Journal of Quantitative Spectroscopy & Radiative Transfer 173 (2016) 65–71

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region

Pushkar Kopparla [a,*], Vijay Natraj [b], Robert Spurr [c], Run-Lie Shia [a], David Crisp [b], Yuk L. Yung [a]

[a] Division of Geological and Planetary Sciences, MC 150-21, California Institute of Technology, Pasadena, CA 91125, USA
[b] Jet Propulsion Laboratory (NASA-JPL), 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[c] RT Solutions, Inc., 9 Channing St, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form
4 December 2015
Accepted 10 January 2016
Available online 16 January 2016

Keywords:
Radiative transfer
Principal component analysis
Performance enhancement
Shortwave broadband
Visible & infrared

ABSTRACT

Accurate radiative transfer (RT) calculations are necessary for many earth–atmosphere applications, from remote sensing retrieval to climate modeling. A Principal Component Analysis (PCA)-based spectral binning method has been shown to provide an order of magnitude increase in computational speed while maintaining an overall accuracy of 0.01% (compared to line-by-line calculations) over narrow spectral bands. In this paper, we have extended the PCA method for RT calculations over the entire shortwave region of the spectrum from 0.3 to 3 microns. The region is divided into 33 spectral fields covering all major gas absorption regimes. We find that the RT performance runtimes are shorter by factors between 10 and 100, while root mean square errors are of order 0.01%.

1. Introduction

The increasing complexity of climate models is a function of the need to include more physical processes and reduce dependence on parameterizations; this is particularly the case for direct and indirect radiative forcing due to clouds and aerosols. Accurate line-by-line (LBL) 1-D multiple scattering (MS) RT models are able to capture the full physics of RT in the shortwave (0.3–1 μm) with simulated radiances showing excellent agreement with observations under cloud free conditions [1]. However such models are computationally very expensive. Therefore, climate models are forced to use fast, approximate RT codes that provide speed as a tradeoff for accuracy. Oreopoulos et al. [2] made an effort to characterize these inaccuracies under the Continual Intercomparison of Radiation Codes program, and they found that the 13 solar RT codes examined tended to overestimate the transmitted and reflected radiances, with large errors ($>10\%$) in the shortwave fluxes at the surface. Another RT model intercomparison experiment that compared 31 RT codes being used in global models, the AeroCom [3], reported that approximate models had biases ranging from -10 to 20% compared to values from LBL models at the top of the atmosphere.

The other typical approaches to dealing with computationally expensive RT in climate modeling are either to update radiative heating calculations only once every few dynamical timesteps or to perform RT on a coarser spatial grid. These approaches are not accurate; they tend to adversely impact model output quite strongly in the presence of clouds and have been shown to affect model climate sensitivity [4]. Pauluis and Emanuel [5] showed that the infrequent RT calculations introduced numerical instabilities; these authors provide a criterion (analogous to the Courant–Friedrichs–Levy criterion) for the maximum allowable time step between two RT calculations.

* Corresponding author.
E-mail address: pkk@gps.caltech.edu (P. Kopparla).

http://dx.doi.org/10.1016/j.jqsrt.2016.01.014
0022-4073/

They also suggest approximate numerical corrections if this criterion is violated due to computational limitations. However, these issues can be avoided if a sufficiently fast and accurate RT model is available.

There have been many approaches to enhance the performance of RT computations, the most widely used being the correlated-k model [6–8]. Other approaches include spectral mapping [9,10], low-stream interpolations [11,12] and low orders of scattering approximations [13]. The idea of using Principal Component Analysis (PCA) for RT was first proposed by Natraj et al. [14] (and later independently by Liu et al. [15]), who reproduced the TOA reflectance spectrum over a small spectral region centered on the O_2 A band at 0.75 μm to accuracies of 0.3% compared to a LBL RT model, while achieving a 10-fold increase in speed. Further development of the PCA model, its expansion to broader spectral regions and the derivation of analytic Jacobians are documented in detail in other papers [16,17].

This paper is organized as follows. In Section 2 we describe the concept of the PCA RT model and its setup. In Section 3, we present runtime and error statistics for TOA radiance calculations, as well as scalings of these parameters for various model settings. In Section 4, we discuss known issues and directions for future model development.

