability. In general, f^F can be any bijective function $f^F : X \to Y$, as long as the optical properties are elements of X. Different functions could be applied to each type of optical property. Similar to Natraj et al. (2005, 2010) and Kopparla et al. (2016), we construct $\mathbf{O}$ using the total gas optical depth profiles $\tau_{i,j}^{(\mathrm{gas})}$. However, instead of adding the single-scattering albedos $\omega_{i,j}$, we use the optical depth due to Rayleigh scattering $\tau_{i,j}^{(\mathrm{Ray})}$. The advantage of this choice is a better reconstruction of the single-scattering albedos as well as aerosol properties, which will be explained in more detail in section A5. To account for wavelength-dependent scattering properties of aerosols, the aerosol scattering coefficient $q_{i,a}^{(\mathrm{sca})}$ (see equation (A3)) for each aerosol type is included. Additionally, we add the aerosol interpolation coefficients c_i (see equation (A4)). Finally, similar to Spurr et al. (2013), we also add the Lambertian surface albedos ρ_i as a bulk property. For different surface models, the surface albedo should be replaced or extended by any number of other spectrally dependent parameters. For the ocean glint Cox-Munk kernel (Cox & Munk, 1954), this parameter would be the relative refractive index of water to air m_i. The order of all of the mentioned quantities in the matrix $\mathbf{O}$ (the column position) is arbitrary as long as the same position of each quantity is used during the reconstruction.

Summarizing, the matrix $\mathbf{F} = f^F(\mathbf{O})$ is written as in equation A12, where the tilde over any property signifies that it has been transformed via the function f^F, that is, $f^F(x) = \tilde{x}$. Thus, the total number of columns in $\mathbf{F}$ is $N_o = 2N_l + N_a + 2$.

$$\mathbf{F} = \begin{pmatrix} \tilde{\tau}_{1,1} & \cdots & \tilde{\tau}_{1,N_l} & \tilde{\tau}_{1,1}^{(\mathrm{Ray})} & \cdots & \tilde{\tau}_{1,N_l}^{(\mathrm{Ray})} & \tilde{q}_{1,1}^{(\mathrm{sca})} & \cdots & \tilde{q}_{1,N_a}^{(\mathrm{sca})} & \tilde{\rho}_1 & \tilde{c}_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \tilde{\tau}_{N_s,1} & \cdots & \tilde{\tau}_{N_s,N_l} & \tilde{\tau}_{N_s,1}^{(\mathrm{Ray})} & \cdots & \tilde{\tau}_{N_s,N_l}^{(\mathrm{Ray})} & \tilde{q}_{N_s,1}^{(\mathrm{sca})} & \cdots & \tilde{q}_{N_s,N_a}^{(\mathrm{sca})} & \tilde{\rho}_{N_s} & \tilde{c}_{N_s} \end{pmatrix}. \tag{A12}$$

Note that contrary to the other properties, c_i can be negative, in which case the transformation into logarithmic space would fail. In this work, we use $f(x) = \ln(x)$ for all elements; however, a constant value of 5 was added to every c_i to avoid the aforementioned issue. This constant value is then subtracted again when calculating the binned optical properties in section A2.

A3. Varying Scattering Properties and Generalization of the Method

An important note shall be made here to the reader. The specific preparation of the aerosol optical properties, described in section A2, is not an integral part of the PCA-based approach itself. In this work, it was chosen to comply with the calculations within the UoL retrieval scheme. For applications in which the varying extinction and scattering properties of aerosols are ignored, the quantities $q^{(\mathrm{sca})}$, $q^{(\mathrm{ext})}$, and c can be omitted in equation (A12). In this case, there is no need to reconstruct the phase function expansion coefficients,

since they are the same for each spectral point. Similarly, if the optical properties depend on additional quantities not mentioned here, then these quantities must be incorporated in **O** in the same way. If the coefficients are not parameterized or interpolated in a similar way as shown here, but rather calculated individually for every wavelength, then the EOF method would need to be reviewed, as the decomposition in the following section potentially proves difficult if N_{mom} is of the order of several hundred. On the other hand, for applications in which N_{mom} is considerably smaller, it is conceivable to directly ingest them into the optical property matrix **O**, thus allowing for a direct reconstruction of $\beta_{j,m}$ without any auxiliary parameters.

Neglecting the wavelength dependence of the scattering properties, if it is part of the (forward) model, will ultimately lead to an overall slope (with wavelength) in the residuals, as seen in Natraj et al. (2005). Contrary to other approaches, such as in O'Dell (2010), no further RT computations are needed to correct for the varying scattering properties. The method can potentially be generalized for other RT models. LIDORT (Spurr, 2008), 2S-ESS (Spurr & Natraj, 2011), and 2OS (Natraj & Spurr, 2007) are models that ingest total atmospheric optical properties (τ, ω, β). Therefore, **O** is set up such that it contains all quantities to exactly calculate the inputs needed by the RT models; no information loss has occurred yet.

It was demonstrated in Efremenko et al. (2014) that the PCA approach can be generalized to other dimensionality reduction techniques, such as local linear embedding methods or discrete linear transforms. They showed for a single example that higher accuracies can potentially be achieved by employing, for example, local linear embedding rather than PCA. As such there is the potential of improving on the results shown in this paper.

A4. Calculation of the Empirical Orthogonal Functions

The principal components are calculated based on the mean-removed covariance matrix **C**, which is constructed by first subtracting the spectral mean of each optical property from **F**, such that

$$\bar{F}_{i,l} = F_{i,l} - \frac{1}{N_s}\sum_{i=1}^{N_s} F_{i,l}. \tag{A13}$$

The covariance matrix **C** is then given by

$$\mathbf{C} = \frac{1}{N_s - 1}\left(\bar{\mathbf{F}}^T\bar{\mathbf{F}}\right). \tag{A14}$$

Note the normalization factor in equation (A14), which was not present in earlier publications. We solve the eigenproblem for the $N_o \times N_o$ matrix **C**:

$$\mathbf{C}\,\mathbf{V}_k = \eta_k \mathbf{V}_k, \tag{A15}$$

with eigenvalues η_k that correspond to eigenvectors $\mathbf{V}_k$. The eigenvectors **V**, also called empirical orthogonal functions (EOFs), are scaled via the square root of their corresponding eigenvalues,

$$\mathbf{W}_k = \sqrt{\eta_k}\,\mathbf{V}_k, \tag{A16}$$

and the PCs **P** are obtained by projecting $\bar{\mathbf{F}}$ onto the EOFs:

$$\mathbf{P}_k = \frac{1}{\eta_k}\bar{\mathbf{F}}\,\mathbf{W}_k. \tag{A17}$$

In practice, since **C** is real and symmetric, we can use the LAPACK solver `DSYEV` or appropriate wrappers, such as NumPy's `eigh` (van der Walt et al., 2011), to compute all $\mathbf{V}_k$ and η_k. In terms of efficiency, the LAPACK routine is able to solve the eigenproblem for a 200×200 matrix in less than a second (Anderson et al., 1992); we can, therefore, neglect its impact on the overall computational effort.

The η_k are subsequently ranked by decreasing values, such that $\mathbf{V}_1$ is the EOF that explains the largest amount of variability of **F**. Generally, we find that the cumulative explained variance for the first three EOFs tends to be >99%. As stated in Natraj et al. (2005), this implies that three to four EOFs are sufficient to reconstruct the optical properties for every spectral point in the bin to high accuracy.

Due to the symmetry of **C**, all Eigenvalues are positive and real. Eigensolvers such as the aforementioned `DSYEV`, however, might return some very small negative Eigenvalues for larger k. Since the magnitude

of these eigenvalues is small and their contribution can be neglected, we set all eigenvalues and thus all corresponding scaled and unscaled eigenvectors and PCs to zero if η_k is sufficiently small:

$$|\eta_k| < 10^{-15} \Rightarrow \mathbf{P}_{i,k} = \mathbf{W}_{l,k} = \mathbf{V}_{l,k} = \eta_k = 0, \ \forall\, i, l. \tag{A18}$$

The eigenvalues satisfying the above condition are usually of high order, such that they would not have been relevant at all for the reconstruction, which is often done using less than ~10 principal components.