2. The PCA RT technique

2.1. RT models

The PCA RT technique is based on two contrasting RT models. First, for accurate MS calculations we use the LIDORT discrete ordinate RT model [18] which includes the treatment of solar-beam incoming attenuation in a spherically curved atmosphere (the pseudo-spherical approximation), and the use of the delta-M scaling approximation for aerosol scattering with sharply peaked forward scattering. We run LIDORT in MS mode with 32 streams (computational quadrature angles; 16 each for upwelling and downwelling polar directions); the single scatter (SS) contribution is not included in the calculations. A full set of LIDORT MS calculations at every wavelength point is computationally expensive; the key to the PCA RT approach is to drastically limit the number of such full-MS calculations to a reduced set of PCA-determined optical profiles that capture the vast majority of optical information for a given wavelength range.

Secondly, fast RT computations are done using a numerically efficient two-stream-exact single scattering (2S-ESS) RT model [19], which comprises two parts. The 2S part is a fast MS calculation based on a single discrete ordinate in each of the up-welling and down-welling directions; the RT calculation is done analytically except for the multi-layer boundary-value problem (which is also solved using a simple and fast pentadiagonal solver rather than typical matrix inversion techniques). The 2S calculation also uses the pseudo-spherical approximation. The ESS part is an accurate spherical-geometry calculation of the singly scattered radiation computed with the complete scattering phase function (not a truncated form based on a limited number of Legendre polynomial expansion coefficients). The 2S-ESS combination is the "fast" RT calculation; in this, the use of the ESS calculation mitigates bias due to the severe phase function truncation inherent in the 2S approximation to MS. The LIDORT-ESS combination provides the most accurate computation of the complete (SS+MS) radiation field; in the following, we will refer to this benchmark computation as the "Exact RT" calculation, against which the accuracy of the PCA RT model is to be compared.

2.2. Formalism

We will briefly summarize the formalism for the PCA RT model here, as the technique continues to evolve. Earlier versions of the procedure can be found in Refs. [14,16,17]. The first step is to partition a given spectral range into a number of appropriately chosen bins, each bin containing spectral points for which the optical properties are broadly "similar". Binning criteria are often determined by similarities in total atmospheric optical depth; we discuss the criteria in Section 2.3. Each bin, which contains a subset of the optical property data, is then subjected to an independent PCA procedure as described below.

Consider a bin with N wavenumbers in an atmosphere stratified into M optically homogeneous layers. Let $\tau_{i,j}$ be the total optical depth in layer i at wavenumber j and $\omega_{i,j}$ be the layer single scattering albedo. The index i goes from 1 to M and j from 1 to N. Let us denote the logarithm of this set of optical properties by $F_{k,j}$ where $F_{[1,M],j}$ are the optical depths and $F_{[M+1,2M],j}$ are the single scattering albedos and the index k goes from 1 to $2M$. The mean-removed covariance matrix C over all layers has elements given by (k and l are layer indices)

$$C_{k,l} = \overline{(F_{k,j} - \overline{F_k})(F_{l,j} - \overline{F_l})} \tag{1}$$

where the overbar denotes a mean-value over all wavenumbers in the bin. The index l also goes from 1 to $2M$. The $2M$ eigenvectors of this covariance matrix form an orthogonal set that is mutually uncorrelated – this is the set of empirical orthogonal functions (ϵ_k). Each EOF has $2M$ components and unit EOFs are scaled by the square root of the variance, making them dimensionally consistent with the corresponding optical properties. This orthogonal set represents a coordinate system onto which the original optical property data set can be mapped. The projection of the original optical properties onto the EOFs gives a set of weights called the principal component scores (PCs, P_k). Again, there are $2M$ PCs and each PC has N components, corresponding to the number of wavenumbers. Thus, we can express all the original information in our new coordinate system as

$$\ln \tau_{k,j} = \ln \overline{\tau_k} + \sum_{i=1}^{M} P_{i,j}\epsilon_{i,k} \tag{2}$$

$$\ln \omega_{k,j} = \ln \overline{\omega_k} + \sum_{i=M+1}^{2M} P_{i,j}\epsilon_{i,k} \tag{3}$$