A5. Preparation of Binned Optical Properties

For every bin, a so-called mean optical state, $\mathbf{O}^{(0)}$ is created by transforming the spectrally averaged $\mathbf{F}$ back into its original space:

$$O_l^{(0)} = f^B\left(\frac{1}{N_s}\sum_{i=1}^{N_s} F_{i,l}\right). \tag{A19}$$

Constructing the total optical depths for the mean state is straight forward, since $\tau_l^{(0)} = O_l^{(0)}$ with $l = 1 \ldots N_l$ (see equation (A12). The other mean properties, such as mean surface albedo, mean Rayleigh optical depth, and the mean auxiliary properties $c^{(0)}$ and $q_a^{(\mathrm{sca},0)}$, are obtained analogously. The single-scattering albedos $\omega_j^{(0)}$, however, are a composite quantity in terms of the contents of $\mathbf{O}$, and they have to be calculated from $\tau_j^{(0)}$, $\tau_j^{(\mathrm{Ray},0)}$ and the mean aerosol scattering coefficient $q_a^{(\mathrm{sca},0)}$ (see equation (A11):

$$\omega_j^{(0)} = \frac{\tau_j^{(\mathrm{Ray},0)} + \sum_{a=1}^{N_a} q_a^{(\mathrm{sca},0)} \cdot \tau_{j,a}^{(\mathrm{aer},755)}}{\tau_j^{(0)}}. \tag{A20}$$

The final quantities needed for the binned calculations are the composite phase function expansion coefficients $\beta_{j,m}^{(0)}$. As explained in section A2, the phase function expansion coefficients related to aerosols $\beta_{j,m}^{(\mathrm{aer})}$ are considered to be wavelength dependent and are calculated for two points near the edges of the spectral band. Through linear interpolation between the given points, the coefficients are obtained for every spectral point in the band. For the mean optical state, we use the mean interpolation factor $c^{(0)}$ to calculate the $\beta_{j,m,a}^{(\mathrm{aer},0)}$ via

$$\beta_{j,m,a}^{(\mathrm{aer},0)} = \beta_{j,m,a}^{(\mathrm{aer,beg})} \cdot (1 - c^{(0)}) + \beta_{j,m,a}^{(\mathrm{aer,end})} \cdot c^{(0)}. \tag{A21}$$

The mean composite phase function expansion coefficients can then be calculated as follows:

$$\beta_{j,m}^{(0)} = \frac{\tau_j^{(\mathrm{Ray},0)} \cdot \beta_{m,j}^{(\mathrm{Ray})} + \sum_{a=1}^{N_a} \beta_{j,m,a}^{(\mathrm{aer},0)} \cdot q_a^{(\mathrm{sca},0)} \cdot \tau_{a,j}^{(\mathrm{aer},755)}}{\tau_j^{(\mathrm{Ray},0)} + q_a^{(\mathrm{sca},0)} \cdot \tau_{a,j}^{(\mathrm{aer},755)}}. \tag{A22}$$

With that, the total mean optical state is fully defined. The binned optical quantities for the perturbed states $+k$ and $-k$ are computed by perturbing the mean optical properties by the scaled EOFs W:

$$O_l^{(\pm k)} = f^B\left(\left[\frac{1}{N_s}\sum_{i=1}^{N_s} F_{i,l}\right] \pm W_{k,l}\right). \tag{A23}$$

From $O_l^{(\pm k)}$, the perturbed properties, $\tau_j^{(\pm k)}$, $\tau_j^{(\mathrm{Ray},\pm k)}$, $q_a^{(\mathrm{sca},\pm k)}$, $\rho^{(\pm k)}$, and $c^{(\pm k)}$, are obtained as before, and the composite quantities $\omega_j^{(\pm k)}$ and $\beta_{j,m}^{(\pm k)}$ are calculated analogously to equations (A20)–(A22).

A6. Binned Calculations and Intensity Reconstruction

For each bin, $2N_{\mathrm{EOF}} + 1$ low-stream and high-stream calculations are performed using LIDORT (I_{LD}), 2S-ESS (I_{2S}) and FO (I_{FO}) and the optical inputs $\mathbf{O}^{(0)}$ and $\mathbf{O}^{(\pm k)}$ that were prepared in section A2. Further optimization can be achieved by choosing the number of EOFs individually for each bin, which could reduce the total number of binned calculations.