Wavenumber information is stored in the PCs, while the

vertical profile is stored in the EOFs. In essence, by choosing wavenumbers with closely related optical properties such as optical depth profiles, it is possible to capture most of the relevant variation in optical properties with a small number of EOFs, usually four or less. Thus, starting with a dataset of dimension $2M \times N$, we have reduced it to a total of $4 \times N$ (PCs) and $4 \times 2M$ (EOFs). The basis for the effectiveness of the PCA technique in RT is that despite the fact that the RT calculations by themselves are highly non-linear, the differences of the 2-Stream (2S) model compared to the multiple stream (LIDORT) model are much more linear. This is even more so if the SS contribution is computed exactly and the 2S and LIDORT models are used only to approximate the multiple scattering. In effect, the variation in the 2S model errors traces the variation in the optical properties. With PCA condensing most of the variation in the optical properties into a small number of EOFs, we can calculate the error (defined as the difference between the 2S and Exact Model radiance or flux) at any one wavenumber and use the principal components to estimate the errors at all other wavenumbers in that bin.

First, we calculate the MS radiances for the mean (over all wavenumbers) optical property profiles with a fully-accurate LIDORT calculation (set up with 32 discrete ordinates) followed by the 2S calculation. Let us define a quantity I_d defined as

$$I_d = \ln(I_{exact}/I_{2S}) \tag{4}$$

The mean optical properties are then perturbed by the magnitude of one scaled EOF in both the positive and negative directions away from the mean and the two MS RT models are run again with these perturbed profiles. Let the new logarithmic ratios be I_d^+ and I_d^-. With respect to this EOF, the first and second order differences are calculated as

$$\delta I_k = \frac{I_d^+ - I_d^-}{2} \tag{5}$$

$$\delta^2 I_k = I_d^+ + I_d^- - 2I_d \tag{6}$$

The procedure is repeated for as many EOFs as necessary to obtain a given level of accuracy (for example, four EOFs are usually enough for radiance accuracies of 0.1% across the whole bin – discussed below). The corrected 2S radiance is thus calculated as

$$I_l = I_l^{2S} \exp\left[I_d + \sum_{k=1}^{4} \delta I_k P_{k,l} + \frac{1}{2}\sum_{k=1}^{4} \delta^2 I_k P_{k,l}^2\right] \tag{7}$$

where l is the wavenumber index.

2.3. Model setup and binning considerations

Wavenumbers in each bin must be chosen such that the optical properties are strongly correlated. This is necessary to reduce the number of EOFs required to attain a given accuracy. If the wavenumbers have optical properties that are very different from each other, we can still capture the variability with a larger number of EOFs, but this has a higher computational cost and defeats the purpose of using PCA.

The entire shortwave region from 0.29 to 3 μm is first divided into 33 spectral "fields" depending on dominant gas absorptions within the field (Table 1). Over each field,

Table 1
Spectral fields as defined by major gas absorptions.

Spectral field	Wavelength range (μm)	No. of spectral points	Gas absorbers
1	0.290–0.340	1001	O_3,NO_2,SO_2, HCHO,BrO
2	0.340–0.400	1200	O_4,NO_2,HCHO, BrO
3	0.400–0.500	2000	NO_2,CHOCHO
4	0.500–0.585	1700	O_3,H_2O,NO_2,O_4
5	0.585–0.605	400	O_3,H_2O,NO_2,O_4
6	0.605–0.630	500	O_3,H_2O,NO_2,O_4
7	0.630–0.650	398	O_3,H_2O,NO_2,O_4
8	0.650–0.681	14,001	H_2O
9	0.681–0.715	14,000	H_2O,O_2
10	0.715–0.752	14,000	H_2O,O_2
11	0.752–0.794	14,000	H_2O,O_2
12	0.794–0.841	14,000	H_2O,O_2
13	0.841–0.894	14,000	H_2O,O_2
14	0.894–0.954	14,000	H_2O,O_2
15	0.954–1.022	14,000	H_2O,O_2
16	1.022–1.101	14,000	H_2O,O_2
17	1.101–1.205	15,701	H_2O,O_2
18	1.205–1.234	20,000	H_2O,O_2
19	1.234–1.560	33,800	H_2O,O_2,CO_2
20	1.560–1.626	26,000	H_2O,O_2,CO_2
21	1.626–1.695	25,000	H_2O,O_2,CO_2,CH_4
22	1.695–1.923	14,000	H_2O,CO_2,CH_4
23	1.924–2.105	45,000	H_2O,O_2,CO_2
24	2.105–2.128	5000	H_2O,O_2,CO_2,N_2O
25	2.128–2.222	20,000	H_2O,CO_2,CH_4, N_2O
26	2.222–2.247	5000	H_2O,O_2,CH_4
27	2.247–2.299	10,000	H_2O,O_2,CH_4, N_2O
28	2.299–2.410	20,000	H_2O,O_2,CH_4,CO
29	2.410–2.439	5000	H_2O,O_2,CO
30	2.439–2.564	20,000	H_2O,N_2O,CO_2, CH_4
31	2.564–2.632	10,000	H_2O,N_2O,CO_2
32	2.632–2.857	30,000	H_2O,O_2,CO_2
33	2.857–3.030	20,000	H_2O,N_2O,CO_2

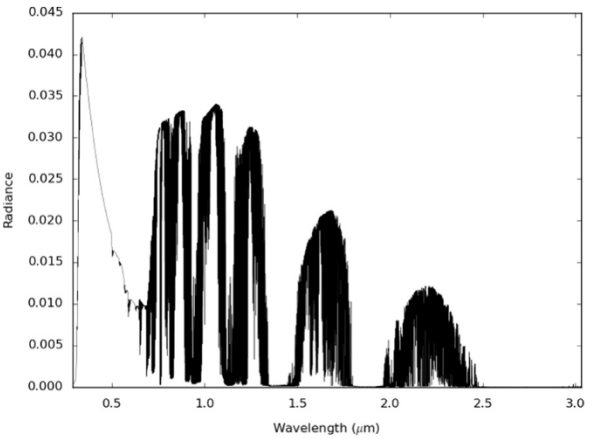

Fig. 1. Top of the Atmosphere (TOA) reflectances from Exact-RT LIDORT for a single profile showing major absorptions. There are 33 spectral fields in all, with 8 in the ultraviolet/visible and the rest in the shortwave infrared (Table 1).