We use the notation in Spurr et al. (2013) and define the differences between high- and low-stream binned calculations in logarithmic space as

$$J^{(0)} = \ln\left[\frac{I_{\mathrm{LD}}(\mathbf{O}^{(0)}) + I_{\mathrm{FO}}(\mathbf{O}^{(0)})}{I_{2S}(\mathbf{O}^{(0)}) + I_{\mathrm{FO}}(\mathbf{O}^{(0)})}\right] \tag{A24}$$

Table A1
Instrument Model Coefficients

Coefficient	O_2 A band	Weak CO_2	Strong CO_2
	Instrument A (OCO-2 like)		
a	4.9×10^{37}	6.0025×10^{36}	1.5625×10^{36}
b	7.0×10^{18}	2.45×10^{18}	1.25×10^{18}
C_{back}	0.00497	0.00671	0.0149
C_{ph}	0.00961	0.00706	0.008
	Instrument B (CarbonSat like)		
SNR_{ref}	150	160	130
L_{ref}	4.2×10^{19}	1.5×10^{19}	3.8×10^{18}
	Instrument C (Sentinel 5 like)		
SNR_{ref}	500	300	100
L_{ref}	4.49×10^{19}	1.15×10^{19}	5.0×10^{18}

Note. Radiances (L, L_{ref}) in ph/s/m^2/sr/µm, C_{back}, C_{ph}, and SNR_{ref} are unitless; a in (ph/s/m^2/sr/µm)^2 and b in ph/s/m^2/sr/µm.

and

$$J^{(\pm k)} = \ln\left[\frac{I_{LD}(\mathbf{O}^{(\pm k)}) + I_{FO}(\mathbf{O}^{(\pm k)})}{I_{2S}(\mathbf{O}^{(\pm k)}) + I_{FO}(\mathbf{O}^{(\pm k)})}\right] \tag{A25}$$

for the mean and perturbed optical states, respectively. Still in logarithmic space, the second-order central difference expansion is then applied, where the PCs transform the result back into the spectral space:

$$C_i = \exp(J_i) \tag{A26}$$

with

$$J_i = J^{(0)} + \sum_{k=1}^{N_{EOF}} \frac{J^{(+k)} - J^{(-k)}}{2} P_{k,i} + \sum_{k=1}^{N_{EOF}} \frac{J^{(+k)} - 2J^{(0)} + J^{(-k)}}{2} P_{k,i}^2. \tag{A27}$$

The correction factors C_i are then used to reconstruct the approximate radiance for each spectral point:

$$I_{approx,i} = \left[I_{FO,i} + I_{2S,i}\right] \cdot C_i. \tag{A28}$$

Here we correct for the entire radiance field, which is referred to as the 3 M mode, rather than correcting the MS field only (2 M mode).

A7. Instrument Noise Models

For section 4.4, we use various noise models to simulate different instruments in order to assess the impact of noise and spectral resolution on the reconstruction accuracy. The noise model for instrument A (OCO-2 like) is a simplified version of the one stated in Eldering et al. (2015):

$$SNR_A = \sqrt{\frac{L^2}{aC_{back}^2 + bC_{ph}^2 L}}, \tag{A29}$$

with L being the radiance value in units of ph/s/m^2/sr/µm. While in Eldering et al. (2015), the coefficients C_{back} and C_{ph} are detector pixel dependent, we have substituted them with a constant value for each band. Similarly, for the CarbonSat-like instrument B, we use the following noise model (Buchwitz et al., 2010),

$$SNR_B = \sqrt{3}\,\frac{SNR_{ref} \cdot \frac{L}{L_{ref}}}{\sqrt{2 + \frac{L}{L_{ref}}}}. \tag{A30}$$

Finally, instrument C was modeled after Sentinel-5-like specifications, with the noise model

$$SNR_C = SNR_{ref}\sqrt{\frac{L}{L_{ref}}}. \tag{A31}$$

The coefficients used in the above noise models are summarized in Table A1.