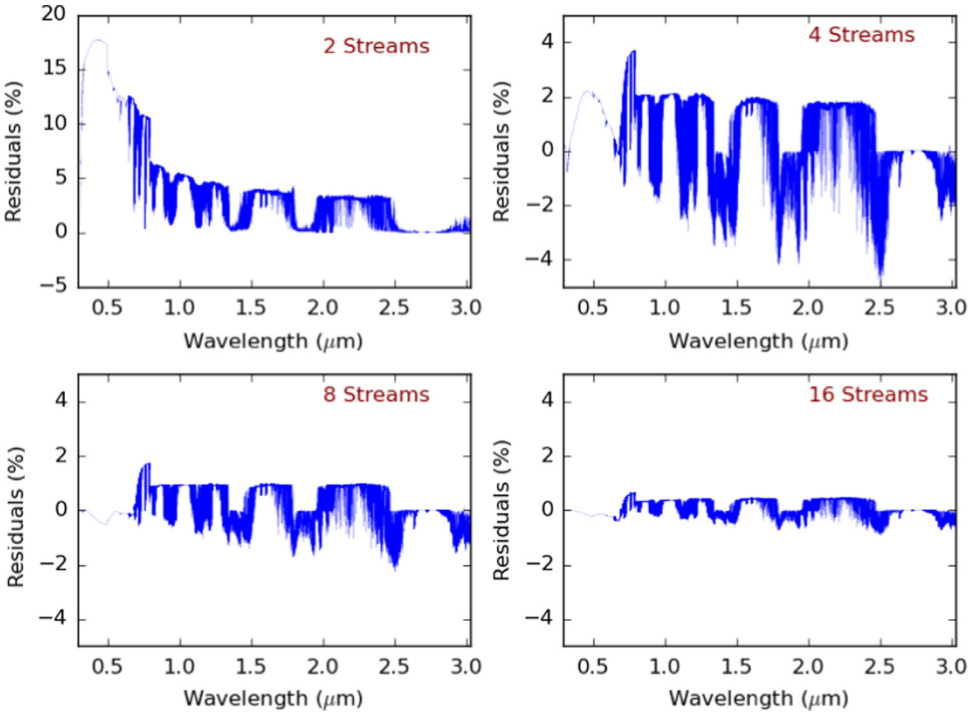

Fig. 2. Residuals (%) of the Exact-RT LIDORT radiances for 2, 4, 8, and 16 streams compared with the 32-streams LIDORT calculations. Optical properties as for Fig. 1.

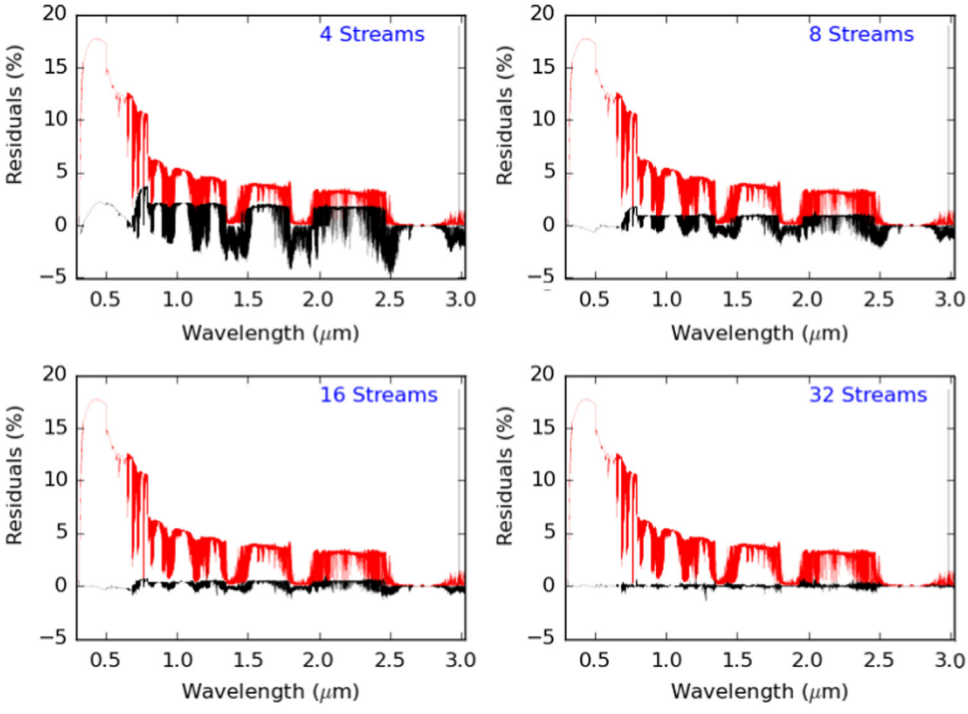

Fig. 3. Residuals (%) for the PCA (black) and 2S (red) models as compared to 32 stream LIDORT over the entire shortwave range. The residuals have been Gaussian smoothed to 0.2 cm^{-1}. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

wavenumbers are grouped into four bins, based on the total column optical depth. We started with the following four bins: (0–0.01), (0.01–0.1), (0.1–1.0) and (1.0–∞). Each of these bins is further bisected by dividing wavenumbers along the median single scattering albedo. Finally, the result is a set of eight wavenumber bins which have the same

absorbers, and somewhat similar total optical depths. The PCA RT model, as described in Sections 2.1 and 2.2, is applied to each spectral field with these initial bins, and the residuals are computed with respect to "Exact Model" using 32-stream LIDORT. Next, bins with large residuals are again bisected with respect to total column optical depth; for instance, the second bin would be split into (0.01–0.05) and (0.05–0.1). This process is repeated until the root mean square residuals are below 0.01% for every spectral field after degrading to typical instrument resolutions (more discussion on this in the results section). We find that, following this procedure, a total of eleven bins provide satisfactory residuals. The upper limits of these bins are $(0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.625, 0.75, 1.0, 5.0, \infty)$. The runtime of the code scales linearly with the number of bins. Note from Table 1 that the number of wavelengths and the spectral resolution vary according to spectral field – finer absorption features require higher resolution.