Acronyms

AOD Aerosol optical depth
CAMS Copernicus Atmosphere Monitoring Service
ECMWF European Centre for Medium-Range Weather Forecasts
EOF Empirical orthogonal function
GOSAT Greenhouse Gases Observing Satellite
IQR Interquartile range
LSI Low-streams interpolation
OD Optical depth
PC Principal component
PCA Principal component analysis
RMS Root-mean-square
SNR Signal-to-noise ratio

SZA Solar zenith angle
TOA Top-of-the-atmosphere

Notation

i Spectral index ($1 \ldots N_s$).
j Layer index ($1 \ldots N_l$).
a Aerosol mixture index ($1 \ldots N_a$).
m Phase function expansion coefficient index ($1 \ldots N_{\text{mom}}$).
l Free index within optical property matrix **F** or **O** ($1 \ldots N_o$).
k EOF, eigenvalue or principal component index. Since the decomposition is based on the covariance matrix **C** with dimension $N_o \times N_o$, this index can vary between $1 \ldots N_o$. For the approximation, in, for example,Equation (A27), the index varies between 1 and the number of EOFs used for the reconstruction, N_{EOF}.

Acknowledgments
P. Somkuti is funded by the ESA Greenhouse Gas Climate Change Initiative (GHG-CCI) as part of a PhD studentship. H. Boesch receives funding by the UK National Centre for Earth Observation (NCEO) as well as GHG-CCI. V. Natraj was supported by the Orbiting Carbon Observatory (OCO-2) Project at the Jet Propulsion Laboratory, California Institute of Technology. P. Kopparla was supported in part by the NASA NNX13AK34G grant to the California Institute of Technology and the OCO-2 Project at the Jet Propulsion Laboratory. We thank the Japanese Aerospace Exploration Agency, National Institute for Environmental Studies, and the Ministry of Environment for the GOSAT data and their continuous support as part of the Joint Research Agreement. GOSAT L1B data are available from the GOSAT data archive service (GDAS, https://data2.gosat.nies.go.jp/index_en.html). ECMWF ERA-Interim data are available through the ECMWF website (http://apps.ecmwf.int/datasets/). THE ESA-CCI land cover classification data can be obtained from https://www.esa-landcover-cci.org/. The radiance residuals as well as the associated XCO_2 errors are available as HDF5 files from http://www.leos.le.ac.uk/data/GHG/GOSAT/pca_method/. The authors would like to thank R. Parker for helpful comments and editing, as well as L. Vogel for assistance regarding the aerosol scheme. We thank the two anonymous reviewers as well as the editor for their feedback and helping to improve the quality of this publication. This research used the ALICE/SPECTRE High Performance Computing Facility at the University of Leicester.

References

Aben, I., Hasekamp, O., & Hartmann, W. (2007). Uncertainties in the space-based measurements of CO_2 columns due to scattering in the Earth's atmosphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *104*(3), 450–459. https://doi.org/10.1016/j.jqsrt.2006.09.013

Anderson, E., Bai, Z., Bischof, C., Demmel, J., Dongarra, J., Croz, J. D., … Sorensen, D. (1992). *Lapack users' guide* (3rd ed.). Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Benner, D. C., Devi, V. M., Sung, K., Brown, L. R., Miller, C. E., Payne, V. H., … Gamache, R. R. (2015). Line parameters including temperature dependences of air- and self-broadened line shapes of $^{12}C_{16}O_2$: 2.06-μm region. *Journal of Molecular Spectroscopy*, *326*, 21–47. https://doi.org/10.1016/j.jms.2016.02.012

Boesch, H., Deutscher, N. M., Warneke, T., Byckling, K., Cogan, A. J., Griffith, D. W. T., … Wang, Z. (2013). HDO/H_2O ratio retrievals from GOSAT. *Atmospheric Measurement Techniques*, *6*(3), 599–612. https://doi.org/10.5194/amt-6-599-2013

Bucholtz, A. (1995). Rayleigh-scattering calculations for the terrestrial atmosphere. *Applied Optics*, *34*(15), 2765–2773. https://doi.org/10.1364/ao.37.000428

Buchwitz, M., Bovensmann, H., Reuter, M., Gerilowski, K., & Burrows, J. P. (2010). Carbon Monitoring Satellite (CarbonSat) Mission Requirements Document (MRD), Institute of Environmental Physics (IUP). Bremen, Germany: University of Bremen. version 3.2.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., … Löscher, A. (2013). Carbon Monitoring Satellite (CarbonSat): Assessment of atmospheric CO_2 and CH_4 retrieval errors by error parameterization. *Atmospheric Measurement Techniques*, *6*(12), 3477–3500. https://doi.org/10.5194/amt-6-3477-2013

Chevallier, F., Bréon, F. M., & Rayner, P. J. (2007). Contribution of the orbiting carbon observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework. *Journal of Geophysical Research*, *112*, D09307. https://doi.org/10.1029/2006JD007375

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J. F. L., … Wunch, D. (2012). Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations. *Journal of Geophysical Research*, *117*, D21301. https://doi.org/10.1029/2012JD018087

Connor, B., Boesch, H., McDuffie, J., Taylor, T., Fu, D., Frankenberg, C., … Jiang, Y. (2016). Quantification of uncertainties in OCO-2 measurements of XCO_2: Simulations and linear error analysis. *Atmospheric Measurement Techniques*, *9*(10), 5227–5238. https://doi.org/10.5194/amt-9-5227-2016

Cox, C., & Munk, W. (1954). Measurement of the roughness of the sea surface from photographs of the Sun's glitter. *Journal of the Optical Society of America*, *44*(11), 838–850. https://doi.org/10.1364/JOSA.44.000838

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., … Wunch, D. (2017). The on-orbit performance of the Orbiting Carbon Observatory-2 (CO-2) instrument and its radiometrically calibrated products. *Atmospheric Measurement Techniques*, *10*(1), 59–81. https://doi.org/10.5194/amt-10-59-2017

Devi, V. M., Benner, D. C., Sung, K., Brown, L. R., Crawford, T. J., Miller, C. E., … Gamache, R. R. (2016). Line parameters including temperature dependences of self- and air-broadened line shapes of 12C16O2: 1.6-μm region. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *177*, 117–144. https://doi.org/10.1016/j.jqsrt.2015.12.020

Drouin, B. J., Benner, D. C., Brown, L. R., Cich, M. J., Crawford, T. J., Devi, V. M., … Yu, S. (2016). Multispectrum analysis of the oxygen A-band. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *186*, 118–138. https://doi.org/10.1016/j.jqsrt.2016.03.037

Efremenko, D., Doicu, A., Loyola, D., & Trautmann, T. (2014). Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *133*, 128–135. https://doi.org/10.1016/j.jqsrt.2013.07.023

Eguchi, N., Yokota, T., & Inoue, G. (2007). Characteristics of cirrus clouds from ICESat/GLAS observations. *Geophysical Research Letters*, *34*, L09810. https://doi.org/10.1029/2007GL029529

Eldering, A., Pollock, R., Lee, R., Rosenberg, R., Oyafuso, F., Crisp, D., … Granat, R. (2015). OCO-2 level 1B algorithm theoretical basis, National Aeronautics and Space Administration. Pasadena, CA: Jet Propulsion Laboratory California Institute of Technology.

European Space Agency (2011). GMES Sentinels 4 and 5 mission requirements document, GMES, Global Monitoring for Environment and Security, Sentinel. EOP-SMA/1507/JL-dr1. Noordwijk.

European Space Agency (2012). Requirements for CO_2 monitoring by Sentinel-5, Final Report of ESA study contract 4000103801. Noordwijk: European Space Agency (ESA). Retrieved from: http://projects.knmi.nl/capacity/

European Space Agency (2016). CCI Land Cover Product User Guide Version 2.6. Retrieved from: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Hasekamp, O. P., & Butz, A. (2008). Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres. *Journal of Geophysical Research*, *113*, D20309. https://doi.org/10.1029/2008JD010379

Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., & Courrèges-Lacoste, G. B. (2012). Requirements for the GMES atmosphere service and ESA's implementation concept: Sentinels-4/-5 and -5p. *Remote Sensing of Environment*, *120*, 58–69. https://doi.org/10.1016/j.rse.2012.01.023

Kopparla, P., Natraj, V., Limpasuvan, D., Spurr, R., Crisp, D., Shia, R.-L., … Yung, Y. (2017). PCA-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning.