This was further tested for a wide range of atmospheric profiles taken from GEOS-CHEM, global 3-D chemical transport model [20]. We used profiles representing locations in North America chosen over a wide range of latitudes and longitudes. The patterns in error residuals were found to be similar for these profile data sets. We also find that some degree of iteration was necessary in order to determine an acceptable binning scheme for a given spectral range. Although the binning selection is done manually at present, it will be advantageous to automate this procedure. This will allow for customizing and minimizing the number of bins over different spectral ranges for better efficiency.

3. Results and discussion

We compare radiances and performance run-times for the fast-PCA and exact-LIDORT models across all shortwave spectral fields. The 11-bin scheme with four EOFs is used. Fig. 1 shows the top of the atmosphere radiances calculated using the Exact Model for a single atmospheric profile from the North America GEOS-CHEM profile set. Figs. 2 and 3 show the residuals comparing the "Exact" and "fast-PCA" models. In addition, we also compare residuals and timing for the full LIDORT calculations in which the number of streams is reduced from 32 down to as low as 4. The original resolution of the RT calculations is of the order $0.05\ \mathrm{cm}^{-1}$; we use a Gaussian smoothing procedure to reduce the spectral resolution to $0.2\ \mathrm{cm}^{-1}$, which is a reasonable value for modern instruments such as OCO-2. As expected, increasing the number of streams improves agreement with the Exact Model. Note particularly that the run-time of the Exact-RT LIDORT model scales as the cube of the number of computational streams, increasing about two orders of magnitude when moving from 4 to 32 streams. However, the PCA model shows only about a 20% increase in run-time over the same range of streams. Thus, when we have scattering aerosols and/or clouds with strongly forward-peaked phase functions that require many streams in the RT, the PCA technique offers a cost-effective way to obtain acceptably accurate results in a reasonable amount of time. We checked the performance of this PCA model on a number of other atmospheric profiles, and found that the magnitude and nature of errors compared to the exact model remain fairly consistent. Errors from two other profiles are shown in Fig. 4.

That said, the PCA model residuals still have scope for improvement. Fig. 5 shows PCA residuals with and without Gaussian smoothing over a small spectral field using 32 stream LIDORT. While residuals are of order 0.1% over most wavenumbers, there are a small number of wavenumbers with errors around 1%. We find that these wavenumbers are located either at the cores of weak absorptions or in the wings of strong absorptions, typically with total column optical depths between 0.1 and 10.

Figs. 6 and 7 show the improvements in runtime of the PCA model versus the Exact Model. We find that run-time is reduced by a factor of 10–100 for different spectral fields. The speed of the PCA model is currently limited by the rate at which optical properties are generated ab initio

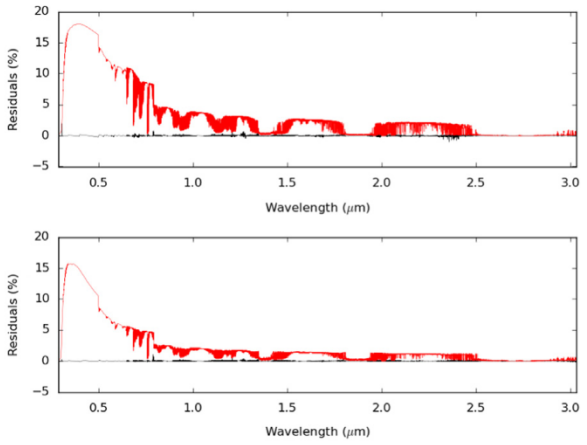

Fig. 4. PCA-RT (black) and 2S (red) residuals against 32-stream LIDORT for two other profiles, as indicated. The general trends in error behavior are the same as in the previous figure. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

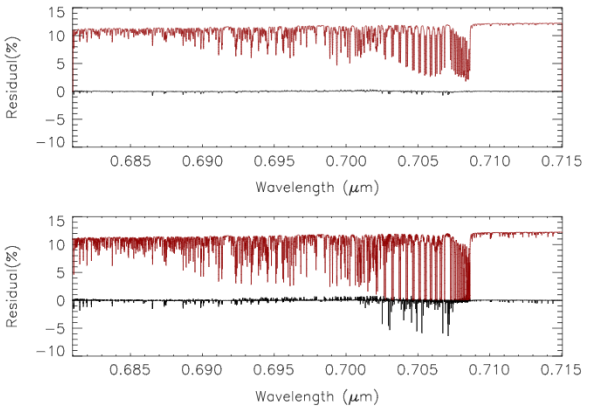

Fig. 5. Residuals of the PCA (black) and 2 stream (red) models as compared to 32 stream LIDORT with (top) and without (bottom) smoothing for spectral field 9, where the dominant gas absorbers are oxygen and water. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

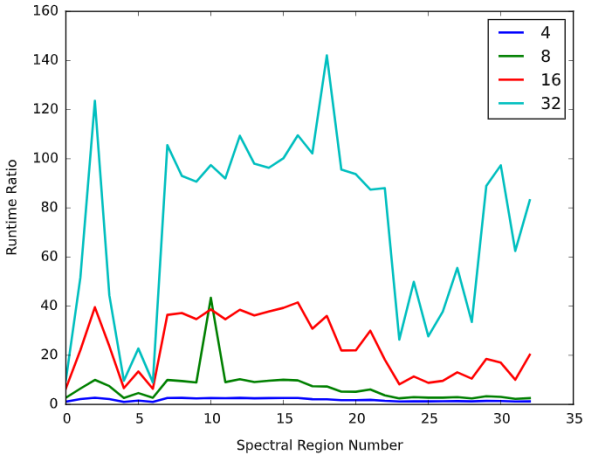

Fig. 6. Ratios of Exact-LIDORT runtimes to those from fast-PC, plotted as a function of spectral field, and for various discrete ordinates as indicated.

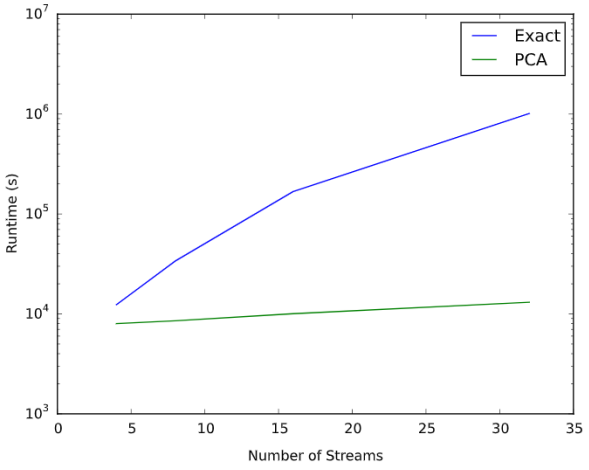

Fig. 7. The sum total of the runtime of LIDORT and PCA model over all spectral fields for one profile with changing number of computational streams. Note from Fig. 3 that increasing the number of streams improves the PCA RT accuracy, with very little time penalty as shown in this figure. Refer to text for further discussion.

from databases, thereby resulting in somewhat arbitrary runtime improvements which are unrelated to the actual RT calculation. Runtime can be further optimized by using tabulated optical property inputs, for instance.

In summary, we have shown that the PCA RT technique is capable of producing accuracy comparable to 32 stream, full LIDORT RT calculations with run-times comparable to those from a 2S model. While the PCA routine works extremely well for $>99.9\%$ of all points on the shortwave spectrum as defined here, a few points continue to show sizeable irreducible errors. Intuitively, these are the wavelengths where the fine details of the gas absorption profile are hard to capture with the PCA. Future work will focus on expanding the PCA-RT technique to the infrared region and to include Jacobians in the analysis and account for polarization. We anticipate that our PCA RT model will enable the application of "full-physics" RT in problems where accuracy and speed could not be simultaneously achieved till now (e.g. climate models or atmospheric retrieval codes).