Kopparla, P., Natraj, V., Spurr, R., Shia, R.-L., Crisp, D., & Yung, Y. L. (2016). A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *173*, 65–71. https://doi.org/10.1016/j.jqsrt.2016.01.014

Kuze, A., Suto, H., Shiomi, K., Kawakami, S., Tanaka, M., Ueda, Y., … Buijs, H. L. (2016). Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space. *Atmospheric Measurement Techniques*, *9*(6), 2445–2461. https://doi.org/10.5194/amt-9-2445-2016

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., … Zhang, H. (2013). Anthropogenic and natural radiative forcing. In Intergovernmental Panel on Climate Change (Ed.), *Climate change 2013—The physical science basis* (pp. 659–740). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

Natraj, V., Jiang, X., Shia, R.-L., Huang, X., Margolis, J. S., & Yung, Y. L. (2005). Application of principal component analysis to high spectral resolution radiative transfer: A case study of the band. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *95*(4), 539–556. https://doi.org/10.1016/j.jqsrt.2004.12.024

Natraj, V., Shia, R. L., & Yung, Y. L. (2010). On the use of principal component analysis to speed up radiative transfer calculations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *111*(5), 810–816. https://doi.org/10.1016/j.jqsrt.2009.11.004

Natraj, V., & Spurr, R. J. D. (2007). A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering-absorbing media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *107*(2), 263–293. https://doi.org/10.1016/j.jqsrt.2007.02.011

Natraj, V., Spurr, R. J. D., Boesch, H., Jiang, Y., & Yung, Y. L. (2007). Evaluation of errors from neglecting polarization in the forward modeling of O_2 A-band measurements from space, with relevance to CO_2 column retrieval from polarization-sensitive instruments. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *103*(2), 245–259. https://doi.org/10.1016/j.jqsrt.2006.02.073

Nelson, R. R., O'Dell, C. W., Taylor, T. E., Mandrake, L., & Smyth, M. (2016). The potential of clear-sky carbon dioxide satellite retrievals. *Atmospheric Measurement Techniques*, *9*(4), 1671–1684. https://doi.org/10.5194/amt-9-1671-2016

O'Dell, C. W. (2010). Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. *Journal of Geophysical Research*, *115*, 1–15. https://doi.org/10.1029/2009JD012803

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., … Wunch, D. (2012). The ACOS CO_2 retrieval algorithm—Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, *5*(1), 99–121. https://doi.org/10.5194/amt-5-99-2012

Rodgers, C. D. (2000). Inverse methods for atmospheric sounding—Theory and practice. In C. D. Rodgers (Ed.), *Series on atmospheric oceanic and planetary physics* (Vol. 2). Oxford: World Scientific Publishing Co. Pte. Ltd. https://doi.org/10.1142/9789812813718

Rodgers, C. D. (2003). Intercomparison of remote sounding instruments. *Journal of Geophysical Research*, *108*, 46–48. https://doi.org/10.1029/2002JD002299

Spurr, R. (2008). LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, *Light Scattering Reviews 3* (pp. 229–275). Berlin: Springer.

Spurr, R., & Natraj, V. (2011). A linearized two-stream radiative transfer code for fast approximation of multiple-scatter fields. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *112*(16), 2630–2637. https://doi.org/10.1016/j.jqsrt.2011.06.014

Spurr, R., Natraj, V., Lerot, C., Van Roozendael, M., & Loyola, D. (2013). Linearization of the principal component analysis method for radiative transfer acceleration: Application to retrieval algorithms and sensitivity studies. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *125*, 1–17. https://doi.org/10.1016/j.jqsrt.2013.04.002

Stamnes, K., Tsay, S. C., Wiscombe, W., & Jayaweera, K. (1988). Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, *27*(12), 2502–2509. https://doi.org/10.1364/AO.27.002502

Taylor, T. E., O'Dell, C. W., Frankenberg, C., Partain, P. T., Cronk, H. Q., Savtchenko, A., … Gunson, M. R. (2016). Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: Validation against collocated MODIS and CALIOP data. *Atmospheric Measurement Techniques*, *9*(3), 973–989. https://doi.org/10.5194/amt-9-973-2016

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engeneering*, *13*, 22–30. https://doi.org/10.1109/MCSE.2011.37