Acknowledgments

We thank members of Prof. Yung's group for useful comments. We also thank the anonymous referee for insightful suggestions that helped improve the manuscript. This research was supported in part by NASA NNX13AK34G grant to the California Institute of Technology and the Orbiting Carbon Observatory (OCO-2) Project at JPL.

References

[1] Mlawer EJ, Brown PD, Clough SA, Harrison LC, Michalsky JJ, Kiedron PW, et al. Comparison of spectral direct and diffuse solar irradiance measurements and calculations for cloud-free conditions. Geophys Res Lett 2000;27(17):2653–6.
[2] Oreopoulos L, Mlawer E, Delamere J, Shippert T, Cole J, Fomin B, et al., The continual intercomparison of radiation codes: results from phase I. J Geophys Res: Atmos 1984–2012; 117: D6.
[3] Randles C, Kinne S, Myhre G, Schulz M, Stier P, Fischer J, et al. Intercomparison of shortwave radiative transfer schemes in global aerosol modeling: results from the AeroCom radiative transfer experiment. Atmos Chem Phys 2013;13(5):2347–79.
[4] Morcrette J-J. On the effects of the temporal and spatial sampling of radiation fields on the ECMWF forecasts and analyses. Mon Weather Rev 2000;128(3):876–87.
[5] Pauluis O, Emanuel K. Numerical instability resulting from infrequent calculation of radiative heating. Mon Weather Rev 2004;132(3):673–86.
[6] Goody R, West R, Chen L, Crisp D. The correlated-k method for radiation calculations in nonhomogeneous atmospheres. J Quant Spectrosc Radiat Transfer 1989;42(6):539–50.
[7] Lacis AA, Oinas V. A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. J Geophys Res: Atmos (1984–2012) 1991;96(D5):9027–63.
[8] Fu Q, Liou K. On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. J Atmos Sci 1992;49(22):2139–56.
[9] Crisp D. Absorption of sunlight by water vapor in cloudy conditions: a partial explanation for the cloud absorption anomaly. Geophys Res Lett 1997;24(5):571–4.
[10] Moncet J-L, Uymin G, Lipton AE, Snell HE. Infrared radiance modeling by optimal spectral sampling. J Atmos Sci 2008;65(12):3917–34.
[11] Duan M, Min Q, Li J. A fast radiative transfer model for simulating high-resolution absorption bands. J Geophys Res: Atmos 1984–2012; 110: D15.
[12] O'Dell CW. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. J Geophys Res: Atmos 1984–2012; 115: D10.
[13] Natraj V, Spurr RJ. A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering-absorbing media. J Quant Spectrosc Radiat Transfer 2007;107(2):263–93.
[14] Natraj V, Jiang X, Shia R-l, Huang X, Margolis JS, Yung YL. Application of principal component analysis to high spectral resolution radiative transfer: a case study of the O2A band. J Quant Spectrosc Radiat Transfer 2005;95(4):539–56.
[15] Liu X, Smith WL, Zhou DK, Larar A. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. Appl Opt 2006;45(1):201–9 URL http://ao.osa.org/abstract.cfm?URI=ao-45-1-201.
[16] Natraj V, Shia R-L, Yung YL. On the use of principal component analysis to speed up radiative transfer calculations. J Quant Spectrosc Radiat Transfer 2010;111(5):810–6.
[17] Spurr R, Natraj V, Lerot C, Roozendael M Van, Loyola D. Linearization of the principal component analysis method for radiative

transfer acceleration: application to retrieval algorithms and sensitivity studies. J Quant Spectrosc Radiat Transfer 2013;125: 1–17. http://dx.doi.org/10.1016/j.jqsrt.2013.04.002.

[18] Spurr R. Lidort and vlidort: linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In: Light scattering reviews, vol. 3. Berlin, Heidelberg: Springer; 2008. p. 229–75.

[19] Spurr R, Natraj V. A linearized two-stream radiative transfer code for fast approximation of multiple-scatter fields. J Quant Spectrosc Radiat Transfer 2011;112(16):2630–7.

[20] Bey I, Jacob DJ, Yantosca RM, Logan JA, Field BD, Fiore AM, et al., Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation.