TECHNISCHE UNIVERSITÄT MÜNCHEN
TUM School of Engineering and Design

# High Performance Processing Algorithms of Satellite Measurements for Retrieval of Trace Gases and Cloud Properties

**Ana del Águila Pérez**

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

**Doktorin der Ingenieurwissenschaften (Dr.-Ing.)**

genehmigten Dissertation.

**Vorsitz:**

Prof. Dr.-Ing. habil. Florian Seitz

**Prüfer*innen der Dissertation:**

1. Priv.-Doz. Dr. Dmitry Efremenko
2. Prof. Dr.-Ing. habil. Richard Hans Georg Bamler
3. Prof. Dr. rer. nat. Vladimir Budack

Die Dissertation wurde am 21.06.2021 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 30.11.2021 angenommen.

# Abstract

Earth Observation (EO) is decisive to monitor and assess the climate change via remote sensing technologies. The satellite-based sensors help to elucidate the factors affecting the Earth's atmosphere with atmospheric composition sensors (ACS). The latter detect and record the spectral information and then, by means of radiative transfer models (RTMs), the composition of the atmosphere can be retrieved in a subsequent step.

The ultimate generation of ACS have achieved extraordinary spatial, temporal and spectral resolutions. Furthermore, the so called hyperspectral sensors contain a vast amount of high spectral resolution data. The challenge of processing hyperspectral data is twofold: on the one hand, it is crucial to boost the performance of the processing algorithms in order to achieve near-real-time (NRT) requirements. With the aim of obtaining high performance computing (HPC), dimensionality reduction or acceleration techniques are employed. On the other hand, after applying an acceleration technique, the accuracy of the retrieved atmospheric parameters should be maintained over the spectral range. The accurate information at fine spectral resolution is crucial to characterize the spectral signatures and hence, to retrieve the truth atmospheric composition.

Through an in-depth investigation of the existing acceleration techniques for RTM computations, this thesis provides improvements to several aspects of well-established acceleration techniques and the design of new methods to boost the processing algorithms performance of benchmark hyperspectral RTMs. A widely-employed dimensionality reduction technique is the principal component analysis (PCA) for forward simulations of the RTMs. In this thesis, a hybrid PCA technique is tested for ultraviolet (UV) region for the TROPOMI instrument. In particular, PCA is applied to the input and output data considering machine learning (ML) techniques to process the RTMs, providing a fast and accurate implementation to enhance the hyperspectral processing chain.

The major contribution of this thesis is the design of a new acceleration technique for hyperspectral RTMs called "Cluster Low-Streams Regression" (CLSR) method. It is based on the reduction of computationally expensive simulations of the fine resolution RTM. Furthermore, the CLSR method is applied to important trace gases such as for $O_3$ in the Hartley-Huggins band, the $O_2$ A- band near 765 nm, the near-infrared water vapour bands ($\rho$-$\sigma$-$\tau$- and the 0.8 $\mu$m bands) and the $CO_2$ weak band for different types of aerosols and clouds, achieving performance enhancements of 1-3 orders of magnitude which clearly outperform other benchmark acceleration techniques.

The performance enhancements for hyperspectral RTMs achieved in this dissertation offer an opportunity to the scientific community, as these techniques have the potential to be implemented in future remote sensing applications and to further optimize the algorithms used for the whole processing chain.

# Zusammenfassung

Die Erdbeobachtung (Earth Observation, EO) stellt ein wichtiges Werkzeug dar, mit dessen Hilfe der Klimawandel durch den Einsatz von Fernerkundungstechnologien überwacht und wissenschaftlich bewertet werden kann. Mit Satelliten-getragenen Instrumenten, den sog. Atmosphärenspektrometern zur Beobachtung der Atmosphärenzusammensetzung (Atmospheric Composition Sensors, ACS), ist es möglich, diejenigen physikalischen Größen näher zu beleuchten, die einen Einfluss auf unsere Erdatmosphäre haben, da sich mit diesen Instrumenten die spektralen Signaturen im Strahlungsfeld der Atmosphäre lokalisieren lassen. Im nachfolgenden Schritt werden dann Strahlungstransfermodelle (Radiative Transfer Models, RTMs) dazu verwendet, aus diesen Beobachtungen die chemische Zusammensetzung der Atmosphäre extrahieren zu können.

Mit der jüngsten Generation von ACS wird inzwischen eine außerordentlich hohe räumliche, zeitliche und spektrale Auflösung erreicht. Ferner wird durch diese sog. hyperspektralen Sensoren eine gewaltige Menge an Beobachtungsdaten bereitgestellt. Somit gibt es zwei große Herausforderungen bei der Verarbeitung dieser hyperspektralen Daten: Einerseits ist es erforderlich, dass die Leistungsfähigkeit der Verarbeitungsalgorithmen in einer Weise gesteigert werden kann, dass sie die Anforderungen an eine nahezu Echtzeit-Prozessierung der Daten erfüllen können. In diesem Kontext werden im Höchstleistungsrechnen (High Performance Computing, HPC) Techniken der Dimensionsreduzierung oder andere Beschleunigungstechniken verwendet. Andererseits sollte nach Anwendung einer bestimmten Beschleunigungstechnik auch die Genauigkeit der extrahierten atmosphärischen Parameter über den verfügbaren Spektralbereich möglichst erhalten bleiben. Letztlich bestimmt die feine spektrale Auflösung den überhaupt extrahierbaren Informationsgehalt, so dass einer genauen Erfassung der spektralen Signaturen eine entscheidende Rolle für eine wahrheitsgetreue Ermittlung der Atmosphärenkomposition zukommt.

Beginnend mit einer detaillierten Gegenüberstellung der in der Literatur existierenden RTM-Beschleunigungsmethoden, werden in der vorliegenden Arbeit mehrere Aspekte von bereits etablierten Beschleunigungstechniken weiter verbessert. So werden ganz neue Methoden entworfen, mit denen die Verarbeitungsgeschwindigkeit von Benchmark Hyperspektral-RTMs weiter gesteigert werden kann. Eine weit verbreitete Technik zur Dimensionsreduzierung im Kontext von RTM Vorwärtssimulationen ist die Hauptkomponentenanalyse (Principal Component Analysis, PCA). In der vorliegenden Arbeit wird eine hybride PCA-Technik zur Verwendung im ultravioletten Spektralbereich von TROPOMI entwickelt und näher untersucht. Im Besonderen wird dabei so vorgegangen, dass die PCA sowohl auf die Eingangs- als auch die Ausgangsdaten angewandt wird, bei gleichzeitigem Einsatz von Techniken des maschinellen Lernens (ML) zur Verarbeitung mit den RTMs. Auf diese Weise lässt sich sowohl eine schnelle als auch präzise Implementierung bewerkstelligen, mit welcher eine Leistungssteigerung der hyperspektralen Verarbeitungskette erreicht werden kann.

Der Hauptbeitrag dieser Dissertation besteht in der Entwicklung einer neuen Beschleunigungstechnik, der sogenannten ,,Cluster Low-Streams Regression" (CLSR) Methode, die für hyperspektrale RTMs eingesetzt werden kann. Die Methode basiert auf der Reduktion der Berechnungszeiten für die rechenintensiven Strahlungstransfersimulationen, wie sie für spektral fein auflösende RTMs erforderlich sind. Weiter wird die CLRS Methode auf eine

Anzahl wichtiger Spurengase angewandt, wie z. B. für $O_3$ in der Hartley-Huggins-Bande, die $O_2$-A-Bande bei 765 nm, die Nahinfrarot-Wasserdampf-Banden (sog. $\rho$-$\sigma$-$\tau$-Banden und das 0.8 $\mu$m-Band) sowie für ein schwaches Absorptionsband von $CO_2$, jeweils bei Vorhandensein verschiedener Aerosol- und Wolkentypen. Es wird gezeigt, dass sich mit der CLRS-Methode Beschleunigungsfaktoren von einer bis zu drei Größenordnungen erzielen lassen, wodurch andere Benchmark-Beschleunigungstechniken klar übertroffen werden.

Die in dieser Dissertation erzielten Steigerungen der Performanz für hyperspektrale RTMs sind für die Wissenschaftsgemeinde von Interesse, da diese Techniken für zukünftige Fernerkundungsanwendungen das Potenzial besitzen, die in der gesamten Prozessierungskette verwendeten Algorithmen weiter optimieren können.

# Contents

# 1 Introduction

## 1.1 Motivation

The Earth's atmosphere is affected by a synergy of natural phenomena and anthropogenic actions that together form a very complex system. Human impact on the Earth's atmosphere has shifted from local to global scale over the last century due to population growth [1]. Understanding and interpreting its influences is a crucial step towards building a more sustainable future. In this endeavour, remote sensing fundamentally helps to elucidate the factors that are affecting the Earth's atmosphere and ultimately the climate change. Therefore, the retrieval of atmospheric constituent's distributions and surface properties from surface-based, airborne and satellite remote sensing instruments plays a vital role in monitoring the Earth's atmosphere and understanding the chemical and physical processes therein [2]. Accordingly, policy makers could not only establish new environmental laws required to minimize the anthropogenic impact on the environment and climate [3], but also obtain scientific evidence on the effectiveness of the measures.

Earth observation (EO) from space, i.e., by satellite composition sensors based remote sensing, is fundamental for globally and continuously monitoring the atmosphere. There are currently more than 150 EO satellites in orbit [4] carrying passive sensors. The passive atmospheric composition sensors (ACS) detect and record the radiance that is emitted or reflected from the Earth's atmosphere or surface in the wide spectral region ($0.3 - 3$ $\mu$m). The information about the atmosphere is then retrieved from the spectral data by using the so called atmospheric processors, i.e., codes which are specifically designed to invert ACS measurements that detect the uniqueness of the spectral signature. There has been significant advances in the capabilities of the existing and new sensors based on a more sophisticated spatial, spectral and temporal resolution [5]. In addition, the hyperspectral sensors that combine information from several spectral bands, have improved significantly due to the recent developments in optics sensors and measurement techniques. Therefore, the new generation of ACS on-board EO satellites, deliver a massive amount of hyperspectral data that has to be stored and analysed. For example, the TROPOsperic Monitoring Instrument (TROPOMI) on board the Copernicus Sentinel 5 Precursor (S5P) satellite [6] has a spatial resolution of two orders of magnitude higher, providing 21 million level 1B (L1B) spectra per day, i.e., almost 8 milliard spectral points [7]. Therefore, extracting the information about geophysical parameters (level-2 data) from spectral radiance distributions (level-1 data) turns out to be a major computational challenge and requires high performance computing (HPC).

The gradual increase of high resolution spectral data presents a twofold challenge. First, that the amount of satellite data is increasing faster than the computational power [8]. Thus, new efficient techniques have to be developed for next generation atmospheric processors to cope with these high efficiency requirements; and second, that it is crucial to achieve the near real time (NRT) requirements for processing the hyperspectral data [9].

Radiative transfer models (RTMs) are key components of the retrieval algorithms to process the hyperspectral data and to ultimately retrieve the atmospheric constituents of the atmosphere. However, the application of the RTMs are the major performance

bottleneck in the processing chain due to the thousands of spectral lines to be retrieved. Accurate and timely simulating of hyperspectral data is a challenging task. The simplest manner to simulate spectra is by accounting for the fine spectral grid or the so called line-by-line (LBL) resolution, so that afterwards the spectra are convolved with the spectral response function (SRF), also referred to as slit function, obtaining smooth spectra which does not change its shape significantly. Nevertheless, the information content of the atmosphere decreases as the slit function becomes wider. Colosimo et al. [10] showed that the simulations with better spectral resolution generally lead to an increase in the total amount of information that can be retrieved. Therefore, it is crucial to be able to obtain as higher spectral resolution data as possible in order to obtain the information about the fine structure of the spectrum. Such high resolution satellite remote sensing observations are extremely useful for diagnosing the impact of the atmospheric constituents on a global scale allowing detection of small-scale sources and increasing the fraction of cloud-free observations.

Dimensionality reduction techniques are an essential part of the big data analysis to overcome the redundancies found in the hyperspectral data and to cope with current and future data flows. In this regard, acceleration techniques for hyperspectral RTMs have to be developed [11]. Essentially, these techniques do not compute spectral radiances in a LBL manner, i.e. with a fine grid spectral resolution, but rather they take into account the interdependency between spectral channels or group them accordingly, thereby reducing the total number of calls to RTMs. This motivates the development of exact and approximate RTMs endowed with acceleration techniques. Thus, the overarching challenge of the thesis is to develop accurate and high performance RTMs which would cope with the current operational requirements of online data processing in the context of atmospheric retrieval codes of trace gases and cloud properties. In this line, disruptive development poses new value-added technology to overcome the difficulties of processing the huge amount of data with current algorithms.

## 1.2 Objectives

It has become a challenging issue to process the vast amount of high resolution spectral data of the new generation of EO sensors. This thesis focuses on the enhanced implementation of hyperspectral processing algorithms in the context of high resolution spectral remote sensing data of the atmosphere. Within this framework, the overarching goal is to investigate several dimensionality reduction techniques for LBL RTMs in different spectral regions and to devise new acceleration methods that computationally improve the state-of-the-art techniques without losing accuracy. Towards this far reaching goal, three essential objectives have been specified in the following:

**Objective 1:** *Gaining an in-depth overview of the existing acceleration techniques for RTM computations.*

This goal serves to explore acceleration techniques for RTMs commonly applied for real purposes. Furthermore, it also helps to understand the relevance of the data reduction methods to improve the efficiency of hyperspectral processing algorithms. To accomplish this goal, a full peer-reviewed review has been written and is presented in Appendix A.2.

**Objective 2:** *Applying established dimensionality reduction techniques for RTM simulations and upgrade them.*

This objective aims at accurately applying well-known dimensionality reduction techniques such as the principal component analysis (PCA) for forward simulations of the RTMs. The ultimate goal is to upgrade the acceleration techniques for spectral bands of

interest and the accurate retrieval of aerosols and clouds. This can be accomplished, for instance, by considering the feasibility of machine learning (ML) techniques to process the RTMs with a hybrid PCA approach. This goal helps to better enhance the hyperspectral processing chain, which plays a vital role in HPC of the RTMs to retrieve atmospheric constituents.

**Objective 3:** *Devising new acceleration techniques to boost the performance of computing high resolution spectral data.*

This objective goes a step further in the HPC of atmospheric composition sensors with the research of acceleration techniques for hyperspectral data. The newly devised techniques ultimately help to overcome the hyperspectral data redundancies and facilitate the fast spectral processing for different atmospheres and spectral bands. They also enrich the classical data-driven algorithms with the implementation of more efficient and fast techniques.

## 1.3 Thesis Structure

This is a cumulative dissertation which includes four full peer-reviewed research articles, one full peer-reviewed review and one peer-reviewed conference paper:

- A. del Águila, D. S. Efremenko, V. Molina García, and J. Xu. Analysis of Two Dimensionality Reduction Techniques for Fast Simulation of the Spectral Radiances in the Hartley-Huggins Band. Atmosphere, 10(3):142, 2019. doi:10.3390/atmos10030142

- A. del Águila, D. S. Efremenko, T. Trautmann. A Review of Dimensionality Reduction Techniques for Processing Hyper-Spectral Optical Signal. Light and Engineering, 27, 3, 85–98, 2019. doi: 10.33383/2019-017

- A. del Águila, D. S. Efremenko, V. Molina García, M. Y. Kataev. Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of $O_2$ A- and $CO_2$ Bands. Remote Sensing, 12(8):1250, 2020. doi: 10.3390/rs12081250

- A. del Águila, D. S. Efremenko. The Cluster Low-Streams Regression Method for Fast Computations of Top-of-the-Atmosphere Radiances in Absorption Bands. 30th International Conference on Computer Graphics and Machine Vision (Graphicon 2020). Part 2, pages short25–1–short25–9, 2020. doi: 10.51130/graphicon-2020-2-4-25

- A. del Águila, D. S. Efremenko. Accuracy Enhancement of the Two-Stream Radiative Transfer Model for Computing Absorption Bands at the Presence of Aerosols. Light and Engineering, 29, 2, 79-86, 2021. doi: 10.33383/2020-078

- A. del Águila, D. S. Efremenko. Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method. Remote Sensing, 13(3):434, 2021. doi: 10.3390/rs13030434

As for the structure of this thesis, Chapter 2 comprises the explanation of the physical concepts of atmospheric radiative transfer. Chapter 3 covers the state-of-the-art existing hyperspectral RTMs related to this dissertation. Chapter 4 attains the work achieved within this thesis by summarizing the above mentioned research items and Chapter 5

provides a summary of achievements, an overview of challenges and future work. The appendices contain the full-publications for which this cumulative dissertation has been organized, the related publications in which I am co-author and the list of abbreviations and symbols.

# 2 Atmospheric Radiative Transfer

Atmospheric radiative transfer (RT) is fundamental for remote sensing applications in the EO framework. In this context, the atmospheric RT describes the propagation of the radiance through the atmosphere by taking into account the absorption, scattering or emission processes of the radiance with the surface and/or the atmospheric features.

In this chapter, the foundations of atmospheric RT are summarized and the current algorithms based on the radiative transfer equation (RTE) are analyzed.

## 2.1 Atmospheric Absorption Bands

The solar spectrum is defined as the distribution of electromagnetic (EM) radiation emitted by the sun as a function of the incident wavelength $\lambda$ on the top-of-the-atmosphere (TOA) [12]. As a result of any particle in motion, the particle has a kinetic energy known as *translational energy* ($W_{\text{tra}}$). The averaged *translational energy* of a single molecule in the three-dimensional space is equal to $k_{\text{B}}\mathcal{B}/2$ where, $k_{\text{B}}$ is the Boltzmann constant and $\mathcal{B}$ is the absolute temperature. Thus, the molecules composed by atoms have a *rotational energy* ($W_{\text{rot}}$) due to its rotation over an axis through its center of gravity. In turn, there exists the *vibrational energy* ($W_{\text{vib}}$) due to the bounding forces of the atoms composing the molecules which permit them to vibrate about their equilibrium positions. These three *energy* types are quantized and it is possible for a molecule to change its energy due to a perturbation of the electrons of which it is composed [13]. The total *energy* ($W_{\text{tot}}$) of a molecule can be expressed as follows:

$$W_{\text{tot}} = W_{\text{tra}} + W_{\text{rot}} + W_{\text{vib}}. \tag{2.1}$$

When the molecules go through transitions from energy levels, the absorption or emission of radiation takes place. Therefore, the spectra are originated from millions of vibrating electrons and molecules that emit and absorb EM radiation in an unique set of energy levels. The absorption bands are areas of the EM spectrum which are absorbed by atmospheric gases (see Figure 2.1) in a specific wavelength range. The EM radiance interacts with objects such as clouds, aerosols or trace gases in an unique combination of wavelengths $\lambda$ which results in an unique spectral signature.

In sum, the atmospheric absorption bands due to electronic transitions of molecular and atomic oxygen and ozone occur mainly in the ultraviolet (UV) region, while those due to the vibrational and rotational transitions of triatomic molecules such as $H_2O$, $O_3$ and $CO_2$ lie in the infrared (IR) region [13]. There is also little absorption in the visible (VIS) region of the solar spectrum.

Satellite measurements consist of acquiring the high resolution spectra of the rotational-vibrational bands of interest. Each absorption band has thousands of absorption lines which are investigated through the high resolution spectroscopic data or hereafter called hyperspectral data. The fine structure of the absorption bands gives information of the composition of the atmosphere. Therefore, it is extremely important to obtain the best possible resolution in order not to lose information about the atmosphere's composition.

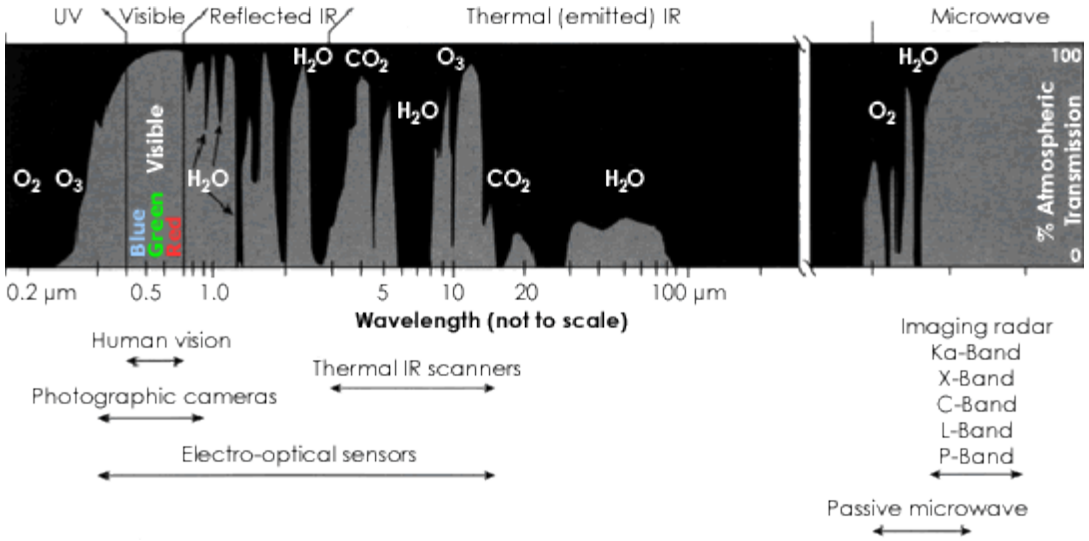


**Figure 2.1:** Diagram of the atmospheric transmission in the EM spectrum and the main absorbing gases [14].

The form of converting spectra into atmospheric composition is by means of RTMs. Determining atmospheric characteristics calls for considerable understanding of radiative transfer. In the following section, a description of the basics of radiative transfer is given.

## 2.2 Physical Assumptions in Radiative Transfer Theory

The RT describes the propagation of EM radiation through the atmosphere taking into account the interactions of the radiance with the atmospheric components and the Earth surface. The physical assumptions in RT theory were based on quantitative predictions rather than rigorous first-principle physical concepts based on Maxwell equations. Lommel [15] made the first treatise of RT in 1889 with a derivation of the physical assumptions that explained the light attenuation through a medium. More recently, in the book of Mishchenko et al. [16], the physical assumptions to RT are explained in detail. The interpretation of the RT theory from the physical point of view is possible to be performed in several ways (see for instance [17, 18, 19, 20, 21, 22] and references therein). The main assumptions under which the RT theory can be applied are summarized as follows [23]:

1. The radiation is assumed to propagate in the form of rays (*ray optics approximation*). Although the wave properties of the light are taken into account when considering scattering by a single particle (e.g. aerosol particle, water droplet or number density fluctuations).

2. The rays are assumed to arrive at a given point from different directions being totally incoherent and can therefore be summed algebraically regardless of the phases and interference terms.

3. The RT theory assumes the measured parameters to be spatially and timely averaged rather than localized values.

4. Radiation has the property of ergodicity, i.e. the average over time for a specific realisation of a random scattering process is equivalent to the ensemble average.

5. The atmospheric medium is assumed to be composed of randomly distributed scatterers, while the interaction between the ray and the medium is statistically independent of the outcome of subsequent scattering events (this is the so-called Markov process).

Following the above physical assumptions, the *radiance* of the RT can be integrated yielding the amount of transferred energy. The simplicity of these physical assumptions for RT leads to the development of efficient numerical models to simulate the scattered radiance. In opposition to this approach, exists the microphysical approach which considers the RT as a propagation of EM waves in the discrete medium consisting of a set of scatterers [19, 20].

## 2.3 The Radiative Transfer Equation

The description of the RT phenomenon mathematically is given by the RTE which can be regarded as a balance equation for the amount of radiant energy going through a medium [24]. The one-dimensional RTE for the scalar radiance $I$ in an homogeneous, isotropic plane-parallel atmosphere reads as follows [25]:

$$\mu \frac{\mathrm{d}I}{\mathrm{d}r}(r, \mathbf{\Omega}) = -k_{\mathrm{ext}} I(r, \mathbf{\Omega}) + \frac{k_{\mathrm{sca}}}{4\pi} \int_{4\pi} P(r, \mathbf{\Omega}, \mathbf{\Omega}') I(r, \mathbf{\Omega}') \, \mathrm{d}\Omega', \tag{2.2}$$

where $\mathbf{\Omega} = (\mu, \varphi)$ is the direction of the light with $\mu = \cos\theta$ the cosine of the polar angle and $\varphi$ the azimuthal angle; the term $k_{\mathrm{ext}}$ and $k_{\mathrm{sca}}$ correspond with the extinction and scattering coefficients, respectively; and $P$ is the single scattering phase function.

For the next step, the Equation 2.2 is transformed to the optical depth ($\tau$) space considering that it is defined as the integral of the extinction coefficient along the path of the light ray:

$$\tau = \int_{r_0}^{r} k_{\mathrm{ext}} \mathrm{d}r'. \tag{2.3}$$

which can be visualized in Figure 2.2. Thus, taking into account that $\mathrm{d}\tau = k_{\mathrm{ext}} \mathrm{d}r$ and that the extinction coefficient is the sum of the absorption and scattering coefficients ($k_{\mathrm{ext}} = k_{\mathrm{sca}} + k_{\mathrm{abs}}$), the RTE reads as follows:

$$\mu \frac{\mathrm{d}I}{\mathrm{d}\tau}(\tau, \mathbf{\Omega}) = -I(\tau, \mathbf{\Omega}) + \frac{\omega(\tau)}{4\pi} \int_{4\pi} P(\tau, \mathbf{\Omega}, \mathbf{\Omega}') I(\tau, \mathbf{\Omega}') \, \mathrm{d}\Omega', \tag{2.4}$$

where $\omega$ is the single scattering albedo (also named SSA) which relates the scattering and extinction coefficients in the form: $\omega = \frac{k_{\mathrm{sca}}}{k_{\mathrm{ext}}}$). The SSA will depend on the medium and the scattering properties of the material with which it interacts.

## 2.4 Solving the Radiative Transfer Equation with the Discrete Ordinate Method

Most of the RTMs are based on the discrete ordinate method, in which the continuous function of angular variables (the radiance in our case) is presented as a set of values along discrete directions (discrete ordinates). The theory of discrete ordinate models has

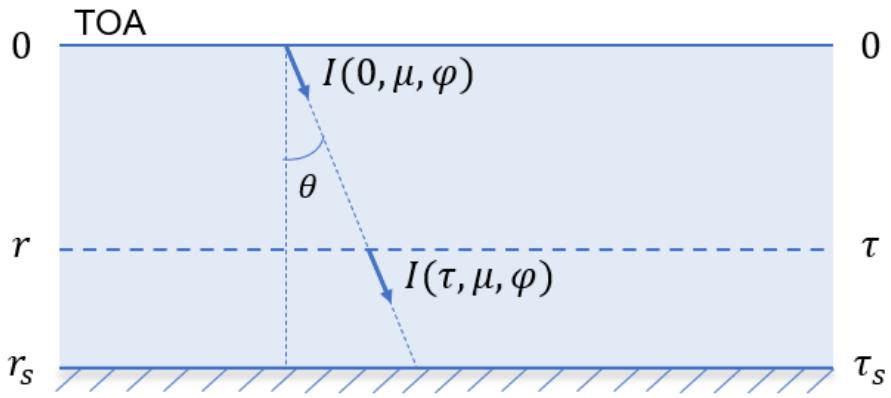


**Figure 2.2:** Scheme of the light attenuation of the radiance passing through a medium.

been described in [26, 27]. Additionally, other studies have applied the discrete ordinate method in combination with the small-angle approximation [28, 29, 30, 31, 32, 33], which computes the total radiance analytically as a sum of small angle parts and numerically by the discrete ordinate method. In this dissertation, we focus on solving the RTE with the discrete ordinate method. The basic steps to derive a numerical solution in the framework of discrete ordinate method can be summarized as follows:

1. Let us consider the cosine azimuthal expansion of the radiance field and the phase function providing an equation for the azimuthal component $I_m(\tau,\mu)$ to factor-out the azimuthal dependency,

$$I(\tau,\mathbf{\Omega}) = \sum_{m=0}^{M} I_m(\tau,\mu)\,\cos m\varphi. \tag{2.5}$$

2. The radiance field is discretized in the $\mu$-domain by considering $N_{\text{do}}$ (discrete ordinates) Gaussian points $\mu$ per hemisphere and corresponding weights $w_i$, where $i = 1, ..., N_{\text{do}}$.

3. An inhomogeneous atmosphere is described as a system of $N$ homogeneous layers (see Figure 2.3): $\tau_1 < \tau_2 < ... < \tau_{N+1}$, where $\tau_1 = 0$ and $\tau_{N+1} = \tau_s$ at the surface boundary condition. A layer $l$ is delimited above by the level $\tau_l$ and below by the level $\tau_{l+1}$. Thus, the geometrical thickness is $\Delta\tau_l = \tau_l - \tau_{l+1}$.

4. The RTE can be rewritten in the discrete space as a system of differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{bmatrix}\mathbf{i}_m^{\uparrow}(\tau)\\ \mathbf{i}_m^{\downarrow}(\tau)\end{bmatrix} = -\mathbf{A}_{ml}\begin{bmatrix}\mathbf{i}_m^{\uparrow}(\tau)\\ \mathbf{i}_m^{\downarrow}(\tau)\end{bmatrix} + \mathbf{b}_{ml}(\tau), \quad \tau_{l+1} \le \tau \le \tau_l, \tag{2.6}$$

   where $\left[\mathbf{i}_m^{\uparrow\downarrow}(\tau)\right]_i = I_m(\tau, \mp\mu_i)$ are the radiance vectors in the discrete ordinate space for the upwelling ($\uparrow$) and downwelling ($\downarrow$) radiances, respectively. $\mathbf{A}_{ml}$ is the layer matrix and $\mathbf{b}_{ml}$ is the source vector (more details can be found in [34]).

The general solution of the linear system of differential equations (Equation 2.6) is a linear combination of all the homogeneous solutions plus the particular solutions for the assumed sources. For a medium with multiple layers, the continuity condition is imposed

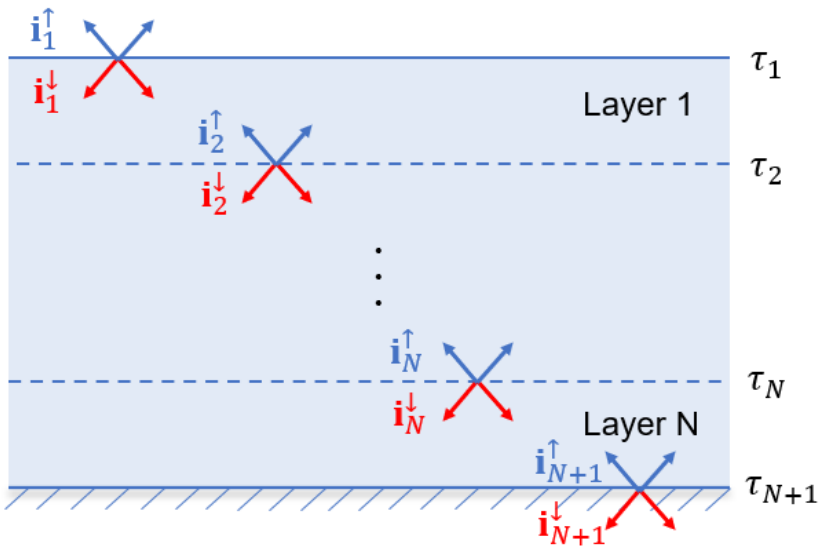


**Figure 2.3:** System of $N$-layers for an inhomogeneous atmosphere. The blue colour of the arrows corresponds with the upwelling direction and the red colour with the downwelling direction at the layer boundaries.

for the radiances across the layer interfaces. An alternative method to solve the system of differential equations is by means of the Discrete Ordinate with Matrix Exponential (DOME) method [26]. In this case, the linear system is treated as a boundary value problem and the layer equation can be derived for each layer. Thus, integrating Equation 2.6, we obtain the following expression which relates the radiances at the boundaries (in the following, we have neglected the azimuthal expansion term ($m$) for simplicity):

$$\begin{bmatrix} \mathbf{i}_{l+1}^{\uparrow} \\ \mathbf{i}_{l+1}^{\downarrow} \end{bmatrix} = \mathrm{e}^{-\mathbf{A}_l \bar{\tau}_l} \begin{bmatrix} \mathbf{i}_{l}^{\uparrow} \\ \mathbf{i}_{l}^{\downarrow} \end{bmatrix} + \mathrm{e}^{\mathbf{A}_l \bar{\tau}_l} \int_0^{\bar{\tau}_l} \mathrm{e}^{\mathbf{A}_l t} \mathbf{b}_l(t) \mathrm{d}t. \tag{2.7}$$

where $t$ is a layer coordinate defined by $t = \tau - \tau_l$ and $0 \leq t \leq \bar{\tau}_l$. Then, by multiplying Equation 2.7 by $\mathrm{e}^{\mathbf{A}_l \bar{\tau}_l}$ we obtain:

$$-\begin{bmatrix} \mathbf{i}_{l}^{\uparrow} \\ \mathbf{i}_{l}^{\downarrow} \end{bmatrix} + \mathrm{e}^{\mathbf{A}_l \bar{\tau}_l} \begin{bmatrix} \mathbf{i}_{l+1}^{\uparrow} \\ \mathbf{i}_{l+1}^{\downarrow} \end{bmatrix} = \int_0^{\bar{\tau}_l} \mathrm{e}^{\mathbf{A}_l t} \mathbf{b}_l \mathrm{d}t. \tag{2.8}$$

In order to evaluate the terms $\mathrm{e}^{\mathbf{A}_l \bar{\tau}_l}$ and $\mathrm{e}^{\mathbf{A}_l t}$, the eigenvalue decomposition method can be applied with DOME [26] by using a spectral decomposition of the layer matrix $\mathbf{A}_l$ as:

$$\mathbf{A}_l = \mathbf{V}_l \Lambda_l \mathbf{V}_l^{-1}, \tag{2.9}$$

where $\mathbf{V}_l$ is the eigenvector matrix and $\Lambda_l$ is the eigenvalue matrix for the $l^{\text{th}}$ layer:

$$\Lambda_l = \mathrm{diag}[\lambda_{l1}, \ldots, \lambda_{lN_{\mathrm{do}}}]. \tag{2.10}$$

Thus, Equation 2.9 would be of the form:

$$\mathrm{e}^{\mathbf{A}_l t} = \mathbf{V}_l \mathrm{e}^{\Lambda_l t} \mathbf{V}_l^{-1}. \tag{2.11}$$

The matrix exponential ($\mathrm{e}^{-\mathbf{A}_l \bar{\tau}_l}$) is called *propagator* because if one-point for boundary condition is given, then the solution from the bottom to the upper boundary can be

propagated. By applying some transformations we can obtain a ready-to-compute system of equations, i.e. by using Equation 2.11 into 2.8 and multiplying by $\mathbf{V}_l^{-1}$ on the left side we obtain the following:

$$-\mathbf{V}_l^{-1}\begin{bmatrix}\mathbf{i}_l^{\uparrow}\\ \mathbf{i}_l^{\downarrow}\end{bmatrix}+\mathrm{e}^{\mathbf{A}_l\bar{\tau}_l}\mathbf{V}_l^{-1}\begin{bmatrix}\mathbf{i}_{l+1}^{\uparrow}\\ \mathbf{i}_{l+1}^{\downarrow}\end{bmatrix}=\int_0^{\bar{\tau}_l}\mathrm{e}^{\mathbf{A}_l t}\mathbf{V}_l^{-1}\mathbf{b}_l\mathrm{d}t. \tag{2.12}$$

This equation is unstable due to the positive exponentials. Therefore, we follow [35] to exclude the exponents with positive powers and in order to find a computationally stable solution for a multilayer atmosphere, we apply certain boundary layer conditions:

- The boundary condition at the TOA ($l = 1$) states that the downwelling diffuse radiation is absent, i.e. $\mathbf{i}_1^{\downarrow} = 0$.

- The surface boundary condition (at $l = N$) states that the upwelling radiance is:

$$\mathbf{i}_N^{\uparrow} = \mathbf{R}_N\mathbf{i}_N^{\downarrow} + \mathbf{p}_N, \tag{2.13}$$

  where $\mathbf{R}$ and $\mathbf{p}$ are the reflectivity matrix and the source term, respectively.

Thus, by rearranging the terms and applying some transformations, we obtain the following system that relates the radiances at the boundary of the layer:

$$\begin{bmatrix}\mathbf{i}_l^{\uparrow}\\ \mathbf{i}_{l+1}^{\downarrow}\end{bmatrix}=\begin{bmatrix}\mathbf{R}_l^{\uparrow}\mathbf{T}_l^{\uparrow}\\ \mathbf{T}_l^{\downarrow}\mathbf{R}_l^{\downarrow}\end{bmatrix}\begin{bmatrix}\mathbf{i}_l^{\downarrow}\\ \mathbf{i}_{l+1}^{\uparrow}\end{bmatrix}+\begin{bmatrix}\mathbf{P}_l^{\uparrow}\\ \mathbf{P}_l^{\downarrow}\end{bmatrix}, \tag{2.14}$$

where $\mathbf{R}_l^{\uparrow\downarrow}$, $\mathbf{T}_l^{\uparrow\downarrow}$ and $\mathbf{P}_l^{\uparrow\downarrow}$ are the reflection, transmission and source terms, respectively. The ready-to-compute formulas can be found in [34] in terms of $\mathbf{V}_l$ and $\Delta_l$.

Although there are several numerical techniques to solve the RTE, we are focused on DOME due to the following reasons:

1. Algorithms based on the discrete ordinate method are generic: DOME provides a numerically stable solution for the arbitrary values of $\tau$, $\omega$ and phase function.

2. It is already implemented in RT models widely used in the scientific and professional community, such as DISORT, libRadtran, SCIATRAN, LIDORT and others.

3. The method is convenient to be used as it has only one parameter that the user can tune (namely, $N_{\mathrm{do}}$).

The DOME solves the RTE and, as other solvers, its performance depends on the number of discrete ordinates ($N_{\mathrm{do}}$) used. Figure 2.4 shows the computational performance for different number of discrete ordinates. It is to be noted that the computational time for computing the eigenvalue solution increases as $\approx N_{\mathrm{do}}^{2.5}$ and the memory as $N_{\mathrm{do}}^2$. Thus, it is pivotal to find efficient techniques to accelerate the retrieval of the RT solutions. Approximate methods are used for that purpose. In the following chapter, an extended description of the fast hyperspectral RTMs state-of-the-art is provided.

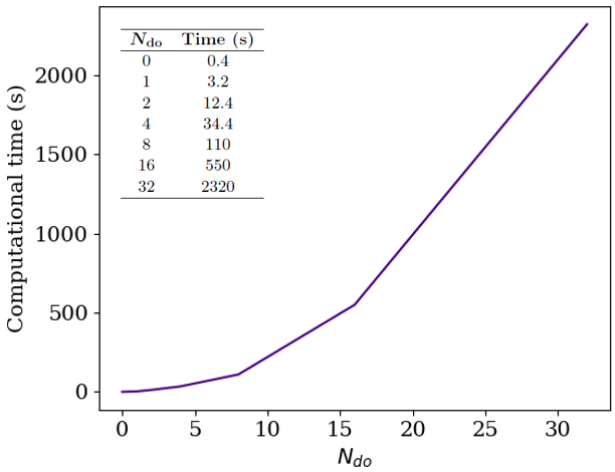


**Figure 2.4:** Computational time (in seconds) vs. number of discrete ordinates ($N_{\mathrm{do}}$) for a RT solver.

# 3 Hyperspectral Radiative Transfer Models

In hyperspectral remote sensing retrieval applications, the RT computations are the bottleneck in the whole processing chain. The techniques of dimensionality reduction in atmospheric sciences exploit the strong interdependency of the hyperspectral data. Therefore, acceleration techniques have been developed for hyperspectral RTMs in order to reduce the monochromatic computations. Essentially, the hyperspectral RTMs techniques do not compute spectral radiances in a LBL framework but rather they take into account the interdependency between spectral channels or group them accordingly, thereby reducing the total number of calls to RTMs [36]. In this chapter, the relevant state-of-the-art of fast hyperspectral RTMs for this thesis are reviewed. Some parts of this chapter are adapted from the full peer-reviewed review of Appendix A.2 published in [7]. This chapter is divided in three main blocks: First, the correlated-$k$ model is presented and the subsequent modifications based on it. Second, the PCA-based RTMs are reviewed from the input and output point of view. Finally, the machine learning based RTMs are summarized with recent examples of its application.

## 3.1 Correlated $k$-distribution Methods

The first approach in atmospheric and astrophysics sciences which avoids time consuming LBL computations was developed by Ambartsumian [37] in 1936. In his aim to explore new speed up RT calculations for the external layers of stars, he noted that the transmission in an homogeneous atmosphere depends on the distribution of the absorption coefficients ($k$) within a spectral interval rather on the spectral dependence of $k$ with the wavelength $\lambda$. Such concept was called the $k$-distribution approach. The method is based on the cumulative frequency distribution of $k$, namely, $G(k)$:

$$G(k) = \int_0^k f(k')dk', \tag{3.1}$$

where $f(k)$ is the distribution function for the absorption coefficients $k$ in a spectral interval $\Delta\lambda$ which makes $G$ lie between 0 and 1. As a matter of fact, the so-called $k$-distribution is the inverse distribution $k(G)$. Since $G(k)$ (Equation 3.1) is a smooth function and $k(\lambda)$ has a strong variation with respect to $\lambda$, i.e. there will be many redundancies between spectral points because they share the same $k(\lambda)$, thus, it is beneficial from the numerical point of view to replace the integration in $\lambda$-space by that in $G$-space in order to reduce the number of computations (less number of quadrature points for numerical integration are required).

Goody et al. [38] extended the $k$-distribution approach to the cases of inhomogeneous, multiple-scattering atmospheres assuming that there is a correlation between $k$-distributions at different pressure levels, and they called it the correlated-$k$ distribution approach. That assumption was shown to be valid when there was strong and weak absorption on the limits. Several modifications of the $k$-distribution method were applied to the thermal IR. Examples of this applications derived in different techniques such as the opacity sampling method [39] or the multidimensional approach to $k$-distribution method

[40], where a general formulation of multidimensional $k$-distributions for the radiative heat transfer is proposed. Also in this spectral region, Fomin has multiple contributions to the field (e.g. [41, 42, 43]). Furthermore, André et al. [40] stated that all $k$-distribution models, i.e. from the most accurate line-by-line up to the simplest correlated or Scaled-$k$ ones, are based on a restricted set of concepts shared by all of them. Several further implementations to the correlated-$k$ method have been utilized over the years. Some of the most interesting are presented in chronological order in the following:

### Exponential-sum fitting transmittance

The exponential sum fitting of transmittances (ESFT) method [44, 45] optimizes the number of $k$ values by fitting the transmission function $T(u)$ for a given spectral range by a sum of exponentials:

$$< T(u) > \sim \sum_{i=1}^{W} w_i e^{-k_i u}, \tag{3.2}$$

being $u$ the quantity of absorber, $W$ the monochromatic wavelengths and $w_i$ the weights (satisfying $\sum_{i=1}^{W} w_i = 1$ ). The latter are used to find the best fit that minimizes the least-squares residuals. This technique was applied to accelerate the fast radiative transfer simulations to: the IR water vapour and $CO_2$ absorption spectrum to compute cooling rates in the atmosphere (e.g. [46]); to the water vapour band in the IR and near-IR (NIR) (e.g. [47]); and to the satellite channels of the five Advanced Very High Resolution Radiometer (AVHRR) (e.g. [48]).

### Spectral mapping

West et al. [49] developed the spectral mapping technique to improve the accuracy of the correlated-$k$ distribution method in the broadband range. Hence, they provide mapping transformations that group together spectral regions with similar values of $k$. Specifically, they tested the method for model atmospheres containing $CO_2$ and $H_2O$ and obtained errors of 1-2% compared to the LBL results when the number of $k$ coefficients was about 100 times less than the number required for the LBL computations. Additionally, the spectral mapping technique takes into account different gas mixtures and spectral variations of the incident solar radiation. The first numerical implementation of this spectral mapping technique was the spectral mapping atmospheric radiative transfer (SMART) model [50].

### $k$-binning approach

Bennartz and Fischer [51] applied a modified version of the $k$-distribution approach where the main differences were:

- They defined a mapping function for the considered spectral interval so that the relation between the approximated absorption coefficients and the associated wavelengths is explicitly fixed for all layers, instead of deriving weights and averaged absorption coefficients in layered atmospheres.
- They introduce the sensor's spectral response function and the spectral variation of the solar constant in the derivation of the relative weights of the single terms of the $k$-distribution fit.

They obtained an improvement of 5-10 factor of the accuracy compared to the $k$-distribution approach and accuracies of 1 and 0.1% for $O_2$ and $H_2O$, respectively. Later on, Bennartz

and Preusker [52] solved the radiative transfer within the $O_2$ A-band using a set of $N$ $k$-binning intervals for the Orbiting Carbon Observatory (OCO) satellite. The idea behind was to subdivide the interval with the highest error in transmission compared to LBL transmittances and process iteratively until reaching an user-defined threshold, as shown in Figure 3.1. Boesche et al. [53] applied a combination of doubling-adding [54] and $k$-binning methods for the $O_2$ A-band obtaining errors of 0.11% and 0.31% for polarized and reflected light, respectively.

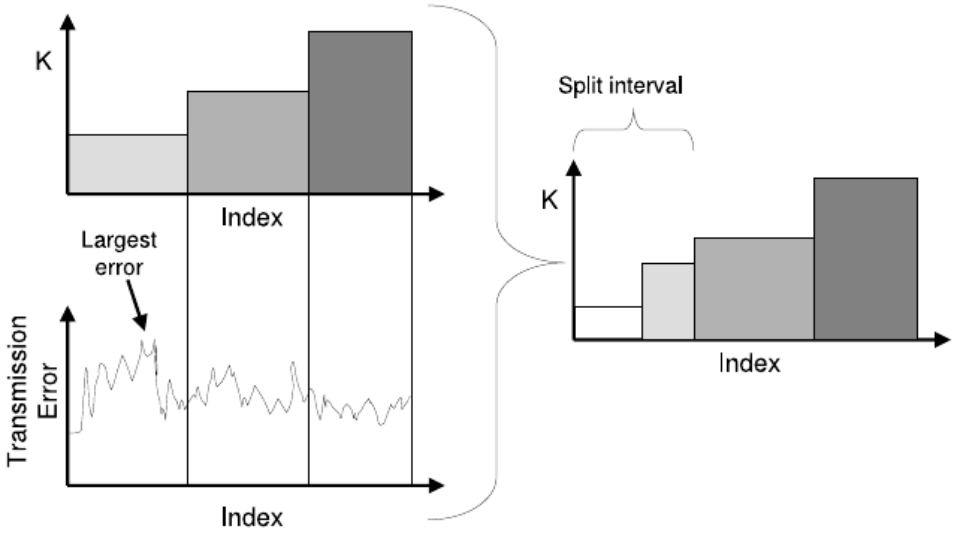


**Figure 3.1:** Schematic representation of the $k$-binning approach [52].

### Double-$k$ method

Duan et al. [55] separated the radiation field into single-scattering component ($I_{\mathrm{SS}}$) and multiple-scattering component ($I_{\mathrm{MS}}$) as a sum:

$$I = I_{\mathrm{SS}}(\lambda) + I_{\mathrm{MS}}(\lambda). \tag{3.3}$$

They performed LBL computations of the single-scattering radiance because it is less computationally expensive although the dependency with the absorption coefficient is less smooth than for the multiple-scattering (see Figure 3.2). The multiple-scattering radiances at specific absorption coefficients $k(\lambda_i)$ were computed with a reduced number of layers and/or streams in a forward RTM [55] which also contributes to a reduction in the computational time. They called this method the double-$k$ method because of the double integration of absorption optical depths to: (1) the total absorption optical depth $k$ and (2) the absorption optical depth from the TOA to the scattering layer $k'$ [56]. Hence, the multiple-scattering contribution is computed as follows:

$$I_{\mathrm{MS}} = g(k) f_k(k'/k), \tag{3.4}$$

where $g$ and $f_k$ are analytical functions whose coefficients are derived from the radiances fitting with different values of $k$ and $k'$. In this form, they account for the vertical distribution of the atmosphere as well as for the distribution of scattering matter. They achieved an accuracy of 0.5% with 28 calculations of the RTM for the $O_2$ A- band and all-sky conditions. This supposed a thousandfold reduction in the forward RTM calculations.

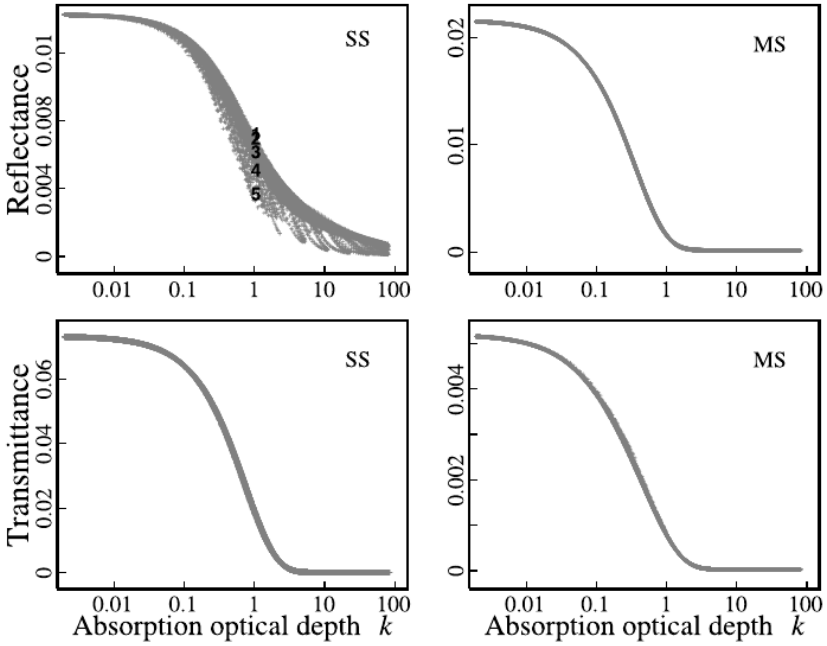


**Figure 3.2:** Multiple- and single-scattering compontents (MS and SS, respectively) of the radiance as a function of the $O_2$ A absorption coefficients for the double-$k$ approach [55].

### Optimal spectral sampling

Moncet et al. [57] designed the Optimal Spectral Sampling (OSS) method as an extension of the ESFT technique for modelling radiances in the IR, microwave and also useful for VIS and UV. The main advantage is that the accuracy with respect to the LBL model is selectable. Further, this technique can be applied for gas mixtures varying with altitude by selecting spectral points rather than the actual absorption cross-sections. The selection of the spectral points and accuracy with respect to the LBL model is explained in these four steps explained in [56]:

1. Select uniformly spaced spectral locations for a sensor channel.
2. Compute the radiances with the LBL model at those spectral locations for atmospheric representative profiles, surface conditions and viewing geometries.
3. Convolve the monochromatic radiances with the instrument slit function.
4. Compute the root mean square difference between the weighted sum of monochromatic radiances and the mean channel radiance for a given node. The optimal weights are obtained by least squares regression technique.

### Linear-$k$ method

Hasekamp and Butz [58] developed a technique called the linear-$k$ method, in order to reduce the number of spectral computations of the multiple scattering radiance ($\mathbf{I_{MS}}$) which depended on the total absorption optical thickness $\tau_{abs}$ and its normalized vertical distribution $\mathbf{n}$. The latter distribution is closely related to the $k$-distribution technique and spectral mapping methods described above. Therefore, the radiance has the form:

$$\mathbf{I_{MS}}(\lambda) = \mathbf{I_{MS}}(\tau_{\mathrm{abs}}(\lambda), \mathbf{n}(\lambda)). \tag{3.5}$$

Accordingly, for a vertical homogeneous atmosphere, the radiance of Equation 3.5, depends smoothly on absorption optical thickness and only a few calculations of $\tau_{\text{abs}}$ are needed. This technique is based on a spectral binning method which accounts for the vertical distribution of the absorption optical thickness by a linear approximation. The errors reported for the three bands of the OCO instrument are below 0.13%, 0.6% and 0.12% for $O_2$ A-, weak $CO_2$ and strong $CO_2$ bands, respectively for a thin case (AOD = 0.3).

### Low streams interpolation method

The low streams interpolation (LSI) method was developed by O'Dell [59]. The approach is similar to the double-$k$ method but it also considers the Stokes components of the radiance. Both methods make a differentiation between high-accuracy RTM and low-accuracy RTM. In the case of the double-$k$ approach, multiple-scattering and single-scattering RTMs are computed, being the latter computed in a LBL manner. However, the LSI method uses the concept of low streams (LS) for the low-accuracy RTMs because two-streams are used for the approximated radiance computations. The parametrization of the radiances as a dependence of the optical components is nevertheless similar to that defined in the double-$k$ approach. The accuracy in the context of the RTM based on discrete ordinates ($N_{\text{do}}$) or streams, is mainly dependent on the number of streams taken for the computations.

O'Dell found a correlation between the gas absorption optical depth $\tau_{\text{gas}}$ and the relative errors. Thus, the relative error was analyzed as:

$$f(\tau_{\text{gas}}) = \frac{I_{\text{LS}}(\tau_{\text{gas}}) - I_{\text{MS}}(\tau_{\text{gas}})}{I_{\text{MS}}(\tau_{\text{gas}})}. \tag{3.6}$$

Figure 3.3 shows an example of the relative errors obtained for the $O_2$ A- band and $CO_2$ band for different atmospheric scenarios.

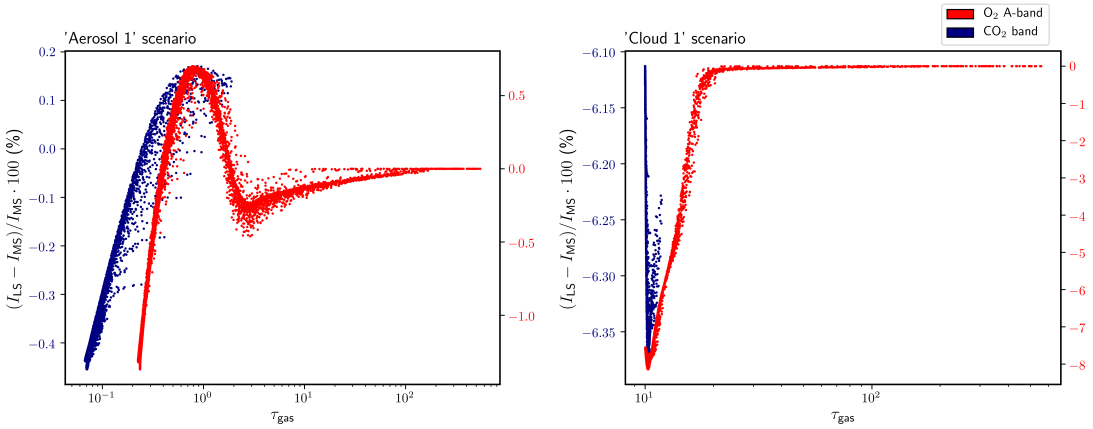


**Figure 3.3:** Relative errors of the two-stream model as a function of the gas optical depth ($\tau_{\text{gas}}$) for the $O_2$ A-band (red) and $CO_2$ band (blue) [60].

However, he encountered a major drawback: the dependence of $f(\tau_{\text{gas}})$ and $\tau_{gas}$ is non-linear and consequently, the application of the regression model requires a binning of the $\tau_{gas}$ values [60]. To overcome this drawback, O'Dell devised a two-dimensional interpolation of radiance errors in spectral bins [59] in terms of $\tau_{gas}$, $\tau'_{gas}$ (gas absorption optical depth from the TOA down to the layer in the atmosphere where the scattering optical depth equals some critical value [59]), which is defined similarly to $k'$ from the

double-$k$ approach [55]; and $\xi$, which corresponds with the multiple scattering error term as the ratio:

$$\xi = \frac{\tau'_{gas}}{\tau_{gas}}. \tag{3.7}$$

Once the errors are calculated, the corrected LBL radiances from the LS radiances are obtained as follows:

$$I = \frac{I_{\mathrm{LS}}}{1 + f(\tau_{\mathrm{gas}})}. \tag{3.8}$$

The only Stokes parameter that we are referring to in this Section is focused on the first Stokes parameter ($I$) for simplicity. For the application to other Stokes parameters, see [59]. As a result of the application of the LSI method, the retrievals from the greenhouse gases observing satellite (GOSAT) and OCO measurements were accelerated by 1-2 orders of magnitude with errors less than 0.1% for most of the atmospheres.

All the correlated-$k$-based methods described above strive to reduce the computational cost of LBL calculations although there are still some drawbacks in their application: (1) we need to assume a perfect correlation over the atmospheric column between the $k$ spectra at each height; (2) including more $k$ terms will improve the accuracy but at a higher computational cost; and (3) in order to capture the variability of the atmosphere, the correlated-$k$ methods have to be called numerous times to provide accurate reference spectra.

## 3.2 Principal Component Analysis (PCA)-based RTMs

The essential elements of the retrieval algorithms in atmospheric composition sensors are the RTMs. They convert the optical parameters of the atmosphere (input space) into the spectral radiances (output space) as explained in Section 2. In fact, the acceleration techniques based on PCA can be broadly classified into two main groups: data reduction of optical parameters and data reduction of spectral radiances. Natraj et al. [11] devised an approach which employs PCA to the RTMs in order to accelerate the LBL computations. Independently, Liu et al. [61] developed a PCA approach for accelerating hyperspectral sensors which was called the principal component-based radiative transfer model (PCRTM). Both studies lie on the application of PCA to the RTMs. Nevertheless, there are two main differences into the application of the PCA approach:

1. The data set to which the PCA is applied is different: Natraj et al. make use of the optical data (input) while Liu et al. uses the spectral data (output).

2. They use a different approach to select the wavelengths to be computed in the reduced data set of RTMs.

The main concept of PCA analysis is extracted from the review [7] (Appendix A.2) and is summarized as follows:

Let $\mathbf{y} = (y(\lambda_1), y(\lambda_2), ..., y(\lambda_W))$, $\mathbf{y} \in \mathbb{R}^{1\times W}$ be a row-vector of atmospheric radiances at $W$ wavelengths $\{\lambda_w\}_{w=1,...,W}$. A set of $S$ spectra are collected in a matrix $\mathbf{Y} \in \mathbb{R}^{S\times W}$ whose $i$-th row entries are $\mathbf{y}_i$. Thus, $\mathbf{y}_i$ can be expressed in the new basis as follows:

$$\mathbf{y}_i = \bar{\mathbf{y}} + \sum_{k=1}^{W} t_{ik}\mathbf{f_k}, \tag{3.9}$$

where $\bar{\mathbf{y}} = \frac{1}{S}\sum_{i=1}^{S}\mathbf{y}_i$ is the average spectrum, $t_{ik}$ is the $k^{th}$ coordinate of vector $\mathbf{y}_i$ in the new basis and $\mathbf{f_k} \in \mathbb{R}^{1\times W}$ is the $k^{th}$ basis vector. The spectrum $\mathbf{y}_i$ can be projected onto the $K$-dimensional subspace ($K < W$) by means of PCA as:

$$\mathbf{y}_i \approx \bar{\mathbf{y}} + \sum_{k=1}^{K} t_{ik}\mathbf{f_k}, \tag{3.10}$$

which can be rewritten in matrix form as:

$$\mathbf{Y} \approx \bar{\mathbf{Y}} + \mathbf{TF}, \tag{3.11}$$

where $\mathbf{Y} \in \mathbb{R}^{S\times W}$, $\mathbf{F} \in \mathbb{R}^{K\times W}$ and $\mathbf{T} \in \mathbb{R}^{S\times K}$ is the matrix whose entries are $\{t_{ik}\}_{i=1,\ldots,S}^{k=1,\ldots,K}$. Hereinafter, the so-called principal components (PCs) or empirical orthogonal functions (EOFs) correspond with the basic vectors $\mathbf{F}$ which are taken as the eigenvectors related to the $K$ most significant eigenvalues of the covariance matrix $\mathrm{cov}(\mathbf{Y},\mathbf{Y}) \in \mathbb{R}^{W\times W}$. Hence, the principal component scores correspond with the coordinates $t_{ik}$ in the new basis and the corresponding matrix $\mathbf{T}$.

The mathematical formulation and physical concepts of the input and output PCA techniques are explained in Chapter 4.2 and [36] (Appendix A.1). Therefore, the main PCA applications to the input and output space for RTMs are adapted from the research article [36] in the following.

### PCA applied to the input optical data space

Natraj et al. [11] aim to reduce the residual errors between the high-accurate and low-accurate RTMs in a reduced space of optical parameters by taking into account the spectrally varying scattering properties of the aerosols. In this regard, the following features apply:

- They use as low-accurate RTM a two-stream model (i.e. $N_{\mathrm{do}} = 1$) while for the high-accurate RTM they use multiple streams (i.e. $N_{\mathrm{do}} \geq 2$) for the RTM.
- The spectral intervals are grouped into bins based on $\tau_{\mathrm{gas}}$ and $\omega$ of the top layer [56].
- The dependency of the correction factor on the optical parameters is modelled by a second order Taylor expansion about the mean value of the optical parameters [62].

Figure 3.4 illustrates the methodology described above for the input data set containing optical parameters of the atmosphere. This method was first applied for simulating the $O_2$ A-band spectra obtaining an accuracy of 0.3% and one order of magnitude improvement in speed [11]. This technique has been applied to other spectral bands such as the $CO_2$ band with accuracies better than 0.1% and more than 50 fold acceleration [63] and for the GOSAT instrument [64] where the radiance residuals did not exceed 0.01% and there was an acceleration of 2 orders of magnitude compared to LBL calculations [65]. On the one hand, Kopparla et al. [66] extended this PCA technique to the wide spectral range (0.3-3000 nm) maintaining an overall accuracy of 0.01%. In this case, the spectral range was divided into 33 spectral regions depending on dominant gas absorptions within the field. On the other hand, for the same region Kopparla et al. [67] introduced a spectral binning approach together with the PCA for atmospheres containing aerosols. In this case, the choice of the bins was sensitive to the complexity of the spectral region and can lead to consecutive subdivisions of bins until the preferred accuracy level is achieved. They obtained residuals of 0.1%. Efremenko et al. [62] parallelized the PCA-based RTM

computations where as much as half of the computational time was due to the PCA itself, showing that no further acceleration of this technique was possible according to Amdahl's law [68]. Other examples are the application of PCA combined with (i) other techniques such as the $k$-distribution (e.g. [69] for cloud properties retrieval) or (ii) look-up tables (LUTs) of corrections to accelerate the RTMs [70]. In the latter research, PCA is applied to the optical measurements in the UV region (Hartley-Huggins band) by the Ozone Monitoring Instrument (OMI) with a specific spectral binning criteria based on the similarities in the logarithm of the total optical depth of gas absorption profiles ($\Gamma_g$). Then, LUTs are generated to correct the RT approximations performed using a scalar RTM with 4 streams and 24 layers. Finally, the accuracy achieved is approximately of 0.03% from PCA computations of the high-accuracy radiances obtaining an acceleration of more than 2 orders of magnitude.

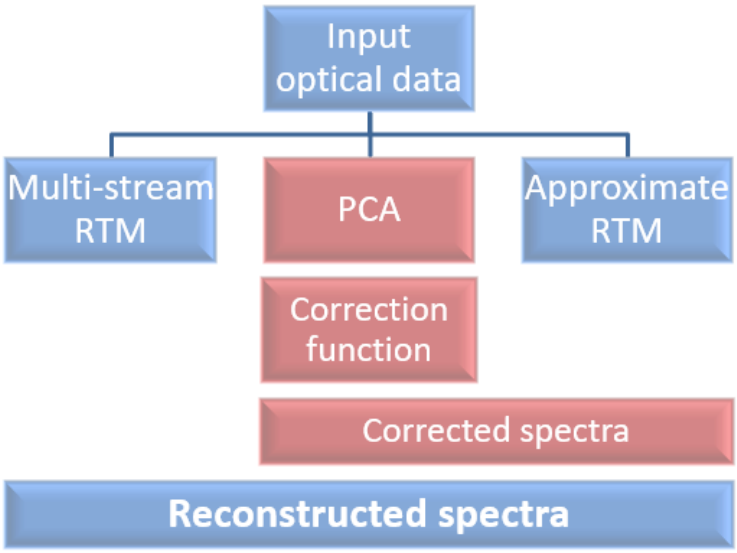


**Figure 3.4:** Schematic representation of the PCA-based RTM for the input optical data set.

### PCA applied to the output spectral data space

The output spectral reduction technique based on PCA is mainly based on mapping the spectral radiances into a lower-dimensional subspace and obtaining a set of EOFs. This approach requires a set of precomputed EOFs which is derived from a training data set of simulated or measured spectra. Then, the full-resolution spectrum can be reproduced by computing only the radiances at selected wavelengths. The choice of the wavelengths is generally different for each approach. For instance, Liu et al. [61] perform a spectral sampling which is illustrated in Figure 3.5 by using a correlation function as follows:

- The correlation coefficients are computed for the radiance values and then converted to vector angles by an arccosine function.
- The spectral data is rearranged according to the magnitudes of the correlation coefficients.
- The monochromatic radiances are selected by choosing predictors with equal distances in the values of the correlation coefficients.

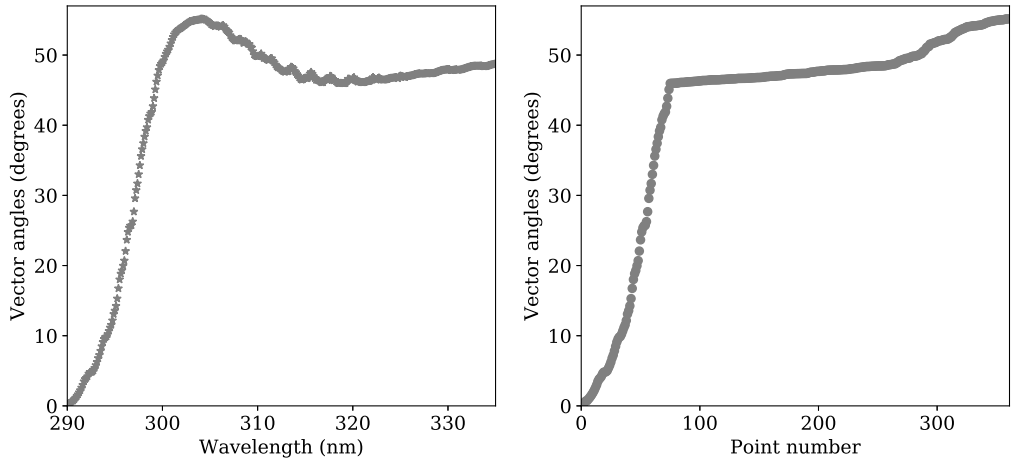


**Figure 3.5:** (**Left**) Arccosine of correlation coefficients as a function of wavelength. (**Right**) Rearranged arccosine of correlation coefficients according to their magnitudes.

Figure 3.6 represents the PCA-based RTMs for output spectral space. The data set of spectra is divided into training data set and validation data set. By applying PCA and the correlation analysis to the training set, the system of EOFs is computed and a subset of spectral points is chosen (spectral sampling), respectively. These two outputs are stored and used for computing PC scores for the validation data set. The spectra in the full wavelength range are restored and the error of this reconstruction can be estimated. If the error is larger than required, the number of generated spectra and the number of principal components are increased. The main output of the training phase are EOFs and spectral sampling, which allow to process new data in the online phase.

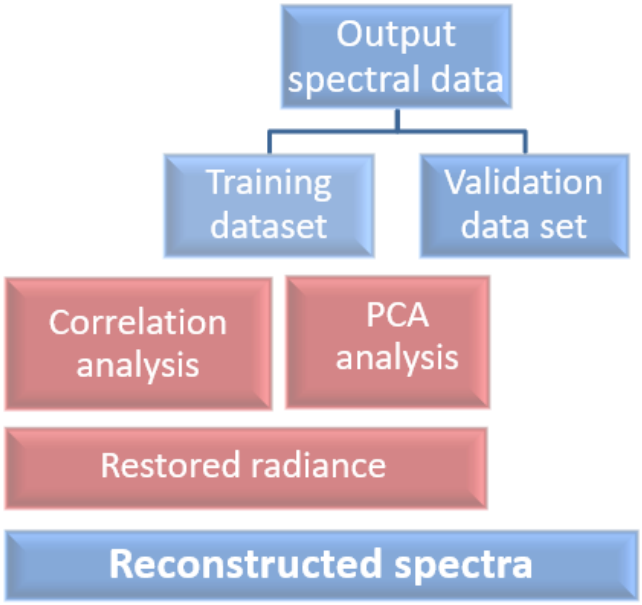


**Figure 3.6:** Schematic representation of the PCA-based RTM for the output spectral data set.

In [61], it was noted that the slit function convolution operator and the PCA are linear. Therefore the PC scores of the convolved spectra are linear functions of monochromatic radiances. Then at the training stage, the corresponding weighting factors of linear de-

pendency are stored together with the system of EOFs and spectral sampling. Authors claim that for an IR spectrum ranging from 650 $cm^{-1}$ to 3000 $cm^{-1}$ this representation for convolved spectra reduces the number of monochromatic computations from a few thousands to a few hundreds. The PCRTM approach proved to be efficient and implemented in different instruments: NAST-I and AIRS [61] and several spectral ranges such as for the solar spectral region (PCRTM-SOLAR) [71] and for the IR to UV-VIS [72] obtaining relevant speed improvements. Moreover, the PCA-based RTM to spectral data has been included in several packages such as the RTTOV fast RTM (PC_RTTOV) by means of a regression scheme [73] which reproduces the LBL radiances in a much higher degree of accuracy at less computational effort. Another application of PCA to the output space is given by Hollstein and Lindstrot [74], where they test several spectral sampling techniques such as equal sampling and random walks for the selection of the radiances. A recent study combines PCA for the input and output space. The method is called spectral data compression (SDCOMP) RTM which simulates hyperspectral resolution radiances in the solar and IR regions. They use PCA twice obtaining relative errors less than 0.2% and approximately three orders of magnitude faster than numerically exact RT computations [75].

## 3.3 Neural Networks (NN)-based RTMs

The first to use neural networks (NN) methods to accelerate the RT computations were Key and Schweiger [76] in 1998 for broadband calculations. Later on, Schawander et al. [77] implemented a NN for the VIS and UV spectral ranges computations. In this case, a RTM is used to calculate radiances at a few selected wavelengths (only 7) at high-accuracy while the NN restores the complete spectrum by learning this task with more than 20000 index values for all kinds of atmospheres. A scheme of the structure of the NN used for modelling RTMs is in Figure 3.7. The input layer had nine neurons consisting of transmittances at seven wavelengths, solar zenith angle and total ozone. The output layer had 153 neurons of high-spectral resolution transmittances. The weights of the NN were refined from the comparison of the errors between the output and the target values. The computational performance of their approach was reduced by a factor of more than 20. Several techniques have also utilized NN for approximating solar radiation (e.g. [78, 79]),

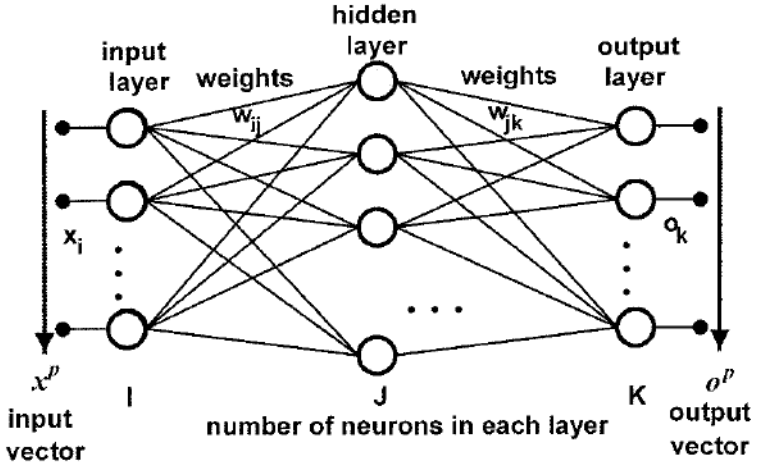


**Figure 3.7:** Scheme the structure of a NN used for RTMs in [77].

to process the EO satellite data [80], or for the geostationary satellite MTSAT-1R and ADEOS-II/GLI [81]. Additionally, [82] reported that the NN reproduced the convolved spectra with a mean absolute error below 1% for the $O_2$ A-band. As for the 350-1050 nm spectral range, [83] reproduced with a NN the data acquired by the picosatellite for remote sensing and Innovative Space Missions (PRISM) imaging spectrometer taking 5 ms per spectrum, while the computations based on monochromatic RTMs generally take 10 min per spectrum (i.e. it provides around three order of magnitude performance enhancement). One of the latest applications of NN for fast hyperspectral RTM simulations is given by Le et al. [84]. They performed computations at a small fraction of hyperspectral wavelengths and extended them across the entire spectral range. The results were compared with PCA model for the OCO-2 simulations, obtaining relative errors less than 0.5% compared to high-accurate computations.

In sum, the computationally expensive RTMs in atmospheric remote sensing can be substituted by NN in order to obtain a multi-fold performance enhancement. However, there are some drawbacks of such an approach as:

- Requirements of extensive computations of the training data sets.
- Restriction to certain atmospheric model with predefined parameters (such as atmospheric layers or fixed slit functions) of the training procedure.
- Loss of generality, i.e., the NN might not reproduce implicit properties of the radiance field such as the symmetry relationship with respect to the incident and viewing polar angles.
- Robustness issue: Large errors might not be captured by the NN-based RTM.

These drawbacks might lead to non-physical results such as the violation of the reciprocity principle for the radiance field [85] which have to be treated carefully.

# 4 Summary of the Work

The objectives of the thesis proposed in Chapter 1, have been accomplished in four journal articles, one review and one conference paper, all of which are peer-reviewed and I am the first author. This chapter provides a summary of the peer-reviewed contributions except for the review since it has been previously summarized in Chapter 3. In the following, the research articles previously published are adapted and summarized in this Chapter as follows:

- Section 4.1 outlines the general considerations of the simulations performed for all the peer-reviewed contributions. Additionally, the software interfaces and the RT code used in the dissertation are integrated in a scheme.
- In Section 4.2, a hybrid PCA technique is developed for fast simulation of the spectral radiances in the Hartley-Huggins band. The content of this section is adapted from the journal article in Appendix A.1.
- In Section 4.3, a new acceleration technique called the Cluster Low-Streams Regression (CLSR) method is implemented for the $O_2$ A- and $CO_2$ bands. This technique is compared with a PCA-based RTM in terms of accuracy and computational performance. The content of this section is adapted from the journal article in Appendix A.3.
- In Section 4.4, the CLSR method is further analysed and tested for other spectral bands and atmospheric scenarios. The content of this section is adapted from the journal articles in Appendix A.4 and A.5.
- In Section 4.5, two modifications of the CLSR method are proposed: (i) to further improve the computational performance (double CLSR method) of the hyperspectral RTMs and (ii) to alternatively estimate aerosols' spectra based on the clear sky spectra. The content of this section is adapted from the journal article in Appendix A.6.

The general structure of the sections comprises a motivation for each study, data overview, methodology, results and conclusions.

## 4.1 General Considerations

This section summarizes the general considerations of this thesis for the spectral simulations. First, the atmospheric and gaseous considerations are presented followed by the RT simulations. Finally, the software interfaces used to simulate the spectra of the atmosphere are related and integrated into a scheme.

#### Atmospheric considerations

The reflected spectral radiances at the TOA are considered for all the studies. In general, for all the simulations the atmosphere is discretized with a step of 1 km between 0 and 25

km, and a step of 2.5 km between 25 km and 50 km, resulting in 35 layers. The chosen altitude grid is used for the purpose of comparing the computational speed of different models. The boundary conditions at the bottom are defined by the Lambertian surface with an albedo of 0.3. The solar zenith angle, the viewing zenith angle and the relative azimuth angle are 45°, 35° and 90°, respectively.

Regarding the simulated atmospheric scenarios, the following possibilities can be found in the research journals:

- *Clear sky:* This scenario corresponds to an atmosphere without clouds and aerosols, i.e. only Rayleigh scattering.
- *Aerosols and clouds:* The simulated aerosols and clouds are obtained from the optical properties of aerosols and clouds (OPAC) database [86].

#### Gaseous considerations

The high spectral resolution computations of the TOA radiances in the Hartley–Huggins (280 – 335 nm), $O_2$ A-band (755 – 775 nm), water vapour (770 – 1000 nm) and $CO_2$ band (1590 – 1620 nm) are simulated. The gaseous absorptions for the $O_2$ A-, water vapour and $CO_2$ bands are computed with the LBL model Py4CAtS [87], while the ozone absorption cross-sections in the Hartley-Huggins band are taken from the high-resolution transmission molecular absorption database (HITRAN) [88]. Hence, the absorption coefficients are pre-computed and stored. Continuum (also referred to as collision-induced absorption (CIA) [89, 90]) contributions to molecular absorption are not taken into account.

The Rayleigh cross-sections and depolarization ratios are computed as in Bodhaine et al. [91], while the pressure and temperature profiles are obtained from the US standard model atmosphere [92]. The computations are performed for the unit solar irradiance at the TOA.

#### Radiative transfer model simulations

Many existing RTMs are based on the discrete ordinate method [93]. The radiative transfer solver used for all simulations is based on the discrete ordinates with matrix exponential (DOME) method [26, 34] as explained in Chapter 2. The model is supplied with the left eigenvector approach [94, 34] based on the scaling transformation [95, 96].

In this method, the number of discrete ordinates (streams) in the polar hemisphere $N_{\mathrm{do}}$ controls the computational time and accuracy. The model is called multi-stream (MS) when $N_{\mathrm{do}} \geq 2$ and low-stream (LS) otherwise. Specifically, the case $N_{\mathrm{do}} = 1$ is the two-stream (TS) model. The simplest RTM is the single-scattering (SS) model, which solves the radiative transfer equation neglecting the integral term [97], i.e. the multiple-scattering events, and the solution can be derived analytically without using the discrete ordinate method. The boundary value problem for the multilayer atmosphere is solved by using the matrix operator method [98], which merges layers into a single layer. The radiance along a viewing direction is computed by using the false discrete ordinate method [99, 94]. Regarding the TS model, two comments are in order:

- the radiation field is found as a sum of the single-scattering solution and the multiple-scattering term;
- in the multiple-scattering computations the delta-M scaling method [100] is applied, while for the single-scattering term the exact phase function (TMS-correction) is used [101].

In addition, when we refer to line-by-line (LBL) computations, we are referring to computations on a fine spectral resolution grid.

### Software interfaces

The software package used in this thesis for the RT solver is PYDOME, which is a Python library for radiative transfer computations implemented in python3 and relies on the NumPy library [102]. The RT code has been developed at the German Aerospace Center (DLR) for the simulation of satellite-based measurements of reflected and scattered solar radiation in the UV and VIS spectral ranges [103]. This tool incorporates the DOME solver ([26, 34]) (see Chapter 2) and other approximate solvers such as the single-scattering and the two-stream solvers. The processing chain of PYDOME has three main steps:

1. Physical adapter
2. Radiative transfer solver
3. Post-processing

In the first step, the physical parameters of the atmosphere are prepared in order to calculate the optical properties to finally be included as input into the code. Thus, physical parameters such as the atmospheric profiles of the temperature, pressure, gas concentration, aerosols, clouds etc. are taken into account for computing the optical properties: optical thickness, single scattering albedo, and phase functions, which are the input of PYDOME. The latter are computed by using the molecular cross-sections. In order to obtain the gaseous absorption parameters, the HITRAN database is used as input of the Py4CAts model [87] to get the LBL gaseous absorption cross-sections. For cases such as ozone absorption in the UV region, the pre-convolved cross-sections from HITRAN can be used directly. The scattering cross-sections are obtained with the model of Bodhaine et al. [91] for Rayleigh scattering. Once the absorption and scattering cross-sections of the atmosphere are computed, the aerosol or cloud optical properties can be included in the model. Therefore, the aerosol absorption/scattering cross-sections are modelled with the OPAC database [86] and by means of an interpolation module, the aerosol absorption/extinction properties for the vertical profiles of the atmosphere are obtained.

For the second step, the RT solver is applied using as input the optical properties generated above. The radiance field is obtained as a function of the location and observation geometry. As previously mentioned, the exact solver employed is the DOME solver based on the 1-D scalar RTM for coupled atmosphere-ocean systems.

The post-processing step depends on the relevant values to be derived for a certain application. For instance, the convolved spectra is required for numerous applications. Hence, the convolved radiance is obtained for a specific space-borne instrument spectral response. Furthermore, the performance of the RT code can be improved with a new module to be included for accelerating the hyperspectral computations.

Figure 4.1 shows a scheme for the integration of the PYDOME solver with the additional software interfaces for computing atmospheric properties.

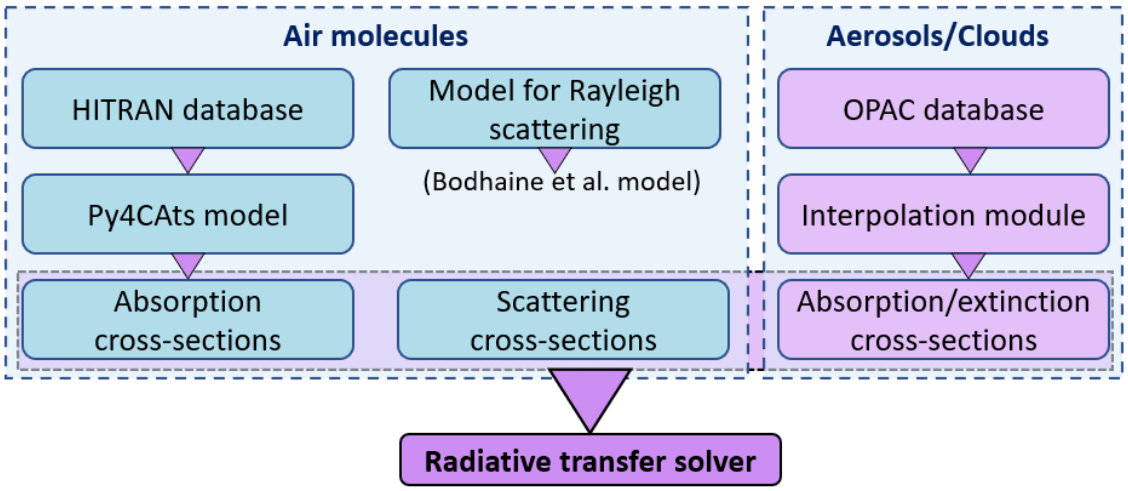


**Figure 4.1:** Diagram of the software interfaces used for the thesis.

## 4.2 Analysis of Two Dimensionality Reduction Techniques: Hybrid PCA Approach

### 4.2.1 Motivation

The new generation of atmospheric composition satellites require efficient dimensionality reduction techniques for fast simulation of the spectral radiances. In this work ([36], cf. Appendix A.1), we investigate the efficiency of two acceleration techniques based on the principal component analysis (PCA). First, we exploit the PCA to map the data set of optical properties (input space) of the atmosphere to a lower-dimensional subspace to finally derive a faster RTM by means of a correction function. Second, the spectral data (output space) are also accelerated by representing the hyperspectral data via a subset of empirical orthogonal functions (EOFs) and a combination of a spectral splitting approach. In this context, we propose a hybrid PCA approach which combines the dimensionality reduction techniques on the input and output data set to improve the hyperspectral data performance in the Hartley-Huggins band for simulating TROPOMI signals in the UV spectral range, which is typically used for ozone retrievals [104].

### 4.2.2 Data overview

The backscatter signal of the TROPOMI instrument and the spectra are computed with a spectral resolution of 0.125 nm in the spectral interval between 290 and 335 nm. The spectra and the first three EOFs of the TROPOMI data are illustrated in Figure 4.2 together with the explained variance. The simulations are performed for a clear sky model atmosphere and a Lambertian surface albedo of 0.1. The total optical property inputs are given by Rayleigh scattering [91] by atmospheric molecules and trace gas absorption. The Brion–Daumont–Malicet cross sections are used for ozone. The atmosphere is discretized into 14 layers. The top of the atmosphere is at 50 km. The solar zenith angle, viewing zenith angle and relative azimuth angle are 45°, 35° and 90°, respectively.

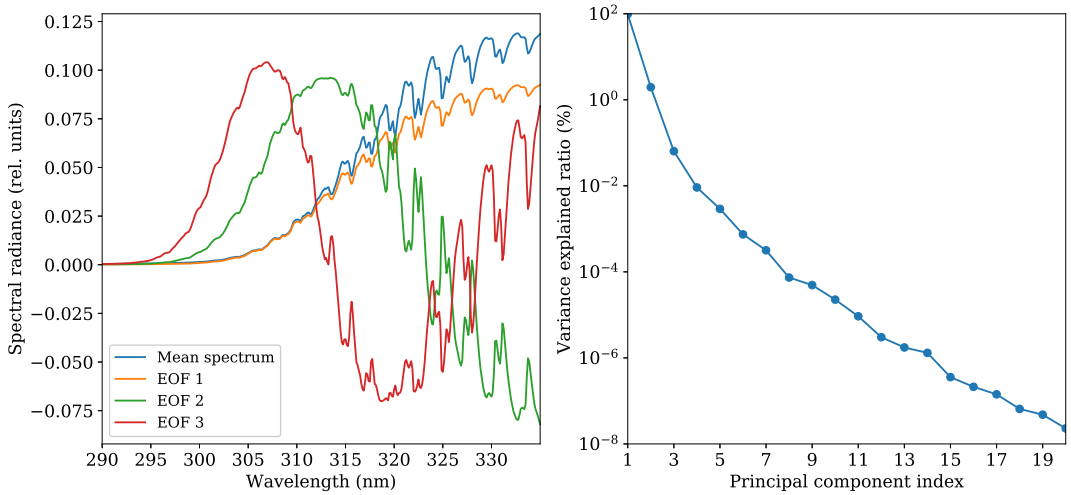


**Figure 4.2:** (**Left**) First three orthogonal functions computed in the Hartley-Huggins band. (**Right**) Explained variance in percentage as a function of the principal component index for the Hartley-Huggins band simulation.

### 4.2.3 Methodology

#### Hybrid PCA-based RTM

The hybrid PCA-based RTM consists in the combined application of PCA for the input and output spaces. A schematic representation of both techniques is given in Figure 4.3. Note that, although both techniques are often called the same (namely "PCA-based RTMs"), they are fundamentally different. In the first case, PCA is applied to a data set of optical parameters, comprising the total optical thicknesses and the single scattering albedos for all given atmospheric layers and wavelengths. In the second case, PCA is applied to the data set of spectral radiances. In the next, both techniques are explained in detail.

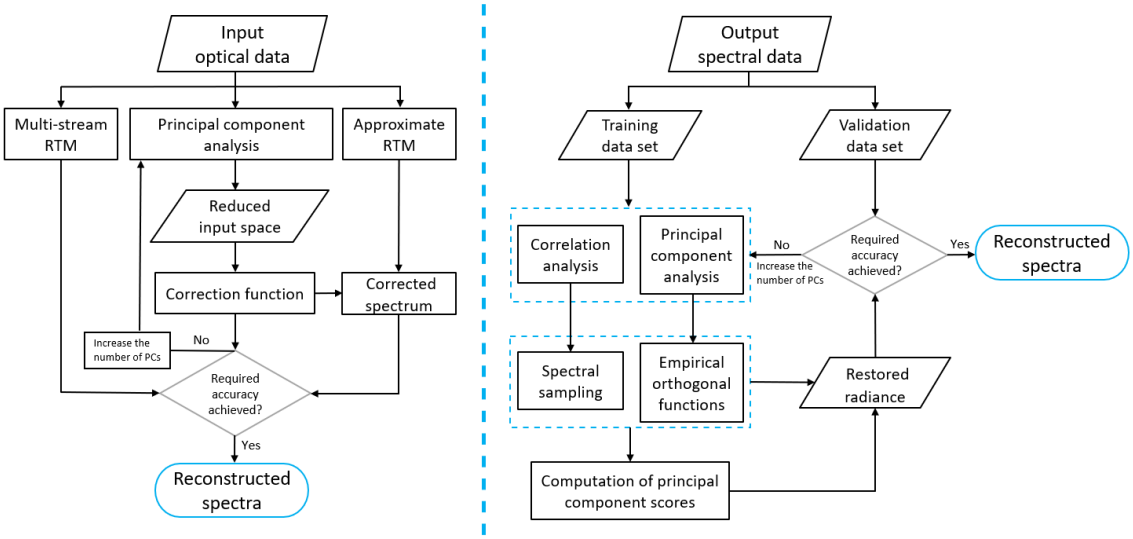


**Figure 4.3:** Schematic representation of the data processing in (**left**) the input space reduction technique with PCA and (**right**) output space reduction technique with precomputed EOFs.

### Input space reduction technique

Considering a discretization of the atmosphere in $L$ layers, we define an $N$-dimensional vector $\mathbf{x}_w$, for each wavelength $\{\lambda_w\}_{w=1,\ldots,W}$, by

$$\mathbf{x}_w^{\mathrm{T}} = [\ln \tau_1(\lambda_w), \ldots, \ln \tau_L(\lambda_w), \ln \omega_1(\lambda_w), \ldots, \ln \omega_L(\lambda_w)], \tag{4.1}$$

where $\tau_k$ and $\omega_k$ are the optical thickness and the single-scattering albedo of the $k$-th layer, respectively, and $N = 2L$. Thus, the vector $\mathbf{x}_w$ encapsulates the wavelength variability of the optical parameters, which are the input parameters of the radiative transfer code. By using PCA, we find an $M$-dimensional subspace spanned by a set of linear independent vectors (empirical orthogonal functions, EOFs) $\{\mathbf{q}_k\}_{k=1}^{M}$ such that the centered (mean-removed) data $\mathbf{x}_w - \bar{\mathbf{x}}$ lie mainly on this subspace, i.e.

$$\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^{M} y_{wk}\mathbf{q}_k, \quad w = 1, \ldots, W, \tag{4.2}$$

where $y_{wk}$ are the principal component (PC) scores.

Let us define a correction function $f(\mathbf{x}_w)$ as follows:

$$f(\mathbf{x}_w) = \ln \frac{I_{\mathrm{MS}}(\mathbf{x}_w)}{I_{\mathrm{LS}}(\mathbf{x}_w)}, \tag{4.3}$$

where $I_{\mathrm{LS}}(\mathbf{x}_w)$ is the radiance provided by the approximate model (here LS stands for "low-streams" and refers to either SS or TS models), and $I_{\mathrm{MS}}(\mathbf{x}_w)$ is the radiance simulated by the MS model. Introducing $\Delta\mathbf{x}_w = \sum_{k=1}^{M} y_{wk}\mathbf{q}_k$, we consider the Taylor expansion of $f(\mathbf{x}_w)$ around $\bar{\mathbf{x}}$ in the direction $\Delta\mathbf{x}_w$ up to fourth order. For computational simplicity, we neglect the mixed directional derivatives and use central differences to approximate the directional derivatives. Since $M < N$ (and in practice $M \ll N$), it leads to a substantial reduction of the computational time. Once the correction function is computed, the results of the approximate RTM can be converted into those of the MS RTM by using Equation (4.3).

### Output space reduction technique

Let $\mathbf{y} = [y(\lambda_1), y(\lambda_2), \ldots, y(\lambda_W)]$ be a row-vector of radiances at $W$ set of wavelengths $\{\lambda_w\}_{w=1,\ldots,W}$. A set of $S$ spectra is assembled into a matrix $\mathbf{Y}$ whose $i$-th row is $\mathbf{y}_i$. Then, a spectrum $\mathbf{y}_i$ can be represented in a new basis system as follows:

$$\mathbf{y}_i = \overline{\mathbf{y}} + \sum_{j=1}^{W} t_{ij}\mathbf{f}_j. \tag{4.4}$$

Here, $\overline{\mathbf{y}} = \sum_{i=1}^{S} \mathbf{y}_i/S$ is the sample mean of the spectra (the average spectrum), $t_{ij}$ is the $j$-th coordinate of the vector $\mathbf{y}_i$ in the new basis system and the $j$-th basis vector is $\mathbf{f}_j = [f_j(\lambda_1), f_j(\lambda_2), \ldots, f_j(\lambda_W)]$. Next, the spectrum $\mathbf{y}_i$ is projected onto the $J$-dimensional subspace ($J < W$) as follows:

$$\mathbf{y}_i \approx \overline{\mathbf{y}} + \sum_{j=1}^{J} t_{ij}\mathbf{f}_j, \tag{4.5}$$

The transformation (4.5) can be done by using PCA [105]. Then the basic vectors $\mathbf{f}_j$ in Equation (4.5) are referred to as "principal components" (PCs) or empirical orthogonal functions (EOFs), while the coordinates $t_{ij}$ in the new coordinate system are called "principal component scores".

In order to reconstruct the full resolution spectrum, the previous theory reveals a linear relationship between PC scores and monochromatic radiances:

$$\mathbf{y}(\lambda) \approx \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda) + \ldots + t_J\mathbf{f}_J(\lambda). \tag{4.6}$$

This approach requires a set of precomputed EOFs which is derived from a training data set of simulated spectra. Hence, for a given set of $J$ EOFs and $J$ spectral points it is possible to obtain a closed linear system of $J$ equations:

$$\begin{cases} \mathbf{y}(\lambda_1) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_1) + \ldots + t_J\mathbf{f}_J(\lambda_1), \\ \mathbf{y}(\lambda_2) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_2) + \ldots + t_J\mathbf{f}_J(\lambda_2), \\ \vdots \\ \mathbf{y}(\lambda_J) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_J) + \ldots + t_J\mathbf{f}_J(\lambda_J). \end{cases} \tag{4.7}$$

The key point here is that the radiance values in $J$ spectral points $\{\lambda_j\}_{j=1,\ldots,J}$ are represented through the same EOFs. Then, by solving Equation (4.7) we obtain the PC scores, and by using Equation (4.5) the full spectrum at $W$ spectral points can be readily restored.

As for the selection of the monochromatic wavelengths, the method proposed is the spectral sampling from Liu et al. [61] (see Chapter 3). Thus, by applying PCA and correlation analysis to the training data set, the system of EOFs is computed and an appropriate subset of spectral points is chosen. These two outputs are stored and used later for computing the PC scores for the validation data set.

### 4.2.4 Results

#### Input space reduction

To reduce the variability of the optical parameters, we divided the spectral range 290–303 nm in disjoint regions. Three cases were considered using the TS and SS models as approximate spectra. Thus, the best agreement with the MS solution was found for the third case (see [36]), in which three sub-intervals are considered: 290-303 nm, 303-321 nm and 321-335 nm. The input space reduction technique is applied to the second and third sub-intervals while for the first one we use the SS model. In Figure 4.4a the spectra of the RTMs are plotted while in Figure 4.4b the corrected TS and SS spectra are shown for the third case. This means that the information about the MS solution is contained in the optical data and can be retrieved by using machine learning algorithms, and the input space reduction technique can be considered as representative of them.

We analyzed the dependency of the errors and computational time in terms of the number of PCs and the order of the Taylor series. We found that the errors decrease when $M$ and the expansion order of the Taylor series increase. However, the computational performance behaves in an opposite way. Therefore, we found the optimal compromise between accuracy and computational performance for $M$=3 and 2 orders of expansion for the TS model.

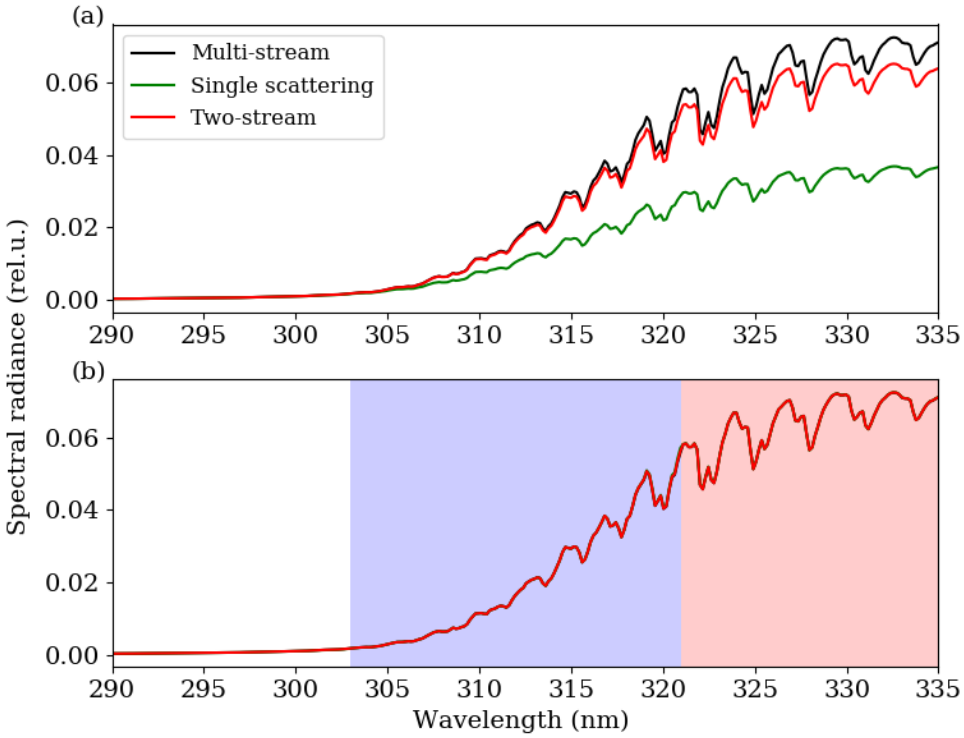


**Figure 4.4:** (**a**) Spectral radiance computed with the single-scattering, two-stream and multi-stream models; (**b**) Spectral radiance after applying the correction within the intervals 303–321 nm (blue area) and 321–335 nm (red area).

### Output space reduction

The efficiency of the output space reduction technique is tested for a data set consisting of $2 \times 10^5$ spectra (see [36]). The data set of spectra is divided into a training data set and a validation data set. By applying PCA to the training data set, the system of EOFs is computed. After applying the spectral sampling, $J = 30$ wavelengths were identified and the EOFs and $J$ wavelengths are saved in the memory and reused for computing PC scores for the validation data set. The spectra at a full resolution are restored using Equation (4.6). Finally, the mean relative error of the spectra is 0.00023% with a standard deviation of 0.12%. Using $J = 30$ spectral points provides a performance enhancement of about 12 times. A higher amount of principal components ($J > 30$) assures a higher accuracy and a more robust result.

### Hybrid PCA approach

The accuracy and computational performance as a result of combining the input and output space reduction techniques is shown in Table 4.1 [36].

**Table 4.1:** Acceleration factor and mean error (in %) for the input space reduction method, the output space reduction method and a combination of both methods.

| | **Input Space Reduction** | **Output Space Reduction** | **Combined Use** |
|---|---|---|---|
| Acceleration factor | 13 | 12 | 18.2 |
| Mean error | 0.05 | 0.00023 | 0.05 |

The output space reduction technique is applied to accelerate the computations of the approximate solution (in this case the TS model) exhibiting that the combined use of both methods (hybrid PCA approach) results in an efficient approach to accelerate the input and output data without losing accuracy.

### 4.2.5 Conclusions

The following main conclusions can be extracted from this study ([36] in Appendix A.1):

- The input optical space reduction technique does not require a precomputed data set and can be regarded as a sort of RTM with ad hoc learning. The output spectral space reduction technique can be applied to real measurements. Both techniques achieve an order of magnitude speed improvement.

- The combination of both techniques is called the hybrid PCA approach, which uses the output space reduction technique to speed up the TS computations in the framework of the input space reduction technique, providing an overall speedup factor of about 20. The hybrid technique is as accurate as the input technique applied separately.

## 4.3 The Cluster Low-Streams Regression (CLSR) Method for Hyperspectral Radiative Transfer Computations

### 4.3.1 Motivation

Current atmospheric composition sensors provide a large amount of hyperspectral resolution data. There are accurate acceleration techniques such as PCA-based RTMs that reduce the number of monochromatic computations. In this line, we developed a new acceleration technique to further improve the state-of-the-art techniques in a study published in [60] that can be found in Appendix A.3. The motivation of developing a new acceleration technique was twofold: first, to find an alternative method which provides faster and accurate results for processing hyperspectral radiative transfer computations and second, we found that the LSI method (see Chapter 3) by O'Dell [59] was subject of improvement. Therefore, we developed a method based on the clustering of the spectral radiances computed with a LS RTM and the regression analysis performed for the LS and MS RTMs within each cluster. Specifically, we found that the dependence between the low- and multi-stream radiances is almost linear (see Figure 4.5) and it seemed reasonable to cluster the spectral points according to the radiance values computed with a low-stream RTM, rather than with the optical properties. Hence, this motivated the development of the Cluster Low-Streams Regression (CLSR) method, in order to decrease the errors of the regression estimates and further accelerate the whole computational performance of the techniques. Moreover, a detailed comparison between the PCA-based RTMs and the CLSR method is provided.

### 4.3.2 Data overview

For this study we consider the reflected spectral radiance at the TOA of two spectral bands. In the $O_2$ A-band the spectral sampling is 0.001 nm in the spectral range 755–775 nm, while in the $CO_2$ band the spectral sampling is 0.0015 nm in the spectral interval 1590–1620 nm. Thus, on each band, 20,000 spectral points are considered.

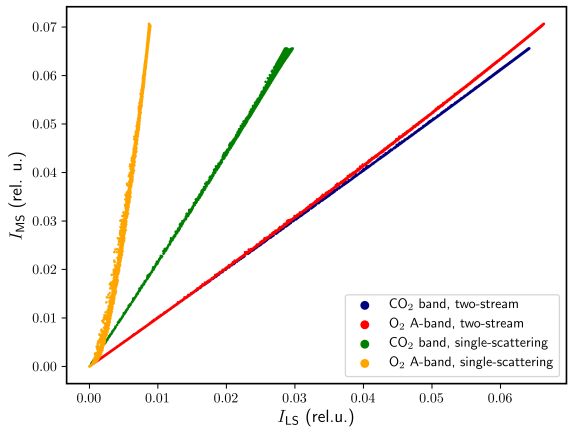


**Figure 4.5:** Radiance computed with the multi-stream model as a function of radiances computed by using the low-stream models (two-stream and single-scattering models) for the $O_2$ A- and $CO_2$ bands. The figure corresponds to the 'Aerosol 2' scenario.

For the simulations, 5 atmospheric scenarios are considered: 'Clear sky', 'Aerosol 1', 'Aerosol 2', 'Cloud 1' and 'Cloud 2'. The 'Clear sky' scenario corresponds to an atmosphere without clouds and aerosols. In the 'Aerosol 1' and 'Aerosol 2' scenarios, the atmosphere contains the clean continental and the polluted continental aerosols taken from the OPAC database [86], respectively. In the 'Cloud 1' and 'Cloud 2' scenarios, a continental clean cumulus cloud with cloud optical depths $\tau = 10$ and $\tau = 20$, respectively. In both cases, the cloud top height is 5 km and the cloud geometrical thickness is 1 km. As an overview of the data, Figure 4.6 shows the spectral radiance for three atmospheric scenarios ('Clear sky', 'Cloud 1' and 'Cloud 2').

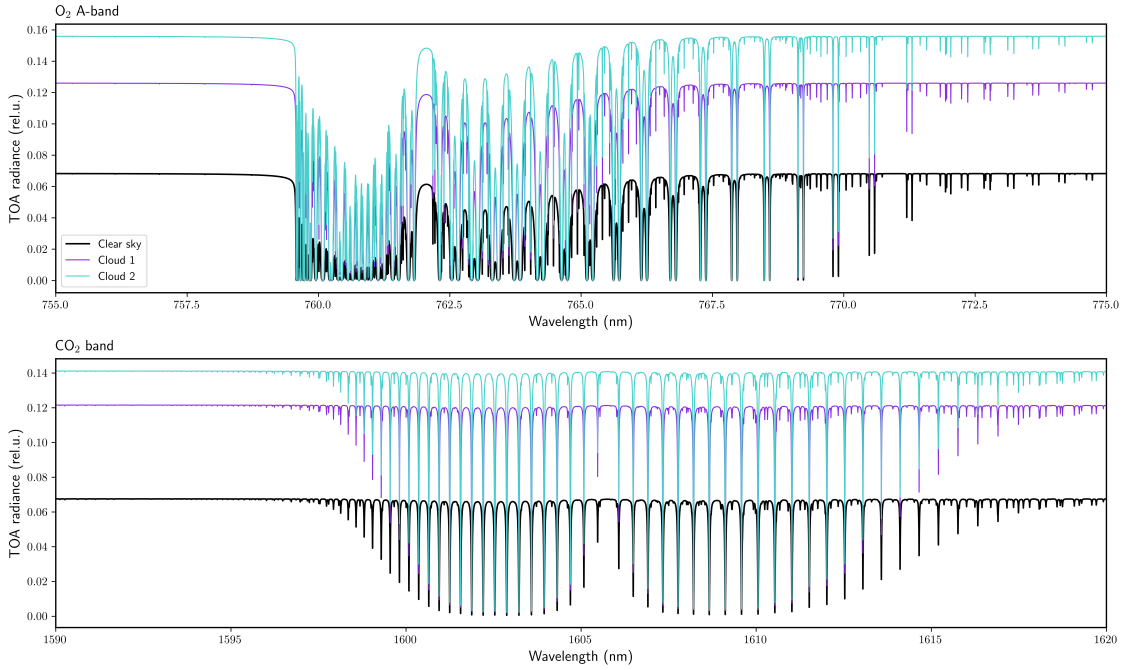


**Figure 4.6:** Radiance spectra computed by using the multi-stream RTM for three atmospheric scenarios: 'Clear sky' (black), 'Cloud 1' (purple) and 'Cloud 2' (blue). The upper panel corresponds to the $O_2$ A-band, while the bottom panel is for the $CO_2$ band.

### 4.3.3 Methodology

#### The CLSR method

The Cluster Low-Streams Regression (CLSR) method [60] can be formulated as follows.

1. Consider a high resolution spectrum $\{I_{\mathrm{LS}}(\lambda_i)\}_{i=1}^{N}$ computed at $N$ spectral points $\{\lambda_i\}_{i=1}^{N}$ by means of a low-stream RTM.

2. Sort the radiance set $\{I_{\mathrm{LS}}(\lambda_i)\}_{i=1}^{N}$ in ascending order, and let $\{\hat{I}_{\mathrm{LS},i}\}_{i=1}^{N}$, with $\hat{I}_{\mathrm{LS},i} \leq \hat{I}_{\mathrm{LS},i+1}$, be the sorted radiance set (Figure 4.7a).

3. Consider $C$ clusters in $\{\hat{I}_{\mathrm{LS},i}\}_{i=1}^{N}$ with $N_C = N/C$ radiance points (Figure 4.7b), and let the $c$ cluster be defined by the radiance set $\{\hat{I}^c_{\mathrm{LS},i}\}_{i=1}^{N_C}$ for $c = 1, ..., C$.

4. Select $n$ equidistant radiance points in the $c$ cluster, i.e. $\{\overline{I}^c_{\mathrm{LS},q}\}_{q=1}^{n}$, and for the corresponding wavelengths compute the multi-stream radiances $\{\overline{I}^c_{\mathrm{MS},q}\}_{q=1}^{n}$ (Figure 4.7c).

5. Assume that in each cluster $c$ we have the linear relationship

$$\hat{I}^c_{\mathrm{MS},i} = \alpha^c \hat{T}^c_i + \beta^c \hat{I}^c_{\mathrm{LS},i} + \gamma^c, \; i = 1, ..., N_C, \tag{4.8}$$

   where $\alpha^c$, $\beta^c$ and $\gamma^c$are the regression coefficients of the $c$-th cluster, and $\hat{T}$ is the corresponding direct transmittance.

6. Compute the regression coefficients $\alpha^c$, $\beta^c$ and $\gamma^c$ as a solution to the least square problem

$$(\alpha^c, \beta^c, \gamma^c) = \arg \min_{\alpha^c, \beta^c, \gamma^c} \sum_{q=1}^{n} \left[\overline{I}^c_{\mathrm{MS},q} - \left(\alpha^c \overline{T}^c_q + \beta^c \overline{I}^c_{\mathrm{LS},q} + \gamma^c\right)\right]^2. \tag{4.9}$$

7. Use the values of $(\alpha^c, \beta^c, \gamma^c)$ found in the previous step to restore the multi-stream radiances $\{\tilde{I}_{\mathrm{MS},i}\}_{i=1}^{N}$ for all the spectral points according to Equation (4.8) (Figure 4.7d).

Here, the "hat" notation $\hat{I}$ refers to the sorted radiances, the "bar" notation $\overline{I}$ refers to the equidistant radiances entering the regression model, and the "tilde" notation $\tilde{I}$ refers to the predicted radiances. Please note that the total number of regression points, and thus the number of calls of the multi-stream model, is $nC$.

#### Efficiency and computational performance estimations

To estimate the accuracy of the acceleration techniques, we consider the residual error for the radiance at each spectral point $\lambda_i$:

$$\Delta I_{\mathrm{res},i} = \frac{\tilde{I}_{\mathrm{MS},i} - I_{\mathrm{MS},i}}{I^{\mathrm{cont}}_{\mathrm{MS}}} \cdot 100, \tag{4.10}$$

where $\tilde{I}_{\mathrm{MS},i}$ is the predicted radiance calculated with either the PCA-based RTM (as explained in 4.2.3) or the CLSR method (cf. Equation (4.8)), while $I^{\mathrm{cont}}_{\mathrm{MS}}$ is the radiance without absorption (i.e., the continuum radiance, which is used to avoid radiance values close to zero in the denominator of Equation (4.10), when strong gas absorption is present [67]). The mean relative error is computed as follows:

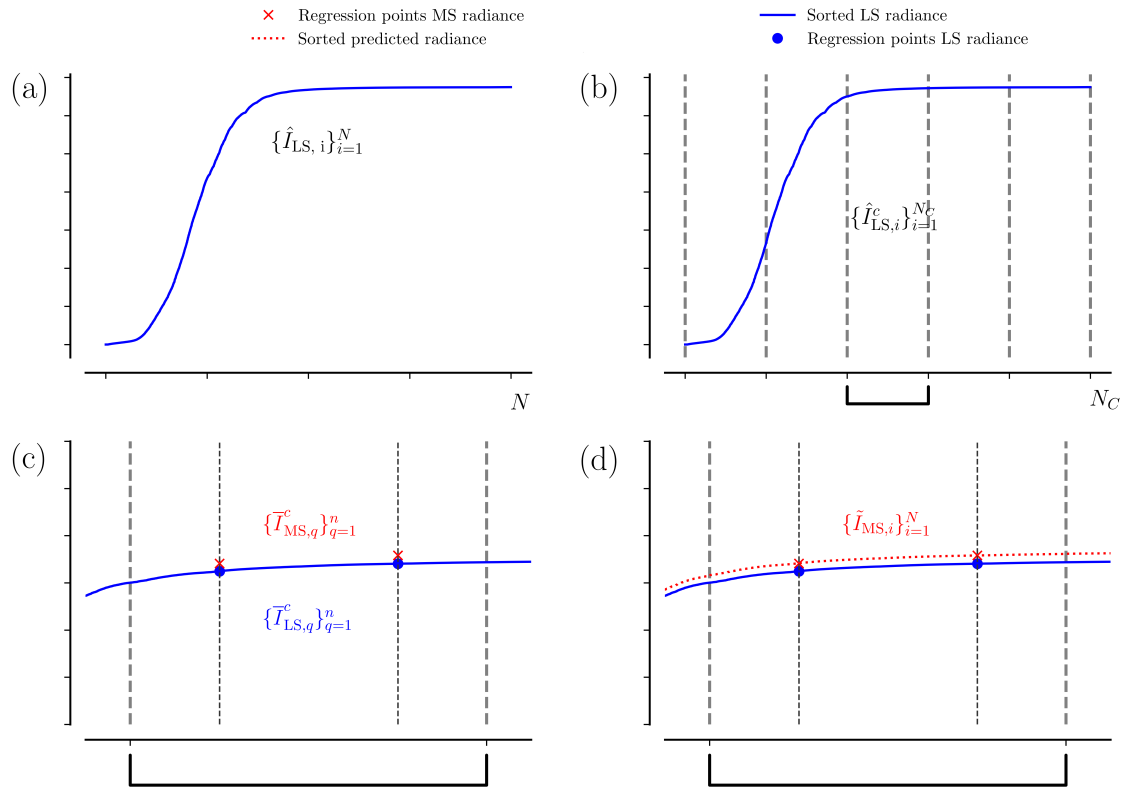


**Figure 4.7:** Scheme of the Cluster Low-Streams Regression (CLSR) method [60]. (**a**) Sorted radiance of the low-stream (LS) model in ascending order (blue line). (**b**) Division of the LS radiance in equal clusters $C$ in the sorted domain. (**c**) Zoom for one cluster and the selected regression points of the multi-stream (MS) radiance (red crosses). (**d**) Reconstruction of the MS spectra: the predicted radiance is computed for all the spectral points (dashed red line).

$$\varepsilon = \frac{\sum_{i=1}^{N} |\Delta I_{\mathrm{res},i}|}{N}. \tag{4.11}$$

Additionally, the median and interquartile range (IQR) of the residual are computed.

To estimate the performance enhancement, we define the speedup factor as the ratio between the computational time of a multi-stream LBL calculation to that of a certain acceleration technique. To exclude the hardware-related factors from our analysis, we estimate the speedup factor for the PCA-based RTM as [65]:

$$S_{\mathrm{PCA}} = \frac{t_{\mathrm{MS}} \times N}{t_{\mathrm{LS}} \times N + t_{\mathrm{PCA}} + (2M+1)(t_{\mathrm{MS}} + t_{\mathrm{LS}})}, \tag{4.12}$$

where $t_{\mathrm{MS}}$ and $t_{\mathrm{LS}}$ are the computational times for a single monochromatic calculation corresponding to the multi- and low-stream RTMs, respectively, while $t_{\mathrm{PCA}}$ is the computational time of the PCA. Please note the PCA-based RTM requires $2M+1$ calls to the multi-stream model. On the other hand, the speedup factor for the CLSR method is defined by

$$S_{\mathrm{CLSR}} = \frac{t_{\mathrm{MS}} \times N}{t_{\mathrm{LS}} \times N + t_{\mathrm{LSM}} \times C + t_{\mathrm{MS}} \times n \times C}, \tag{4.13}$$

### Estimation of the required parameters for the acceleration techniques

In order to select the optimal number of PCs, we compute the explained variance ratio of the optical data, which accounts for the variance associated with a given number of PCs. The results in Figure 4.8**(left)** show that in all scenarios almost 99% of the optical data variance can be explained within the first four PCs. To select the number of clusters and regression points for the CLSR method, we estimate the mean error (cf. Equation (4.11)). As an example, we illustrate in Figure 4.8**(right)** the mean errors for the 'Aerosol 2' scenario. The results show that 4–5 clusters and 3–5 regression points guarantee a small mean error.

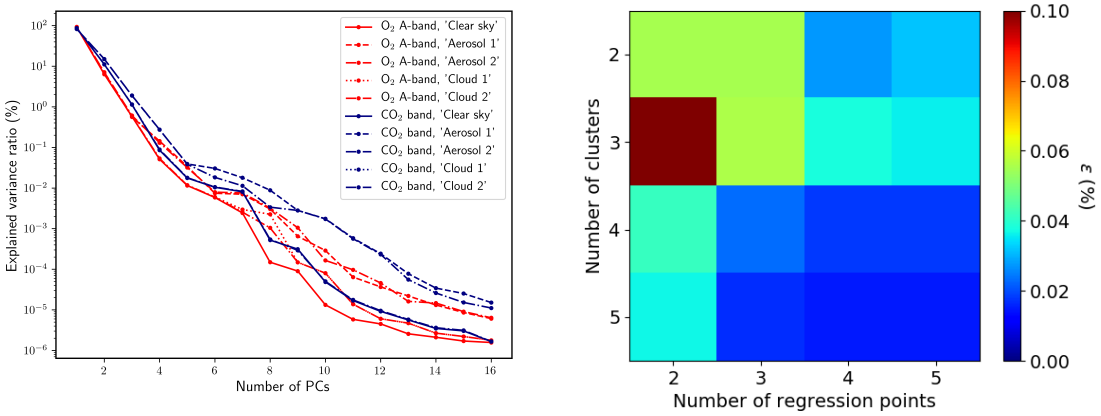


**Figure 4.8: (Left)** Explained variance ratio in percentage as a function of the number of PCs for all the atmospheric scenarios: 'Clear sky', 'Aerosol 1', 'Aerosol 2', 'Cloud 1' and 'Cloud 2'; and the two spectral bands $O_2$ A- and $CO_2$ bands; **(Right)** Dependence of the number of clusters and number of regression points with the mean error in percentage for the $CO_2$ band for the 'Aerosol 2' scenario. The low-stream model used is the two-stream model.

## 4.3.4 Results: PCA vs. CLSR

### Accuracy evaluation

The residuals (Equation (4.10)) and interquartile ranges corresponding to the PCA-based and CLSR methods, as well as for the TS and SS models, are shown in Figure 4.9 [60]. The main results that can be extracted from the Figure are that: (i) the TS model with the PCA-based and CLSR methods yields accurate results, (ii) the efficiency of the PCA-based method decreases when increasing the optical thickness, and (iii) both the TS and the SS models with the CLSR method provide reasonable accuracies. Furthermore, the IQRs help us to identify outliers. Thus, the smaller IQRs mean that the acceleration technique effectively captures the full range of the TOA radiance variations. In general, lower IQRs are more frequent for the CLSR method compared to the PCA method.

In sum, the residuals for the TS model in the $O_2$ A-band are above 1% for the PCA-based method and below 0.01% for the CLSR method, while they yield comparable accuracies in the $CO_2$ band (residuals are below 0.1% for the PCA-based method and below 0.01% for the CLSR method). On the other hand, the residuals for the SS model with the PCA-based method are higher than those corresponding to the TS model while, the residuals of the TS and SS models with the CLSR method are comparable (they are below 0.2% in the $O_2$ A-band and below 0.1% in $CO_2$ band).

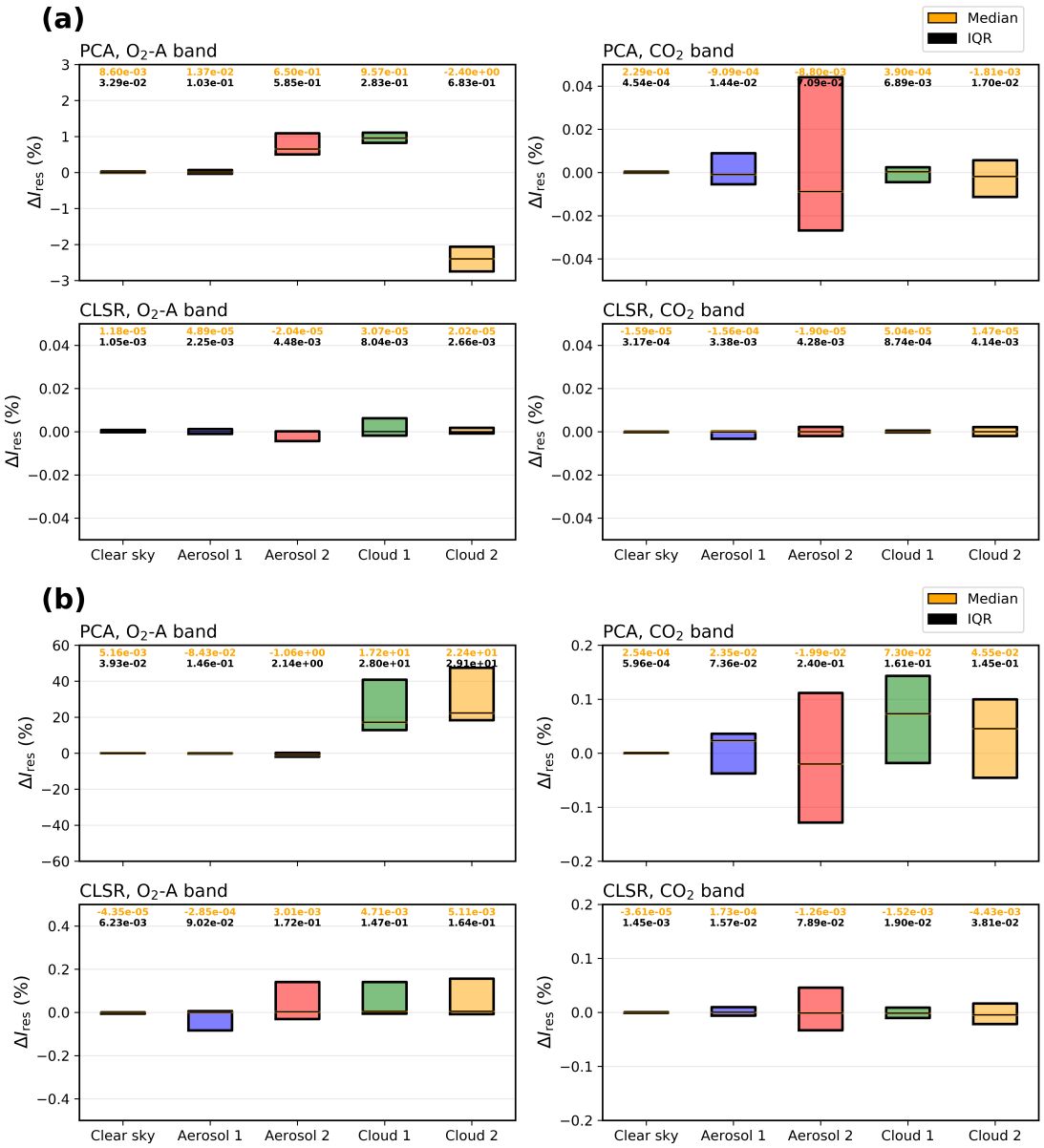


**Figure 4.9:** Comparison of the residuals for the methods PCA and CLSR for all the atmospheric scenarios (grey: 'Clear sky'; blue: 'Aerosol 1'; red: 'Aerosol 2'; green: 'Cloud 1'; yellow: 'Cloud 2') and gases ($O_2$ A- and $CO_2$ bands), when the LS model is (**a**) the TS model or (**b**) the SS model. Note the differences in scales for the PCA technique for the $O_2$ A-band with the rest of cases. The orange values on top of each box indicate the median values and the black values correspond to the IQR value.

## Computational performance

The computational efficiency is compared for the CLSR method against the PCA-based method. The speedup factors computed by using Equations (4.12) and (4.13) for the PCA-based and CLSR methods, respectively, are given in Table 4.2 [60].

The efficiencies of the TS model with the CLRS and the PCA-based method are comparable. In contrast, the computational performance of the SS model with the CLSR method is much higher than that with the PCA-based method due to the neglect of multiple scattering computations. Comparing the speedup factors of this study with those of other

**Table 4.2:** Speedup factor of the PCA-based ($S_{\mathrm{PCA}}$) and CLSR methods with: the two-stream model ($S^{\mathrm{TS}}_{\mathrm{CLSR}}$) and single-scattering model ($S^{\mathrm{SS}}_{\mathrm{CLSR}}$).

| $S_{\mathrm{PCA}}$ | $S^{\mathrm{TS}}_{\mathrm{CLSR}}$ | $S^{\mathrm{SS}}_{\mathrm{CLSR}}$ |
|---|---|---|
| 534 | 505 | 1294 |

authors (e.g. [59]), we find that our values are of the order of their speedup factors and one order of magnitude higher when considering the SS model for the CLSR technique.

#### Convolved spectra

The high resolution spectra in the $O_2$ A-band are convolved with the slit functions corresponding to Global Ozone Monitoring Experiment (GOME)-2 and TROPOMI instruments, while the radiance spectra in the $CO_2$ band are convolved with the GOSAT slit function. In this paper [60], slit functions are modelled with a Gaussian function. The corresponding full widths at half maximum (FWHM) are listed in Table 4.3. The FWHM considered for the $O_2$ A-band are based on pre-launch calibrations [106] and for the $CO_2$ band on [107].

**Table 4.3:** Spectral ranges and FWHM of the Gaussian slit functions of the instruments used in the study: TROPOMI, GOME-2 and GOSAT.

| Instrument | Spectral Range | FWHM |
|---|---|---|
| TROPOMI | 710–775 nm | 0.183 nm |
| GOME-2 | 590–790 nm | 0.51 nm |
| GOSAT | 1.56–1.69 $\mu$m | 0.2 cm$^{-1}$ |

The examples of convolved spectra corresponding to the 'Cloud 1' scenario are shown in Figure 4.10. Tables 4.4 and 4.5 show the mean relative errors for the PCA-based ($\varepsilon_{\mathrm{PCA}}$) and CLSR ($\varepsilon_{\mathrm{CLSR}}$) methods for the $O_2$ A-band and the $CO_2$ band, respectively. In addition, the residuals for non-convolved spectra are shown for comparison. For the clear sky and aerosol scenarios the accuracies of both methods are comparable, while for cloud scenarios the CLSR method is more accurate. Please note the residuals estimated for the convolved spectra are very close to those for the non-convolved ones and hence, the value of residuals according to Equation (4.10) is robust.

### 4.3.5 Conclusions

The main results obtained from this study published in [60] (cf. Appendix A.3) are:

- In general, the CLSR method shows more accurate results than the PCA-based RTMs. However, the PCA-based binned approaches and the CLSR method are comparable in terms of accuracy.
- The CLSR method requires more calls to the MS RTM than the PCA-based model although the speedup factors are very similar, with slightly better results for the CLSR method.
- The CLSR method can be used either in conjuction with the TS or the SS model providing a performance enhancement of almost two orders of magnitude while keeping the maximum error below 0.1%.

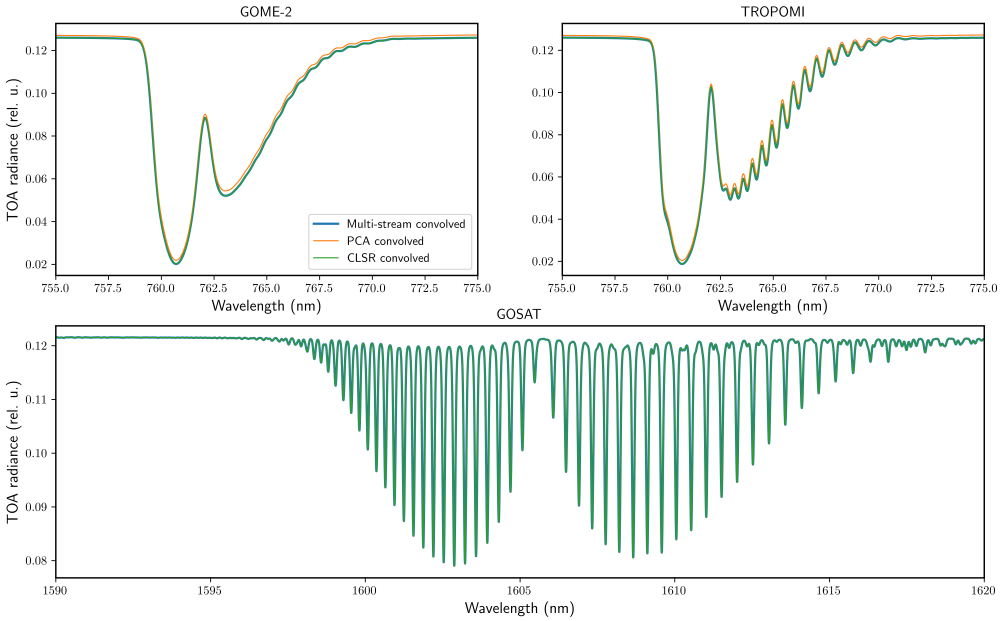


**Figure 4.10:** Convolved spectra for the multi-stream model and the two acceleration methods for the 'Cloud 1' scenario using PCA and CLSR methods for the sensors GOME-2, TROPOMI and GOSAT. For GOME-2 and TROPOMI the $O_2$ A-band spectra are convolved, while for GOSAT the $CO_2$ spectra are convolved.

**Table 4.4:** Mean relative error $\varepsilon$ for the convolved spectra compared with the multi-stream spectra for the PCA-based and CLSR methods and for the different atmospheric scenarios considered for the $O_2$ A-band. In all cases, the low-stream model considered is the two-stream model. The instruments analyzed are GOME-2 and TROPOMI, which are compared with the non-convolved values.

| | **$O_2$ A-Band** | | | | | |
|---|---|---|---|---|---|---|
| **Scenario** | **GOME-2** | | **TROPOMI** | | **Non-Convolved** | |
| | $\varepsilon_{\mathbf{PCA}}(\%)$ | $\varepsilon_{\mathbf{CLSR}}(\%)$ | $\varepsilon_{\mathbf{PCA}}(\%)$ | $\varepsilon_{\mathbf{CLSR}}(\%)$ | $\varepsilon_{\mathbf{PCA}}(\%)$ | $\varepsilon_{\mathbf{CLSR}}(\%)$ |
| Clear sky | 0.021 | 0.004 | 0.021 | 0.004 | 0.022 | 0.006 |
| Aerosol 1 | 0.049 | 0.007 | 0.050 | 0.007 | 0.061 | 0.011 |
| Aerosol 2 | 0.837 | 0.019 | 0.837 | 0.019 | 0.856 | 0.026 |
| Cloud 1 | 1.23 | 0.011 | 1.23 | 0.011 | 1.24 | 0.017 |
| Cloud 2 | 2.92 | 0.006 | 2.92 | 0.006 | 2.93 | 0.009 |

## 4.4 Application of the CLSR Method to other Spectral Bands and Aerosols

### 4.4.1 Motivation

The idea of the CLSR method is to perform LBL computations by using a fast TS RTM and then to refine results by using the correlation model for the TS and reference MS RTMs. This approach was successfully applied to the $O_2$ A-band and the weak $CO_2$ band for several atmospheric scenarios in [60], such as clear sky, aerosols and clouds. The results were compared with the PCA-based RTMs showing an improvement over the last

**Table 4.5:** Mean relative error $\varepsilon$ for the convolved spectra compared with the MS spectra for the PCA-based and CLSR methods and for the different atmospheric scenarios considered for the $CO_2$ band. In all cases, the LS model considered is the TS model. The instrument analyzed is GOSAT, which is compared with the non-convolved values.

| **$CO_2$ Band** | | | | |
|---|---|---|---|---|
| **Scenario** | **GOSAT** | | **Non-Convolved** | |
| | $\varepsilon_{\mathbf{PCA}}(\%)$ | $\varepsilon_{\mathbf{CLSR}}(\%)$ | $\varepsilon_{\mathbf{PCA}}(\%)$ | $\varepsilon_{\mathbf{CLSR}}(\%)$ |
| Clear sky | 0.0003 | 0.0003 | 0.0006 | 0.0006 |
| Aerosol 1 | 0.0075 | 0.0067 | 0.010 | 0.012 |
| Aerosol 2 | 0.035 | 0.012 | 0.044 | 0.017 |
| Cloud 1 | 0.011 | 0.007 | 0.013 | 0.008 |
| Cloud 2 | 0.011 | 0.005 | 0.016 | 0.006 |

in terms of accuracy. Therefore, we have extended in the studies published in [108] and [109] (refer to Appendix A.4 and A.5, respectively) the CLSR method to other spectral bands: Hartley-Huggins and water vapour for several atmospheric models containing different aerosol types in order to test the applicability of the method to a higher number of atmospheres and spectral ranges.

### 4.4.2 Data overview

The study is focused on the UV region with the Hartley-Huggins band, and in the IR region with the $O_2$ A-, water vapour and $CO_2$ bands. Table 4.6 summarizes the spectral bands examined with their corresponding spectral range, spectral resolution and number of spectral points to be simulated.

**Table 4.6:** Spectral ranges, resolutions and number of spectral points for the absorption bands used in this study.

| **Band** | **Spectral range (nm)** | **Spectral resolution (nm)** | **Number of spectral points** |
|---|---|---|---|
| Hartley–Huggins | 280–335 | 0.18 | 300 |
| $O_2$ A | 755–775 | 0.0010 | 20,000 |
| Water vapour | 770–1000 | 0.0058 | 40,000 |
| $CO_2$ | 1590–1620 | 0.0015 | 20,000 |

For modelling the aerosol properties, the OPAC database [86] is used. The aerosol optical thickness (AOD) and the single scattering albedo (SSA) are summarized in Table 4.7 (extracted from [108, 109]).

**Table 4.7:** Values of AOD (SSA) at the middle of the spectral range for the spectral bands and aerosol types considered.

| **Aerosol type** | **Hartley-Huggins** | **$O_2$A** | **Water vapour** | **$CO_2$** |
|---|---|---|---|---|
| Tropospheric | 0.133 (0.950) | 0.018 (0.947) | 0.015 (0.942) | 0.010 (0.936) |
| Continental clean | 0.42 (0.959) | 0.20 (0.962) | 0.17 (0.959) | 0.08 (0.958) |
| Urban | 4.28 (0.940) | 0.46 (0.935) | 0.35 (0.929) | 0.16 (0.913) |
| Desert | 0.71 (0.932) | 0.20 (0.953) | 0.20 (0.950) | 0.19 (0.945) |
| Continental polluted | 2.4 (0.951) | 1.2 (0.960) | 0.9 (0.957) | 0.4 (0.950) |

Figure 4.11 [109] shows an overview of the LBL radiances of two aerosols for the MS and TS RTMs for the different absorption bands. Note that both pairs of spectra have a similar spectral behaviour and this allows to establish a regression model between the low- and multi-stream radiances, which is subject to the CLSR method considered further.

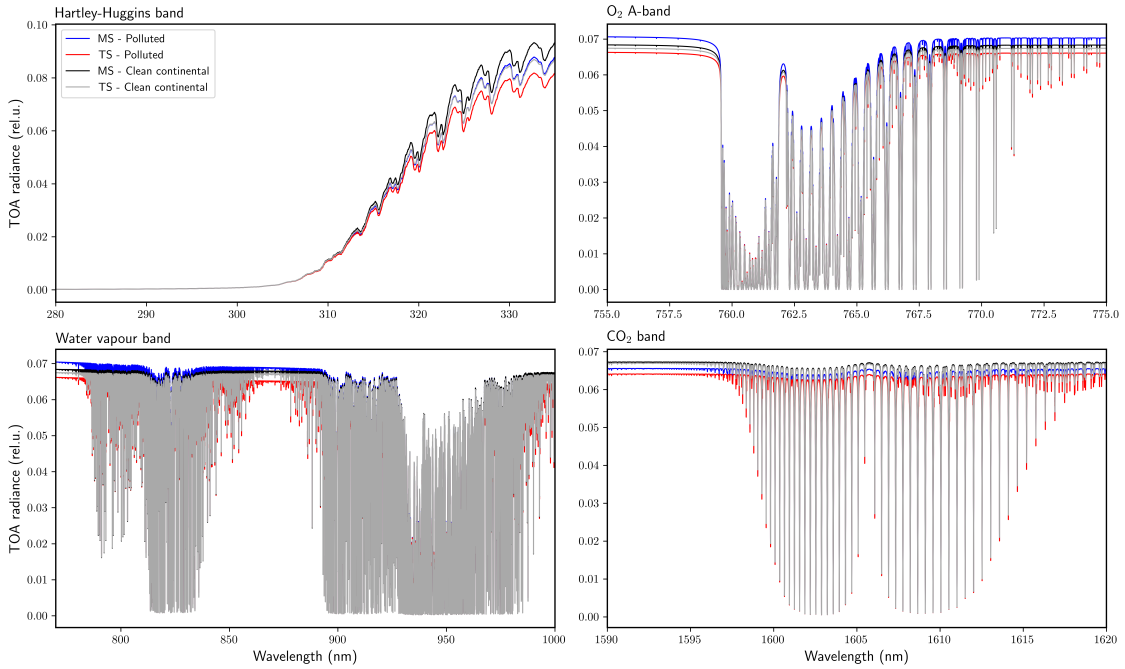


**Figure 4.11:** TOA radiances computed for the absorption bands: Hartley- Huggins, $O_2$ A-, water vapour and $CO_2$ bands for two aerosol cases: (blue) polluted aerosol and (red) clean continental aerosol. Solid lines correspond to the MS RTM while dashed lines correspond to the TS RTM.

### 4.4.3 Methodology

The CLSR method is described in detail in [60]. Therefore, only the most important Equation of the CLSR method is shown in matrix form. Using the same notation as in Section 4.3, we can rewrite Equation 4.8 as follows:

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{X}, \tag{4.14}$$

where

$$\begin{cases} \mathbf{Y} = \left[\hat{I}^c_{\mathrm{MS},i}\right], \\ \mathbf{A} = \left[\alpha^c, \beta^c, \gamma^c\right], \\ \mathbf{X} = \left[\hat{T}^c_i, \hat{I}^c_{\mathrm{TS},i}, 1\right]. \end{cases} \tag{4.15}$$

Finally, we find the regression coefficients as a solution to the following least squares problem:

$$\mathbf{A} = \arg\min_{\mathbf{A}} \sum_{q=1}^{n} \left[\overline{I}^c_{\mathrm{MS},q} - \mathbf{Y}\right]^2. \tag{4.16}$$

Finally, we can restore the spectra of the MS radiances $\{\tilde{I}_{\mathrm{MS},i}\}_{i=1}^{N}$.

### 4.4.4 Accuracy performance

#### Spectral residuals of the CLSR method for several atmospheric scenarios

The residuals (cf. Equation (4.10)) are computed for different numbers of regression points per cluster in the case of the CLSR method. Figure 4.12 shows these residuals for the clear sky and polluted aerosol cases (see [108]).

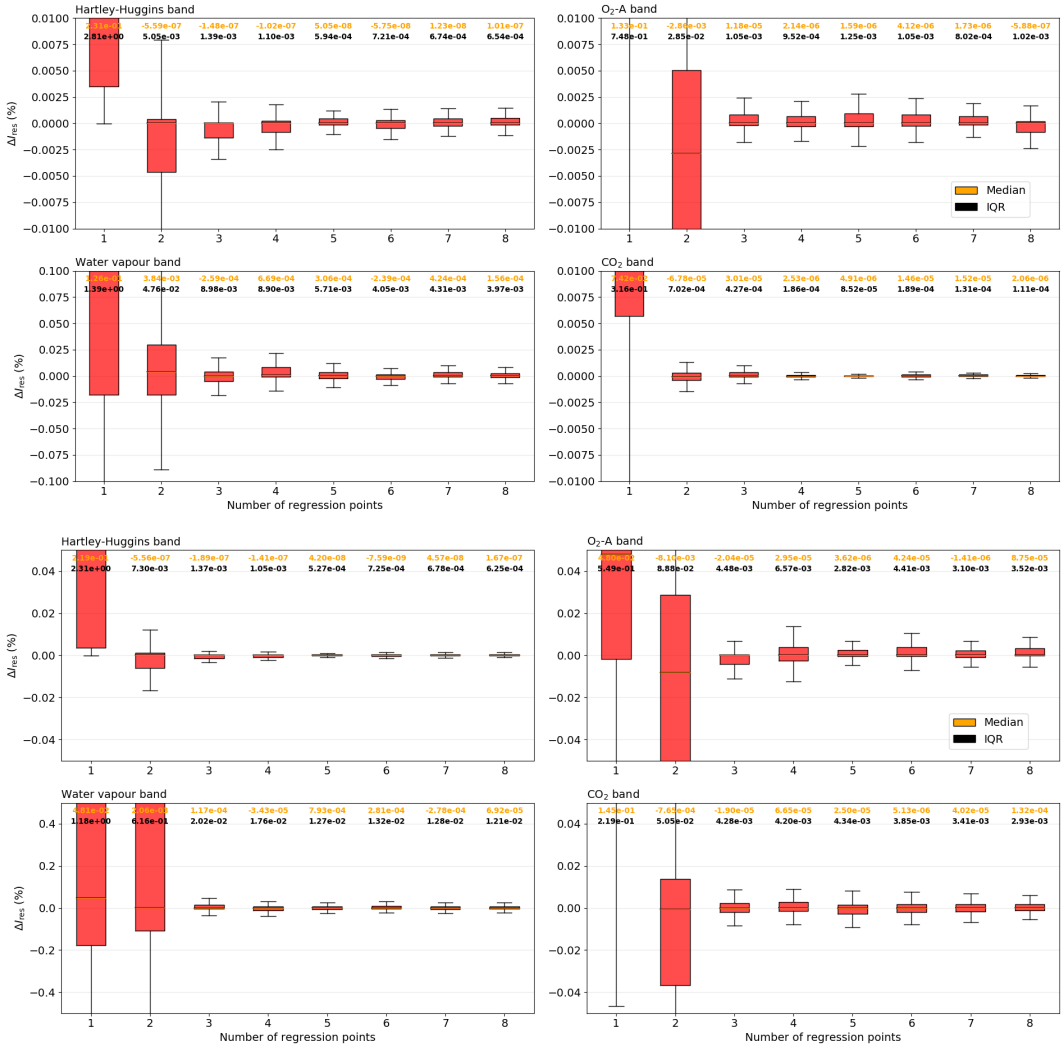


**Figure 4.12:** Box plots of the residuals for the CLSR method for **(a)** the clear sky scenario and **(b)** the polluted aerosol in the following absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The orange and black values on top of each box indicate the median values and the IQR values, respectively.

From Figure 4.12 we can infer that the residuals gradually decrease with the number of spectral points. In fact, they are significantly reduced when switching from 1–2 to 3 regression points. Therefore, the median values remain almost constant from 3 regression points. This trend is identical to the one found in [60] for different atmospheric scenarios. Note that the scale of residuals for the water vapour band is one order of magnitude higher than for the Hartley-Huggins, $O_2$ A- and $CO_2$ bands.

**Accuracy of the two-stream model equipped with the CLSR method for aerosol cases**

Figure 4.13 and Figure 4.14 (see [109]) show the probability density and the cumulative probability distribution of residuals, respectively, for four bands and 5 aerosol models. Note that most of the highest probabilities of the residuals are found below 0.005% for the Hartley-Huggins, $O_2$ A- and $CO_2$ bands. However, the probability density of the water vapour band shows a wider distribution than for the other bands. From the cumulative probability distributions, the following conclusions are drawn:

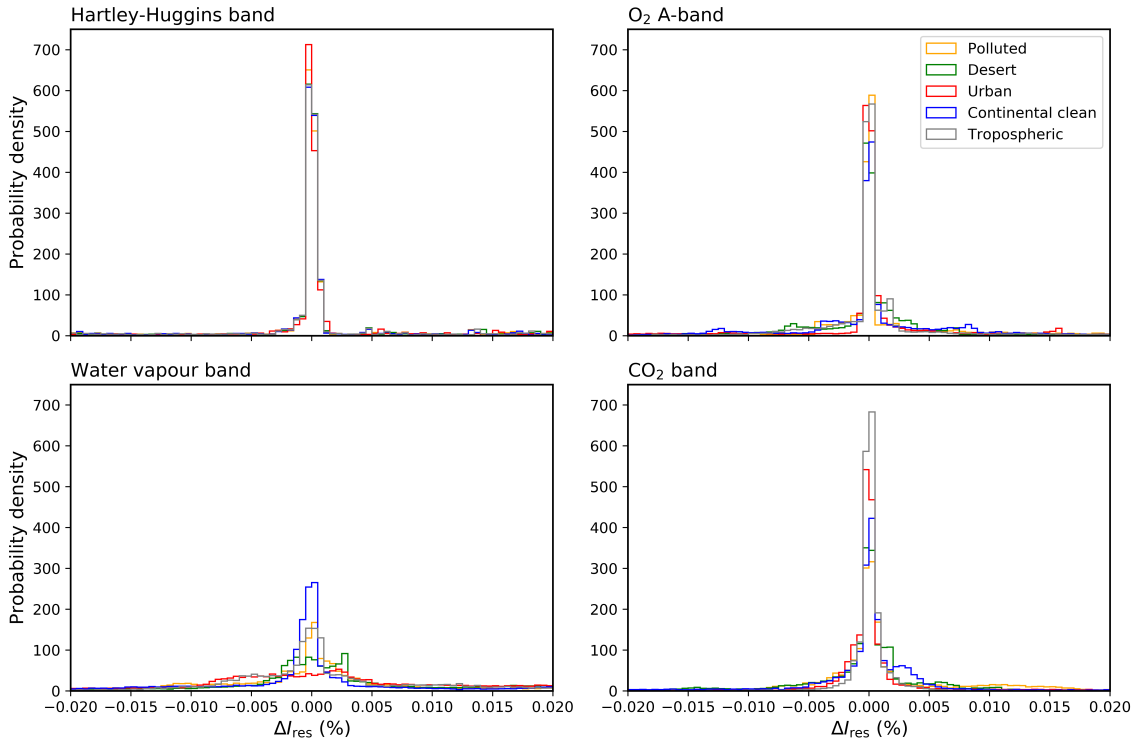


**Figure 4.13:** Probability density of the residuals for the CLSR method for the following absorption bands: Hartley- Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The color of the lines represents the type of aerosol: (grey) tropospheric; (blue) continental clean; (red): urban; (green): desert; (yellow): polluted.

- For the Hartley-Huggins band and for all aerosol types, the 90% of the residuals are below 0.05%.
- For the $O_2$ A-band, the 90% of the probability for the urban and tropospheric aerosol are below 0.02% while the for the other aerosols, the residuals are higher.
- For the $CO_2$ band and the tropospheric aerosol, most of their residuals are below 0.01%. Regarding the continental clean and desert aerosols, 90% of the residuals are below 0.03% and for urban and polluted aerosols are below 0.05%. The reason these latter aerosols present higher errors might be that they have higher AODs and this influences their values. This also occurs for the water vapour band and those specific aerosols.
- Similarly to the other spectral bands, more than 80% of the residuals in the water vapour band presents an accuracy below 0.05% for all aerosols. However, their influence on the water vapour band is more pronounced.

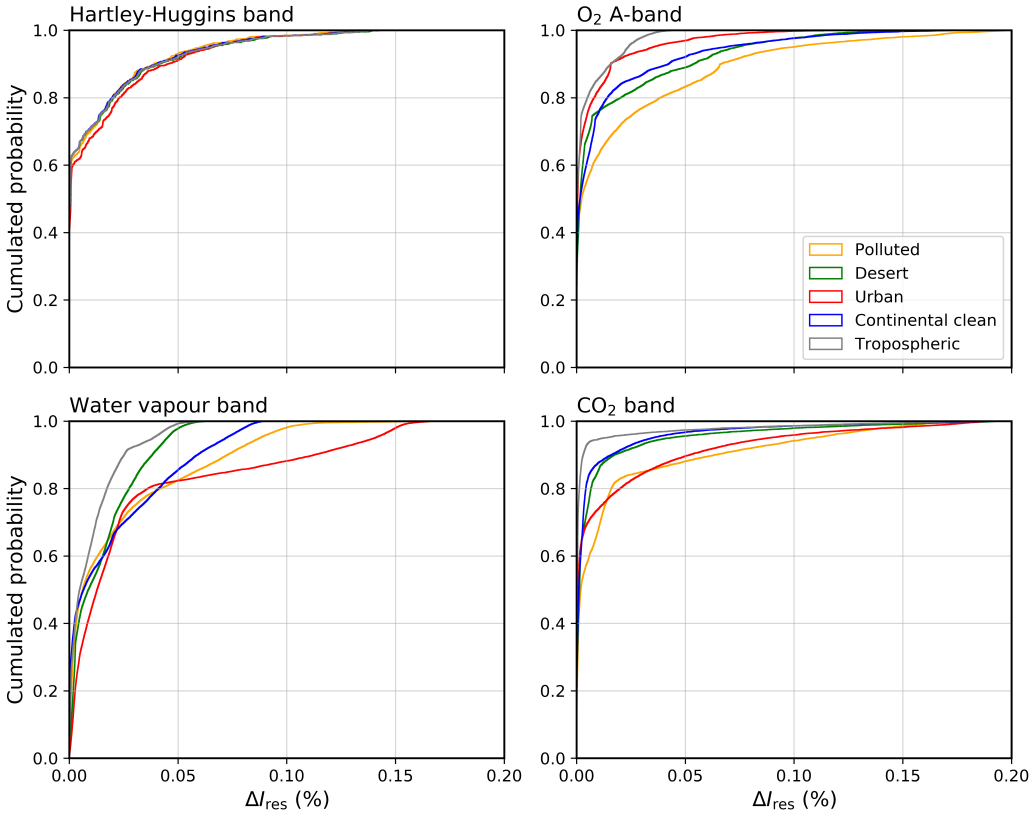


**Figure 4.14:** Cumulative probability distribution of the residuals for the CLSR method for the following absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The color of the lines represents the type of aerosol: (grey) tropospheric; (blue) continental clean; (red) urban; (green) desert; (yellow) polluted.

### 4.4.5 Computational performance

Table 4.8 shows the number of calls to TS and MS RTMs and the speedup factors with respect to the multi-stream LBL simulations for the Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands [109].

**Table 4.8:** Summary of number of calls, computational time and speedup factors for the Hartley-Huggins, $O_2$A-, $CO_2$ and water vapour bands with LBL and CLSR methods.

| | **Hartley-Huggins** | | **$O_2$ A- and $CO_2$** | | **Water vapour** | |
|---|---|---|---|---|---|---|
| | **LBL** | **CLSR** | **LBL** | **CLSR** | **LBL** | **CLSR** |
| Number of calls to MS RTM | 300 | 20 | 20000 | 20 | 40000 | 20 |
| Number of calls to TS RTM | — | 300 | — | 20000 | — | 40000 |
| Speedup factor | — | 15 | — | 1000 | — | 2000 |

We have used 5 clusters and 4 regression points for the simulations of the CLSR method for all the spectral bands. The same number of CLSR method computations applied to the different absorption bands (with different number of spectral points), provide different speedup factors. For instance, the speedup factor for the Hartley-Huggins band is around 15, while for the $O_2$ A- and $CO_2$ bands, the speedup is of 1000. Presumably, the water vapour band has more spectral points in the LBL RTM so that with 20 calls to the MS RTM, we obtain a much superior speed of 2000 compared with the LBL.

### 4.4.6 Conclusions

The conclusions of these studies published in [108, 109] (Appendix A.4 and A.5, respectively) can be summarized in the following points:

- The accuracy of the TS model can be enhanced by using the CLSR method, which exploits the linear relationship between the TS and MS RTMs.
- The efficiency of the approach has been tested for computing the TOA radiances in four spectral bands at the presence of aerosols. The error of the computations is generally below 0.05% and robust with varying aerosol properties.
- The number of calls to the time-consuming MS RTM has been reduced by 1-3 orders of magnitude, depending on the spectral band.

## 4.5 Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method

### 4.5.1 Motivation

The performance enhancement due to the CLSR method is about two orders of magnitude. However, that performance can be further enhanced with a modification of the CLSR method. Specifically, by applying the CLSR twice, once to reconstruct the TS spectra by means of the SS approximation and the second, to reconstruct the MS RTM by means of the TS spectra obtained in step one. Another modification to the CLSR method consists of providing efficient computations for atmospheric scenarios containing aerosols. In particular, using the clear sky conditions for computing the aerosol spectra in the framework of the CLSR method. The results are published in [110] and can be found in Appendix A.6.

### 4.5.2 Data overview

To check the efficiency of the proposed modifications of the CLSR method, we consider high spectral resolution computations of the TOA radiances in the Hartley–Huggins (315 nm), $O_2$ A-band (760 nm), water vapour (885 nm) and $CO_2$ band (1610 nm). The optical properties of aerosols are computed by using the OPAC database [86]. The following aerosol types are considered: tropospheric, continental clean, urban, desert and continental polluted. For this study, clouds are not taken into account. The values of the AOD, SSA and asymmetry factor (g) are summarized in Figure 4.15.

### 4.5.3 Methodology

#### The CLSR vs. the double CLSR method

In the double CLSR method, the TS spectra are computed also by applying the CLSR technique. In this case, as an approximate model, we use the SS RTM. Thus, the algorithm can be described as follows:

**Step 1**: We compute the LBL spectra $\{I_{\rm SS}(\lambda_i)\}_{i=1}^N$ by using the SS RTM and apply sorting and clustering to the space of SS radiances. Assuming a regression model between SS and TS radiances within each cluster $z$, we obtain

$$\text{Step 1}: \hat{I}^z_{{\rm TS},i} = a^z \hat{T}^z_i + b^z \hat{I}^z_{{\rm SS},i} + d^z,\ i = 1, ..., N_Z, \tag{4.17}$$

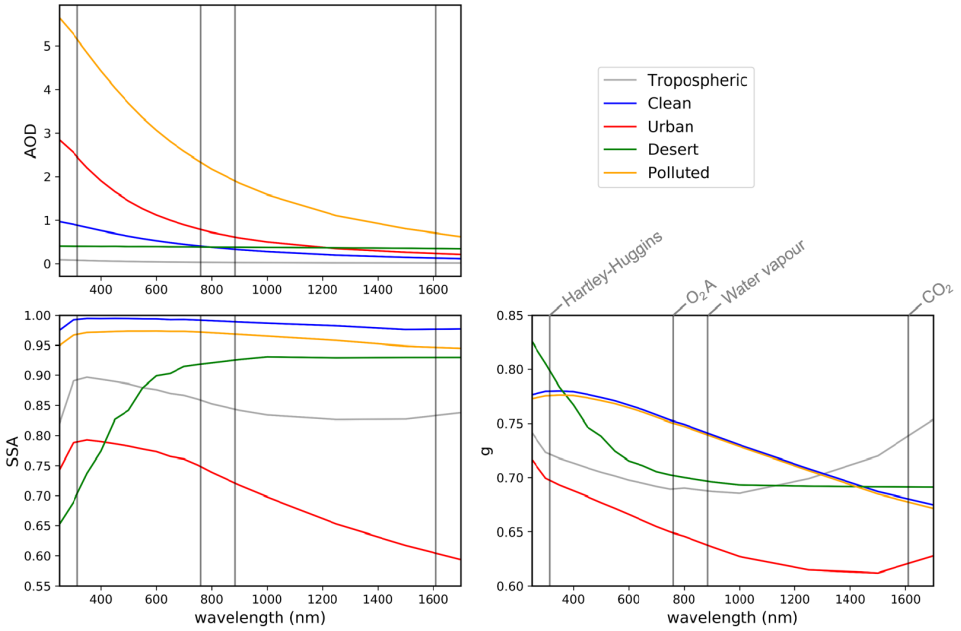


**Figure 4.15:** Overview of the optical properties (AOD, SSA and g) obtained from the OPAC database for the aerosol types: tropospheric, clean continental, urban, desert and polluted. Each vertical line corresponds with the middle wavelength of the spectral bands in order: Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands.

with number of radiance points $N_Z = N/Z$ and cluster index $z$. The regression coefficients $[a^z, b^z, d^z]$ are found as a solution of the following least squares problem

$$\mathbf{A} = \arg\min_{\mathbf{A}} \sum_{q=1}^{n} \left[\overline{I}^z_{\mathrm{TS},q} - \mathbf{Y}\right]^2, \tag{4.18}$$

where $\mathbf{A} = [a^z, b^z, d^z]$ and $\mathbf{Y} = \left[\hat{I}^z_{\mathrm{TS},i}\right]$. By knowing the regression coefficients, the TS spectra $\{\tilde{I}_{\mathrm{TS},i}\}_{i=1}^N$ can be restored from $\{I_{\mathrm{SS}}(\lambda_i)\}_{i=1}^N$ at high spectral resolution.

**Step 2**: We apply the CLSR method as described in Section 4.3 using the TS spectra computed in Step 1.

Figure 4.16 shows a schematic representation of the CLSR and double CLSR methods [110]. Note that the TS spectra derived at Step 1 by using the CLSR method differ from those computed by the TS RTM in a LBL manner. However, the possible bias obtained in the double CLSR method at Step 1 is removed by the regression model at Step 2.

### Improvement to aerosol schemes

To compute the spectrum in the case of the atmosphere with aerosol, the CLSR and double CLSR methods can be applied [110]. For this case, the regression model (Equation (4.14)) is used, in which

$$\mathbf{Y} = [I^{\mathrm{aer}}_{\mathrm{MS}}] \tag{4.19}$$

and the original $\mathbf{X}$- matrix is substituted by

$$\mathbf{X}_0 = [T, I^{\mathrm{aer}}_{\mathrm{TS}}, 1], \tag{4.20}$$

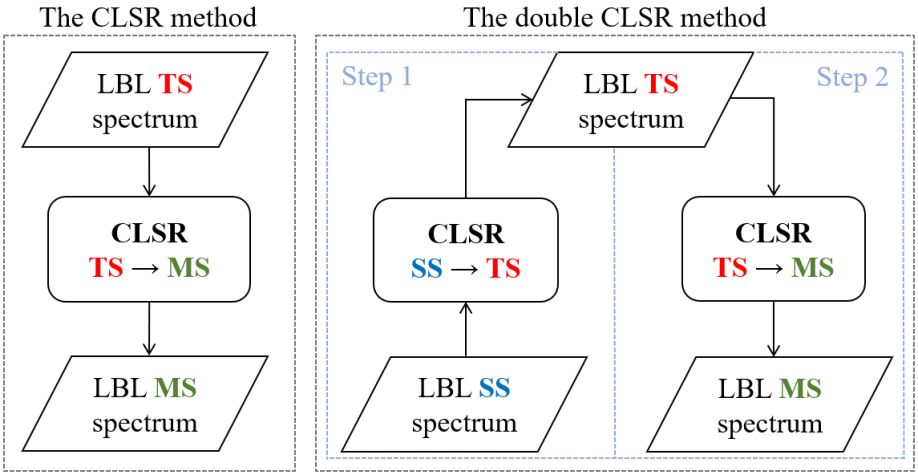


**Figure 4.16:** Scheme of the CLSR method vs. the double CLSR method.

where the upper index 'aer' explicitly indicates that the computations are performed for the aerosol case. In this regard, alternative formulations of the CLSR method can be considered. For instance, taking the **X**- matrix as

$$\mathbf{X}_1 = \left[T, I_{\mathrm{MS}}^{\mathrm{clear}}, 1\right], \tag{4.21}$$

we obtain a method, which converts the MS clear sky spectra into spectra corresponding to the aerosol conditions. The upper index 'clear' indicates that the computations are performed for the clear sky case. The possible benefit of such a scheme is that the clear sky spectra can be precomputed and stored in LUTs and perform the computations offline, while the computations for actual aerosol properties can be performed online.

Alternatively, we consider the **X**-matrix in the following form:

$$\mathbf{X}_2 = \left[T, I_{\mathrm{TS}}^{\mathrm{clear}} - I_{\mathrm{TS}}^{\mathrm{aer}}, I_{\mathrm{MS}}^{\mathrm{clear}}, 1\right]. \tag{4.22}$$

In this case, the regression model is supplied with the first order perturbation computed by using the TS RTM. As a matter of fact, in this case, we do not expect performance enhancement compared to the $\mathbf{X}_0$-scheme. The question is if the error can be reduced by involving precomputed LUTs for clear sky cases as compared to the original $\mathbf{X}_0$-scheme. Note that for all these cases, the MS RTM for the aerosol scenarios is called for a few spectral points.

### 4.5.4 Results

#### Accuracy results: single vs. double CLSR

Figure 4.17 shows the probability density function of the spectral residuals $\Delta I_{\mathrm{res}}$ (cf. Equation 4.10) for both methods and the four spectral bands. The main conclusions that can be drawn from the figure are the following:

- More than 70% and 60% of the residuals are below 0.01% for the single and double CLSR methods, respectively, for all bands, with the exception of the water vapour band.

- The residuals of the water vapour band present a wider distribution in comparison with the other spectral bands.

- The probability densities are almost indistinguishable for both acceleration methods, demonstrating that both techniques provide accurate results among the different spectral bands.

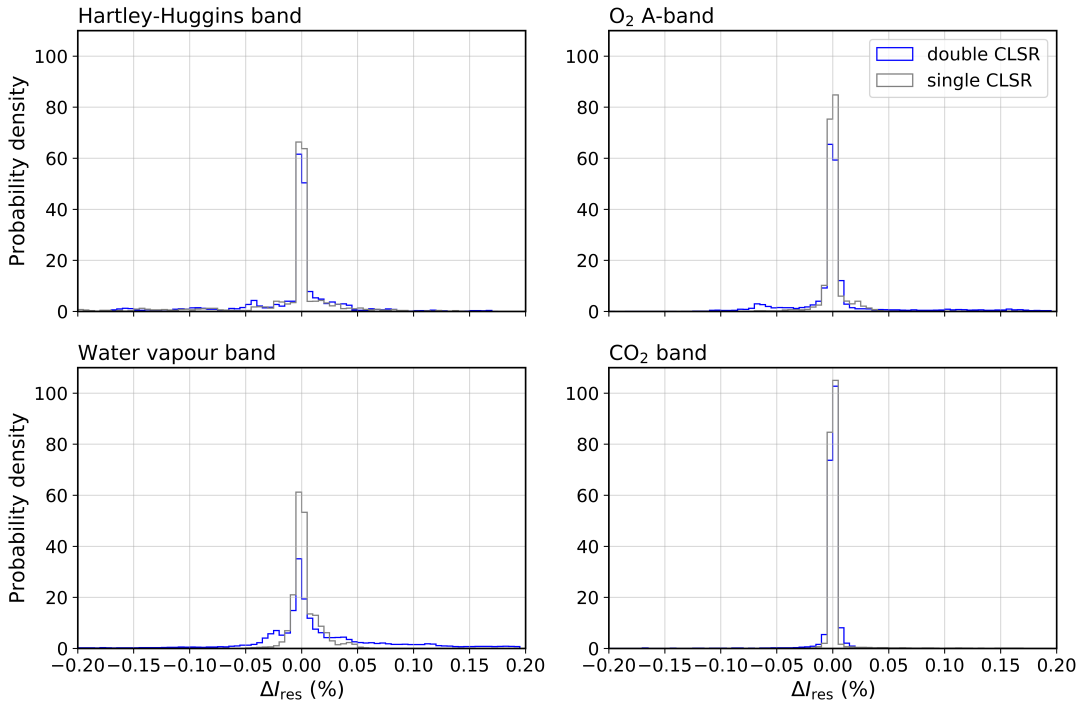


**Figure 4.17:** Probability density function of the residuals for the single CLSR (grey) and the double CLSR (blue) methods for the Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands and for the tropospheric aerosol case.

It can be observed that the accuracy of the CLSR method is slightly higher than that of the double CLSR method. This result can be expected, since the TS spectra used in the double CLSR method are approximate and obtained from the SS spectrum (see Figure 4.16).

Figure 4.18 shows the cumulated probability functions of the CLSR and double CLSR methods for all spectral bands. Over 90% residuals are less than 0.05% in the case of the Hartley–Huggins band and 0.01% in the case of the $CO_2$ band. Higher differences can be seen between the CLSR and double CLSR methods for the $O_2$A- and water vapour bands. For these bands, over 90% of the CLSR residuals are less than 0.025%. Meanwhile, the double CLSR provides slightly larger errors: over 60% and 80% of the residuals are less than 0.05% for the $O_2$ A- and water vapour band, respectively. However, these residual values are still low and of the same order as those obtained by using the PCA-based RTMs. For instance, in Liu et al. [75] PCA was applied to optical parameters and spectral radiances yielding an error lower than 0.2% in the solar region (775–920 nm). Kopparla et al. [67] combined the PCA technique for optical parameters and the spectral binning for accurate computations in the case of aerosols. The residuals were below 0.01% for the $O_2$ A-band, which are of the same order as our results.

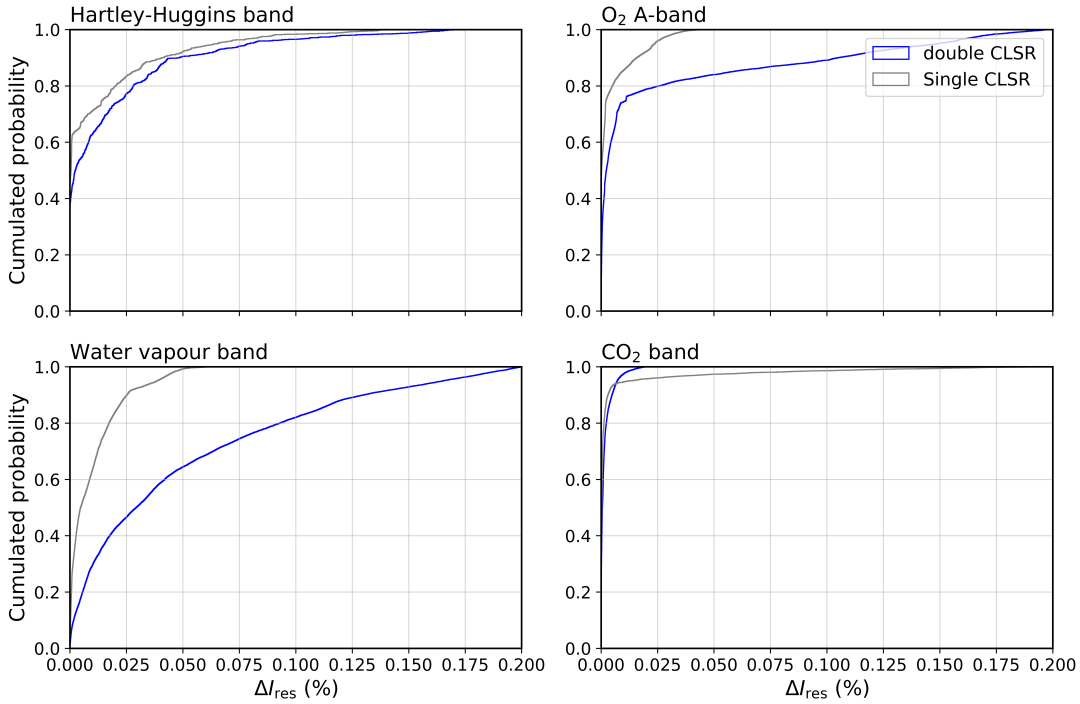


**Figure 4.18:** Same as Figure 4.17, but for the cumulated probability of the residuals.

### Computational performance: single vs. double CLSR

In this section, we analyse the computational performance of the single and double CLSR methods. Tables 4.9–4.11 show the number of calls to the SS, TS and MS RTMs, computational times and corresponding acceleration factors with respect to the MS LBL simulations (see [110]).

For the single CLSR method, 5 clusters and 4 regression points per cluster are used. Hence, computations for each absorption band involve 20 MS RTM calls, while the number of TS RTM calls is equal to the number of spectral points in the high resolution LBL computations (i.e., 300, 20,000 and 40,000 calls to the TS RTM for the Hartley–Huggins, $O_2$ A- and $CO_2$ and water vapour band, respectively). As can be seen in Tables 4.9–4.11, the TS RTM imposes the computational burden of the single CLSR method for the $O_2$ A-, $CO_2$ and water vapour bands, consuming up to 70% of the whole computation time.

At the first step of the double CLSR method, 8 clusters and 4 regression points are used. Thus, the TS RTM is called for 32 spectral points. In the case of the double CLSR, the SS RTM is utilized for the LBL computations, resulting in an additional performance enhancement by 2 times for the $O_2$A- and $CO_2$ bands and 3 times for the water vapour band. We note that in the double CLSR, the computational burden corresponds to the MS RTM, while the computation times related to the TS and SS RTMs are three and two orders of magnitude lower than those of the MS RTM, respectively. In the case of the Hartley–Huggins band, the computational burden is still due to the MS RTM [36] and the double CLSR does not further improve the performance.

As a final remark, the accuracy is crucial to determine the number of calls needed for the CLSR methods, and we could improve it by increasing the number of calls to the RTM models. However, this would add a computational burden to the simulations, while providing little improvement in the accuracy. Several tests have been performed by

**Table 4.9:** Summary of number of calls, computational time and acceleration factors for the Hartley–Huggins band. The computational times marked in red indicate the computational burden.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 300 | 20 | 20 |
| | TS | 0 | 300 | 32 |
| | SS | 0 | 0 | 300 |
| Computation time (s) | MS | 35 | 2.32 | 2.32 |
| | TS | 0 | 0.048 | 0.005 |
| | SS | 0 | 0 | 0.006 |
| Total computational time (s) | | 35 | 2.37 | 2.33 |
| Acceleration factor | | – | 14.8 | 15.0 |

**Table 4.10:** Same as for Table 4.9 but for the $O_2$A- and $CO_2$ bands.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 20,000 | 20 | 20 |
| | TS | 0 | 20,000 | 32 |
| | SS | 0 | 0 | 20,000 |
| Computation time (s) | MS | 2320 | 2.32 | 2.32 |
| | TS | 0 | 3.2 | 0.005 |
| | SS | 0 | 0 | 0.4 |
| Total computational time (s) | | 2320 | 5.52 | 2.725 |
| Acceleration factor | | – | 420 | 850 |

**Table 4.11:** Same as for Table 4.9 but for the water vapour band.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 40,000 | 20 | 20 |
| | TS | 0 | 40,000 | 32 |
| | SS | 0 | 0 | 40,000 |
| Computation time (s) | MS | 4640 | 2.32 | 2.32 |
| | TS | 0 | 6.4 | 0.005 |
| | SS | 0 | 0 | 0.8 |
| Total computational time (s) | | 4640 | 8.72 | 3.13 |
| Acceleration factor | | – | 532 | 1482 |

increasing the number of calls to the SS RTM or TS RTM but the errors are of the same order of magnitude as the actual values.

#### Further improvements to aerosol schemes

The efficiency of the CLSR method for various configurations outlined in Section 4.5.3 is examined for several OPAC aerosol models. The results obtained using $\mathbf{X}_1$ and $\mathbf{X}_2$ (corresponding to Equations (4.21) and (4.22), respectively) are compared against the original CLSR method in which the matrix $\mathbf{X}_0$ (Equation (4.20)) is used. The mean absolute relative errors are shown in Figure 4.19 [110].

In general, mean relative absolute errors are below 0.05% for all aerosol types and bands when using the original $\mathbf{X}_0$-configuration for the CLSR method. However, these errors are slightly higher for the water vapour band due to its higher spectral complexity.

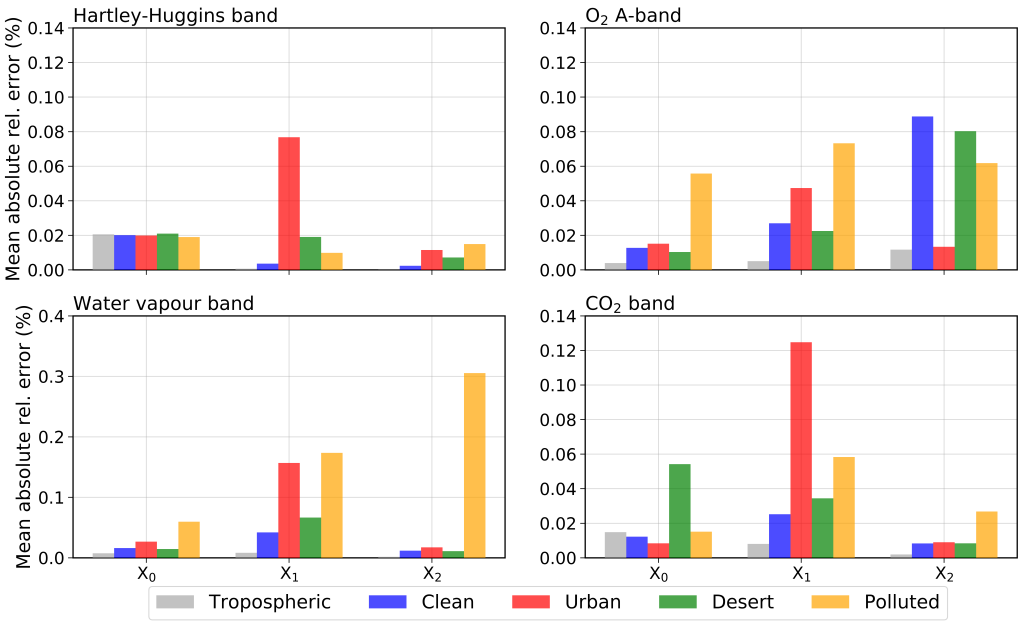


**Figure 4.19:** Mean absolute relative error in % for the four spectral bands and all aerosol types. Each group of bars corresponds to one of the $\mathbf{X}_i$-matrix ($i = 0, 1, 2$) (Equations (4.20)–(4.22)) used in the CLSR method.

In sum, we present two alternative configurations to the original $\mathbf{X}_0$, in order to obtain more accurate results for: (1) low aerosol loading conditions with the $\mathbf{X}_1$ configuration for all bands; and (2) the Hartley–Huggins, water vapour and $CO_2$ bands with the $\mathbf{X}_2$ configuration.

### Combined application of single vs. double CLSR for aerosol scenarios

The single and double CLSR methods are tested for the full set of aerosol models and $\mathbf{X}_i$-configurations. The probability density functions of the residuals for the tropospheric aerosol model are shown in Figures 4.20 and 4.21.

The conclusions drawn from the comparison of the two figures can be summarized as follows:

- The residual distributions of the single CLSR method are narrower than those of the double CLSR method, meaning that the single CLSR method is more accurate. However, in general, the residuals are below 0.01% for both methods and all spectral bands, except for the water vapour band, where the residual distributions are slightly wider and still below 0.05%. The distributions are not biased.

- For the Hartley–Huggins, $O_2$ A- and $CO_2$ bands, the residuals are below 0.05% for both single and double CLSR methods. Regarding the water vapour band, the residuals are below 0.05% and 0.1% for the single and double CLSR method, respectively. Similar accuracies were achieved in Kopparla et al. [67] for the water vapour band using the PCA-based RTM.

- In the case of the low aerosol load, the probability density functions are similar for all $\mathbf{X}_i$-configurations. However, as the aerosol load increases, the residual distributions for the $\mathbf{X}_1$ configuration provided by the single and double CLSR methods sometimes become biased for the water vapour band.

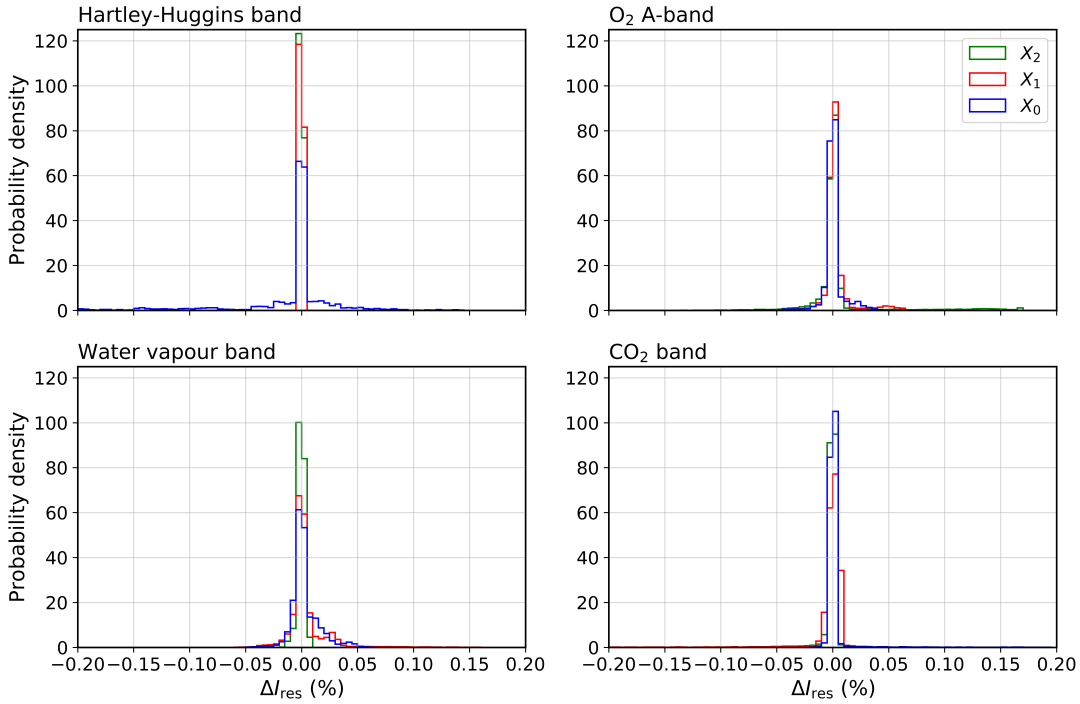


**Figure 4.20:** Probability density of the residuals for the single CLSR method and for several matrix configurations: $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$. Each plot represents the absorption bands: Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The case presented corresponds to the tropospheric aerosol.

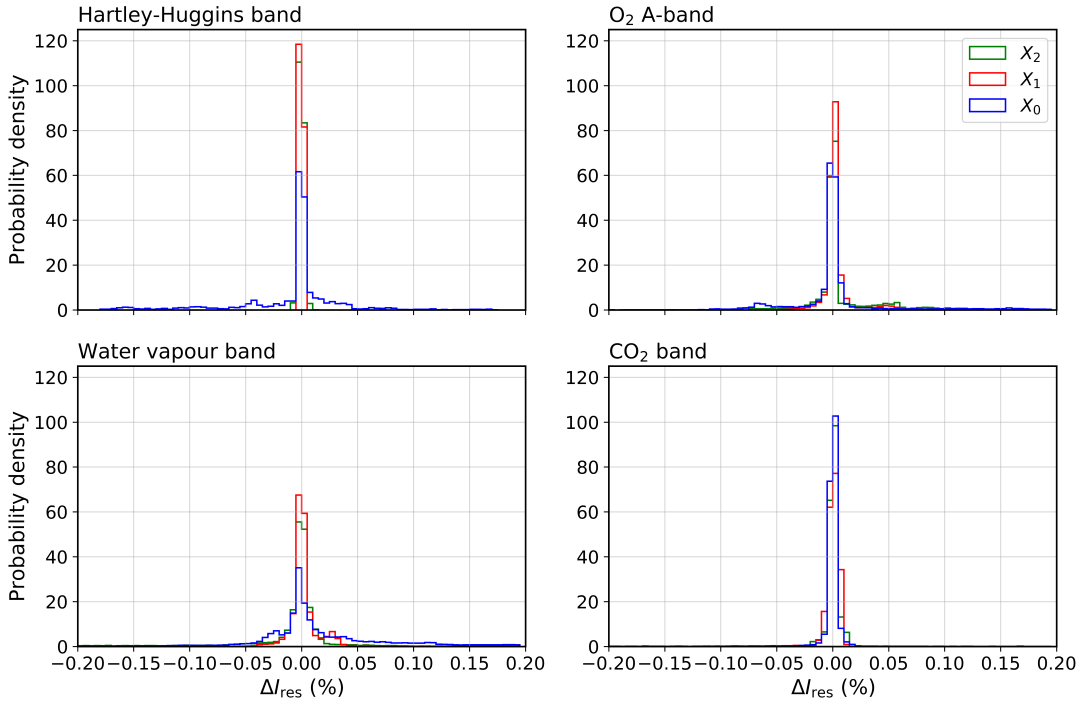


**Figure 4.21:** Same as Figure 4.20, but for the double CLSR method.

### 4.5.5 Conclusions

The major conclusions that can be drawn from the study published in [110] (refer to Appendix A.6) are:

- The double CLSR further improves the computational performance of the original CLSR method by two ($O_2$ A- and $CO_2$ bands) and three times (water vapour band), yet keeping the error below 0.05% for all spectral points on a high resolution grid.
- The second modification of the CLSR method allows to reproduce the LBL spectra with low aerosol load based on the information of the clear sky spectra. With such a configuration we obtained absolute relative errors below 0.05% for all spectral bands while enhancing the computational performance by three orders of magnitude.
- The new configurations of the second modification have been tested for both CLSR and double CLSR methods, revealing similar probability density functions.

# 5 Conclusions

## 5.1 Summary

This thesis aims to research and design modern acceleration techniques for hyperspectral RTMs with potential for remote sensing applications. To this end, the overarching objective proposed is to investigate several dimensionality reduction techniques for LBL RTMs in different spectral regions and to devise new acceleration methods that computationally improve the state-of-the-art techniques without losing accuracy. In this framework, the specific objectives were pointed out in Chapter 1 which have been achieved on the basis of the work specified in Chapter 4. This thesis provides significant contributions to the performance enhancement for atmospheric composition sensors. In summary, the major achievements of this thesis are outlined:

- The main dimensionality reduction techniques for processing hyperspectral RTMs are reviewed and compiled in the full-peer review in Appendix A.2. It has been shown that the concept of dimensionality reduction in the framework of hyperspectral RTMs implicitly entails addressing the strong interdependencies of hyperspectral data. In addition, it has been also shown that numerous RTMs exploit PCA for hyperspectral modelling which are called PCA-based RTMs. Nevertheless, the specific features of the algorithms are based on different principles, i.e., in some studies the PCA is applied to the input optical data and in others to the output spectral data although in the end they are all referred to as "PCA-based". Following this line of research, we identified the need to investigate further how to combine these techniques, which led us to the next point.

- An innovative hybrid PCA-based RTM approach has been developed for the Hartley-Huggins band for the TROPOMI instrument. The approach consists of applying twice the PCA method: (1) to the input optical data and (2) to the output spectral data of the sensor. The novel addition to the PCA method is the combined application for the input and output data together with the application of ML techniques to the latter. Thus, the number of computations has been reduced for both input and output and the performance enhancement is upgraded. This hybrid approach was devised for the UV spectral region and other independent studies (e.g. [75]) have used it to extend to the wide spectral range.

- A new acceleration technique called CLSR method has been developed. It is based on the clustering of the spectral radiances computed with a low-stream RTM (i.e., an approximate RTM which is either single-scattering or two-streams) and regression analysis performed to the low-stream and multi-stream RTMs. The method has shown to provide a performance enhancement of 1-2 orders of magnitude with accuracies below 0.01% for two important spectral bands: $O_2$ A- and $CO_2$ bands and for several atmospheric scenarios including clouds and aerosols. These values outperform the state-of-the-art acceleration techniques, which is an asset to be considered for the future development of atmospheric sensors.

- Two modifications have been implemented in the CLSR method to (1) boost its performance and (2) include variations for aerosols. For the first modification, the CLSR method was applied twice for accelerating the approximate LBL RTM. Additionally, the double CLSR method was extended to other spectral bands such as the Hartley-Huggins and water vapour bands, obtaining a performance enhancement in the latter case of 3 orders of magnitude while maintaining the errors below 0.05% for all spectral points. The second modification of the CLSR method permits to reproduce LBL spectra with low aerosol load based on the clear sky spectra information. This implies that if we would have the clear sky spectra precomputed, the aerosol spectra can be retrieved significantly faster.

Thus, in this thesis a set of new acceleration techniques has been developed which helps to boost the performance of RTMs in the hyperspectral processing chain.

## 5.2 Discussion

This section compares the newly developed acceleration techniques within this thesis with other available acceleration techniques in the literature. For the comparison, the acceleration factors of each method are gathered in Table 5.1 for the different spectral regions. In this analysis, we consider the studies covering absorption bands in the 280–3000 nm spectral range. The table is part of the analysis for the full-peer reviewed article of Appendix A.6. This comparison aims to highlight the variability of acceleration factors when the techniques are applied to different spectral bands. Thus, the main conclusions that can be drawn from the comparison are as follows:

- For simulations in the Hartley–Huggins band, PCA techniques, linear embedding methods (LEM) and double CLSR have been applied. The double CLSR does not further improve the performance, since the computational burden is due to the MS RTM computations (see Section 4.5). The highest acceleration factor is provided by the hybrid PCA-based RTM method described in [36], in which PCA is applied to both optical parameters and spectral radiances. The performance enhancement in this case is up to 18 times.

- There are several studies in which fast RTMs for the $O_2$ A- and $CO_2$ bands (either weak or strong) have been designed. In general, all considered techniques provide acceleration factors of about 2 orders of magnitude. For example, those based on artificial NN [84] provide an acceleration factor of 250. In contrast, the CLSR and double CLSR methods provide accelerations of 505 and 850, respectively, for the named bands, which places them among the highest values.

- The water vapour band represents a challenge for acceleration techniques due to its complicated spectral structure. Therefore, the accuracy of the acceleration techniques is generally lower than for the $O_2$ A-band. For the water vapour band, the double CLSR method provides an acceleration factor of about 3 orders of magnitude (acceleration factor of 1500), while the $k$-distribution [42] and PCA-based RTMs [67] achieve lower acceleration factors, of one order of magnitude. Therefore, the double CLSR method provides superior acceleration factors for the water vapour band compared to the state-of-the-art techniques.

We can note that the acceleration factors are ambiguous. As mentioned in previous chapters, the acceleration factors of the RTMs depend not only on the acceleration technique but also on the number of discrete ordinates $N_{\mathrm{do}}$ used for the reference RTM. In

**Table 5.1:** Acceleration techniques for RTMs compared with those of this thesis (bottom). Each row corresponds with an acceleration method, its corresponding spectral region, acceleration factor (computed with respect to the LBL model) and reference. Note that the some of the acceleration factors are given in one or two order of magnitude indicated as 10× and 100×, respectively.

| Acceleration Technique | Band/Spectral Region | Acceleration Factor | Reference |
|---|---|---|---|
| $k$-distribution | $H_2O$, $CO_2$ , $O_3$, and $O_2$ | 10× * | Fomin [42] |
| double-$k$ approach | $O_2$ A | 1000 | Duan [55] |
| LSI | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 45 [b], 210 [c] | O'Dell [59] |
| PCA | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 50 | Natraj et al. [63] |
| PCA | 290–340 nm | 10 | Spurr et al. [111] |
| LEM | 325–335 nm | 10 | Efremenko et al. [62] |
| PCA | 325–335 nm | 2 | Efremenko et al. [112] |
| PCA | 300–3000 nm | 10× | Kopparla et al. [67] |
| PCA | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 100× | Somkuti et al. [65] |
| $k$-distribution + PCA | $O_2$ A | 342 | Molina García et al. [113] |
| NN | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 250 [d] | Le et al. [84] |
| LEM | $NO_2$ (425–450 nm) | 12 [e] | Doicu et al. [114] |
| SDCOMP [a] | 750–920 nm | 1000 [d] | Liu et al. [75] |
| PCA [a] | Hartley-Huggins | 18 | del Águila et al. [36] |
| CLSR | $O_2$ A, $CO_2$ weak | 505 | del Águila et al. [60] |
| | Hartley-Huggins | 15 | del Águila et al. [110] |
| double CLSR | $O_2$ A, $CO_2$ weak | 850 | del Águila et al. [110] |
| | Water vapour | 1500 | del Águila et al. [110] |

[a] PCA is applied twice, (1) to the optical properties and (2) the radiance data set.
[b] Nadir observations.
[c] Glint observations.
[d] The relative computational efficiency with respect to the accurate simulations is considered.
[e] Relative to the $k$-distribution method.
* It is estimated from the information found in the reference.

turn, the required number of $N_{\mathrm{do}}$ depends on aerosol properties, surface properties, geometry and the required accuracy. Therefore, different number of streams or atmospheric parameters are used for the reference RTM in the different studies and it is important to state the number of $N_{\mathrm{do}}$ taken into account on each study.

It is worth noting in this section, the advantages and disadvantages of the acceleration methods of this thesis as well as other aspects to consider:

- The main advantage of the CLSR method over the $k$-distribution approach is that in the former cross-validation can be used, i.e., for adding an extra point to the cluster it is only necessary to add a new exact computation into the CLSR method and if the accuracy is acceptable for the specific use, then the solution is found. If this it not the case, one can simply add more points to the cluster and perform a new iteration.

- The acceleration factors of the PCA-based models and the single CLSR method are very similar, with slightly better results for the CLSR method although the difference is not significant. One advantage is that the CLSR method does not have to deal with the eigenvalue problem as the PCA-based RTMs do.

- One annotation of the acceleration techniques developed in the thesis is that they are tested for specific spectral bands instead of taking the wide spectral range. This is not a problem since most of the remote sensing applications are focused on specific spectral ranges or windows. However, an unified acceleration technique with application to the whole spectral range would be desirable for theoretical applications.

In addition, there should be tunable parameters of the acceleration method in order to obtain a balance between accuracy and acceleration necessary for the specific application.

- In an initial step in the development of the acceleration techniques of this thesis, the exact LBL RTM computations have to be computed a priori. However, once the acceleration methods are tested for several atmospheric scenarios and spectral bands, this can be automatized and no a priori exact spectra would be needed. Indeed, this is the ultimate purpose of these acceleration methods.

- The integration of the acceleration techniques into PYDOME RT code would be a future step. In this case, it is of great importance to take into account the architecture of the RTM, i.e., there are some RTMs such as MODerate resolution atmospheric TRANsmission (MODTRAN), that have build-in acceleration techniques. Alternatively, the acceleration technique could be implemented separately from the open-source code. The latter case would be integrated into PYDOME for validation purposes.

## 5.3 Outlook

The newly developed acceleration RTMs have different potential applications. The techniques can be easily integrated into future sensors and be the reference for the development of future algorithms. Some of the potential practical benefits of the accelerations methods described in the previous section are as follows:

- For efficient and fast LUT generation, the CLSR method can be used in order to reduce the number of calls to the computationally expensive LBL RTM. Therefore, the single and double CLSR methods could potentially be incorporated into the generation of LUT radiance fields in a fine resolution grid.

- To reduce the size of LUTs and interpolation errors, it seems beneficial to implement the CLSR method in conjunction with gradient-based LUT generators [115, 116].

- The results of accelerations obtained with the both CLSR methods have the potential to be applied in NRT applications of aerosol computations or aerosol retrieval algorithms.

Apart from the benefits encountered for the acceleration methods, there are a few potential topics worth investigating further for the development of future fast hyperspectral RTMs and related applications for EO sensors. The main items are outlined in the following:

- Computing the Stokes parameters is of great interest in order to have full information of the spectra. Future studies might take into account the polarization by means of the vectorized solution of the RTE.

- In order to enhance the acceleration of the approximate RTM, a more efficient method could replace it. This could be carried out by using the asymptotic radiative transfer theory for optically thick media and the diffuse approximation instead of the two-stream RTM. Further, this replacement would be of great interest to be applied for the CLSR method for modelling the Stokes parameters.

- Combining several acceleration techniques would be an interesting topic of research. The combined use of the CLSR method, PCA-based methods and correlated $k$-distribution would be a worthy direction of research to see if the hybrid fast RTM methods have any impact on the acceleration.

- Coupling the CLSR method with ML techniques within the hyperspectral RTMs framework would be worth exploring for future works. Recent studies have shown that the accuracy of hyperspectral RT simulations can be improved by applying ML techniques. In this framework, NNs are used as universal approximators which replace time-consuming RTMs. Thus, an item of relevance for future investigations would be to replace the regression model of the CLSR method with a NN. A significant reduction in the number of calls to the MS RTM would be expected with the consequent acceleration.

# Bibliography


[1] A. D. Barnosky, E. A. Hadly, J. Bascompte, E. L. Berlow, J. H. Brown, M. Fortelius, W. M. Getz, J. Harte, A. Hastings, P. A. Marquet, N. D. Martinez, A. Mooers, P. Roopnarine, G. Vermeij, J. W. Williams, R. Gillespie, J. Kitzes, C. Marshall, N. Matzke, D. P. Mindell, E. Revilla, and A. B. Smith. Approaching a state shift in earth's biosphere. *Nature*, 486(7401):52–58, 2012. doi:10.1038/nature11018.

[2] J. P. Burrows, P. Borrell, and U. Platt, editors. *The Remote Sensing of Tropospheric Composition from Space*. Springer Berlin Heidelberg, 2011. doi:10.1007/978-3-642-14791-3.

[3] K. Urwin and A. Jordan. Does public policy support or undermine climate change adaptation? exploring policy interplay across different scales of governance. *Global Environmental Change*, 18(1):180–191, 2008. doi:10.1016/j.gloenvcha.2007.08.002.

[4] A. Tatem, S. Goetz, and S. Hay. Fifty years of earth-observation satellites. *American Scientist*, 96(5):390, 2008. doi:10.1511/2008.74.390.

[5] W. Fu, J. Ma, P. Chen, and F. Chen. *Remote Sensing Satellites for Digital Earth*, pages 55–123. Springer Singapore, Singapore, 2020. doi:10.1007/978-981-32-9915-3_3.

[6] J. Veefkind, I. Aben, K. McMullan, H. Forster, J. de Vries, G. Otter, J. Claas, H. Eskes, J. de Haan, Q. Kleipool, and et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, 120:70–83, 2012. doi:10.1016/j.rse.2011.09.027.

[7] A. del Águila, D. S. Efremenko, and T. Trautmann. A review of dimensionality reduction techniques for processing hyper-spectral optical signal. *Light & Engineering*, pages 85–98, 2019. doi:10.33383/2019-017.

[8] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie. Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51:47–60, 2015. doi:10.1016/j.future.2014.10.029.

[9] D. G. Loyola, J. Xu, K.-P. Heue, and W. Zimmer. Applying FP_ILM to the retrieval of geometry-dependent effective lambertian equivalent reflectivity (GE_LER) daily maps from UVN satellite measurements. *Atmospheric Measurement Techniques*, 13(2):985–999, 2020. doi:10.5194/amt-13-985-2020.

[10] S. F. Colosimo, V. Natraj, S. P. Sander, and J. Stutz. A sensitivity study on the retrieval of aerosol vertical profiles using the oxygen a-band. *Atmospheric Measurement Techniques*, 9(4):1889–1905, 2016. doi:10.5194/amt-9-1889-2016.

[11] V. Natraj, X. Jiang, R. Shia, X. Huang, J. Margolis, and Y. Yung. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the O2A-band. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 95(4):539–556, 2005. doi:10.1016/j.jqsrt.2004.12.024.

[12] Chapter 2 - solar radiation at the top of the atmosphere. In K. Liou, editor, *An Introduction to Atmospheric Radiation*, volume 84 of *International Geophysics*, pages 37–64. Academic Press, 2002. doi:10.1016/S0074-6142(02)80017-1.

[13] Chapter 3 - absorption and scattering of solar radiation in the atmosphere. In K. Liou, editor, *An Introduction to Atmospheric Radiation*, volume 84 of *International Geophysics*, pages 65–115. Academic Press, 2002. doi:10.1016/S0074-6142(02)80018-3.

[14] Diagram of atmospheric windows. https://earthobservatory.nasa.gov/ContentFeature/RemoteSensingAtmosphere/Images/windows.gif. Accessed: 01 June 2021.

[15] E. Lommel. Die photometrie der diffusen zurückwerfung. *Annalen der Physik*, 272(2):473–502, 1889. doi:10.1002/andp.18892720211.

[16] M. I. Mishchenko, L. D. Travis, and A. A. Lacis. *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*. Cambridge University Press, Cambridge, 2006.

[17] L. S. Dolin. Propagation of a narrow beam of light in a medium with strongly anisotropic scattering. *Soviet Radiophysics*, 9(1):40–47, 1967. doi:10.1007/bf01108374.

[18] L. Apresyan and Y. A. Kravtsov. Radiation transfer theory: Statistical and wave aspects. 1996.

[19] A. Doicu and M. I. Mishchenko. Overview of methods for deriving the radiative transfer theory from the maxwell equations. i: Approach based on the far-field foldy equations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 220:123–139, 2018. doi:10.1016/j.jqsrt.2018.09.004.

[20] A. Doicu and M. I. Mishchenko. An overview of methods for deriving the radiative transfer theory from the maxwell equations. II: Approach based on the dyson and bethe–salpeter equations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 224:25–36, 2019. doi:10.1016/j.jqsrt.2018.10.032.

[21] J. B. Aizenberg and V. P. Budak. The science of light engineering, fields of application and theoretical foundations. *Light & Engineering*, pages 4–6, 2018. doi:10.33383/2018-045.

[22] J. Aizenberg and V. Budak. Light science is not only science of lighting: theoretical bases and application area. In *Proceeding of the 29th Quadrennial Session of the CIE*. International Commission on Illumination, CIE, 2019. doi:10.25039/x46.2019.po120.

[23] D. Efremenko and A. Kokhanovsky. Light scattering, absorption, extinction, and propagation in the terrestrial atmosphere. In *Foundations of Atmospheric Remote Sensing*, pages 77–147. Springer International Publishing, 2021. doi:10.1007/978-3-030-66745-0_3.

[24] D. Efremenko and A. Kokhanovsky. Radiative transfer models. In *Foundations of Atmospheric Remote Sensing*, pages 149–232. Springer International Publishing, 2021. doi:10.1007/978-3-030-66745-0_4.

[25] S. Chandrasekhar. *Radiative transfer*. London: Oxford University Press, 1950.

[26] A. Doicu and T. Trautmann. Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(1-2):146–158, 2009. doi:10.1016/j.jqsrt.2008.09.014.

[27] K. Stamnes and J. J. Stamnes, editors. *Radiative Transfer in Coupled Environmental Systems*. Wiley-VCH Verlag GmbH & Co. KGaA, 2015. doi:10.1002/9783527696604.

[28] A. A. Kokhanovsky. Small-angle approximations of the radiative transfer theory. *Journal of Physics D: Applied Physics*, 30(20):2837–2840, 1997. doi:10.1088/0022-3727/30/20/009.

[29] A. A. Kokhanovsky. *Radiative Transfer*, pages 113–206. Springer Netherlands, Dordrecht, 2006. doi:10.1007/1-4020-4020-2_3.

[30] V. P. Budak, D. A. Klyuykov, and S. V. Korkin. Convergence acceleration of radiative transfer equation solution at strongly anisotropic scattering. In *Light Scattering Reviews 5*, pages 147–203. Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-10336-0_5.

[31] V. P. Budak, D. A. Klyuykov, and S. V. Korkin. Complete matrix solution of radiative transfer equation for PILE of horizontally homogeneous slabs. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(7):1141–1148, 2011. doi:10.1016/j.jqsrt.2010.08.028.

[32] V. P. Budak, D. S. Efremenko, and O. V. Shagalov. Efficiency of algorithm for solution of vector radiative transfer equation in turbid medium slab. *Journal of Physics: Conference Series*, 369:012021, 2012. doi:10.1088/1742-6596/369/1/012021.

[33] V. P. Afanas'ev, A. Y. Basov, V. P. Budak, D. S. Efremenko, and A. A. Kokhanovsky. Analysis of the discrete theory of radiative transfer in the coupled "ocean–atmosphere" system: Current status, problems and development prospects. *Journal of Marine Science and Engineering*, 8(3):202, 2020. doi:10.3390/jmse8030202.

[34] D. S. Efremenko, V. M. García, S. G. García, and A. Doicu. A review of the matrix-exponential formalism in radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 196:17–45, 2017. doi:10.1016/j.jqsrt.2017.02.015.

[35] V. P. Budak, G. A. Kaloshin, O. V. Shagalov, and V. S. Zheltov. Numerical modeling of the radiative transfer in a turbid medium using the synthetic iteration. *Optics Express*, 23(15):A829, 2015. doi:10.1364/oe.23.00a829.

[36] A. del Águila, D. S. Efremenko, V. Molina García, and J. Xu. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the hartley-huggins band. *Atmosphere*, 10(3):142, 2019. doi:10.3390/atmos10030142.

[37] V. Ambartsumian. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars. *Publications of the Astronomical Observatory of Leningrad State University*, 6:7–18, 1936.

[38] R. Goody, R. West, L. Chen, and D. Crisp. The correlated k-method for radiation calculations in nonhomogeneous atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 42(6):539–550, 1989. doi:10.1016/0022-4073(89)90044-7.

[39] C. Helling and U. G. Jorgensen. Optimizing the opacity sampling method. *Astronomy and Astrophysics*, 337:477–486, 1998.

[40] F. André, L. Hou, M. Roger, and R. Vaillon. The multispectral gas radiation modeling: A new theoretical framework based on a multidimensional approach to k-distribution methods. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 147:178–195, 2014. doi:10.1016/j.jqsrt.2014.05.021.

[41] B. A. Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 1. FKDM, fast k-distribution model for the longwave. *Journal of Geophysical Research*, 109(D2), 2004. doi:10.1029/2003jd003802.

[42] B. A. Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. *Journal of Geophysical Research*, 110(D2), 2005. doi:10.1029/2004jd005163.

[43] V. A. Falaleeva and B. A. Fomin. Overcoming spectroscopic challenges in direct problems of satellite sounding of the atmosphere. *Atmospheric and Oceanic Optics*, 30(1):1–6, 2017. doi:10.1134/s1024856017010055.

[44] G. E. Hunt and I. P. Grant. Discrete space theory of radiative transfer and its application to problems in planetary atmospheres. *Journal of the Atmospheric Sciences*, 26(5):963–972, 1969. doi:10.1175/1520-0469(1969)026<0963:dstort>2.0.co;2.

[45] W. Wiscombe and J. Evans. Exponential-sum fitting of radiative transmission functions. *Journal of Computational Physics*, 24(4):416–444, 1977. doi:10.1016/0021-9991(77)90031-6.

[46] M.-D. Chou, W. L. Ridgway, and M. M.-H. Yan. One-parameter scaling and exponential-sum fitting for water vapor and CO2 infrared transmission functions. *Journal of the Atmospheric Sciences*, 50(14):2294–2303, 1993. doi:10.1175/1520-0469(1993)050<2294:opsaes>2.0.co;2.

[47] Y. Mano. Modified esft method for application to atmospheric radiation. *Papers in Meteorology and Geophysics*, 46(1):1–8, 1995. doi:10.2467/mripapers.46.1.

[48] D. P. Kratz. The correlated k-distribution technique as applied to the AVHRR channels. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 53(5):501–517, 1995. doi:10.1016/0022-4073(95)90050-0.

[49] R. West, D. Crisp, and L. Chen. Mapping transformations for broadband atmospheric radiation calculations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 43(3):191–199, 1990. doi:10.1016/0022-4073(90)90051-7.

[50] V. S. Meadows and D. Crisp. Ground-based near-infrared observations of the venus nightside: The thermal structure and water abundance near the surface. *Journal of Geophysical Research: Planets*, 101(E2):4595–4622, 1996. doi:10.1029/95je03567.

[51] R. Bennartz and J. Fischer. A modified k-distribution approach applied to narrow band water vapour and oxygen absorption estimates in the near infrared, 2000.

[52] R. Bennartz and R. Preusker. k-binning: a new approach to simulate narrow band satellite channels in layered atmospheres with variable gas absorption. http://web.gps.caltech.edu/~vijay/Papers/RT%20Fundamentals/preusker-bennartz-06.pdf, 2006. Accessed: 5 May 2020.

[53] E. Boesche, P. Stammes, R. Preusker, R. Bennartz, W. Knap, and J. Fischer. Polarization of skylight in the o_2a band: effects of aerosol properties. *Applied Optics*, 47(19):3467, 2008. doi:10.1364/ao.47.003467.

[54] J. F. de Haan, P. B. Bosma, and J. W. Hovenier. The adding method for multiple scattering calculations of polarized light. *Astronomy and Astrophysics*, 183(2):371–391, 1987.

[55] M. Duan. A fast radiative transfer model for simulating high-resolution absorption bands. *Journal of Geophysical Research*, 110(D15), 2005. doi:10.1029/2004jd005590.

[56] V. Natraj. A review of fast radiative transfer techniques. In *Light Scattering Reviews 8*, pages 475–504. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-32106-1_10.

[57] J.-L. Moncet, G. Uymin, A. E. Lipton, and H. E. Snell. Infrared radiance modeling by optimal spectral sampling. *Journal of the Atmospheric Sciences*, 65(12):3917–3934, 2008. doi:10.1175/2008jas2711.1.

[58] O. P. Hasekamp and A. Butz. Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres. *Journal of Geophysical Research*, 113(D20), 2008. doi:10.1029/2008jd010379.

[59] C. W. O'Dell. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. *Journal of Geophysical Research*, 115(D10), 2010. doi:10.1029/2009jd012803.

[60] A. del Águila, D. S. Efremenko, V. Molina García, and M. Y. Kataev. Cluster low-streams regression method for hyperspectral radiative transfer computations: Cases of O2 A- and CO2 bands. *Remote Sensing*, 12(8):1250, 2020. doi:10.3390/rs12081250.

[61] X. Liu, W. Smith, D. Zhou, and A. Larar. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics*, 45(1):201–208, 2006. doi:10.1364/AO.45.000201.

[62] D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. 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, 2014. doi:10.1016/j.jqsrt.2013.07.023.

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

[64] A. Kuze, H. Suto, K. Shiomi, S. Kawakami, M. Tanaka, Y. Ueda, A. Deguchi, J. Yoshida, Y. Yamamoto, F. Kataoka, T. E. Taylor, and H. L. Buijs. Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space. *Atmospheric Measurement Techniques*, 9(6):2445–2461, 2016. doi:10.5194/amt-9-2445-2016.

[65] P. Somkuti, H. Boesch, V. Natraj, and P. Kopparla. Application of a PCA-based fast radiative transfer model to $xco_2$ retrievals in the shortwave infrared. *Journal of Geophysical Research: Atmospheres*, 122(19):10477–10496. doi:10.1002/2017JD027013.

[66] P. Kopparla, V. Natraj, R. Spurr, R. Shia, D. Crisp, and Y. Yung. 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, 2016. doi:10.1016/j.jqsrt.2016.01.014.

[67] P. Kopparla, V. Natraj, D. Limpasuvan, R. Spurr, D. Crisp, R.-L. Shia, P. Somkuti, and Y. L. Yung. Pca-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 198:104–111, 2017. doi:10.1016/j.jqsrt.2017.05.005.

[68] G. M. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the Spring Joint Computer Conference*, AFIPS '67 (Spring), pages 483–485, New York, NY, USA, 1967. ACM.

[69] V. Molina García, S. Sasi, D. S. Efremenko, A. Doicu, and D. Loyola. Linearized radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 213:241–251, 2018. doi:10.1016/j.jqsrt.2018.03.008.

[70] J. Bak, X. Liu, R. Spurr, K. Yang, C. R. Nowlan, C. C. Miller, G. G. Abad, and K. Chance. Radiative transfer acceleration based on the principal component analysis and look-up table of corrections: Optimization and application to UV ozone profile retrievals. 2020. doi:10.5194/amt-2020-349.

[71] X. Liu, Q. Yang, H. Li, Z. Jin, W. Wu, S. Kizer, D. K. Zhou, and P. Yang. Development of a fast and accurate PCRTM radiative transfer model in the solar spectral region. *Applied Optics*, 55(29):8236, 2016. doi:10.1364/ao.55.008236.

[72] X. Liu, W. Wu, and Q. Yang. Multiple scattering principal component-based radiative transfer model (pcrtm) from far ir to uv-vis. In *AGU Fall Meeting Abstracts*, volume 2017, pages A23A–2309, 2017.

[73] M. Matricardi. A principal component based version of the RTTOV fast radiative transfer model. *Quarterly Journal of the Royal Meteorological Society*, 136:1823–1835, 2010. doi:10.1002/qj.680.

[74] A. Hollstein and R. Lindstrot. Fast reconstruction of hyperspectral radiative transfer simulations by using small spectral subsets: application to the oxygen A

band. *Atmospheric Measurement Techniques*, 7(2):599–607, 2014. doi:10.5194/amt-7-599-2014.

[75] C. Liu, B. Yao, V. Natraj, P. Kopparla, F. Weng, T. Le, R.-L. Shia, and Y. L. Yung. A spectral data compression (SDCOMP) radiative transfer model for high-spectral-resolution radiation simulations. *Journal of the Atmospheric Sciences*, 77(6):2055–2066, 2020. doi:10.1175/jas-d-19-0238.1.

[76] J. R. Key and A. J. Schweiger. Tools for atmospheric radiative transfer: Streamer and FluxNet. *Computers & Geosciences*, 24(5):443–451, 1998. doi:10.1016/s0098-3004(97)00130-1.

[77] H. Schwander, A. Kaifel, A. Ruggaber, and P. Koepke. Spectral radiative-transfer modeling with minimized computation time by use of a neural-network technique. *Applied Optics*, 40(3):331, 2001. doi:10.1364/ao.40.000331.

[78] A. S. Dorvlo, J. A. Jervase, and A. Al-Lawati. Solar radiation estimation using artificial neural networks. *Applied Energy*, 71(4):307–319, 2002. doi:10.1016/s0306-2619(02)00016-8.

[79] L. F. Zarzalejo, L. Ramirez, and J. Polo. Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index. *Energy*, 30(9):1685–1697, 2005. doi:10.1016/j.energy.2004.04.047.

[80] D. G. Loyola. Applications of neural network methods to the processing of earth observation satellite data. *Neural Networks*, 19(2):168–177, 2006. doi:10.1016/j.neunet.2006.01.010.

[81] H. Takenaka, T. Y. Nakajima, A. Higurashi, A. Higuchi, T. Takamura, R. T. Pinker, and T. Nakajima. Estimation of solar radiation using a neural network based on radiative transfer. *Journal of Geophysical Research*, 116(D8), 2011. doi:10.1029/2009jd013337.

[82] D. G. Loyola, S. G. García, R. Lutz, A. Argyrouli, F. Romahn, R. J. D. Spurr, M. Pedergnana, A. Doicu, V. M. García, and O. Schüssler. The operational cloud retrieval algorithms from TROPOMI on board sentinel-5 precursor. *Atmospheric Measurement Techniques*, 11(1):409–427, 2018. doi:10.5194/amt-11-409-2018.

[83] B. D. Bue, D. R. Thompson, S. Deshpande, M. Eastwood, R. O. Green, V. Natraj, T. Mullen, and M. Parente. Neural network radiative transfer for imaging spectroscopy. *Atmospheric Measurement Techniques*, 12(4):2567–2578, 2019. doi:10.5194/amt-12-2567-2019.

[84] T. Le, C. Liu, B. Yao, V. Natraj, and Y. L. Yung. Application of machine learning to hyperspectral radiative transfer simulations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 246:106928, 2020. doi:10.1016/j.jqsrt.2020.106928.

[85] L. D. Girolamo. Reciprocity principle applicable to reflected radiance measurements and the searchlight problem. *Applied Optics*, 38(15):3196, 1999. doi:10.1364/ao.38.003196.

[86] M. Hess, P. Koepke, and I. Schult. Optical properties of aerosols and clouds: The software package OPAC. *Bulletin of the American Meteorological Society*, 79(5):831–844, 1998. doi:10.1175/1520-0477(1998)079<0831:opoaac>2.0.co;2.

[87] F. Schreier, S. Gimeno García, P. Hochstaffl, and S. Städt. Py4cats—PYthon for computational ATmospheric spectroscopy. *Atmosphere*, 10(5):262, 2019. doi:10.3390/atmos10050262.

[88] I. Gordon, L. Rothman, C. Hill, R. Kochanov, Y. Tan, P. Bernath, M. Birk, V. Boudon, A. Campargue, K. Chance, B. Drouin, J.-M. Flaud, R. Gamache, J. Hodges, D. Jacquemart, V. Perevalov, A. Perrin, K. Shine, M.-A. Smith, J. Tennyson, G. Toon, H. Tran, V. Tyuterev, A. Barbe, A. Császár, V. Devi, T. Furtenbacher, J. Harrison, J.-M. Hartmann, A. Jolly, T. Johnson, T. Karman, I. Kleiner, A. Kyuberis, J. Loos, O. Lyulin, S. Massie, S. Mikhailenko, N. Moazzen-Ahmadi, H. Müller, O. Naumenko, A. Nikitin, O. Polyansky, M. Rey, M. Rotger, S. Sharpe, K. Sung, E. Starikova, S. Tashkun, J. V. Auwera, G. Wagner, J. Wilzewski, P. Wcisło, S. Yu, and E. Zak. The HITRAN2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 203:3–69, 2017. doi:10.1016/j.jqsrt.2017.06.038.

[89] C. Richard, I. Gordon, L. Rothman, M. Abel, L. Frommhold, M. Gustafsson, J.-M. Hartmann, C. Hermans, W. Lafferty, G. Orton, K. Smith, and H. Tran. New section of the HITRAN database: Collision-induced absorption (CIA). *Journal of Quantitative Spectroscopy and Radiative Transfer*, 113(11):1276–1285, 2012. doi:10.1016/j.jqsrt.2011.11.004.

[90] E. J. Mlawer, V. H. Payne, J.-L. Moncet, J. S. Delamere, M. J. Alvarado, and D. C. Tobin. Development and recent evaluation of the MT_CKD model of continuum absorption. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1968):2520–2556, 2012. doi:10.1098/rsta.2011.0295.

[91] B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser. On Rayleigh optical depth calculations. *Journal of Atmospheric and Oceanic Technology*, 16(11):1854–1861, 1999. doi:10.1175/1520-0426(1999)016<1854:orodc>2.0.co;2.

[92] G. Anderson, S. Clough, F. Kneizys, J. Chetwynd, and E. Shettle. *AFGL Atmospheric Constituent Profiles (0-120 km)*. Air Force Geophysics Laboratory, Hanscom Air Force Base, MA, USA, 1986.

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

[94] D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. Acceleration techniques for the discrete ordinate method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 114:73–81, 2013. doi:10.1016/j.jqsrt.2012.08.014.

[95] P. C. Waterman. Matrix-exponential description of radiative transfer. *Journal of the Optical Society of America*, 71(4):410, 1981. doi:10.1364/josa.71.000410.

[96] T. Nakajima and M. Tanaka. Matrix formulations for the transfer of solar radiation in a plane-parallel scattering atmosphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 35(1):13–21, 1986. doi:10.1016/0022-4073(86)90088-9.

[97] V. P. Afanas'ev, D. S. Efremenko, and A. V. Lubenchenko. On the application of the invariant embedding method and the radiative transfer equation codes for

surface state analysis. In A. A. Kokhanovsky, editor, *Light Scattering Reviews 8: Radiative transfer and light scattering*, pages 363–423. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-32106-1_8.

[98] J. Fischer and H. Grassl. Radiative transfer in an atmosphere-ocean system: an azimuthally dependent matrix-operator approach. *Applied Optics*, 23(7):1032, 1984. doi:10.1364/ao.23.001032.

[99] E. Chalhoub and R. Garcia. The equivalence between two techniques of angular interpolation for the discrete-ordinates method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 64(5):517–535, 2000. doi:10.1016/s0022-4073(99)00134-x.

[100] W. Wiscombe. The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *Journal of the Atmospheric Sciences*, 34(9):1408–1422, 1977. doi:10.1175/1520-0469(1977)034<1408:TDMRYA>2.0.CO;2.

[101] T. Nakajima and M. Tanaka. Algorithms for radiative intensity calculations in moderately thick atmos using a truncation approximation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 40(1):51–69, 1988. doi:10.1016/0022-4073(88)90031-3.

[102] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'ıo, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi:10.1038/s41586-020-2649-2.

[103] Pydome: python library for radiative transfer computations. https://elib.dlr.de/128047/. Accessed: 12 April 2021.

[104] V. Roozendael, R. Spurr, D. Loyola, C. Lerot, D. Balis, J. Lambert, W. Zimmer, J. Gent, J. Van Geffen, M. Koukouli, J. Granville, A. Doicu, C. Fayt, and C. Zehner. Sixteen years of GOME/ERS2 total ozone data: the new direct-fitting GOME Data Processor (GDP) Version 5: I. algorithm description. *Journal of Geophysical Research: Atmospheres*, 117:D03305(1–18), 2012. doi:10.1029/2011JD016471.

[105] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi:10.1080/14786440109462720.

[106] S. Beirle, J. Lampel, C. Lerot, H. Sihler, and T. Wagner. Parameterizing the instrumental spectral response function and its changes by a super-Gaussian and its derivatives. *Atmospheric Measurement Techniques*, 10(2):581–598, 2017. doi:10.5194/amt-10-581-2017.

[107] GOSAT spectral resolution. http://www.gosat-2.nies.go.jp/about/spacecraft_and_instruments/. Accessed: 17 March 2020.

[108] A. del Aguila and D. Efremenko. The cluster low-streams regression method for fast computations of top-of-the-atmosphere radiances in absorption bands. *Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020). Part 2*, pages short25–1–short25–9, 2020. doi:10.51130/graphicon-2020-2-4-25.

[109] A. del Águila and D. S. Efremenko. Accuracy enhancement of the two-stream radiative transfer model for computing absorption bands at the presence of aerosols. *Light & Engineering*, pages 79–86, 2021. doi:10.33383/2020-078.

[110] A. del Águila and D. S. Efremenko. Fast hyper-spectral radiative transfer model based on the double cluster low-streams regression method. *Remote Sensing*, 13(3):434, 2021. doi:10.3390/rs13030434.

[111] R. Spurr, V. Natraj, C. Lerot, M. V. Roozendael, and D. Loyola. 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, 2013. doi:10.1016/j.jqsrt.2013.04.002.

[112] D. S. Efremenko, D. G. Loyola, R. J. Spurr, and A. Doicu. Acceleration of radiative transfer model calculations for the retrieval of trace gases under cloudy conditions. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 135:58–65, 2014. doi:10.1016/j.jqsrt.2013.11.014.

[113] V. Molina García, S. Sasi, D. Efremenko, A. Doicu, and D. Loyola. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 213:228–240, 2018. doi:10.1016/j.jqsrt.2018.03.014.

[114] A. Doicu, D. Efremenko, and T. Trautmann. A spectral acceleration approach for the spherical harmonics discrete ordinate method. *Remote Sensing*, 12(22):3703, 2020. doi:10.3390/rs12223703.

[115] J. Vicent, J. Verrelst, J. Rivera-Caicedo, N. Sabater, J. Munoz-Mari, G. Camps-Valls, and J. Moreno. Statistical learning for end-to-end simulations. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018. doi:10.1109/igarss.2018.8518366.

[116] J. Vicent Servera, L. Alonso, L. Martino, N. Sabater, J. Verrelst, G. Camps-Valls, and J. Moreno. Gradient-based automatic lookup table generator for radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):1040–1048, 2019. doi:10.1109/tgrs.2018.2864517.

# Acknowledgements

This thesis has been possible thanks to the support and encouragement of a great deal of people during my life including my family, friends, supervisors and colleagues.

I would first like to thank my supervisor, PD Dr. Dmitry Efremenko, for giving me the opportunity to work on this thesis. You didn't doubt for any second to take me as your PhD student and that trust has been a strong support over the years. Your immense knowledge and invaluable expertise have inspired and encouraged me during my thesis, resulting in a fruitful journey as you have been an outstanding role model. I will always appreciate your constant dedication, full availability and your high-quality advice as a researcher. This thesis would not have been possible without your supervision.

The same sincere gratitude goes to Prof. Dr. Richard Bamler, for his invaluable scientific and professional guidance and to Dr. Adrian Doicu, who has been a strong support for the development of my thesis thanks to his profound mathematical knowledge and our fruitful discussions in radiative transfer theory. I would also like to thank Prof. Dr. Thomas Trautmann, which has always encouraged me through my thesis and has been available for scientific and valuable discussions. His wisdom and integrity have been a great example to me.

This thesis would not have been possible without the financial support of the German Academic Exchange Service (DAAD) and the Remote Sensing Technology Institute (IMF) at DLR that have provided me with the resources to perform my work. Even in the difficult times of the pandemic I have always found support and great help, for which I will be eternally grateful.

I would like to acknowledge all my colleagues from the Atmospheric Processors (ATP) department for sharing their expertise and for providing valuable feedback. I would like to thank the Passive Sensing team leader Dr. Diego Loyola for academic discussions and advice. Many thanks goes to Nathalie, for taking care of the administrative matters and for her kindness. I'm sincerely grateful for all the delightful moments shared with all my colleagues during coffees, cakes and gatherings.

A special mention goes to Víctor, who has helped me so much during all these years. From the very beginning you have been a great support with the thesis and I have learned a lot from our discussions. I would also like to express my warmest gratitude to Sruthy, my office mate and my friend. You have been life changing and we have shared so many good moments that I will always carry with me.

Going to DLR was a strong motivation from the very first moment thanks to the good friends I made and with whom I shared many lunches, coffees and great times that have always cheered me up, specially Athina, Fabian, Mario, Özge, Song, Sruthy and Víctor.

Finally, I would like to give my deepest gratitude to my friends and family, who have been a strong pillar for this thesis and have influenced and shaped me into the person I am today. I would like to make a special mention of those who have influenced me the most:

To my grandfather, who has been an example of positivity and how to handle life's difficult situations with courage. Today he would be proud of me as I have always been proud of him.

To my sister, Elena, who means intelligence and love to me. You are my best friend and I admire your strength and attitude to face life. Thank you for teaching me to be a better person with you as my inspiration.

To Alex, the person who I admire everyday and who has supported and loved me over the years, no matter the distance or the circumstances. Having you around means laughing, security, critical thinking, endless conversations and love. I have to say that always being able to count on you is the best possible gift I have ever had, for which I feel enormously lucky.

I would like to dedicate this thesis to my parents Ana and Francisco, for raising me as a caring and hard-working person with their daily example. You are the best role models I could have had, always supportive, willing to give wise advice and making me aware that every effort has its rewards. For as long as I can remember you have accompanied me in every step, in every decision, always encouraging me to seek the best future and letting me choose. I will always be grateful for your unconditional support.

# Appendices

# A Publications

## A.1 Publication I: Analysis of Two Dimensionality Reduction Techniques for Fast Simulation of the Spectral Radiances in the Hartley-Huggins Band

### Reference

A. del Águila, D. S. Efremenko, V. Molina García, and J. Xu. Analysis of Two Dimensionality Reduction Techniques for Fast Simulation of the Spectral Radiances in the Hartley-Huggins Band. Atmosphere, 10(3):142, 2019. doi:10.3390/atmos10030142

### Copyright



### Abstract

The new generation of atmospheric composition sensors such as TROPOMI is capable of providing spectra of high spatial and spectral resolution. To process this vast amount of spectral information, fast radiative transfer models (RTMs) are required. In this regard, we analyzed the efficiency of two acceleration techniques based on the principal component analysis (PCA) for simulating the Hartley-Huggins band spectra. In the first one, the PCA is used to map the data set of optical properties of the atmosphere to a lower-dimensional subspace, in which the correction function for an approximate but fast RTM is derived. The second technique is based on the dimensionality reduction of the data set of spectral radiances. Once the empirical orthogonal functions are found, the whole spectrum can be reconstructed by performing radiative transfer computations only for a specific subset of spectral points. We considered a clear-sky atmosphere where the optical properties are defined by Rayleigh scattering and trace gas absorption. Clouds can be integrated into the model as Lambertian reflectors. High computational performance is achieved by combining both techniques without losing accuracy. We found that for the Hartley-Huggins band, the combined use of these techniques yields an accuracy better than 0.05% while the speedup factor is about 20. This innovative combination of both PCA-based techniques can be applied in future works as an efficient approach for simulating the spectral radiances in other spectral regions.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko; V.M.G: Víctor Molina García; J.X.: Jian Xu)
As stated in the article, the author contributions are the following: The TROPOMI data

generation was done by J.X., the investigation and simulations were done by A.d.Á. The methodology development was done by D.S.E. and the validation by V.M.G. The formal analysis was carried out by A.d.Á. and D.S.E. The creation of the figures, tables and schemes was done by A.d.Á. and writing the manuscript was done by A.d.Á and D.S.E. with reviews and editing by V.M.G. and supervision of D.S.E.

The overall own contribution of A.d.Á for the publication in A.1. is estimated at 76 %, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.1).

**Table A.1:** Criteria and estimated own contribution for Appendix A.1

| **Criteria** | **Estimated contribution** |
|---|---|
| Conceptualization | 50% |
| Implementation | 90% |
| Analysis and discussion | 80% |
| Figures and tables | 100% |
| Manuscript writing | 60% |
| **Total** | **76** % |

*atmosphere*

MDPI

*Article*

# Analysis of Two Dimensionality Reduction Techniques for Fast Simulation of the Spectral Radiances in the Hartley-Huggins Band

**Ana del Águila *, Dmitry S. Efremenko, Víctor Molina García and Jian Xu**

Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany; Dmitry.Efremenko@dlr.de (D.S.E.); Victor.MolinaGarcia@dlr.de (V.M.G.); Jian.Xu@dlr.de (J.X.)

* Correspondence: Ana.delAguilaPerez@dlr.de; Tel.: +49-8153-28-1983

Received: 22 February 2019; Accepted: 13 March 2019; Published: 16 March 2019


**Abstract:** The new generation of atmospheric composition sensors such as TROPOMI is capable of providing spectra of high spatial and spectral resolution. To process this vast amount of spectral information, fast radiative transfer models (RTMs) are required. In this regard, we analyzed the efficiency of two acceleration techniques based on the principal component analysis (PCA) for simulating the Hartley-Huggins band spectra. In the first one, the PCA is used to map the data set of optical properties of the atmosphere to a lower-dimensional subspace, in which the correction function for an approximate but fast RTM is derived. The second technique is based on the dimensionality reduction of the data set of spectral radiances. Once the empirical orthogonal functions are found, the whole spectrum can be reconstructed by performing radiative transfer computations only for a specific subset of spectral points. We considered a clear-sky atmosphere where the optical properties are defined by Rayleigh scattering and trace gas absorption. Clouds can be integrated into the model as Lambertian reflectors. High computational performance is achieved by combining both techniques without losing accuracy. We found that for the Hartley-Huggins band, the combined use of these techniques yields an accuracy better than 0.05% while the speedup factor is about 20. This innovative combination of both PCA-based techniques can be applied in future works as an efficient approach for simulating the spectral radiances in other spectral regions.

**Keywords:** PCA; data-driven algorithms; Hartley-Huggins band; dimensionality reduction techniques; two-stream; multi-stream

## 1. Introduction

The new generation of atmospheric composition sensors such as the TROPospheric Ozone Monitoring Instrument (TROPOMI) delivers a great amount of data. Extracting the information about geophysical parameters (e.g., trace gas concentrations) from spectral radiances requires high-performance computing. The key component of retrieval algorithms is the radiative transfer models (RTMs), which convert optical parameters of the atmosphere (i.e., input space) into spectral radiances (i.e., output space). In hyper-spectral remote sensing retrieval applications, the radiative transfer computations are the bottleneck in the whole processing chain. In this regard, acceleration techniques for hyper-spectral RTMs have to be developed [1]. The correlated *k*-distribution [2], the exponential-sum fitting transmittance [3], the radiance sampling method [4], and the optimal spectral mapping method [5] are examples of them. Essentially, these techniques do not compute spectral radiances in a line-by-line manner, but rather they take into account the interdependency between spectral channels or group them accordingly, thereby reducing the total number of calls to RTMs.

In the RTM development, there has been a rapid rise in the use of the data reduction concept, e.g., the principal component analysis (PCA). The acceleration techniques based on PCA can be broadly categorized into two groups: data reduction of optical parameters (input space reduction) and data reduction of spectral radiances (output space reduction).

The input space reduction technique was proposed by Natraj et al. [6], and its main idea is to reproduce the residual between the accurate and the approximate RTMs in the reduced space of optical parameters. This approach has the following attributes:

- the approximate model is a two-stream radiative transfer model, while the accurate model is a multi-stream radiative transfer model;
- the PCA is used to reduce the dimensionality of the optical parameters of the atmospheric system;
- the dependency of the correction factor on the optical parameters is modeled by a second-order Taylor expansion about the mean value of the optical parameters.

Previously, this approach was applied for simulating the spectra in the $O_2$ A-band [6], the Huggins band [7], and the $CO_2$ bands [8,9]. Kopparla et al. [10] applied a similar approach to modeling the radiances in the wide spectral range (0.3–3000 nm). In all cases, authors reported that the root-mean-square errors of the computed radiances are about 0.3%, yet achieving almost a 10-fold increase in speed.

In the output space reduction technique, the principal component analysis (PCA) is used to map the spectral radiances into a lower-dimensional subspace and to obtain a set of empirical orthogonal functions. The spectrum at full spectral resolution can be reproduced by computing the radiances only at certain wavelengths. Such an approach is used in several fast RTMs (e.g., PCRTM [11], RTTOV [12] and others [13,14]) and retrieval algorithms [15,16]. To select the most representative spectral subsets, the spectral sampling methods are used (see, e.g., [11,14] and references therein).

Note that, although both techniques are often called the same (namely "PCA-based RTMs"), they are fundamentally different. In the first case, PCA is applied to a data set of optical parameters, comprising the total optical thicknesses and the single scattering albedos for all given atmospheric layers and wavelengths. In the second case, PCA is applied to the data set of spectral radiances.

In the current study, we investigate the efficiency of the dimensionality reduction technique of optical parameters in the Hartley-Huggins band and improve upon the methods to maximize the advantages of the data reduction concept. First, we examine the efficiency of the input space reduction technique for simulating TROPOMI signals in the ultraviolet spectral range, which is typically used for ozone retrieval [17]. As approximate models, we consider the two-stream model and the single-scattering model. The spectral splitting is applied to improve the accuracy of the method. Second, we analyze the effect of including higher-order terms in the Taylor expansion of the correction function. Third, we apply the output space reduction technique to the data set of synthetic TROPOMI spectra and analyze the possible benefits of it. Fourth, we show how the input space reduction technique can be used in conjunction with the output space reduction method. The paper is organized as follows. In Section 2 we explain the input space reduction technique. The output space reduction and the spectral sampling techniques are described in Section 3. In Section 4, we evaluate the performance of these methods and discuss the combined usage of them.

## 2. Input Space Reduction Technique

### *2.1. Radiative Transfer Models*

Many existing RTMs are based on the discrete ordinate method [18]. The number of streams (discrete ordinates) in the polar hemisphere, $N_{\mathrm{do}}$, is an important parameter controlling the computational time and accuracy. Thus, RTMs are called "multi-stream" if $N_{\mathrm{do}} \geq 2$ and "two-stream" (TS) if $N_{\mathrm{do}} = 1$. The simplest analytical RTM is the single-scattering (SS) model, in which multiple scattering events are neglected. The TS and the SS RTMs are considerably faster than the multi-stream

models [7]. However, they suffer from a lack of accuracy. TS and SS models are taken in this study as approximate models for fast computations of the spectral radiances. Then the multi-stream model is applied to correct them.

In this study, the multi-stream model is based on the discrete ordinate method with matrix exponential (DOME) [19–21] with 16 streams per hemisphere. Regarding the two-stream model, two comments are in order:

1. the radiation field is found as a sum of the single-scattering solution and the multiple scattering term;
2. in the multiple scattering computations the delta-M truncation method [3] is applied, while for the single-scattering term the exact phase function is used [22].

### 2.2. Correction Function in the Reduced Input Space

Considering a discretization of the atmosphere in $L$ layers, we define an $N$-dimensional vector $\mathbf{x}_w$, for each wavelength $\{\lambda_w\}_{w=1,\ldots,W}$, by

$$\mathbf{x}_w^{\mathrm{T}} = \left[\ln \tau_1\left(\lambda_w\right), \ldots, \ln \tau_L\left(\lambda_w\right), \ln \omega_1\left(\lambda_w\right), \ldots, \ln \omega_L\left(\lambda_w\right)\right], \tag{1}$$

where $\tau_k$ and $\omega_k$ are the optical thickness and the single-scattering albedo of the $k$-th layer, respectively, and $N = 2L$. Thus, the vector $\mathbf{x}_w$ encapsulates the wavelength variability of the optical parameters, which are the input parameters of the radiative transfer code. By using PCA, we find an $M$-dimensional subspace spanned by a set of linear independent vectors (empirical orthogonal functions) $\{\mathbf{q}_k\}_{k=1}^{M}$ such that the centered (mean-removed) data $\mathbf{x}_w - \bar{\mathbf{x}}$ lie mainly on this subspace, i.e.,

$$\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^{M} y_{wk}\mathbf{q}_k, \quad w = 1, \ldots, W, \tag{2}$$

where $y_{wk}$ are the principal component (PC) scores.

Let us define a correction function $f\left(\mathbf{x}_w\right)$ as follows:

$$f\left(\mathbf{x}_w\right) = \ln \frac{L\left(\mathbf{x}_w\right)}{L_{\mathrm{A}}\left(\mathbf{x}_w\right)}, \tag{3}$$

where $L_{\mathrm{A}}\left(\mathbf{x}_w\right)$ is the radiance provided by the approximate model (here A stands for "approximate" and refers to SS and TS models), and $L\left(\mathbf{x}_w\right)$ is the radiance simulated by the multi-stream model. Introducing $\Delta\mathbf{x}_w = \sum_{k=1}^{M} y_{wk}\mathbf{q}_k$, we consider the Taylor expansion of $f\left(\mathbf{x}_w\right)$ around $\bar{\mathbf{x}}$ in the direction $\Delta\mathbf{x}_w$ up to fourth order. For computational simplicity, we neglect the mixed directional derivatives and use central differences to approximate the directional derivatives. Since $M < N$ (and in practice $M \ll N$), it leads to a substantial reduction of the computational time. The final result for the correction function reads as

$$\begin{aligned} f\left(\mathbf{x}_w\right) \approx\ & f\left(\bar{\mathbf{x}}\right) + \frac{1}{2}\sum_{k=1}^{M}\left(f_k^{+} - f_k^{-}\right) y_{wk} + \frac{1}{2}\sum_{k=1}^{M}\left(f_k^{+} - 2f\left(\bar{\mathbf{x}}\right) + f_k^{-}\right) y_{wk}^2 + \\ & + \frac{1}{12}\sum_{k=1}^{M}\left(f_k^{+2} - 2f_k^{+} + 2f_k^{-} - f_k^{-2}\right) y_{wk}^3 + \\ & + \frac{1}{24}\sum_{k=1}^{M}\left(f_k^{+2} - 4f_k^{+} + 6f\left(\bar{\mathbf{x}}\right) - 4f_k^{-} + 2f_k^{-2}\right) y_{wk}^4, \end{aligned} \tag{4}$$

where $f_k^{\pm n} = f\left(\bar{\mathbf{x}} \pm n\mathbf{q}_k\right)$, being $h$ either $k$ or $k+1$. Once the correction function is computed, the results of the approximate RTM can be converted into those of the multi-stream RTM by using (3). A schematic representation of the input data processing is shown in Figure 1

To improve the accuracy of the input space reduction technique, PCA can be combined with the clustering of the data space, involving a two-step procedure [23]: (1) clustering of the input data space into disjoint regions (groups of wavelengths), and (2) dimensionality reduction within each region by using PCA. Thus, each region (group of wavelengths) is characterized by its own orthogonal basis, and so, by its own set of correction factors. If $P$ is the number of disjoint regions, then $(2M+1)P$ calls of the multi- and two-stream models are required to compute the correction function.

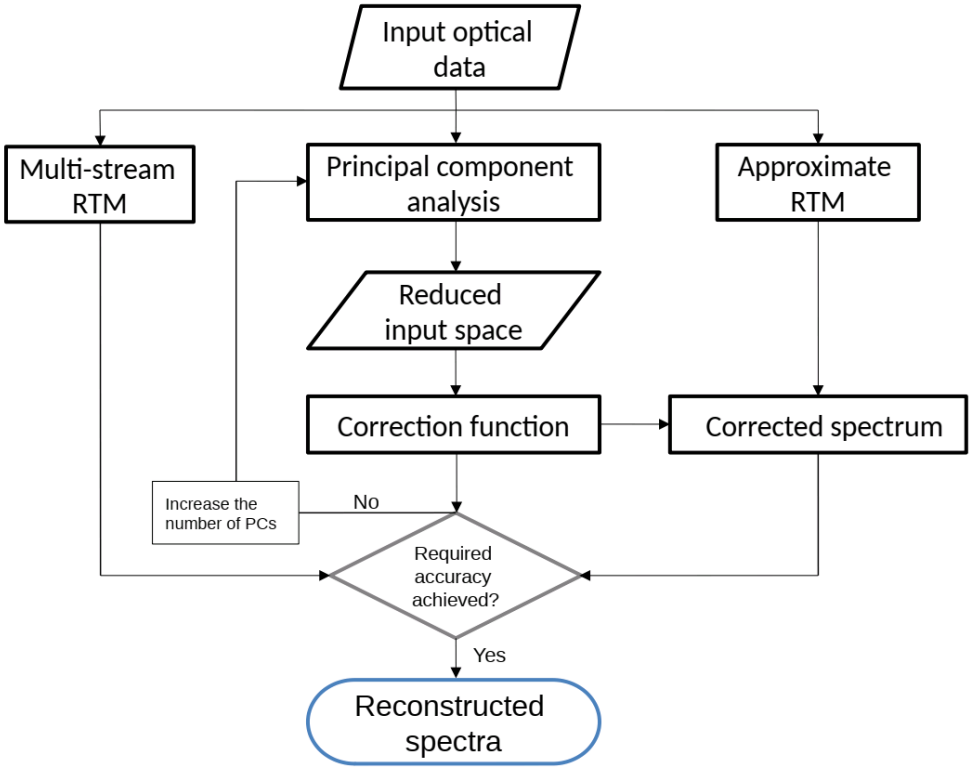


**Figure 1.** Schematic representation of the data processing in the input space technique with principal component analysis (PCA).

## 3. Output Space Reduction Technique

### *3.1. PCA Description*

Let $\mathbf{y} = [y(\lambda_1), y(\lambda_2), \ldots, y(\lambda_W)]$ be a row-vector of radiances at $W$ wavelengths $\{\lambda_w\}_{w=1,\ldots,W}$. A set of $S$ spectra is assembled into a matrix $\mathbf{Y}$ whose $i$-th row is $\mathbf{y}_i$. Then, a spectrum $\mathbf{y}_i$ can be represented in a new basis system as follows:

$$\mathbf{y}_i = \overline{\mathbf{y}} + \sum_{j=1}^{W} t_{ij}\mathbf{f}_j. \tag{5}$$

Here, $\overline{\mathbf{y}} = \sum_{i=1}^{S} \mathbf{y}_i / S$ is the sample mean of the spectra (the average spectrum), $t_{ij}$ is the $j$-th coordinate of the vector $\mathbf{y}_i$ in the new basis system and $\mathbf{f}_j = [f_j(\lambda_1), f_j(\lambda_2), \ldots, f_j(\lambda_W)]$ is the $j$-th basis vector. Next, the spectrum $\mathbf{y}_i$ is projected onto the $J$-dimensional subspace ($J < W$) as follows:

$$\mathbf{y}_i \approx \overline{\mathbf{y}} + \sum_{j=1}^{J} t_{ij}\mathbf{f}_j, \tag{6}$$

or in matrix form:

$$\mathbf{Y} \approx \bar{\mathbf{Y}} + \mathbf{TF}, \tag{7}$$

where $\overline{\mathbf{Y}} = \{\overline{\mathbf{y}}, \ldots, \overline{\mathbf{y}}\}$, $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_J\}^{\mathrm{T}}$ and $\mathbf{T}$ is the matrix whose entries are $\{t_{ij}\}_{i=1,\ldots,S}^{j=1,\ldots,J}$. The superscript $^{\mathrm{T}}$ stands for "transpose". The transformation (6) can be done by using PCA [24]. Then the basic vectors $\mathbf{f}_j$ in (6) are referred to as "principal components" (PCs) or empirical orthogonal functions (EOFs), while the coordinates $t_{ij}$ in the new coordinate system and the corresponding matrix $\mathbf{T}$ are called "principal component scores". Considering a spectral decomposition of the covariance matrix $\mathrm{cov}(\mathbf{Y}, \mathbf{Y}) \equiv \mathbf{C}_{\mathrm{Y}}$ :

$$\mathbf{C}_{\mathrm{Y}} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{\mathrm{T}}, \tag{8}$$

where $\mathbf{E}$ is the eigenvector matrix and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues, the EOFs are taken as $J$ eigenvectors related to the $J$ most significant eigenvalues. The PC scores are then computed as follows:

$$\mathbf{T} = (\mathbf{Y} - \bar{\mathbf{Y}})\,\mathbf{F}^{\mathrm{T}}. \tag{9}$$

### *3.2. Reconstruction of the Full Resolution Spectrum*

Let us consider a method of reconstruction of the full resolution spectrum containing $W$ spectral points using $J$ monochromatic radiances, given $J < W$. Assume that we have precomputed a data set of spectra. By applying PCA to it, EOFs are obtained. The theory of PCA discussed in Section 3.1, reveals a linear relationship between PC scores and monochromatic radiances:

$$\mathbf{y}(\lambda) \approx \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda) + \ldots + t_J\mathbf{f}_J(\lambda). \tag{10}$$

This approach requires a set of precomputed EOFs which is derived from a training data set of simulated spectra. Hence, for a given set of $J$ EOFs and $J$ spectral points, it is possible to obtain a closed linear system of $J$ equations:

$$\begin{cases} \mathbf{y}(\lambda_1) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_1) + \ldots + t_J\mathbf{f}_J(\lambda_1), \\ \mathbf{y}(\lambda_2) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_2) + \ldots + t_J\mathbf{f}_J(\lambda_2), \\ \vdots \\ \mathbf{y}(\lambda_J) = \bar{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_J) + \ldots + t_J\mathbf{f}_J(\lambda_J). \end{cases} \tag{11}$$

The key point here is that the radiance values in $J$ spectral points $\{\lambda_j\}_{j=1,\ldots,J}$ are represented through the same EOFs. Then, by solving (11) we obtain the PC scores, and by using (6) the full spectrum at $W$ spectral points can be readily restored.

The number of principal components $J$ can either be tuned empirically or chosen by using semi-empirical rules (see e.g., [25]). Essentially, the number $J$ depends on the desired level of variance to be captured by the principal components.

### *3.3. Spectral Sampling*

Regarding the choice of wavelengths, in [11] a method is proposed for selecting the location of monochromatic wavelengths by using a correlation function. This method involves the following steps:

1. the correlation coefficients are computed for the radiance values and then converted to vector angles by an arccosine function;
2. the spectral data are rearranged according to the magnitudes of the correlation coefficients;
3. the monochromatic radiances are selected with equal distances in the space of correlation coefficients.

In [14], alternative techniques such as the equal sampling method and the optimization technique are considered.

To summarize, the output space reduction technique can be used in the following way. For the input data set containing optical parameters of the atmosphere for a set of wavelengths, the monochromatic radiative transfer solver is called for each wavelength. That produces a data set of spectra which is divided into a training data set and a validation data set. By applying PCA and the correlation analysis to the training set, the system of EOFs is computed and an appropriate subset of spectral points is chosen (that is the spectral sampling procedure). These two outputs are stored and used later for computing the PC scores for the validation data set. The spectra at the full spectral resolution are restored using (10) and the error of this reconstruction is estimated. If the error is larger than required, then the number of preserved principal components $J$ is increased. The schematic representation of the output space reduction technique is shown in Figure 2.

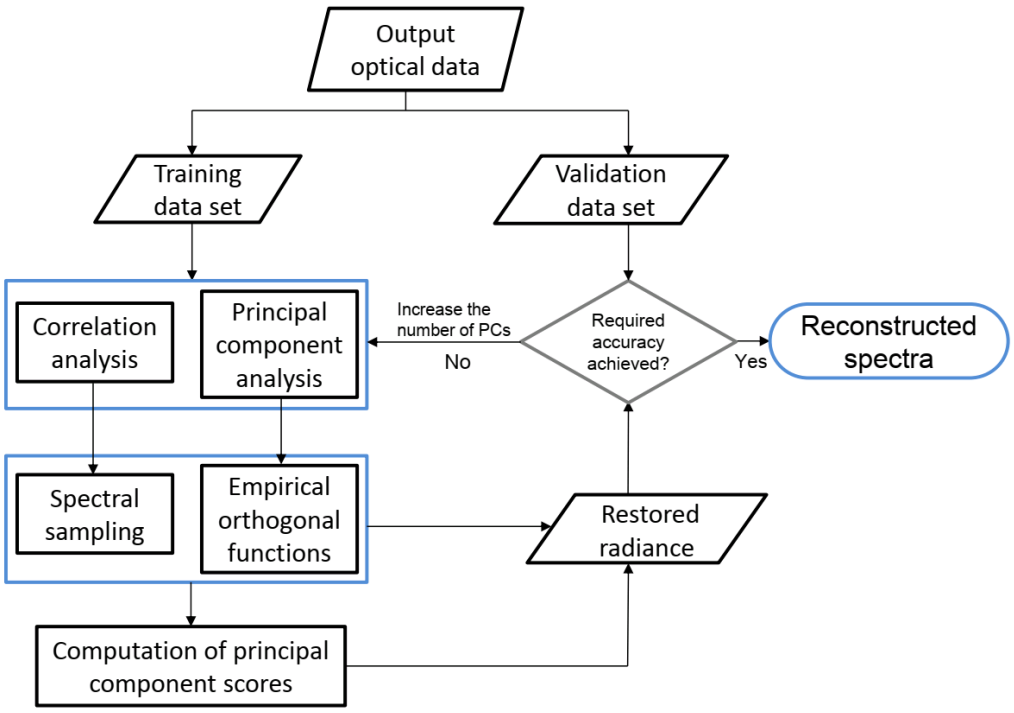


**Figure 2.** Schematic representation of the data processing in the output space reduction technique with precomputed empirical orthogonal functions.

## 4. Results

### *4.1. Dimensionality Reduction of the Optical Parameters in the Hartley-Huggins Band*

In this section, we simulate the backscatter signal of the TROPOMI instrument. The spectra are computed with a spectral resolution of 0.125 nm in the spectral interval between 290 and 335 nm, incorporating the longer part of the Hartley band (200–310 nm) and the shorter part of the Huggins band (320–360 nm). Thus, each spectrum contains 361 spectral points.

The simulations are performed for a clear sky model atmosphere and a Lambertian surface albedo of 0.1. The total optical property inputs are given by Rayleigh scattering [26] by atmospheric molecules and trace gas absorption. Additional cloud properties are neglected in the simulated spectra. To represent cloudy scenes, the effective scene approximation [27] can be used, assuming clouds as Lambertian reflectors. In this case, the cloud top is treated as a reflecting surface characterized by internal surface closure [28]. The Brion–Daumont–Malicet cross sections are used for ozone. Regarding

other trace gas species in the spectral range of interest, $SO_2$ and $NO_2$ absorptions are very weak compared to that of ozone, and thus both species are not included in the model. The atmosphere is discretized into 14 layers. The top of the atmosphere is at 50 km. The solar zenith angle, viewing zenith angle and relative azimuth angle are 45°, 35° and 90°, respectively. The simulations are performed on a personal computer with 16 GB RAM and processor Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz.

Firstly, we compute the radiances of the three RTMs. Figure 3a shows the simulated spectra computed with the multi-stream, TS and SS models. As expected, the TS and SS radiances are inaccurate with maximum errors of about 10% and 30%, respectively. Since the predicted radiances of TS and SS models are inaccurate, three steps are followed in order to improve the accuracy and the computational time: (1) the full spectrum is divided into disjoint regions to apply PCA to each region; (2) a correction function is computed for the different number of PCs and spectral regions; (3) the Taylor expansion up to fourth-order is computed for each spectral region.

Thus, following step (1), we distinguish between three cases depending on the division of the spectral regions:

- Case 1: considering the whole spectral range of 290–335 nm.
- Case 2: considering two intervals of 290–303 nm and 303–335 nm.
- Case 3: considering three intervals: 290–303 nm, 303–321 nm and 321–335 nm.

Figure 3b illustrates the spectra when this method is applied to the whole spectral range of 290–335 nm (Case 1). Although the accuracy of the TS model is significantly improved, the mean relative errors are still high (reaching 52% and 2.8% for the SS and TS models, respectively). Note that, due to the high variability in the optical properties over the whole spectral range, the use of input space reduction leads to worse results for the SS model as compared to the TS model.

To reduce the variability of the optical parameters, we compute the spectral range 290–303 nm just by using the single-scattering model (without applying PCA) since the mean error was below 1%. The remaining spectral interval 303–335 nm is computed as before (Case 2). The results are shown in Figure 3c. The gray region corresponds to the domain in which the dimensionality reduction of optical parameters is applied. In this case, the errors are below 3% for the SS model and 0.8% for the TS model.

Finally, in Case 3, three sub-intervals are considered. The single-scattering model was applied to the spectral range 290–303 nm as in Case 2. The input space reduction technique was applied to the remaining two sub-intervals: 303–321 nm and 321–335 nm. Figure 3d shows good agreement between the multi-stream model and the corrected SS/TS models. Note that the accuracy and the computational time strongly depend on the number of preserved principal components ($M$) and the order of the Taylor expansion, as indicated in Table 1 for Case 3. Here the main conclusion is that the error decreases when $M$ and the expansion order of the Taylor series increase and vice versa.

In all the considered cases, the TS model is more accurate than the SS model. This result is expected since the TS model includes the SS model and estimates the multiple scattering term (approximately, though). The accuracy of the SS and TS models is significantly improved by the correction procedure in (3). This means that the information about the multi-stream solution is contained in the optical data and can be retrieved by using machine learning algorithms, and the input space reduction technique can be considered as representative of them. Here we are confronted with the *ad hoc* learning, i.e., the algorithm extracts the most informative part of the data and predicts the correct solution using the computations in the reduced data space.

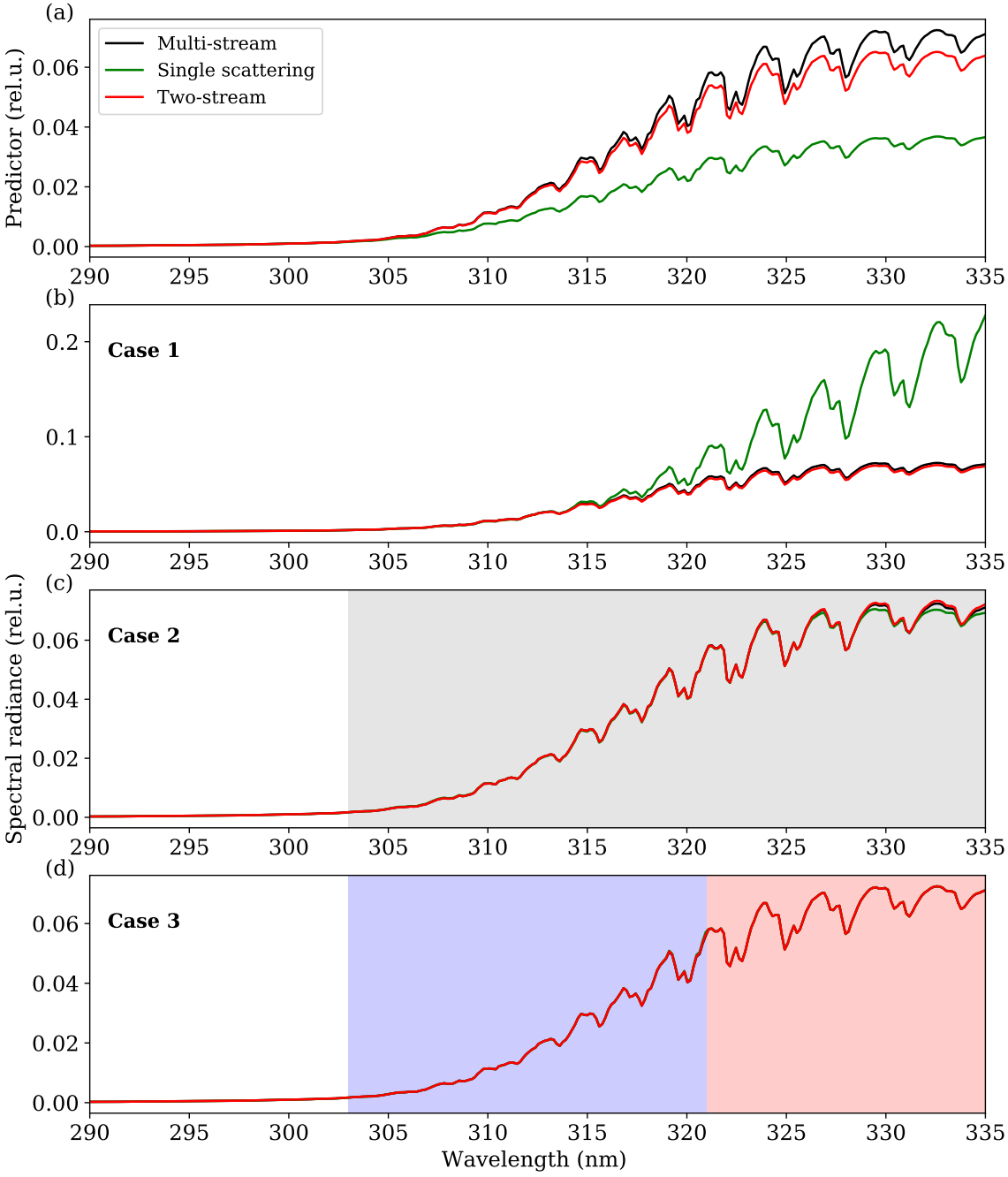


**Figure 3.** (**a**) Spectral radiance computed by the single-scattering (SS), two-stream (TS) and multi-stream models; (**b**) Spectral radiance after applying the correction to the whole spectra (Case 1); (**c**) Spectral radiance after applying the correction within the interval 303–335 nm (Case 2); (**d**) Spectral radiance after applying the correction within the intervals 303–321 nm and 321–335 nm (Case 3). The different spectral regions are separated by shaded areas: the gray area refers to the interval 303–335 nm; the blue area refers to the interval 303–321 nm, and the red area refers to the interval 321–335 nm. For estimating the correction function, $M = 4$ principal components and Taylor expansion up to the second term were used.

The explained variances of the first PC score in the spectral region 303–321 nm and 321–335 nm are 98.32% and 99.76% respectively, while those for the second PC score in the same regions are 1.63% and 0.21%, respectively. The correction functions are highly correlated with the first PC score and depend smoothly on the first PC score, as shown in Figure 4. However, as it follows from Table 1,

the use of just one PC score leads to a loss in accuracy. Therefore, the correction function is highly sensitive to the high order PC scores and the classical methods of estimating the appropriate number of PC scores may fail.

**Table 1.** The mean error in percentage (%) for the single-scattering and two-stream models, in Case 3: for 303–321 nm and 321–335 nm, depending on the number of principal components (PCs) ($M$) and derivatives (expansion order) computed.

| M | Expansion Order | Single-Scattering | | Two-Stream | |
|---|---|---|---|---|---|
| | | 303–321 nm | 321–335 nm | 303–321 nm | 321–335 nm |
| 1 | 1 | 3.02 | 1.19 | 0.67 | 0.088 |
| | 2 | 1.43 | 1.07 | 0.19 | 0.081 |
| | 3 | 1.07 | 1.10 | 0.19 | 0.087 |
| | 4 | 1.11 | 1.06 | 0.18 | 0.068 |
| 2 | 1 | 2.40 | 0.41 | 0.67 | 0.088 |
| | 2 | 1.19 | 0.24 | 0.09 | 0.082 |
| | 3 | 1.09 | 0.25 | 0.10 | 0.087 |
| | 4 | 0.59 | 0.18 | 0.11 | 0.069 |
| 3 | 1 | 2.35 | 0.33 | 0.67 | 0.042 |
| | 2 | 1.03 | 0.15 | 0.09 | 0.036 |
| | 3 | 1.02 | 0.18 | 0.10 | 0.046 |
| | 4 | 0.46 | 0.14 | 0.10 | 0.038 |
| 4 | 1 | 2.34 | 0.29 | 0.67 | 0.034 |
| | 2 | 1.03 | 0.11 | 0.09 | 0.029 |
| | 3 | 1.02 | 0.13 | 0.10 | 0.039 |
| | 4 | 0.46 | 0.11 | 0.11 | 0.034 |

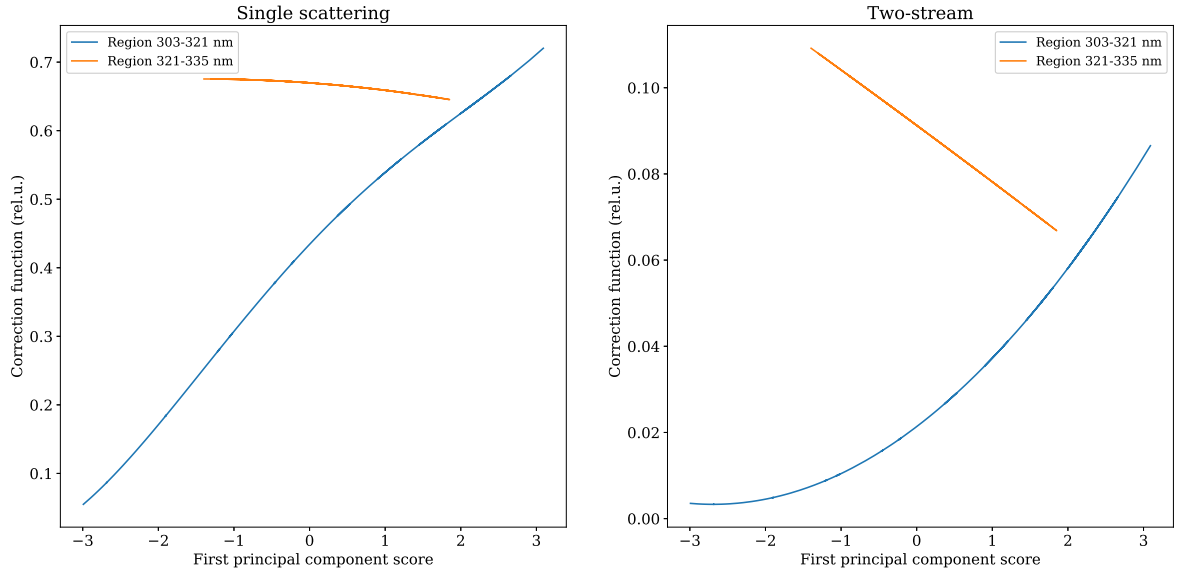


**Figure 4.** Corrector vs. the first principal component score for (**left**) the SS model and (**right**) the TS model. The plots correspond to Case 3 for the spectral regions 303–321 nm and 321–335 nm.

The computational times are listed in Table 2 for Case 3 since it provides the most accurate spectral radiances among the three cases. Two conclusions follow from the table:

1. the increase in the number of PC scores results in the increase of the computational time. However, this increase is not significant (0.05 s per PC score, i.e., ≈1% from the total computational time). Therefore, it is recommended to choose $M \geq 3$.

2. the TS model is 40 times slower than the SS model for simulating the approximate spectra. However, the overall computational times differ by a factor of 2. Given that the TS model is more accurate than the SS model (as it has been shown in Table 1), it is recommended to use the TS model.

Note that the computational time for the multi-stream model called for each spectral point is ∼6.8 s. Thus, the input space reduction technique provides the performance enhancement of about 10 times depending on the model used, the number of PC scores, $M$, and the expansion order of the Taylor series.

**Table 2.** Computational time in seconds for computing the approximate spectrum ($L_A$), corrected radiances ($L$), and the total time (Total) for each model in Case 3. In addition, the number of calls to multiple scattering (MS) radiative transfer model (RTM) is indicated. The computations depend on the number of PCAs ($M$) and the order of Taylor expansion.

| $M$ | Order | Single-Scattering | | | | Two-Stream | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $L_A$ | $L$ | Total | Number of MS Calls | $L_A$ | $L$ | Total | Number of MS Calls |
| 1 | 1 | 0.0043 | 0.227 | 0.231 | 3 | 0.175 | 0.268 | 0.434 | 6 |
| | 2 | 0.0043 | 0.251 | 0.255 | 3 | 0.175 | 0.297 | 0.477 | 6 |
| | 3 | 0.0043 | 0.416 | 0.420 | 5 | 0.175 | 0.494 | 0.674 | 10 |
| | 4 | 0.0043 | 0.418 | 0.422 | 5 | 0.175 | 0.493 | 0.673 | 10 |
| 2 | 1 | 0.0043 | 0.276 | 0.280 | 5 | 0.175 | 0.320 | 0.485 | 10 |
| | 2 | 0.0043 | 0.306 | 0.310 | 5 | 0.175 | 0.354 | 0.533 | 10 |
| | 3 | 0.0043 | 0.526 | 0.531 | 9 | 0.175 | 0.605 | 0.784 | 18 |
| | 4 | 0.0043 | 0.514 | 0.519 | 9 | 0.175 | 0.595 | 0.775 | 18 |
| 3 | 1 | 0.0043 | 0.325 | 0.329 | 7 | 0.175 | 0.370 | 0.536 | 14 |
| | 2 | 0.0043 | 0.350 | 0.354 | 7 | 0.175 | 0.288 | 0.577 | 14 |
| | 3 | 0.0043 | 0.619 | 0.623 | 13 | 0.175 | 0.700 | 0.877 | 26 |
| | 4 | 0.0043 | 0.627 | 0.632 | 13 | 0.175 | 0.711 | 0.891 | 26 |
| 4 | 1 | 0.0043 | 0.363 | 0.367 | 9 | 0.175 | 0.409 | 0.572 | 18 |
| | 2 | 0.0043 | 0.401 | 0.406 | 9 | 0.175 | 0.452 | 0.629 | 18 |
| | 3 | 0.0043 | 0.711 | 0.715 | 17 | 0.175 | 0.794 | 0.970 | 34 |
| | 4 | 0.0043 | 0.726 | 0.730 | 17 | 0.175 | 0.813 | 0.991 | 34 |

### *4.2. Principal Component Analysis of the Data Set of Spectral Radiances*

In this section, we examine the efficiency of the output space reduction technique. For that purpose, we generate a data set consisting of $2 \times 10^5$ spectra. The following parameters are varied for the generation of radiance spectra: the solar zenith angle, the viewing zenith angle, the relative azimuth angle, the surface albedo, the ozone total column, the surface height, and the temperature. The smart sampling technique [29] based on Halton sequences [30] is employed to optimally cover the multi-dimensional input space and to concurrently minimize the number of samples generated to represent the output space.

The data set of spectra is divided into a training data set and a validation data set (see Figure 2). By applying PCA to the training data set, the system of EOFs is computed (see Figure 5 left). The right plot in Figure 5 shows that almost 99.9% of the variance in the output data can be explained just with $J = 2$ principal components.

Figure 6 illustrates the spectral sampling procedure [11]. The left panel in Figure 6 shows an example of the arccosine function of the correlation coefficients as a function of the wavelength, while the right panel in Figure 6 shows the resulting plot with rearranging the arccosine function of the correlation coefficients. By using this technique, $J = 30$ wavelengths are identified.

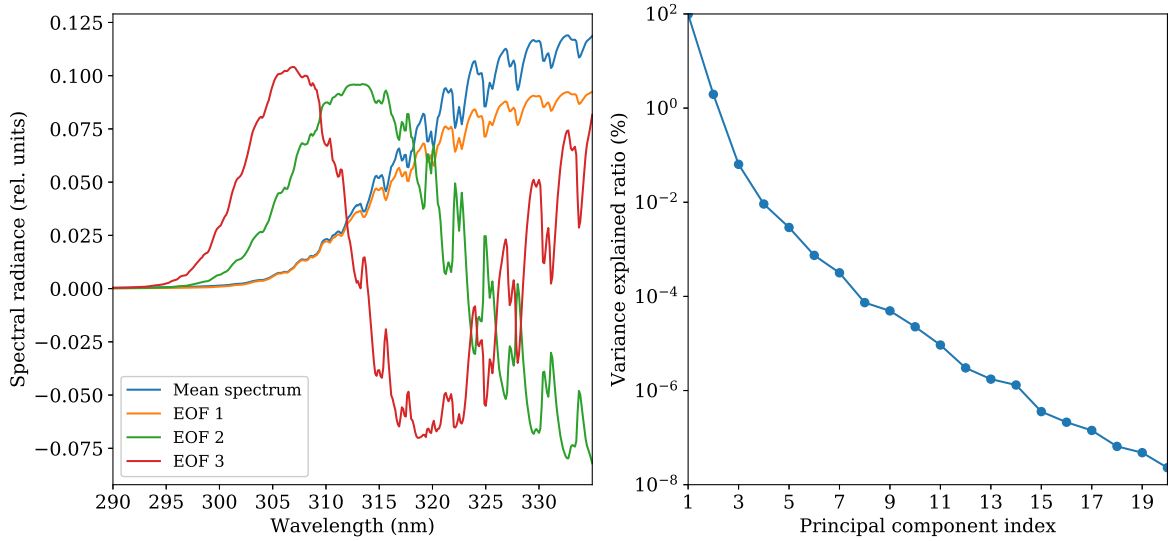


**Figure 5.** (**Left panel**): First three orthogonal functions computed in the Hartley-Huggins band. (**Right panel**): explained variance in percentage as a function of the principal component index for the Hartley-Huggins band simulation.

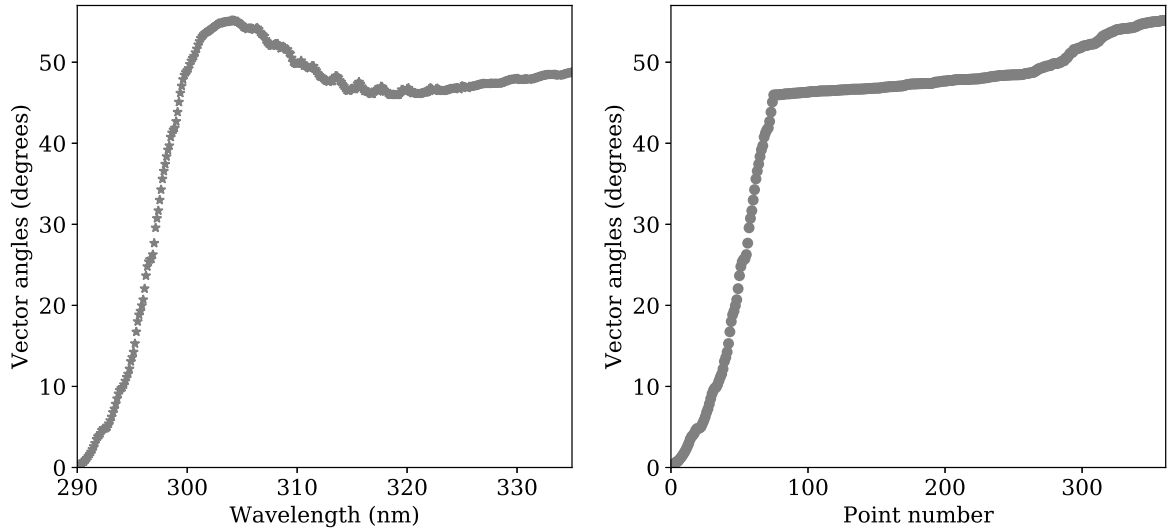


**Figure 6.** (**Left panel**): Arccosine of correlation coefficients as a function of wavelength. (**Right panel**): Rearranged arccosine of correlation coefficients according to their magnitudes.

Next, the EOFs and $J$ wavelengths are saved in the memory and reused for computing PC scores for the validation data set. The spectra at a full resolution are restored using Equation (10). Figure 7 shows one example of the restored spectrum in the Hartley-Huggins band. The histogram of the relative error of the restored spectra from the validation data set is shown in Figure 8. The mean relative error of the spectra is 0.00023% with a standard deviation of 0.12%. Using $J = 30$ spectral points provides a performance enhancement of about 12 times. A higher amount of principal components ($J > 30$) assures a higher accuracy and a more robust result.

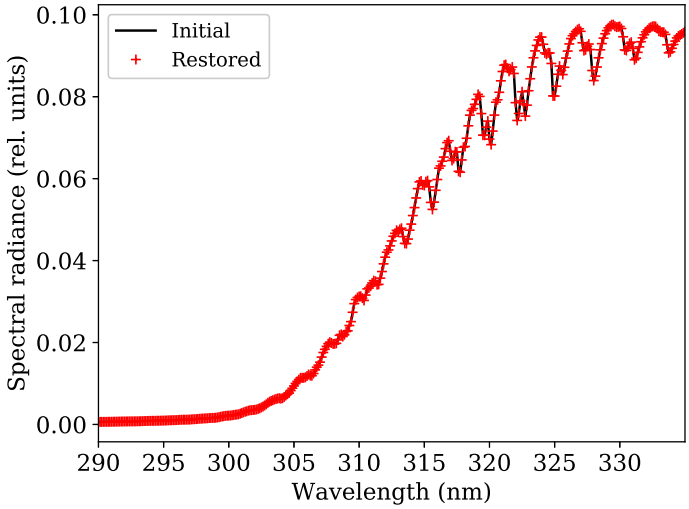


**Figure 7.** Example of one initial spectrum in the Hartley-Huggins band and the restored spectra with the empirical orthogonal function method.

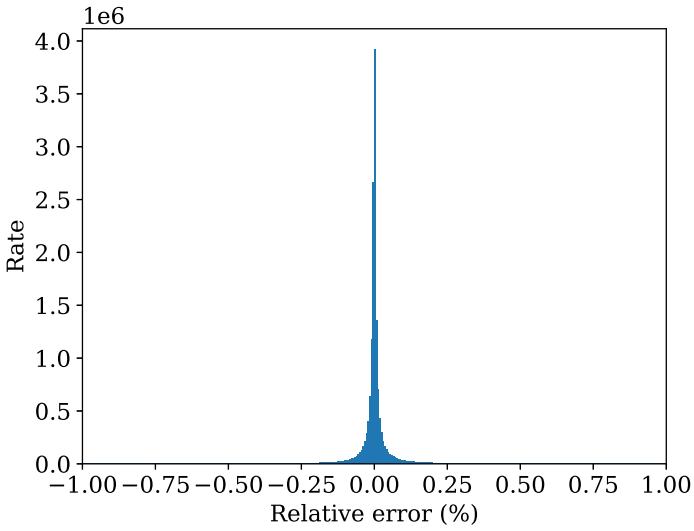


**Figure 8.** A histogram of the relative error of the restored spectra from the validation data set.

### *4.3. Combined Use of Input and Output Space Reduction Techniques*

In this section, we examine the combined use of the input and output space reduction techniques. The input space reduction technique is used with the parameters corresponding to the case in Table 2 marked with a gray color, namely:

- the number of preserved principal components is $M = 3$,
- the correction function is expanded in Taylor series up to the second order,
- the two-stream model is used for computing the approximate solution; and for that, the two-stream model is called for each spectral point (i.e., $W = 361$ times).

In this case, the computations of the TS solutions require ∼30% of the total computational time.

The output space reduction technique is applied to accelerate the computations of the approximate solution. For that, the data set of the spectra computed by using the two-stream solution has been generated and analysis similar to that from Section 4.2 is conducted. To ensure a robust result, $J = 30$ spectral points are chosen. Table 3 summarizes the acceleration factors and mean errors for both methods when they are applied separately and jointly.

The input space reduction technique provides a speedup factor of about 13. The output space reduction technique accelerates the two-stream computations by 12 times, practically making this part negligible in terms of overall computational time. The resulting acceleration factor is about 18. Note that the combined use of two techniques does not compromise the robustness of the scheme keeping the error below 0.05%.

**Table 3.** Acceleration factor and mean error for the input space reduction method, the output space reduction method and a combination of both methods for the best configuration (i.e., three PC scores and the second order expansion).

| | Input Space Reduction | Output Space Reduction | Combined Use |
|---|---|---|---|
| Acceleration factor | 13 | 12 | 18.2 |
| Mean error | 0.05 | 0.00023 | 0.05 |

Hence, the results presented in Table 3 exhibit that the combined used of both methods results in an efficient approach to accelerate the input and output data without losing accuracy.

## 5. Summary

In this work, we have examined in detail the efficiency of two acceleration techniques based on the principal component analysis (PCA) for simulating the spectral radiances in the Hartley-Huggins band. The first technique is based on the input space reduction, in which the correction function for approximate radiative transfer models is applied. The second technique deals with the analysis of the precomputed data sets of spectra. By estimating the set of empirical orthogonal functions and most relevant (representative) spectral points, it is possible to reproduce a spectrum at full spectral resolution by performing radiative transfer simulations only for a subset of wavelengths.

The first technique does not require a precomputed data set and therefore can be regarded as a sort of RTM with *ad hoc* learning. The second technique can be applied to the real measurements, and therefore, the instrumental features like detector degradation can be easily taken into account. It has been shown that both techniques achieve an order of magnitude speed improvement.

An innovative method which combines these PCA-based techniques has been introduced to increase the computational efficiency of radiative transfer models. The output space reduction technique is used to speed up the two-stream computations in the framework of the input space reduction technique, providing an overall speedup factor of about 20. The combined method is just as accurate as the input space reduction technique, applied separately.

In our future works, we plan to examine the efficiency of the proposed approach for other spectral bands and to investigate the overall performance enhancement of the ozone retrieval algorithms.

**Author Contributions:** Data curation, J.X.; Investigation and simulations, A.d.Á.; Methodology, D.S.E.; Validation, V.M.G.; Formal analysis, A.d.Á. and D.S.E.; Writing—original draft preparation, A.d.Á. and D.S.E.; Writing—review and editing, V.M.G.; Visualization, A.d.Á.; Supervision, D.S.E.

**Funding:** This work was funded by the DLR/DAAD Research Fellowships 2018 and 2015 (grants no. 57424731 and 57186656 respectively) organized by the German Academic Exchange Service (DAAD) and the German Aerospace Center (DLR).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| DOME | Discrete Ordinate Method with Matrix Exponential |
| EOF | Empirical Orthogonal Function |
| PCA | Principal Component Analysis |

| | |
|---|---|
| PC | Principal Component |
| RTM | Radiative Transfer Model |
| SS | Single Scattering |
| TROPOMI | TROPospheric Ozone Monitoring Instrument |
| TS | Two-Stream |

## References


1. Natraj, V. A review of fast radiative transfer techniques. In *Light Scattering Reviews 8*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 475–504. doi:10.1007/978-3-642-32106-1_10.
2. Goody, R.; West, R.; Chen, L.; Crisp, D. The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **1989**, *42*, 539–550. doi:10.1016/0022-4073(89)90044-7.
3. Wiscombe, W. The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *J. Atmos. Sci.* **1977**, *34*, 1408–1422. doi:10.1175/1520-0469(1977)034<1408:TDMRYA>2.0.CO;2.
4. Tjemkes, S.A.; Schmetz, J. Synthetic satellite radiances using the radiance sampling method. *J. Geophys. Res. Atmos.* **1997**, *102*, 1807–1818. doi:10.1029/96jd02684.
5. West, R.; Crisp, D.; Chen, L. Mapping transformations for broadband atmospheric radiation calculations. *J. Quant. Spectrosc. Radiat. Transf.* **1990**, *43*, 191–199. doi:10.1016/0022-4073(90)90051-7.
6. Natraj, V.; Jiang, X.; Shia, R.; Huang, X.; Margolis, J.; Yung, Y. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the O2A-band. *J. Quantit. Spectrosc. Radiat. Transf.* **2005**, *95*, 539–556. doi:10.1016/j.jqsrt.2004.12.024.
7. Efremenko, D.S.; Loyola, D.G.; Doicu, A.; Spurr, R.J. Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method. *Comput. Phys. Commun.* **2014**, *185*, 3079–3089. doi:10.1016/j.cpc.2014.07.018.
8. Natraj, V.; Shia, R.; Yung, Y. On the use of principal component analysis to speed up radiative transfer calculations. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 810–816. doi:10.1016/j.jqsrt.2009.11.004.
9. Somkuti, P.; Boesch, H.; Natraj, V.; Kopparla, P. Application of a PCA-Based Fast Radiative Transfer Model to XCO2 Retrievals in the Shortwave Infrared. *J. Geophys. Res. Atmos.* **2017**, *122*, 10477–10496. doi:10.1002/2017JD027013.
10. Kopparla, P.; Natraj, V.; Spurr, R.; Shia, R.; Crisp, D.; Yung, Y. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *173*, 65–71. doi:10.1016/j.jqsrt.2016.01.014.
11. Liu, X.; Smith, W.; Zhou, D.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Appl. Opt.* **2006**, *45*, 201–208. doi:10.1364/AO.45.000201.
12. Matricardi, M. A principal component based version of the RTTOV fast radiative transfer model. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 1823–1835. doi:10.1002/qj.680.
13. Hurley, P.D.; Oliver, S.; Farrah, D.; Wang, L.; Efstathiou, A. Principal component analysis and radiative transfer modelling of Spitzer Infrared Spectrograph spectra of ultraluminous infrared galaxies. *Mon. Not. R. Astronom. Soc.* **2012**, *424*, 2069–2078. doi:10.1111/j.1365-2966.2012.21352.x.
14. Hollstein, A.; Lindstrot, R. Fast reconstruction of hyperspectral radiative transfer simulations by using small spectral subsets: Application to the oxygen A band. *Atmos. Meas. Tech.* **2014**, *7*, 599–607. doi:10.5194/amt-7-599-2014.
15. Efremenko, D.S.; Loyola, D.G.; Hedelt, P.; Spurr, R.J.D. Volcanic $SO_2$ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm. *Int. J. Remote Sens.* **2017**, *38*, 1–27. doi:10.1080/01431161.2017.1348644.
16. Xu, J.; Schussler, O.; Rodriguez, D.L.; Romahn, F.; Doicu, A. A Novel Ozone Profile Shape Retrieval Using Full-Physics Inverse Learning Machine (FP-ILM). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5442–5457. doi:10.1109/jstars.2017.2740168.

17. Roozendael, V.; Spurr, R.; Loyola, D.; Lerot, C.; Balis, D.; Lambert, J.; Zimmer, W.; Gent, J.; Van Geffen, J.; Koukouli, M.; et al. Sixteen years of GOME/ERS2 total ozone data: the new direct-fitting GOME Data Processor (GDP) Version 5: I. algorithm description. *J. Geophys. Res. Atmos.* **2012**, *117*, D03305. doi:10.1029/2011JD016471.
18. Stamnes, K.; Tsay, S.; Wiscombe, W.; Jayaweera, K. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Appl. Opt.* **1988**, *12*, 2502–2509. doi:10.1364/AO.27.002502.
19. Doicu, A.; Trautmann, T. Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. *J. Quant. Spectrosc. Radiat. Transf.* **2009**, *110*, 146–158. doi:10.1016/j.jqsrt.2008.09.014.
20. Efremenko, D.S.; García, V.M.; García, S.G.; Doicu, A. A review of the matrix-exponential formalism in radiative transfer. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *196*, 17–45. doi:10.1016/j.jqsrt.2017.02.015.
21. Molina García, V.; Sasi, S.; Efremenko, D.; Doicu, A.; Loyola, D. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *213*, 228–240. doi:10.1016/j.jqsrt.2018.03.014.
22. Nakajima, T.; Tanaka, M. Algorithms for radiative intensity calculations in moderately thick atmos using a truncation approximation. *J. Quant. Spectrosc. Radiat. Transf.* **1988**, *40*, 51–69. doi:10.1016/0022-4073(88)90031-3.
23. Kerschen, G.; Golinval, J. Non-linear generalization of principal component analysis: From a global to a local approach. *J. Sound Vib.* **2002**, *254*, 867–876. doi:10.1006/jsvi.2001.4129.
24. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinburgh Dublin Philos. Mag J. Sci.* **1901**, *2*, 559–572. doi:10.1080/14786440109462720.
25. MacArthur, R.H. On the relative abundance of bird species. *Proc. Natl. Acad. Sci. USA* **1957**, *43*, 293–295.
26. Bodhaine, B.A.; Wood, N.B.; Dutton, E.G.; Slusser, J.R. On rayleigh optical depth calculations. *J. Atmos. Ocean. Technol.* **1999**, *16*, 1854–1861. doi:10.1175/1520-0426(1999)016<1854:ORODC>2.0.CO;2.
27. Coldewey-Egbers, M.; Weber, M.; Lamsal, L.N.; de Beek, R.; Buchwitz, M.; Burrows, J.P. Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach. *Atmos. Chem. Phys.* **2005**, *5*, 1015–1025. doi:10.5194/acp-5-1015-2005.
28. Lerot, C.; Van Roozendael, M.; Lambert, J.C.; Granville, J.; van Gent, J.; Loyola, D.; Spurr, R. The GODFIT algorithm: A direct fitting approach to improve the accuracy of total ozone measurements from GOME. *Int. J. Remote Sens.* **2010**, *31*, 543–550. doi:10.1080/01431160902893576.
29. Loyola, D.G.; Pedergnana, M.; Gimeno García, S. Smart sampling and incremental function learning for very large high dimensional data. *Neural Netw.* **2016**, *78*, 75–87. doi:10.1016/j.neunet.2015.09.001.
30. Halton, J.H. Algorithm 247: Radical-inverse quasi-random point sequence. *Commun. ACM* **1964**, *7*, 701–702. doi:10.1145/355588.365104.


## A.2 Publication II: A Review of Dimensionality Reduction Techniques for Processing Hyper-Spectral Optical Signal

### Reference

A. del Águila, D. S. Efremenko and T. Trautmann. A Review of Dimensionality Reduction Techniques for Processing Hyper-Spectral Optical Signal. Light and Engineering, 27, 3, 85–98, 2019. doi: 10.33383/2019-017

### Copyright



### Abstract

Hyper-spectral sensors take measurements in the narrow contiguous bands across the electromagnetic spectrum. Usually, the goal is to detect a certain object or a component of the medium with unique spectral signatures. In particular, the hyper-spectral measurements are used in atmospheric remote sensing to detect trace gases. To improve the efficiency of hyper-spectral processing algorithms, data reduction methods are applied. This paper outlines the dimensionality reduction techniques in the context of hyper-spectral remote sensing of the atmosphere. The dimensionality reduction excludes redundant information from the data and currently is the integral part of high-performance radiation transfer models. In this survey, it is shown how the principal component analysis can be applied for spectral radiance modelling and retrieval of atmospheric constituents, thereby speeding up the data processing by orders of magnitude. The discussed techniques are generic and can be readily applied for solving atmospheric as well as material science problems.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko; T.T.: Thomas Trautmann)
The conceptualization of the study was proposed by D.S.E, A.d.Á and T.T. The design and realization of the study was developed by D.S.E and A.d.Á. The analysis, investigation and discussion was done with the help of D.S.E, A.d.Á and T.T. The visualization was mostly done by D.S.E and A.d.Á. helped with the figures. Regarding the manuscript writing, was done by A.d.Á and D.S.E. with corrections of T.T.

The overall own contribution of A.d.Á for the publication in Appendix A.2. is estimated at 53 %, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.2).

**Table A.2:** Criteria and estimated own contribution for Appendix A.2

| Criteria | Estimated contribution |
|---|---|
| Conceptualization | 45% |
| Implementation | 60% |
| Analysis and discussion | 50% |
| Figures and tables | 60% |
| Manuscript writing | 50% |
| **Total** | **53** % |

*Light & Engineering*
*Vol. 27, No. 3, pp. 85–98, 2019* *https://doi.org/10.33383/2019 - 017*

# A REVIEW OF DIMENSIONALITY REDUCTION TECHNIQUES FOR PROCESSING HYPER-SPECTRAL OPTICAL SIGNAL

Ana del Águila, Dmitry S. Efremenko, and Thomas Trautmann

*Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung (IMF), 82234 Oberpfaffenhofen, Germany*
*E-mail: dmitry.efremenko@dlr.de*

**ABSTRACT**

Hyper-spectral sensors take measurements in the narrow contiguous bands across the electromagnetic spectrum. Usually, the goal is to detect a certain object or a component of the medium with unique spectral signatures. In particular, the hyper-spectral measurements are used in atmospheric remote sensing to detect trace gases. To improve the efficiency of hyper-spectral processing algorithms, data reduction methods are applied. This paper outlines the dimensionality reduction techniques in the context of hyper-spectral remote sensing of the atmosphere. The dimensionality reduction excludes redundant information from the data and currently is the integral part of high-performance radiation transfer models. In this survey, it is shown how the principal component analysis can be applied for spectral radiance modelling and retrieval of atmospheric constituents, thereby speeding up the data processing by orders of magnitude. The discussed techniques are generic and can be readily applied for solving atmospheric as well as material science problems.

**Keywords:** passive remote sensing, hyper-spectral data, principal component analysis, full-physics machine learning, trace gas retrieval

## 1. INTRODUCTION

Hyper-spectral sensors record the transmitted or reflected radiance in the narrow contiguous bands across the electromagnetic spectrum. The goal is to detect a certain object or a component of the medium, which has a unique spectral signature, i.e. a fingerprint. Hyper-spectral imaging has emerged as one of the most powerful technologies in various fields including astronomy, mineralogy, agriculture, medicine and chemistry. For instance, hyper-spectral data (sometimes referred to as hypercube data or as an image cube) are used in astronomy and Earth remote sensing to create a spatially-resolved spectral image allowing more accurate recognition and classification of the objects in the instrument field of view. At the same time, significant data storage and computational power are required to process the hyper-spectral information and to retrieve a certain parameter of the scattering medium.

In this survey we are focused on hyper-spectral remote sensing of the atmosphere. The passive atmospheric composition sensors (ACS) detect and record the radiance reflected by the Earth atmosphere in the ultraviolet (UV), visible (VIS), and thermal infrared (IR) regions. The information about the atmosphere is then retrieved from the spectral data by using the so called atmospheric processors, i.e. codes which are specifically designed to invert ACS measurements [1]. Extracting the information about geophysical parameters (level-2 data) from spectral radiance distributions (level-1 data) turns out to be a major computational challenge and requires high performance computing (HPC) [2].

The recent developments in optics, sensor design and measurement techniques significantly improve the characteristics of hyper-spectral ACS, such as the spatial resolution and the signal-to-noise

**Table. Characteristics of Atmospheric Composition Sensors**

| **Instrument** | GOME | GOME-2 | TROPOMI |
|---|---|---|---|
| **Platform** | ERS-2 | MetOp (A, B, C) | Sentinel 5 Precursor |
| **Spatial resolution (km$^2$)** | 320x40 | 80x40 | 7x3.5 |
| **Amount of level-1 data (TB per year)** | 0.8 | 4.2 | 240 |
| **Operational** | 1995–2011 | 2006-present | 2017-present |

ratio. Table shows a comparison between previous generation instruments like the Global Ozone Monitoring Experiment (GOME and GOME-2) and the newest, the TROPOspheric Monitoring Instrument (TROPOMI) [3] on board of the Copernicus Sentinel 5 Precursor (S5P) satellite. The spatial resolution of TROPOMI is two orders of magnitude higher, providing 21 million level-1B spectra per day, i.e. almost 8 milliards spectral points, while the signal-to-noise ratio in the UV/VIS channels reaches the values of about 1500. Fig. 1 shows the example of a retrieved map of tropospheric nitrogen dioxide ($NO_2$) from S5P measurements (the data is freely available at https://s5phub.copernicus.eu/dhus). Observe that air pollution emitted by big cities and shipping lanes is clearly visible. With the high resolution data it is possible to detect air pollution over individual cities as well as to locate where pollutants are being emitted, and so, identifying pollution hotspots. Such high resolution satellite remote sensing observations are extremely useful for diagnosing the impact of atmospheric constituents on a global scale, in particular, allowing detection of small-scale sources, and increasing the fraction of cloud-free observations. However, the high spatial resolution of the state-of-the-art ACS results in very challenging data volumes to be processed – 240 TB (terabyte) per year of level-1 data.

In fact, the amount of satellite data increases faster than the computational power [4]. The remote sensing data is recognized as Big Data [5] since it satisfies Doug Laney's 3V criterion: significant growth in the volume, velocity and variety. New efficient techniques have to be developed for next generation atmospheric processors to cope with these high efficiency requirements.

The radiative transfer modelling (RTM) is the key component and the major performance bottle-neck in the atmospheric processors. Furthermore, the hyper-spectral RTMs involve a hierarchy of nested computational loops [6] as shown in the pseudo-code in Fig. 2. Recent surveys such as those provided by V. Natraj [7] and D. Efremenko et al. [8, 9] showed that a significant performance enhancement can be achieved by optimizing the framework in which the radiative transfer solver is called rather than accelerating the RTM solver itself. In fact, the efficiency of monochromatic radiative transfer solvers hardly can be further improved [10]. Several attempts have been made to optimize loops over ground pixels and geometry (see, e.g., [11, 12, 13, 14] and references therein). Currently the loop over wavelengths (which expresses the hyper-spectral processing) remains the most computationally demanding part.

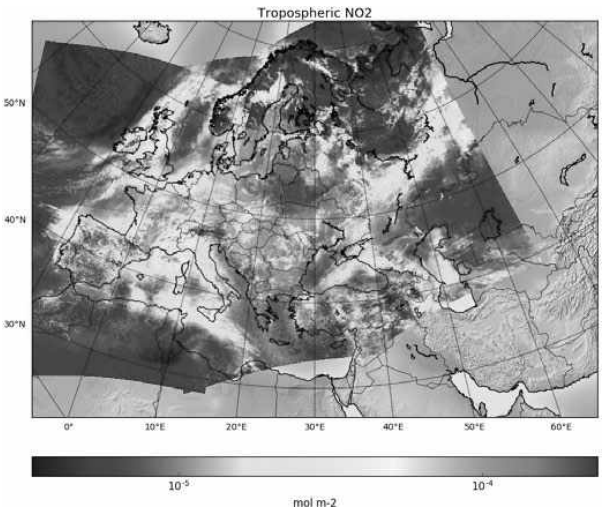


Fig. 1. Example of Sentinel-5P tropospheric nitrogen dioxide ($NO_2$) measurements on 1.04.2019 (data is freely available at https://s5phub.copernicus.eu/dhus)

The essential part of the Big Data analysis is the dimensionality reduc-

```
for each ground pixel:
   for each wavelength: #<--loop for hyper-spectral processing
      for each geometry:
         call Radiative_transfer_solver;
```

Fig. 2. Hierarchy of computational loops, in which the radiative transfer solver is called

tion procedures which are related in the context of RTMs to the loop over wavelengths. The principal component analysis (PCA) is one of the famous representatives of them. PCA was proposed in 1901 by K. Pearson [15] and today has become an integral part of hyper-spectral RTMs.

In this regard, many efforts have been made to develop hyper-spectral RTMs which explicitly take into account the interdependency and statistical relations between level-1 and level-2 data [16]. The motivation of this survey is to present basic concepts of dimensionality reduction for design of atmospheric processors in a systematic way and to put in one context recent developments in this field.

## 2. DIMENSIONALITY REDUCTION

### 2.1. Heritage from the k-Distribution Technique

The techniques of dimensionality reduction in atmospheric science and astrophysics were preceded by methods which exploit a strong interdependency in the hyper-spectral data, although do not use PCA explicitly. Ambartsumian [17] noted that the transmission within a spectral interval does not dependent on the line-by-line (LBL) variation of the absorption coefficient $k$ with respect to wavelength $\lambda$, but rather on the distribution of absorption coefficient within the spectral interval. Such concept is called the *k*-distribution approach. It is based on the cumulative frequency distribution of $k$, namely, $G(k)$; as a matter of fact, the inverse distribution $k(G)$ is the *k*-distribution function. Since $G(k)$ is a smooth function and $k(\lambda)$ has a strong variation with respect to $\lambda$, it is beneficial from the numerical point of view to replace the integration in $\lambda$-space by that in $G$-space (less number of quadrature points for numerical integration is required). This method was extended by Goody et al. [18] to the cases of inhomogeneous atmosphere assuming that there is a correlation between *k*-distributions at different pressure levels, and that is the correlated-$k$ distribution method. Other techniques (e.g. exponential sum fitting [19], spectral mapping [20], *k*-binning approach [21], opacity sampling method [22], multi-dimensional *k*-distribution method [23], and fast *k*-distribution models [24, 25]) use similar ideas although based on a more elaborative mathematical basis. In [26], a modification of the k-distribution technique was considered, in which the integration is performed in the original $\lambda$ -space. In [27], an empirical procedure based on uniform spectral grids was proposed for choosing the most representative spectral points in LBL computations. Together, the studies related to the *k*-distribution function outline that the hyper-spectral radiances can be modelled by much fewer number of monochromatic computations than required by the LBL-framework, thereby providing a basis to dimensionality reduction of the problem.

### 2.2. Basic Concept of Dimensionality Reduction

In simple words, dimensionality reduction means representing the initial data set with less number of parameters than it is initially represented. It can be considered as one of the lossy data compression paradigms [28]. Dimensionality reduction is crucial for stable and high-performance processing of spectral measurements. It excludes redundant information from the initial dataset, reduces the number of independent parameters and improves the efficiency of machine learning.

There is a distinction between linear and non-linear techniques for dimensionality reduction. A more detailed review can be found in [29, 30, 31, 32] and references therein. Linear and non-linear techniques have been inter-compared in [33]. Results of these numerical experiments reveal that non-linear techniques perform well on selected artificial tasks. However, they hardly outperform PCA on real-world tasks. Similar conclusions were reported in [34], where several methods for dimensionality reduction were inter-compared in the context of accelerating radiative transfer performance. Bearing in mind that no obviously superior method has emerged in the benchmarking studies (increasingly time-consuming and sophisticated dimensionality

reduction techniques lead to more accurate results, and vice versa) our analysis will be limited with the classical PCA.

### 2.3. Principal Component Analysis

Although the dimensionality reduction techniques are well-known and covered by many statistical libraries (e.g., scikit-learn [35] for Python), we make a short mathematical exposition to put the above considerations in a proper context. For clarity, we specify sizes of matrices using the notation $\in \mathbb{R}^{\text{rows}\times\text{columns}}$. Let $\mathbf{y} = \left(y(\lambda_1), y(\lambda_2), ..., y(\lambda_W)\right)$, $\mathbf{y} \in R^{1\times W}$, be a row-vector of atmospheric radiances at $W$ wavelengths $\{\lambda_w\}_{w=1,...,W}$. A set of $S$ spectra are assembled into a matrix $\mathbf{Y} \in R^{S\times W}$ whose $i$-th row is $\mathbf{y}_i$. Then, $\mathbf{y}_i$ can be represented in a new basis system as follows:

$$\mathbf{y}_i = \overline{\mathbf{y}} + \sum_{k=1}^{W} t_{ik}\mathbf{f}_k.$$

Here, $\overline{\mathbf{y}} = \sum_{i=1}^{S} \mathbf{y}_i / S$, $\overline{\mathbf{y}} \in R^{1\times W}$ is the sample mean of the spectra (the average spectrum), $t_{ik}$ is the $k^{\text{th}}$ coordinate of the vector $\mathbf{y}_i$ in the new basis system and $\mathbf{f}_k = \left(f_k(\lambda_1), f_k(\lambda_2), ..., f_k(\lambda_W)\right) \in R^{1\times W}$ is the $k^{\text{th}}$ basis vector. Noting that high-dimensional real data are often situated on or near a lower-dimensional manifold, the spectrum $\mathbf{y}_i$ can be projected onto the $K$-dimensional subspace ($K < W$) as follows:

$$\mathbf{y}_i \approx \overline{\mathbf{y}} + \sum_{k=1}^{K} t_{ik}\mathbf{f}_k, \tag{1}$$

or in matrix form for the initial dataset:

$$\mathbf{Y} \approx \overline{\mathbf{Y}} + \mathbf{TF},$$

where $\overline{\mathbf{Y}} = \left\{\overline{\mathbf{y}}, \ldots, \overline{\mathbf{y}}\right\} \in R^{S\times W}$,

$\mathbf{F} = \left\{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_K\right\}^{\mathrm{T}} \in R^{K\times W}$, $\mathbf{T} \in R^{S\times K}$ is the matrix whose entries are $\{t_{ik}\}_{i=1,...,S}^{k=1,...,K}$. Hereinafter the superscript $^{\mathrm{T}}$ stands for ''transpose''. The transformation (1) can be done using dimensionality reduction techniques, such as PCA [15]. In the latter, basic vectors $\mathbf{f}_k$ in (1) are referred to as ''principal components'' (PCs) or empirical orthogonal functions (EOFs) and are taken as $K$ eigenvectors related to the $K$ most significant eigenvalues of the covariance matrix $\text{cov}(\mathbf{Y}, \mathbf{Y}) \in R^{W\times W}$. The coordinates $t_{ik}$ in the new coordinates system and the corresponding matrix $\mathbf{T}$ are called ''principal component scores''.

## 3. PRINCIPAL COMPONENT-BASED RADIATIVE TRANSFER MODEL FOR HYPER-SPECTRAL SIGNALS

The most conceptually simple approach uses the training data set of spectra in order to establish a set of EOFs by using PCA and then to restore hyper-spectral signal in $W$ spectral points by using $K$ monochromatic radiances. Naturally, we have $K<W$.

The theory of PCA briefly discussed in the previous section reveals a linear relationship between PC scores and monochromatic radiances:

$$\mathbf{y}(\lambda) = \overline{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda) + ... + t_K\mathbf{f}_K(\lambda). \tag{2}$$

Hence, for a given set of $K$ EOFs and $K$ spectral points it is possible to obtain a closed linear system of $K$ equations:

$$\begin{cases} \mathbf{y}(\lambda_1) = \overline{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_1) + ... + t_K\mathbf{f}_K(\lambda_1), \\ \mathbf{y}(\lambda_2) = \overline{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_2) + ... + t_K\mathbf{f}_K(\lambda_2), \\ ... \\ \mathbf{y}(\lambda_K) = \overline{\mathbf{y}}(\lambda) + t_1\mathbf{f}_1(\lambda_K) + ... + t_K\mathbf{f}_K(\lambda_K). \end{cases} \tag{3}$$

The key point here is that the radiance values in $K$ spectral points are represented through the same EOFs. Then, by solving (3) we obtain PC scores $t_1, ...t_K$, and, by using (2), the full spectrum in $W$ spectral points can be readily restored.

This approach requires a set of precomputed EOFs which is derived from a training data set of simulated or measured spectra. Fig. 3 shows the example of the PCA applied to the dataset of spectra computed in the Hartley-Huggins band used for ozone retrieval. The data set consists of $10^5$ spectra. The following parameters are varied for the generation of reflectance spectra: the solar zenith angle, the viewing zenith angle, the relative azimuthally

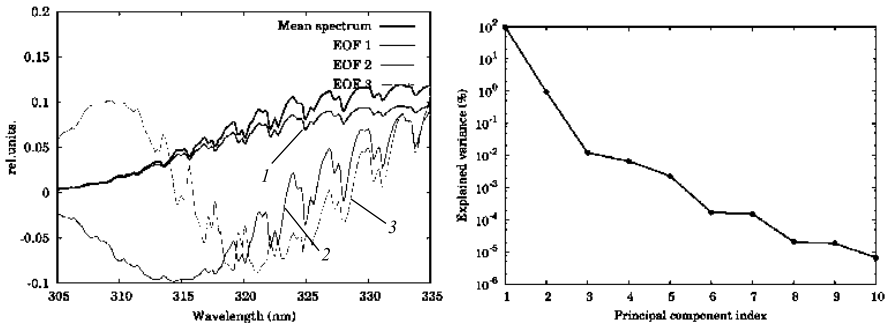


Fig. 3. (Left) Mean spectrum and first three empirical orthogonal functions computed in the Huggins band; (right) explained variance in percentage as a function of the principal component index

angle, the surface albedo, the ozone total column, the surface height, and the temperature. The right plot in Fig. 3 shows that almost 99.9 % of the variance in the data can be explained just with 5 principal components.

To obtain best efficiency, the parameters of this scheme, such as $K$ and a set of chosen wavelengths, have to be tuned empirically. The number $K$ depends on the desired level of variance to be captured by the principal components. Several semi-empirical rules have been proposed for the optimal number of principal components (e.g. the broken stick model [36]). However, there is no universal rule for the selection of $K$; the choice is application-specific.

Regarding the choice of wavelengths, in [37] a method is proposed for selecting the location of monochromatic wavelengths by using a correlation function. In particular, this method involves the following steps:

– The correlation coefficients are computed for the radiance values and then converted to vector angles by an arccosine function;

– The spectral data is rearranged according to the magnitudes of the correlation coefficients;

– The monochromatic radiances are selected by choosing predictors with equal distances in the values of the correlation coefficients.

The schematic representation of the principal component-based hyper-spectral radiative transfer model is shown in Fig. 4. For the input data set containing optical parameters of the atmosphere for a set of wavelengths, the monochromatic radiative transfer solver is called. To obtain the most representative dataset, the smart sampling method [38] is recommended based on Halton sequences [39]. That produces a data set of spectra, which is divided into training data set and validation data set. By applying PCA and the correlation analysis to the training set, the system of EOFs is computed and a subset of spectral points is chosen (spectral sampling), respectively. These two outputs are stored and used for computing PC scores for the validation data set. The spectra in the full wavelength range are restored using Eq. (2) and the error of this reconstruction can be estimated. If the error is larger than required, the number of generated spectra and the number of principal components) are increased. The main output of the training phase are empirical orthogonal functions and spectral sampling (marked with red in Fig. 4), which allow to process new data in the online phase (as shown in Fig. 5).

In [37], it was noted that the slit function convolution operator and the PCA are linear. Therefore the PC scores of the *convolved* spectra are linear functions of monochromatic radiances. Then at the training stage, the corresponding weighting factors of linear dependency are stored together with the system of EOFs and spectral sampling. Authors claim that for an infrared spectrum ranging from 650 $cm^{-1}$ to 3000 $cm^{-1}$ this representation for convolved spectra reduces the number of monochromatic computations from a few thousands to a few hundreds.

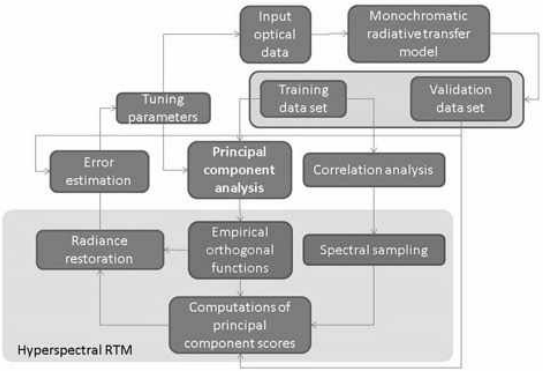


Fig. 4. Schematic representation of the PCA-based radiative transfer model with precomputed empirical orthogonal functions – offline phase


Fig. 5 The same as in Fig. 4, but online phase

The presented approach proved to be efficient and implemented in several packages (e.g. PCTRM [37], RTTOV [40] and others [41, 42]). The main drawback is that it requires time consuming computations of the training data set.

## 4. PRINCIPAL COMPONENT ANALYSIS IN THE FRAMEWORK OF DIFFERENTIAL OPTICAL ABSORPTION SPECTROSOPY

In this section we consider the application of the PCA in the framework of the differential optical absorption spectroscopy (DOAS) [43]. DOAS is widely used to retrieve the trace gas content. The main advantages of this technique are the simplicity and robustness as it is able to filter out the influence of factors, which are not taken into account properly in the RTM. The main idea of DOAS consists in the fact that the absorption caused by gases leads to a strong spectral signatures in the spectral radiances, while the influence of multiple scattering and other factors is usually smooth in the wavelength space. Following [44], the radiance reflected by the atmosphere consisting of $N_g$ gas species with absorption coefficients $\sigma_{abs,g}$ is represented using the weak absorption Beer-Lambert law as

$$\ln y(\lambda) = -\sum_{g=1}^{N_g} S_g \sigma_{abs,g}(\lambda) - P(\lambda) - \mathrm{RRS}(\lambda), \quad (4)$$

where $S_g$ is the number density of gas $g$ along the optical path (also referred to as the slant column density), $P(\lambda)$ is the polynomial term which represents the impact of Rayleigh and aerosol/cloud scattering as well as the surface reflectance, while RRS is a term representing the rotational Raman scattering [45, 46]. In the conventional DOAS, $S_g$ is retrieved through least squares fitting that minimizes the residual between the measured (left part of Eq. (4)) and simulated (right part of Eq. (4)) radiance spectra. Then $S_g$ in converted into the vertical column density ($\Omega_g$) though the air mass factor (AMF). The latter is computed at a single wavelength assuming a prescribed vertical profile of gas $g$ [47, 48].

In [49] the modification of the DOAS approach was proposed for retrieving the $SO_2$ total column. The PCA is applied for the measured spectra in regions with no significant $SO_2$, e.g. the equatorial Pacific:

$$\ln \mathbf{y}(\lambda) = \overline{\ln \mathbf{y}}(\lambda) + \sum_{i=1}^{K} t_k \mathbf{f}_k(\lambda). \quad (5)$$

In this way, the EOFs capture the variability of the data caused by physical processes (i.e. Rayleigh and Raman scattering and ozone absorption). In addition, the features of the instrument (e.g. the instrumental degradation, the slit function and measurement artefacts) are implicitly accounted for by EOFs. That is the training phase. Then, for polluted regions with $SO_2$, representation (5) will produce a residual which is associated with $SO_2$ content. Thus,

$$\ln \mathbf{y}(\lambda) = \overline{\ln \mathbf{y}}(\lambda) + \sum_{i=1}^{K} t_k \mathbf{f}_k(\lambda) + \Omega_{\mathrm{SO}_2} \frac{\partial \ln \mathbf{y}(\lambda)}{\partial \Omega_{\mathrm{SO}_2}}, \quad (6)$$

where $\Omega_{\mathrm{SO}_2}$ is the $SO_2$ vertical column density. The derivative in the last term can be estimated either by finite differences or by using linearized radiative transfer models [50, 51]. Then $\Omega_{\mathrm{SO}_2}$ can be readily retrieved from Eq. (6).

This method has been applied to the Ozone Monitoring Instrument (OMI) [52] data in the spectral range (310–340) nm. As the high order principal components represent the noise rather than a useful signal, the truncation over the principal components also acts as a filter. To reconstruct the spectral radiances, at least 20–30 principal components were required while in the presence of relatively strong $SO_2$ signals that number could be reduced to 8. Authors claim that the noise in the data was decreased by factor of 2 thereby providing greater sensitivity to anthropogenic sources of $SO_2$.

So far, there are no reports of applying the similar approach to other trace gases. One reason for that is the difficulty of obtaining the system of EOFs for regions without a certain trace gas. The second reason is that, strictly speaking, representation (6) is approximate. For $SO_2$ it works correctly and the residual is associated with the $SO_2$ signal. For other trace gases representation (6) might be not valid and more elaborated approach is required.

## 5. DIMENSIONALITY REDUCTION IN RTMS WITH MACHINE LEARNING

### 5.1. General Consideration

Following [53], the inverse problem is solved by reducing it to an exercise in optimization. The main idea behind this method is to find the state vector that minimizes the residual between simulated data and measurements. A non-linear inverse problem is solved iteratively [54, 55]. Assuming an *a priori* state vector **x**, a non-linear forward model is linearized about **x**. Then, the linearized model can be easily inverted and a new estimation for the state vector can be found. This iterative approach is widely used for trace gases retrieval as well as for estimating aerosol and cloud properties [56]. However this inversion method is very time-consuming, due to repeated calls to complex radiative transfer forward models that simulate radiances and Jacobians (i.e. matrices of the first-order partial derivatives of spectral radiances with respect to **x**), and subsequent inversion of relatively large matrices. These considerations motivate the development of alternative inversion techniques for remote sensing real-time applications, which are based on machine learning and therefore sometimes referred to as full-physics inverse learning machines (FP-ILM) [57].

### 5.2. The Concept of Learning Machines for Atmospheric Retrievals

Machine learning algorithms do not consider the optimization problem explicitly. Rather, they *learn* from a given dataset and make predictions regarding parameters of interest. Conceptually, the machine learning algorithm consists of a training phase, wherein the inversion operator is obtained using synthetic data generated by the radiative transfer model, which expresses the "full-physics" component, and an operational phase, in which the inversion operator is applied to real measurements. Here the main advantage over the classical optimization approach is that the time-consuming training phase involving complex radiative transfer modelling is performed off-line; the inverse operator itself is robust and computationally simple.

Fig. 6 is a schematic representation of the possible implementation of the learning machine. During the training phase, a training dataset is computed using a full-physics forward model, which

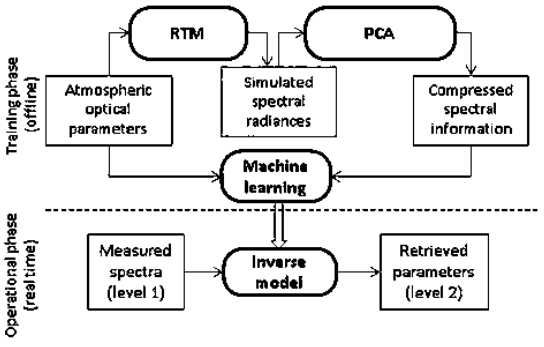


Fig. 6. Schematic representation of the machine learning retrieval algorithm which exploits the dimensionality reduction of the spectral radiances

in our case is the radiative transfer model. In order to capture the essential features of the simulated data and to avoid "over-dimensionality" (the so-called Hughes effect [58]), the simulated spectral data are compressed using an appropriate dimensionality-reduction technique. The mapping between the dimensionality-reduced spectral simulations and the parameter of interest is captured via machine learning.

### 5.3. Machine Learning Based on the Linear Regression Schemes

In the retrieval algorithms based on linear regression, the following representation for the retrieved parameter $x$ is exploited:

$$x = c + \sum_{w=1}^{W} l_w y(\lambda_w),$$

where $c$ is the linear offset and $l_w$ are the regression coefficients. The principal component regression (PCR) method employs the linear regression model between $x$ and the principal component scores of the spectral radiance:

$$x = c + \sum_{k=1}^{K} l_k t_k;$$

As $K << W$, the dimension of the linear regression model (and the corresponding inverse problem) is reduced. Moreover, since the instrument noise does not affect PC scores of low order, the whole inversion scheme is more stable.

For noisy data, the set of eigenvectors $\mathbf{F}$ must be computed for the matrix $\mathbf{C}_Y + \mathbf{C}_e$, rather than for $\mathbf{C}_Y$, where $\mathbf{C}_e$ is the noise covariance matrix. In this case, the PC scores for the noisy data are cor-

related and are therefore called ''projected principal components'' [59]. If the statistics of the noise is unknown, the noise covariance matrix can be estimated by making some assumptions (e.g., Gaussian noise) or by using the following approximation $\mathbf{C}_e \approx \alpha \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\alpha$ is the regularization parameter. This procedure reduces the impact of high-order principal components.

The kernel ridge regression (KRR) algorithm [60] generalizes the PCR method; KRR has been used for predicting atmospheric profiles from the IASI (the infrared atmospheric sounding interferometer) instrument [61]. One drawback of the PCR and KRR models is that the basis vectors $\mathbf{F}$ characterize the measurements $\mathbf{Y}$, while information contained in $\mathbf{X}$ is not taken into account. An alternative model that gets round this drawback is the partial least squares regression (PLSR) [62]. In [62, 64], it was shown that PLSR leads to model-fitting with fewer PCs than required with PCR. In its turn, the PLSR approach can be generalized to the case when we are retrieving a set of correlated parameters (e.g., the temperature profile) rather than a single variable $x$. The corresponding method is then referred to as canonical correlations [65]. The use of canonical correlations in atmospheric sciences applications is summarized in [66].

The approach based on the PCR has been successfully applied for solving the problems of volcanic plume-height retrieval from GOME-2 [67] and TROPOMI measurements [68], as well as $CO_2$ retrieval from GOSAT measurements [69, 70].

## 6. DIMENSIONALITY REDUCTION OF INPUT OPTICAL DATA

### 6.1. Spectra Simulation

An efficient technique using the dimensionality reduction of the optical data has been proposed in [71]. This method relies on the local linearization of the radiative transfer model with respect to input parameters using finite differences. To reduce the number of radiative transfer model calls for estimating finite difference values, the linearization is done in the reduced data space. The method can be summarized as follows. We introduce a correction function as follows:

$$Q(\lambda_w) = \ln[y(\lambda_w) / y_a(\lambda_w)]. \tag{7}$$

Here $y$ is the radiance computed with a full radiative transfer model, while $y_a$ is the radiance computed with an approximate model (e.g., the two-stream model). Then, for the atmosphere consisting of $L$ layers, we consider a state vector $\mathbf{x}_w \in R^{2L+1}$ containing optical parameters for all layers, i.e.,

$$\mathbf{x}_w^T = \begin{bmatrix} \sigma_{\text{abs},1}(\lambda_w),...,\sigma_{\text{abs},L}(\lambda_w), \\ \sigma_{\text{sct},1}(\lambda_w),...,\sigma_{\text{sct},L}(\lambda_w),\rho(\lambda_w) \end{bmatrix},$$

where $\sigma_{\text{sct},i}$ and $\sigma_{\text{abs},i}$ are is the scattering coefficient and the absorption coefficient in the $i^{\text{th}}$ layer, respectively, while $\rho$ is the surface albedo. Thus, the wavelength variability of the optical parameters, representing the radiative transfer code input parameters, is encapsulated in the vector $\mathbf{x_w}$. Note, that the phase function is assumed to be constant within a given spectral interval and therefore not included in the vector $\mathbf{x_w}$. By applying the PCA to $\{\mathbf{x}_w\}_{w=1}^{W}$, we obtain

$$\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^{K} t_{wk}\mathbf{f}_k, \quad \bar{\mathbf{x}} = (1/W)\sum_{w=1}^{W}\mathbf{x}_w.$$

Fig. 7 shows the results of PCA for input optical data in the Huggins band (315–335 nm) and O2A band (755–775 nm). Optical data is taken from [72]. Note that four principal components are sufficient to capture 99.9 % variability of the datasets.

Now, let us assume that $Q(\mathbf{x}_w)$ can be approximated sufficiently well by its Taylor expansion around $\bar{\mathbf{x}}$, that is,

$$Q(\mathbf{x}_w) \approx Q(\bar{\mathbf{x}} + \Delta\mathbf{x}_w) \approx$$
$$\approx Q(\bar{\mathbf{x}}) + \Delta\mathbf{x}_w^T \nabla Q(\bar{\mathbf{x}}) + \frac{1}{2}\Delta\mathbf{x}_w^T \nabla^2 Q(\bar{\mathbf{x}})\Delta\mathbf{x}_w, \tag{8}$$

where $\nabla Q$ and $\nabla^2 Q$ are the gradient and the Hessian of $Q$, respectively. By using central differences to approximate the first and the second-order directional derivatives in (8), we obtain

$$Q(\mathbf{x}_w) \approx Q(\bar{\mathbf{x}}) + \frac{1}{2}\sum_{k=1}^{K}\left[Q(\bar{\mathbf{x}} + \mathbf{f}_k) - Q(\bar{\mathbf{x}} - \mathbf{f}_k)\right]t_{wk}$$
$$+ \frac{1}{2}\sum_{k=1}^{K}\left[Q(\bar{\mathbf{x}} + \mathbf{f}_k) - 2Q(\bar{\mathbf{x}}) + Q(\bar{\mathbf{x}} - \mathbf{f}_k)\right]t_{wk}^2. \tag{9}$$

From (9) and (7) it is apparent that the computation of the correction factor requires $2K+1$ calls of the full- and two-stream models. Note that if we estimated the correction function using finite differences in the initial data space, that would require $2L+1$ calls of the full- and two-stream models. As a result and taking into account that usually $K \ll L$, we are led to a substantial reduction of the computational time.

This approach has been applied for simulating the spectra in the O2A band [71], the Huggins band [34], and $CO_2$ bands [73, 74]. Kopparla et al [75] applied the similar approach for modelling the radiances in the UV/Vis/NIR spectral range (0.3–3000) nm. In all cases authors reported that the root mean square errors of the computed radiances are of order 0.01 %, yet achieving almost a 10-fold increase in speed. The big advantage of this method is that unlike previously considered techniques, this one does not require precomputed databases of spectra.

In [76], the efficiency of input and output space dimensionality reduction techniques was analyzed for simulating the Hartley-Huggins band. The hybrid usage of these techniques was proposed. The output space reduction and the spectral sampling methods are applied to the two-stream solution by using corresponding lookup tables, while multi-stream solution computations are performed within the input data reduction framework, described in this Section. It was found that the combined use of these techniques yields accuracy better than 0.05 % while the speedup factor is about 20.

### 6.2. Retrieval in the Reduced Input Data Space

Since the atmospheric retrieval problem is severely ill-posed, a physically correct result can be obtained only by using a regularization procedure. The latter takes into account some *a priori* information. In this context, dimensionality reduction of the input data space can be regarded as a special type of regularization, i.e. the retrieved parameters should obey a certain dependency reproduced by a chosen set of EOFs.

Timofeyev et al. [77] applied the dimensionality reduction technique to parameterize the aerosol extinction coefficient for incorporation into the inversion algorithm, in which the corresponding PC scores rather than aerosol extinction dependence were retrieved. The system of EOFs was defined for

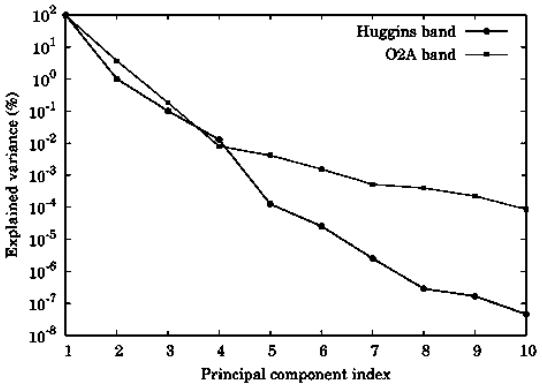


Fig. 7. The explained variance in percentage as a function of the principal component index in the Huggins band and the O2A band

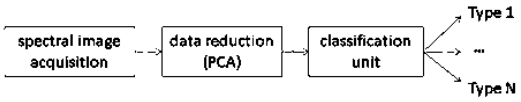


Fig. 8. A schematic representation of the PCA-based classification algorithms applied to the hyperspectral data

a dataset of aerosol extinction coefficients computed on the base of Mie theory [78] and algorithms for particle ensembles. Finally, in [79] the dimensionality reduction is performed in the input (temperature and humidity profiles) and output spaces (spectral radiances), while artificial neural networks are used to establish the interdependency between PC scores. Since the number of independent parameters is reduced, such scheme is more robust and efficient than the conventional one.

## 7. DIMENSIONALITY REDUCTION OF HYPER-SPECTRAL DATA IN CLASSIFICATION PROBLEMS

Dimensionality reduction plays an important role in classification algorithms applied to the hyper-spectral data. On-line anomaly detection and object recognition in remote sensing imagery is extremely important for forest fire and volcanic activity monitoring. In such kind of applications we are confronted with the classification problem. In a reduced data space, the number of variables used in classification is smaller, yet their value is higher than that in the original space. Therefore, the classification algorithms (e.g. *K*-nearest neighbour) are more robust and efficient [80, 81]. PCA can be used to visualize the hyper-spectral

data on a 2D plane, thereby identifying regions with certain features. The concept of combined usage of PCA with classifiers is illustrated in Fig. 8. Such an approach is used not only in space-borne data processing, but also in other fields, such as material science [82, 83], tobacco industry [84] and food production [85].

## 8. CONCLUSIONS

In this review, several techniques of hyper-spectral data processing have been considered. They are all based on the dimensionality reduction procedure. It has been shown that the principal component analysis can be utilized in several ways for hyper-spectral modelling. Therefore it seems that the nomenclature "PCA-based radiative transfer model" is not appropriate since it does not characterize the specific features of the algorithm (e.g., papers [40, 49, 75] present absolutely different models, although all of them are "PCA-based").

It has been shown that the concept of dimensionality reduction gives the framework for formulating hyper-spectral RTM that directly takes into account a strong interdependency in the hyper-spectral data. Further research needs to examine more closely how to combine techniques outlined in this review. For instance in [86], a hybrid approach comprising the correlated-*k* method and the dimensionality reduction of the input data has been described. Such models are extremely important for processing remote sensing Big Data in the current missions, and becoming a mainstream in the development of next generation atmospheric processors.

The discussed principles of data reduction of hyper-spectral data are generic and can be applied in various applications, including material science and electron spectroscopy due to the similar methodologies adopted in these fields [87]. In addition, PCA is a perspective tool for analysing hyper-spectral optical data in medicine. In particular, recent studies have shown that the dimensionality reduction of data cubes can improve the recognition and classification algorithms, which would be extremely important for optical early disease diagnostics [88].

## ACKNOWLEDGEMENTS

The authors are grateful to the Editor-in-Chief of the journal "Light and Engineering" professor Vladimir P. Budak for inviting to write this survey, as well as to the anonymous reviewer for careful reading of the manuscript and valuable suggestions.

## REFERENCES

1. D.G. Loyola, S. Gimeno Garcia, R. Lutz, A. Argyrouli, F. Romahn, R.J.D. Spurr, M. Pedergnana, A. Doicu, V. Molina Garcá, and O. Schüssler. The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 precursor. *Atmospheric Measurement Techniques*, 2018,11(1), pp.409–427.

2. C.A. Lee, S.D. Gasster, A. Plaza, C. Chang, and B. Huang. Recent developments in high performance computing for remote sensing: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 20114(3), pp.508–527.

3. J.P. Veefkind, I. Aben, K. McMullan, H. Forster, J. de Vries, G. Otter, J. Claas, H.J. Eskes, J.F. de Haan, Q. Kleipool, and et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, 2012, 120, pp.70–83.

4. Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie. Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 2015, 51, pp.47–60.

5. P. Liu. A survey of remote-sensing big data. *Frontiers in Environmental Science*, 2015, 3, 45p.

6. V.M. Roozendael, R. Spurr, D. Loyola, C. Lerot, D. Balis, J.C. Lambert, W. Zimmer, J. Gent, J. Van Geffen, M.E. Koukouli, J. Granville, A. Doicu, C. Fayt, and C. Zehner. Sixteen years of GOME/ERS2 total ozone data: the new direct-fitting GOME Data Processor (GDP) Version 5: I. algorithm description. *J Geophys Res: Atmospheres*, 2012,117: D03305, pp.1–18.

7. V. Natraj. A review of fast radiative transfer techniques. In A.A. Kokhanovsky, editor, *Light scattering reviews*, volume 8, pages 475–504. Springer Berlin Heidelberg, 2013.

8. D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. Acceleration techniques for the discrete ordinate method. *J Quant Spectrosc Radiat Transfer*, 2013,114, pp.73–81.

9. D.S. Efremenko, D. Loyola, A. Doicu, and T. Trautmann. Data-intensive computing in radiative transfer modelling. In P. Soille and P.G. Marchetti, editors, *Proc. of the 2016 conference on Big Data from Space (BiDS16)*, Santa Cruz de Tenerife, Spain, 2016, pp. 188–191.

10. V.P. Budak, G.A. Kaloshin, O.V. Shagalov, and V.S. Zheltov. Numerical modeling of the radiative transfer in a turbid medium using the synthetic iteration. *Opt. Express*, 23(15): A829, 2015.

11. D.S. Efremenko, D.G. Loyola, A. Doicu, and R.J.D. Spurr. Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method. *Computer Physics Communications*, 2014, 185(12), pp.3079–3089.

12. D.S. Efremenko, D. Loyola, R.J.D. Spurr, and A. Doicu. Acceleration of radiative transfer model calculations for the retrieval of trace gases under cloudy conditions. *J Quant Spectrosc Radiat Transfer*, 2014,135, pp.58–65.

13. R.J.D. Spurr. VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. *J Quant Spectrosc Radiat Transfer*, 2006, 102(2), pp.316–342.

14. R.J.D. Spurr, T.P. Kurosu, and K.V. Chance. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. *J Quant Spectrosc Radiat Transfer*, 2001, 68(6), pp.689–735.

15. K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil Mag*, 19012, 6, pp.559–572.

16. R. D. Morris, A. Kottas, M. Taddy, R. Furfaro, and B.D. Ganapol. A statistical framework for the sensitivity analysis of radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, 2008,46(12), pp.4062–4074.

17. V.A. Ambartzumyan. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars. *Publ. Obs. Astron. Univ. Leningrad*, 1936, 6, pp.7–18.

18. R.M. Goody, R. West, L. Chen, and D. Crisp. The correlated k-method for radiation calculations in nonhomogeneous atmosphere. *J Quant Spectrosc Radiat Transfer*, 1989, 42, 6, pp.539–550.

19. W.J. Wiscombe and J.W. Evans. Exponential-sum fitting of radiative transmission functions. *J Comput Phys*, 1997, 24, 4, pp.416–444.

20. R. West, D. Crisp, and L. Chen. Mapping transformations for broadband atmospheric radiation calculation. *J Quant Spectrosc Radiat Transfer*, 1990, 43,3, pp.191–199.

21. E. Boesche, P. Stammes, R. Preusker, R. Bennartz, W. Knap, and J. Fischer. Polarization of skylight in the O2A band: effects of aerosol properties. *Applied Optics*, 2008, 47, 19, p. 3467.

22. Christiane Helling and Uffe Grae Jorgensen. Optimizing the opacity sampling method. *Astronomy and Astrophysics*, 1998, 337, pp.477–486.

23. Frédéric André, Longfeng Hou, Maxime Roger, and Rodolphe Vaillon. The multispectral gas radiation modeling: A new theoretical framework based on a multidimensional approach to k-distribution methods. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2014,147, pp.178–195.

24. Boris A. Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 1. FKDM, fast k-distribution model for the longwave. *Journal of Geophysical Research*, 109(D2), 2004.

25. Boris Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. *Journal of Geophysical Research*, 110(D2), 2005.

26. V. A. Falaleeva and B.A. Fomin. Overcoming spectroscopic challenges in direct problems of satellite sounding of the atmosphere. *Atmospheric and Oceanic Optics*, 2017, 30(1), pp.1–6.

27. B.A. Fomin. Effective interpolation technique for line-by-line calculations of radiation absorption in gases. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 1995, 53(6), pp.663–669.

28. S. Najmabadi, P. Offenhäuser, M. Hamann, G. Jajnabalkya, F. Hempert, C. Glass, and S. Simon. Analyzing the effect and performance of lossy compression on aeroacoustic simulation of gas injector. *Computation*, 2017, 5(4):24.

29. S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(22), pp.2323–2326.

30. M.A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 1991, 37(2), pp.233–243.

31. A.N. Gorban, B. Kégl, D.C. Wunsch, and A.Y. Zinovyev, editors. *Principal Manifolds for Data Visualization and Dimension Reduction*. Springer Berlin Heidelberg, 2008.

32. I.K. Fodor. *A Survey of Dimension Reduction Techniques*. Office of Scientific and Technical Information (OSTI), 2002.

33. L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. *Dimensionality Reduction: A Comparative Review*. Tilburg University Technical Report, TiCC-TR2009–005, 2009.

34. D.S. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone re-

trievals. *J Quant Spectrosc Radiat Transfer*, 2014, 133, pp.128–135.

35. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011,12, pp.2825–2830.

36. Robert H. MacArthur. On the relative abundance of bird species. *Proc Natl Acad Sci USA*, 1957, 43(3), pp.293–295.

37. X Liu, W.L. Smith, D.K. Zhou, and A. Larar. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics*, 2006, 45(1), pp.201–208.

38. D. G. Loyola, M. Pedergnana, and S. Gimeno Garcia. Smart sampling and incremental function learning for very large high dimensional data. *Neural Networks*, 2016, 78, pp.75–87.

39. J. H. Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Commun ACM*, 1964, 7(12), pp,701–702.

40. M. Matricardi. A principal component based version of the RTTOV fast radiative transfer model. *Quarterly Journal of the Royal Meteorological Society*, 2010, 136, pp.1823–1835.

41. P. D. Hurley, S. Oliver, D. Farrah, L. Wang, and A. Efstathiou. Principal component analysis and radiative transfer modelling of Spitzer Infrared spectrograph spectra of ultraluminous infrared galaxies. *Monthly Notices of the Royal Astronomical Society*, 2012, 424(3), pp.2069–2078.

42. A. Hollstein and R. Lindstrot. Fast reconstruction of hyperspectral radiative transfer simulations by using small spectral subsets: application to the oxygen A band. *Atmospheric Measurement Techniques*, 2014, 7(2), pp.599–607.

43. U. Platt. Differential optical absorption spectroscopy (DOAS). *Chem Anal Series*, 1994, 127, pp.27–83.

44. U. Platt and J. Stutz. *Differential Optical Absorption Spectroscopy: Principles and Applications*. Springer-Verlag, Berlin, Heidelberg, 2008.

45. G.W. Kattawar, A.T. Young, and T.J. Humphreys. Inelastic scattering in planetary atmospheres. I. The Ring effect, without aerosols. *Astrophys J*, 1981, 243, pp.1049–1057.

46. M. Vountas, V.V. Rozanov, and J.P. Burrows. Ring effect: Impact of rotational Raman scattering on radiative transfer in earth's atmosphere. *J Quant Spectrosc Radiat Transfer*, 1998, 60(6), pp.943–961.

47. James Slusser, Kyle Hammond, Arve Kylling, Knut Stamnes, Lori Perliski, Arne Dahlback, Donald Anderson, and Robert DeMajistre. Comparison of air mass computations. *Journal of Geophysical Research: Atmospheres*, 1996,101(D5), pp. 9315–9321.

48. Lori M. Perliski and Susan Solomon. On the evaluation of air mass factors for atmospheric near-ultraviolet and visible absorption spectroscopy. *Journal of Geophysical Research*, 1993, 98(D6), p.10363.

49. Can Li, Joanna Joiner, Nickolay A. Krotkov, and Pawan K. Bhartia. A fast and sensitive new satellite SO2 retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument. *Geophysical Research Letters*, 2013, 40(23), pp. 6314–6318.

50. R.J.D. Spurr. LIDORT and VLIDORT. Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In A.A. Kokhanovsky, editor, *Light scattering reviews*, 2008, V. 3, pp. 229–275.

51. A. Doicu and T. Trautmann. Two linearization methods for atmospheric remote sensing. *J Quant Spectrosc Radiat Transfer*, 2009, 110(8), pp. 477–490.

52. P.F. Levelt, G.H.J. van den Oord, M.R. Dobber, A. Malkki, Huib Visser, Johan de Vries, P. Stammes, J.O.V. Lundell, and H. Saari. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(5), pp.1093–1101.

53. A.N. Tikhonov and V.Y. Arsenin. *Solution of Ill-Posed Problems*. Winston, 1977.

54. C.D. Rodgers. *Inverse methods for atmospheric sounding: Theory and Practice*. Wolrd Scientific Publishing, 2000.

55. A. Doicu, T. Trautmann, and F. Schreier. *Numerical Regularization for Atmospheric Inverse Problems*. Springer, 2010.

56. D. S. Efremenko, O. Schüssler, A. Doicu, and D. Loyola. A stochastic cloud model for cloud and ozone retrievals from UV measurements. *J Quant Spectrosc Radiat Transfer*, November 2016,184, pp.167–179.

57. J. Xu, O. Schussler, D.G. Loyola Rodriguez, F. Romahn, and A. Doicu. A novel ozone profile shape retrieval using full-physics inverse learning machine (FP-ILM). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(12), pp.5442–5457.

58. G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inform. Theory*, 1968, 14(1), pp.55–63.

59. W.J. Blackwell and F.W. Chen. *Neural Networks in Atmospheric Remote Sensing*. Lexington, 2009.

60. John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, NY, USA, 2004.

61. G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet. Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data. *IEEE Transactions on Geoscience and Remote Sensing*, May 2012 50(5), pp.1759–1769.

62. Roman Rosipal and Nicole Krämer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection*, pages 34–51. Springer Science + Business Media, 2006.

63. Peter D. Wentzell and Lorenzo Vega Montoto. Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and Intelligent Laboratory Systems*, 2003, 65(2), pp.257–279.

64. Saikat Maitra and Jun Yan. Principal component analysis and partial least squares: Two dimension reduction techniques for regression. In *Discussion Papers: 2008 Discussion Paper Program – Applying Multivariate Statistical Models*, 2008, pp.79–90, Casualty actuarial society.

65. H. Hotelling. The most predictable criterion. *Journal of Educational Psychology*, 1935, 26, pp.139–142.

66. D. Wilks. *Statistical Methods in the Atmospheric Sciences, 3rd Edition*. New York: Elsevier, 2011.

67. D.S. Efremenko, D.G. Loyola R., P. Hedelt, and R.J.D. Spurr. Volcanic SO2 plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm. *International Journal of Remote Sensing*, 2017, 38(sup1), pp.1–27.

68. Pascal Hedelt, Dmitry S. Efremenko, Diego G. Loyola, Robert Spurr, and Lieven Clarisse. SO2 layer height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM. *Atmospheric Measurement Techniques Discussions*, Feb. 2019, pp.1–23.

69. M.Yu. Kataev, A.K. Lukyanov, and A.A. Bekerov. Modification of the empirical orthogonal functions method for solving the inverse task of retrieving of the CO2 total content from satellite data. *Journal of Siberian Federal University. Engineering & Technologies*, 2018,11(1), pp.77–85.

70. M.Y. Kataev and A.K. Lukyanov. Empirical orthogonal functions and its modification in the task of retrieving of the total amount CO2 and CH4 with help of satellite Fourier transform spectrometer GOSAT (TANSO-FTS). In G.G. Matvienko and O.A. Romanovskii, editors, *22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. SPIE-Intl Soc Optical Eng, 2016.

71. V. Natraj, X. Jiang, R.L. Shia, X. Huang, J.S. Margolis, and Y.L. Yung. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the O2A-band. *J Quant Spectrosc Radiat Transfer*, 2005, 95(4), pp.539–556.

72. I.E. Gordon, L.S. Rothman, C. Hill, R.V. Kochanov, Y. Tan, P.F. Bernath, M. Birk, V. Boudon, A. Campargue, K.V. Chance, B.J. Drouin, J.-M. Flaud, R.R. Gamache, J.T. Hodges, D. Jacquemart, V.I. Perevalov, A. Perrin, K.P. Shine, M.-A.H. Smith, J. Tennyson, G.C. Toon, H. Tran, V.G. Tyuterev, A. Barbe, A.G. Császár, V.M. Devi, T. Furtenbacher, J.J. Harrison, J.-M. Hartmann, A. Jolly, T.J. Johnson, T. Karman, I. Kleiner, A.A. Kyuberis, J. Loos, O.M. Lyulin, S.T. Massie, S.N. Mikhailenko, N. Moazzen-Ahmadi, H.S.P. Müller, O.V. Naumenko, A.V. Nikitin, O.L. Polyansky, M. Rey, M. Rotger, S.W. Sharpe, K. Sung, E. Starikova, S.A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Wciso, S. Yu, and E.J. Zak. The HITRAN2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2017, 203, pp.3–69.

73. P. Somkuti, H. Boesch, V. Natraj, and P. Kopparla. Application of a PCA-based fast radiative transfer model to XCO2 retrievals in the shortwave infrared. *Journal of Geophysical Research: Atmospheres*, 122(19), pp.10,477–10,496.

74. V. Natraj, R.L. Shia, and Y.L. Yung. On the use of principal component analysis to speed up radiative transfer calculations. *J Quant Spectrosc Radiat Transfer*, 2010, 111(5), pp.810–816.

75. P. Kopparla, V. Natraj, R. Spurr, R. Shia, D. Crisp, and Y.L. Yung. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2016, 173, pp. 65–71.

76. Ana del Águila, Dmitry Efremenko, Víctor Molina García, and Jian Xu. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band. *Atmosphere*, Mar. 2019,10(3), p.142.

77. Y. Timofeyev, A. Polyakov, H. Steele, and M. Newchurch. Optimal eigenanalysis for the treatment of aerosols in the retrieval of atmospheric composition from transmission measurements. *Appl. Opt.*, July 2003,42 (15), p.2635.

78. G. Mie. Beitraege zur optik trueber medien, speziell kolloidaler metalloesungen. *Annalen der Physik*, 1908, 330(3), pp.377–445.

79. A. V. Polyakov, Yu.M. Timofeev, and Ya.A. Virolainen. Using artificial neural networks in the temperature and humidity sounding of the atmosphere. *Izvestiya, Atmospheric and Oceanic Physics*, 2014, 50(3), pp.330–336.

80. Michael T. Eismann, Joseph Meola, and Russell C. Hardie. Hyperspectral change detection in the presence of diurnal and seasonal variations. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(1), pp. 237–249.

81. Dandan Ma, Yuan Yuan, and Qi Wang. Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation. *Remote Sensing*, May 2018,10(5), p.745.

82. David B Brough, Daniel Wheeler, and Surya R. Kalidindi. Materials knowledge systems in python – a data science framework for accelerated development of hierarchical materials. *Integrating Materials and Manufacturing Innovation*, Mar.2017, 6(1), pp. 36–53.

83. Sai Kiranmayee Samudrala, Prasanna Venkataraman Balachandran, Jaroslaw Zola, Krishna Rajan, and Baskar Ganapathysubramanian. A software framework for data dimensionality reduction: application to chemical crystallography. *Integrating Materials and Manufacturing Innovation*, 3(1), Jun. 2014.

84. Pilar B. Garcá-Allende, Olga M. Conde, Ana M. Cubillas, César Jáuregui, and José M. López-Higuera. New raw material discrimination system based on a spatial optical spectroscopy technique. *Sensors and Actuators A: Physical*, Apr. 2007, 135(2), pp. 605–612.

85. Juan Xing, Cédric Bravo, Pál T. Jancsók, Herman Ramon, and Josse De Baerdemaeker. Detecting bruises on 'golden delicious' apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering*, Jan 2005, 90(1), pp.27–36.

86. V. Molina García, S. Sasi, D.S. Efremenko, A. Doicu, and D. Loyola. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2018, 213, pp. 228–240.

87. V.P. Afanas'ev, V.P. Budak, D.S. Efremenko, and P.S. Kaplya. Application of the photometric theory of the radiance field in the problems of electron scattering. *Light & Engineering*, 2019, 27(2), pp. 88–96.

88. F. Vasefi, N. MacKinnon, and D.L. Farkas. Hyperspectral and multispectral imaging in dermatology. In *Imaging in Dermatology*, Elsevier, 2016, pp. 187–201.

***Ana del Águila***
graduated in Physics from the Granada University (UGR) in 2015. From 2016-2018 she worked as an early-stage researcher at the National Institute for Aerospace Technology (INTA) in Spain. At present, she is doing the Ph.D. in the German Aerospace Centre (DLR) with a DAAD/DLR scholarship. Her scientific interests are in-situ atmospheric aerosols, lidar systems, remote sensing, radiate transfer and Big Data analysis

***Dmitry S. Efremenko***
graduated from the Moscow Power Engineering institute (MPEI) in 2009. He received his Ph.D. degree from the Moscow State University in 2011 and the habilitation degree from MPEI in 2017. Since 2011 he works as a research scientist at the German Aerospace Centre (DLR). His scientific interests include radiate transfer, remote sensing, and Big Data analysis

***Thomas Trautmann***,
Prof., Dr., graduated with a diploma in meteorology at the Johannes Gutenberg University at Mainz (JGU) in 1985. From the JGU he received a Ph.D. degree in 1989, a habilitation degree in 1997, and in 2003 he was appointed as applicant Prof. at the University of Leipzig. Since 2003 he works as head of department Atmospheric Processors at the German Aerospace Center (DLR). His scientific interests include radiate transfer, electromagnetic scattering and atmospheric remote sensing

## A.3 Publication III: Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of $O_2$ A- and $CO_2$ Bands

### Reference

A. del Águila, D. S. Efremenko, V. Molina García, M. Y. Kataev. Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of $O_2$ A- and $CO_2$ Bands. Remote Sensing, 12(8):1250, 2020. doi: 10.3390/rs12081250

### Copyright



### Abstract

Current atmospheric composition sensors provide a large amount of high spectral resolution data. The accurate processing of this data employs time-consuming line-by-line (LBL) radiative transfer models (RTMs). In this paper, we describe a method to accelerate hyperspectral radiative transfer models based on the clustering of the spectral radiances computed with a low-stream RTM and the regression analysis performed for the low-stream and multi-stream RTMs within each cluster. This approach, which we refer to as the Cluster Low-Streams Regression (CLSR) method, is applied for computing the radiance spectra in the O2 A-band at 760 nm and the CO2 band at 1610 nm for five atmospheric scenarios. The CLSR method is also compared with the principal component analysis (PCA)-based RTM, showing an improvement in terms of accuracy and computational performance over PCA-based RTMs. As low-stream models, the two-stream and the single-scattering RTMs are considered. We show that the error of this approach is modulated by the optical thickness of the atmosphere. Nevertheless, the CLSR method provides a performance enhancement of almost two orders of magnitude compared to the LBL model, while the error of the technique is below 0.1both bands.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko; V.M.G: Víctor Molina García; M.Y.K.: M. Y. Kataev)
As stated in the publication, the author contributions are the following: the methodology of the study was developed by A.d.Á. and D.S.E. while the software and implementations were carried out by A.d.Á. and V.M.G. The validation of the study was done by D.S.E. and M.Y.K. The data visualization and the tables were done by A.d.Á. while the supervision and funding acquisition were done by D.S.E.

The overall own contribution of A.d.Á for the publication in Appendix A.3. is estimated at 85%, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.3).

**Table A.3:** Criteria and estimated own contribution for Appendix A.3

| **Criteria** | **Estimated contribution** |
|---|---|
| Conceptualization | 70% |
| Implementation | 90% |
| Analysis and discussion | 80% |
| Figures and tables | 100% |
| Manuscript writing | 85% |
| **Total** | **85** % |

*remote sensing*

MDPI

*Article*

# Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of $O_2$ A- and $CO_2$ Bands

**Ana del Águila [1,*], Dmitry S. Efremenko [1], Víctor Molina García [1] and Michael Yu. Kataev [2]**

[1] Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany; dmitry.efremenko@dlr.de (D.S.E.); Victor.MolinaGarcia@dlr.de (V.M.G.)

[2] Department of Control Systems, Tomsk State University of Control Systems and Radioelectronics, Tomsk 634050, Russia; kmy@asu.tusur.ru

* Correspondence: Ana.delAguilaPerez@dlr.de; Tel.: +49-8153-28-1983

Received: 19 March 2020; Accepted: 10 April 2020; Published: 15 April 2020


**Abstract:** Current atmospheric composition sensors provide a large amount of high spectral resolution data. The accurate processing of this data employs time-consuming line-by-line (LBL) radiative transfer models (RTMs). In this paper, we describe a method to accelerate hyperspectral radiative transfer models based on the clustering of the spectral radiances computed with a low-stream RTM and the regression analysis performed for the low-stream and multi-stream RTMs within each cluster. This approach, which we refer to as the Cluster Low-Streams Regression (CLSR) method, is applied for computing the radiance spectra in the $O_2$ A-band at 760 nm and the $CO_2$ band at 1610 nm for five atmospheric scenarios. The CLSR method is also compared with the principal component analysis (PCA)-based RTM, showing an improvement in terms of accuracy and computational performance over PCA-based RTMs. As low-stream models, the two-stream and the single-scattering RTMs are considered. We show that the error of this approach is modulated by the optical thickness of the atmosphere. Nevertheless, the CLSR method provides a performance enhancement of almost two orders of magnitude compared to the LBL model, while the error of the technique is below 0.1% for both bands.

**Keywords:** hyperspectral data; fast radiative transfer models; acceleration techniques; regression; $O_2$ A-band; $CO_2$ band; GOSAT; TROPOMI

## 1. Introduction

Radiative transfer models (RTM) are a key part of the remote sensing algorithms, which are used to retrieve atmospheric parameters from Earth observation data. The new atmospheric composition sensors with high spatial and spectral resolution require accurate and efficient computational algorithms. An accurate method to simulate spectral radiances in the absorption bands is the line-by-line (LBL) approach [1]. However, due to the high spectral variability of the gas absorption coefficient $k$ in the absorption bands, the LBL-approach is very time-consuming because it requires up to several thousands of monochromatic computations per absorption band. In this regard, specifically designed hyperspectral RTMs are required. They reduce the number of monochromatic computations considerably without compromising accuracy. The basic idea behind the hyperspectral RTMs is to get rid of redundancies in hyperspectral data. This principle can be traced back to Ambartsumian [2], who noted that the transmission within a spectral interval does not depend on the LBL variation of $k$, but rather on the distribution of values of $k$ within the spectral interval. This idea was used in the $k$-distribution method [3–6], in which the wavelengths are grouped into a smaller number of bins with similar values of $k$, and the LBL calculations are replaced by a smaller set of radiative transfer

simulations. In [7,8], the transmission function for a given spectral interval is fitted by a sum of exponentials, while the corresponding fitting coefficients are computed from a reduced number of monochromatic computations. A similar approach is described in Moncet et al. [9], where the fitting weights and the most representative wavelengths are chosen appropriately.

The state-of-the-art hyperspectral RTMs employ dimensionality reduction techniques such as principal component analysis (PCA). In [10,11], PCA is applied to the spectral radiance data to establish a set of empirical orthogonal functions (EOFs), so that an arbitrary spectrum at full spectral resolution can be reconstructed as a weighted sum of EOFs. The weights are found by performing monochromatic simulations at a reduced number of wavelengths. To accelerate the computations in the $O_2$ A-band, Natraj et al. [12] proposed a fundamentally different PCA-based radiative transfer model, in which the dimensionality of the optical properties data is reduced. A two-stream radiative transfer model was used as an approximate model, and the dependency of the corresponding correction factor on the optical parameters was modeled by a second-order Taylor expansion about the mean value of the optical parameters in the reduced optical data space. This approach was extended to other dimensionality reduction techniques [13] and spectral ranges [14–16]; moreover, it was implemented in conjunction with PCA for spectral radiances [17] and with the *k*-distribution method [18]. The errors of these approaches are usually below 0.1% for the spectral radiances, while the performance enhancement may reach several orders of magnitude depending on the spectral region and the required level of accuracy.

In Efremenko et al. [19] it was shown that, after parallelizing the PCA-based RTM computations, as much as half of the computational time is due to the PCA itself; consequently, according to Amdahl's Law [20], no further acceleration of a PCA-based RTM is possible. In this regard, it is of great interest to develop acceleration techniques for hyperspectral RTM computations that do not exploit PCA. A possible candidate, which deserves this purpose, is the low streams interpolation method proposed by O'Dell [21]. In this method, (i) each spectral point is assigned to a specific bin according to the values of the gas optical depth, (ii) an approximate two-stream model is used to compute the radiances at all spectral points, and (iii) the two-stream radiances are corrected to the multi-stream radiances by calling a multi-stream model only once for each bin. In this paper, we propose the so-called Cluster Low-Streams Regression (CLSR) method, in which (i) wavelengths are grouped according to the radiance values obtained with the approximate model (rather than the optical properties), and (ii) the single-scattering and the two-stream models are used as approximate models. The CLSR method is tested against the PCA-based RTM for the $O_2$ A-band at 760 nm and the weak $CO_2$ band at 1610 nm. These spectral bands are of great interest for retrieving aerosol and cloud parameters [22], as well as $CO_2$ concentrations [23].

The paper is organized as follows. In Section 2, the methodology is presented, including the descriptions of the PCA-based RTM adopted in our research and the CLSR method. In Section 3, the accuracy and efficiency of the CLSR method for several atmospheric scenarios in the $O_2$ A- and weak $CO_2$ absorption bands are analyzed and compared with the PCA-based RTM. In addition, the accuracy of the radiance spectra convolved with the slit functions corresponding to GOME-2 [24], TROPOMI on board Sentinel 5-P [25] and GOSAT [26] instruments is examined. The paper concludes with a summary.

## 2. Methodology

### *2.1. Reference Radiative Transfer Model*

As a reference RTM, the scalar method of discrete ordinates with matrix exponential (DOME) [27] is used. The model is supplied with the left eigenvector approach [28–30] based on the scaling transformation [31,32]. In this method, the number of discrete ordinates (streams) in the polar hemisphere $N_{\mathrm{do}}$ controls the computational time and accuracy. The model is called multi-stream (MS) when $N_{\mathrm{do}} \geq 2$ and low-stream (LS) otherwise. Specifically, the case $N_{\mathrm{do}} = 1$ is the two-stream

model. The simplest RTM is the single-scattering model, which solves the radiative transfer equation neglecting the integral term [33].

The gaseous absorptions for the $O_2$ A- and $CO_2$ bands are computed with the LBL model Py4CAtS [34], while the gas absorption cross-sections are taken from the HITRAN 2016 database [35]. The uncertainties in the spectroscopic parameters are not considered because their role is irrelevant for this study [36]. Continuum (also referred to as collision-induced absorption (CIA) [37,38]) contributions to molecular absorption are not taken into account. We note that the CIA process gives a broad and smooth contribution [39] and hence, does not impose difficulties for the regression techniques considered in this study.

### 2.2. Study Cases

We consider the reflected spectral radiance at the top of the atmosphere (TOA). In our simulations, the atmosphere is discretized into 35 layers with a step of 1 km between 0 and 25 km, and a step of 2.5 km between 25 km and 50 km. We assume a Lambertian surface with an albedo of 0.3. The solar zenith angle, the viewing zenith angle and the relative azimuth angle are 45°, 35° and 90°, respectively.

In the $O_2$ A-band the spectral sampling is 0.001 nm in the spectral range 755–775 nm, while in the $CO_2$ band the spectral sampling is 0.0015 nm in the spectral interval 1590–1620 nm. Thus, on each band, 20,000 spectral points are considered. The Rayleigh cross-sections and depolarization ratios are computed as in [40], while the pressure and temperature profiles correspond to the US standard model atmosphere [41]. The computations are performed for the unit solar irradiance at the TOA.

We define 5 atmospheric scenarios: 'Clear sky', 'Aerosol 1', 'Aerosol 2', 'Cloud 1' and 'Cloud 2'. The 'Clear sky' scenario corresponds to an atmosphere without clouds and aerosols. In the 'Aerosol 1' and 'Aerosol 2' scenarios, the atmosphere contains the clean continental and the polluted continental aerosols taken from the OPAC database [42], respectively. The aerosol optical depths at the middle of the absorption bands are shown in Table 1. In the 'Cloud 1' and 'Cloud 2' scenarios, a continental clean cumulus cloud with a modified Gamma size distribution [42,43]:

$$p(a) \propto a^{\alpha} \exp\left[-\frac{\alpha}{\gamma}\left(\frac{a}{a_{\mathrm{mod}}}\right)^{\gamma}\right],$$

is considered. The size distribution parameters are $a_{\mathrm{mod}} = 4.8$ µm, $\alpha = 5$ and $\gamma = 2.16$, the droplet size ranges between 0.02 and 50.0 µm, and the cloud optical depths are $\tau = 10$ ('Cloud 1') and $\tau = 20$ ('Cloud 2'). In both cases, the cloud-top height is 5 km and the cloud geometrical thickness is 1 km.

The number of streams considered for the multi-stream model is the same for all scenarios. We might need a different number of streams for different scenarios, but in order to simplify and ensure convergence of relative errors for all cases, we use 32 streams.

**Table 1.** Aerosol optical thickness used in the simulations for $O_2$ A- and $CO_2$ bands at the middle of the corresponding absorption band.

| | **$O_2$ A-Band (760 nm)** | **$CO_2$ Band (1610 nm)** |
|---|---|---|
| Aerosol 1 | 0.2 | 0.08 |
| Aerosol 2 | 1.2 | 0.41 |

### 2.3. PCA-Based RTM

Since a PCA-based RTM can be implemented in several ways (e.g., [12,13,15–17,44,45]), we describe below the main features of the PCA-based RTM used in this paper. The idea of the method is to refine the two-stream solution by a correction function estimated in the reduced space of the optical properties. Considering a discretization of the atmosphere in $L$ layers, we define a $2L$-dimensional vector $\mathbf{x}_w$ for each wavelength $\{\lambda_w\}_{w=1}^{W}$, by

$$\mathbf{x}_w^{\mathrm{T}} = [\ln\tau_1(\lambda_w), \ldots, \ln\tau_L(\lambda_w), \ln\omega_1(\lambda_w), \ldots, \ln\omega_L(\lambda_w)], \quad (1)$$

where $\tau_k$ and $\omega_k$ are the optical thickness and the single-scattering albedo of the $k$-th layer, respectively. Thus, the vector $\mathbf{x}_w$ encapsulates the wavelength variability of the optical parameters, which are the input parameters of the radiative transfer code. By applying PCA, a $M$-dimensional subspace is found which is spanned by a set of linear independent vectors (empirical orthogonal functions) $\{\mathbf{q}_k\}_{k=1}^{M}$ such that the centered (mean-removed) data $\mathbf{x}_w - \bar{\mathbf{x}}$ lie mainly on this subspace, i.e.,

$$\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^{M} y_{wk}\mathbf{q}_k, \quad w = 1, \ldots, W, \tag{2}$$

where $y_{wk}$ are the principal component (PC) scores.

Following Natraj et al. [12], we introduce a correction function

$$f(\mathbf{x}_w) = \ln \frac{I_{\mathrm{MS}}(\mathbf{x}_w)}{I_{\mathrm{LS}}(\mathbf{x}_w)}, \tag{3}$$

where $I_{\mathrm{LS}}(\mathbf{x}_w)$ is the radiance provided by the low-stream RTM, and $I_{\mathrm{MS}}(\mathbf{x}_w)$ is the radiance simulated by the multi-stream model. Setting $\Delta\mathbf{x}_w = \sum_{k=1}^{M} y_{wk}\mathbf{q}_k$, we consider a second-order Taylor expansion of $f(\mathbf{x}_w)$ around $\bar{\mathbf{x}}$ in the direction $\Delta\mathbf{x}_w$. Approximating the directional derivatives with central differences (see [13,17] for mathematical details) we obtain

$$f(\mathbf{x}_w) \approx f(\bar{\mathbf{x}}) + \frac{1}{2}\sum_{k=1}^{M}[f(\bar{\mathbf{x}}+\mathbf{q}_k) - f(\bar{\mathbf{x}}-\mathbf{q}_k)]y_{wk} + \frac{1}{2}\sum_{k=1}^{M}[f(\bar{\mathbf{x}}+\mathbf{q}_k) - 2f(\bar{\mathbf{x}}) + f(\bar{\mathbf{x}}-\mathbf{q}_k)]y_{wk}^2. \tag{4}$$

Please note, since $M < 2L$ (and in practice, $M \ll 2L$), the estimation of the correction function in the reduced data space is much faster than in the original space. Also note that the total number of calls of the multi-stream RTM is $2M + 1$. Once $f(\mathbf{x}_w)$ is computed, the low-stream radiances can be converted into the multi-stream radiances by using Equation (3). Finally, the predicted multi-stream radiance $\widetilde{I}_{\mathrm{MS}}$ can be expressed as follows:

$$\widetilde{I}_{\mathrm{MS}} = I_{\mathrm{LS}} \exp[f(\mathbf{x}_w)]. \tag{5}$$

In del Águila et al. [17], it was shown that a higher order expansion of $f(\mathbf{x}_w)$ in Equation (4) does not substantially improve the accuracy of the solution. In our simulations, an appropriate value for $M$ is found iteratively, i.e., $M$ is increased until the approximation error is sufficiently small.

### *2.4. Regression Relationship between Multi-Stream and Low-Stream Models*

The LBL computations can be accelerated by finding a regression relationship between the multi-stream radiances and the optical parameters. In O'Dell [21], the correction function of the low-stream radiance as a function of the gas optical thickness $\tau_{\mathrm{gas}}$, i.e., the relative error

$$f(\tau_{\mathrm{gas}}) = \frac{I_{\mathrm{LS}}(\tau_{\mathrm{gas}}) - I_{\mathrm{MS}}(\tau_{\mathrm{gas}})}{I_{\mathrm{MS}}(\tau_{\mathrm{gas}})},$$

was analyzed (Figure 1). The idea is to compute $f(\tau_{\mathrm{gas}})$ at certain values of $\tau_{\mathrm{gas}}$, and then to use a regression model for computing $f(\tau_{\mathrm{gas}})$ at the remaining values of $\tau_{\mathrm{gas}}$. However, for the present application, such an approach has two major drawbacks:

1. the dependence of $f(\tau_{\mathrm{gas}})$ on $\tau_{\mathrm{gas}}$ is non-linear and, therefore, the application of the regression model requires a binning of the $\tau_{\mathrm{gas}}$ values;
2. in a mathematical sense, $f(\tau_{\mathrm{gas}})$ is not a function (for a value of $\tau_{\mathrm{gas}}$, there are several values of $f(\tau_{\mathrm{gas}})$) and therefore, even for a fine binning, the regression model will be not accurate.

To overcome these drawbacks, O'Dell [21] devised a regression model based on the so-called adjusted gas optical thickness, defined as the gas optical thickness from the TOA down to the layer in the atmosphere where the scattering optical depth equals some critical value.

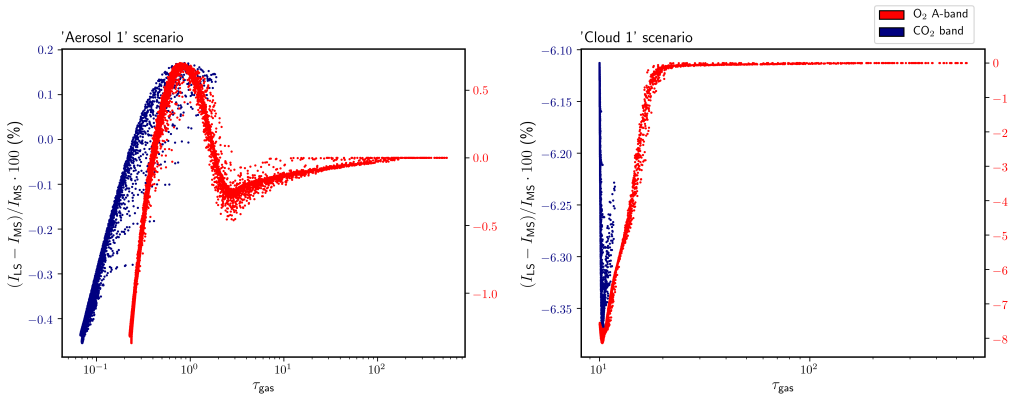


**Figure 1.** Relative errors of the two-stream model as a function of the gas optical depth ($\tau_{gas}$) for the $O_2$ A-band (red) and $CO_2$ band (blue). Left panel corresponds to the 'Aerosol 1' scenario, while right panel shows the results for the 'Cloud 1' scenario. Note that left Y axis refer to $CO_2$ while right Y axis refer to $O_2$ A-band.

A similar idea, which we exploit in the following, is to use a regression model for the radiances computed by a low-stream RTM rather than in the $\tau_{gas}$-space. Figure 2 shows a comparison between the multi- and low-stream radiances for the 'Aerosol 2' scenario. Although the errors of the low-stream RTM are significant (e.g., they may reach 200%), the low- and multi-stream radiances have a similar spectral behaviour. Moreover, as shown in Figure 3, the dependence between the low- and multi-stream radiances is almost linear. In this respect, it seems reasonable to cluster the spectral points according to the radiance values computed with a low-stream RTM (rather than according to the optical properties) and to establish a regression model between the low- and multi-stream radiances within each cluster. In order to decrease the errors of the regression estimates, an additional wavelength-dependent parameter will be included in the regression model: the direct transmittance $T$.

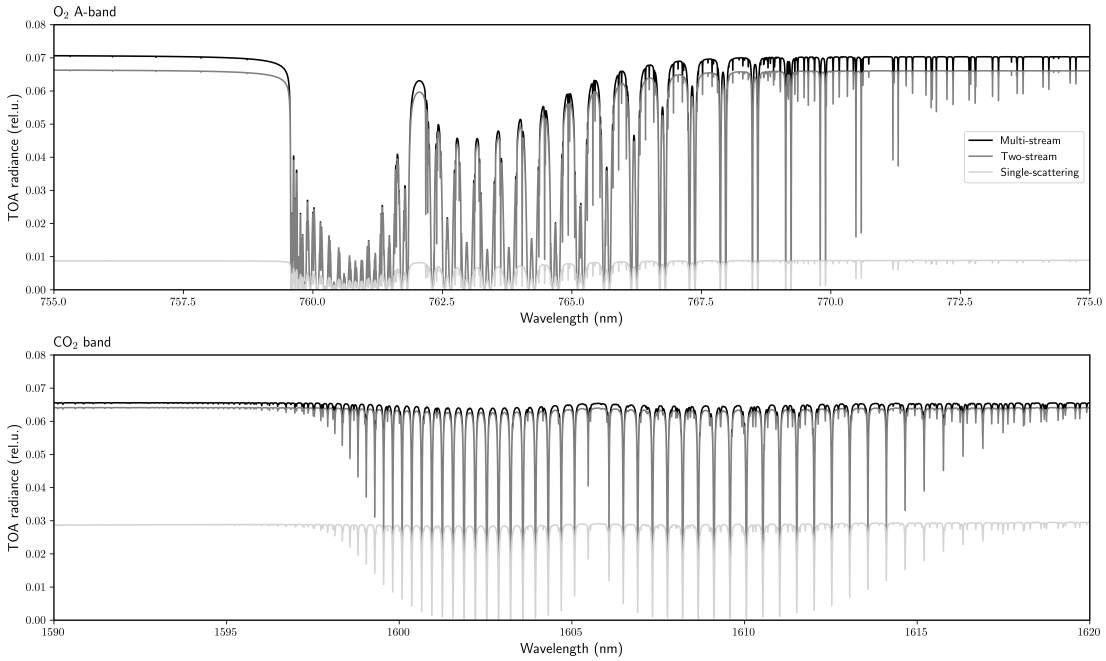


**Figure 2.** Spectral radiances for the 'Aerosol 2' scenario computed by using the multi-stream (black), the two-stream (gray) and the single-scattering (light gray) RTMs: (upper panel) $O_2$ A-band, (lower panel) $CO_2$ band.

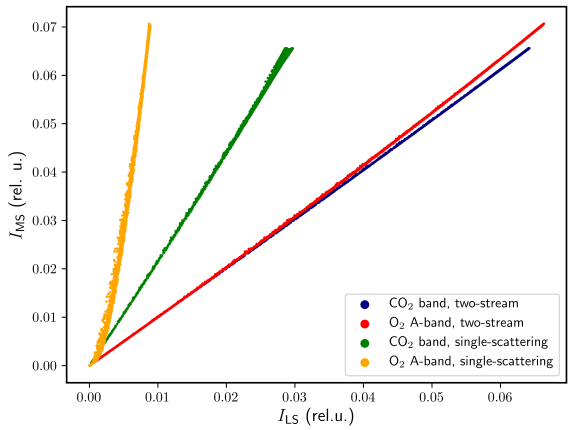


**Figure 3.** Radiance computed with the multi-stream model as a function of radiances computed by using the low-stream models (two-stream and single-scattering models) for the $O_2$ A- and $CO_2$ bands. The figure corresponds to the 'Aerosol 2' scenario.

### *2.5. Cluster Low-Streams Regression Method*

The Cluster Low-Streams Regression (CLSR) method can be formulated as follows.

1. Consider a high-resolution spectrum $\{I_{\mathrm{LS}}(\lambda_i)\}_{i=1}^{N}$ computed at $N$ spectral points $\{\lambda_i\}_{i=1}^{N}$ by means of a low-stream RTM.
2. Sort the radiance set $\{I_{\mathrm{LS}}(\lambda_i)\}_{i=1}^{N}$ in ascending order, and let $\{\hat{I}_{\mathrm{LS},i}\}_{i=1}^{N}$, with $\hat{I}_{\mathrm{LS},i} \leq \hat{I}_{\mathrm{LS},i+1}$, be the sorted radiance set (Figure 4a).
3. Consider $C$ clusters in $\{\hat{I}_{\mathrm{LS},i}\}_{i=1}^{N}$ with $N_C = N/C$ radiance points (Figure 4b), and let the $c$ cluster be defined by the radiance set $\{\hat{I}^c_{\mathrm{LS},i}\}_{i=1}^{N_C}$ for $c = 1, ..., C$.
4. Select $n$ equidistant radiance points in the $c$ cluster, i.e., $\{\bar{I}^c_{\mathrm{LS},q}\}_{q=1}^{n}$, and for the corresponding wavelengths compute the multi-stream radiances $\{\bar{I}^c_{\mathrm{MS},q}\}_{q=1}^{n}$ (Figure 4c).
5. Assume that in each cluster $c$ we have the linear relationship

$$\hat{I}^c_{\mathrm{MS},i} = \alpha^c \hat{T}^c_i + \beta^c \hat{I}^c_{\mathrm{LS},i} + \gamma^c, \; i = 1, ..., N_C, \tag{6}$$

where $\alpha^c$, $\beta^c$ and $\gamma^c$are the regression coefficients of the $c$-th cluster, and $\hat{T}$ is the corresponding direct transmittance.

6. Compute the regression coefficients $\alpha^c$, $\beta^c$ and $\gamma^c$ as a solution to the least square problem

$$(\alpha^c, \beta^c, \gamma^c) = \arg \min_{\alpha^c, \beta^c, \gamma^c} \sum_{q=1}^{n} \left[ \bar{I}^c_{\mathrm{MS},q} - \left( \alpha^c \bar{T}^c_q + \beta^c \bar{I}^c_{\mathrm{LS},q} + \gamma^c \right) \right]^2. \tag{7}$$

7. Use the values of $(\alpha^c, \beta^c, \gamma^c)$ found in the previous step to restore the multi-stream radiances $\{\tilde{I}_{\mathrm{MS},i}\}_{i=1}^{N}$ for all the spectral points according to Equation (6) (Figure 4d).

Here, the "hat" notation $\hat{I}$ refers to the sorted radiances, the "bar" notation $\bar{I}$ refers to the equidistant radiances entering the regression model, and the "tilde" notation $\tilde{I}$ refers to the predicted radiances. Please note that the total number of regression points, and thus the number of calls of the multi-stream model, is $nC$.

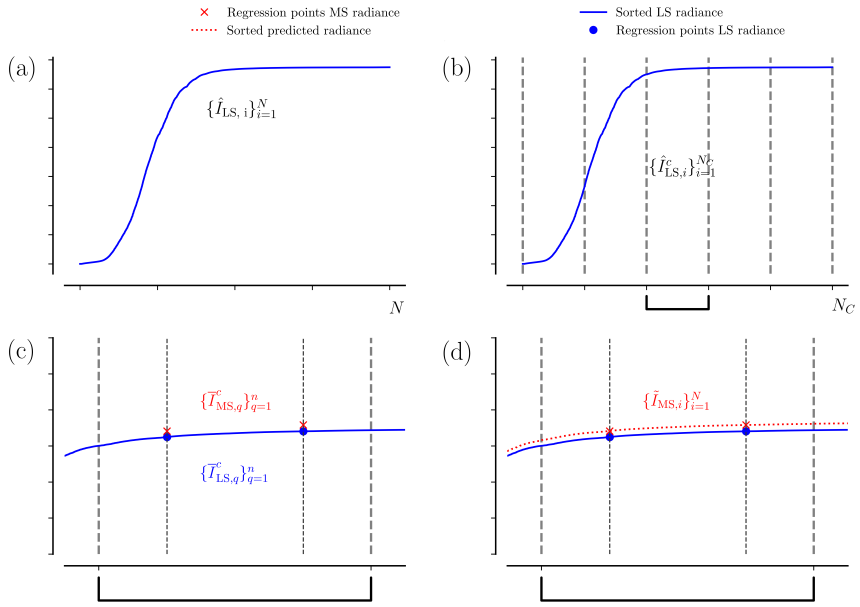


**Figure 4.** Scheme of the Cluster Low-Streams Regression (CLSR) method. (**a**) Sorted radiance of the low-stream (LS) model in ascending order (blue line). (**b**) Division of the LS radiance in equal clusters $C$ in the sorted domain. (**c**) Zoom for one cluster and the selected regression points of the multi-stream (MS) radiance (red crosses). (**d**) Reconstruction of the MS spectra: the predicted radiance is computed for all the spectral points (dashed red line).

### *2.6. Efficiency and Computational Performance Estimation*

To estimate the accuracy of the acceleration techniques, we consider the residual error for the radiance at each spectral point $\lambda_i$:

$$\Delta I_{\text{res},i} = \frac{\tilde{I}_{\text{MS},i} - I_{\text{MS},i}}{I_{\text{MS}}^{\text{cont}}} \cdot 100, \tag{8}$$

where $\tilde{I}_{\text{MS},i}$ is the predicted radiance calculated with either the PCA-based RTM (cf. Equation (5)) or the CLSR method (cf. Equation (6)), while $I_{\text{MS}}^{\text{cont}}$ is the radiance without absorption (i.e., the continuum radiance, which is used to avoid radiance values close to zero in the denominator of Equation (8), when strong gas absorption is present [15]). The mean relative error is computed as follows:

$$\varepsilon = \frac{\sum_{i=1}^{N} |\Delta I_{\text{res},i}|}{N}. \tag{9}$$

Additionally, the median and interquartile range (IQR) of the residual are computed. Essentially, these parameters serve as robust metrics to evaluate the variability of the residuals in the corresponding spectral band, thereby with a lower sensitivity to single outliers [16].

To estimate the performance enhancement, we define the speedup factor as the ratio between the computational time of a multi-stream LBL calculation to that of a certain acceleration technique. To exclude the hardware-related factors from our analysis, we estimate the speedup factor for the PCA-based RTM as [16]

$$S_{\text{PCA}} = \frac{t_{\text{MS}} \times N}{t_{\text{LS}} \times N + t_{\text{PCA}} + (2M+1)(t_{\text{MS}} + t_{\text{LS}})}, \tag{10}$$

where $t_{\text{MS}}$ and $t_{\text{LS}}$ are the computational times for a single monochromatic calculation corresponding to the multi- and low-stream RTMs, respectively, while $t_{\text{PCA}}$ is the computational time of the PCA.

Thus, the numerator in Equation (10) is the time required for the LBL computations. Please note the PCA-based RTM requires $2M + 1$ calls to the multi-stream model.

The speedup factor for the CLSR method is defined by

$$S_{\mathrm{CLSR}} = \frac{t_{\mathrm{MS}} \times N}{t_{\mathrm{LS}} \times N + t_{\mathrm{LSM}} \times C + t_{\mathrm{MS}} \times n \times C}, \tag{11}$$

where $t_{\mathrm{LSM}}$ is the computational time needed for the least-squares method according to Equation (7). Recall that the total number of regression points and the corresponding calls of the multi-stream model is $nC$.

## 3. Results and Discussion

### *3.1. Performance of the Low- and Multi-Streams Models*

In this section, we assess the performance of the low- and multi-stream RTMs before applying the acceleration techniques. The spectral radiances for three atmospheric scenarios ('Clear sky', 'Cloud 1' and 'Cloud 2') are illustrated in Figure 5.

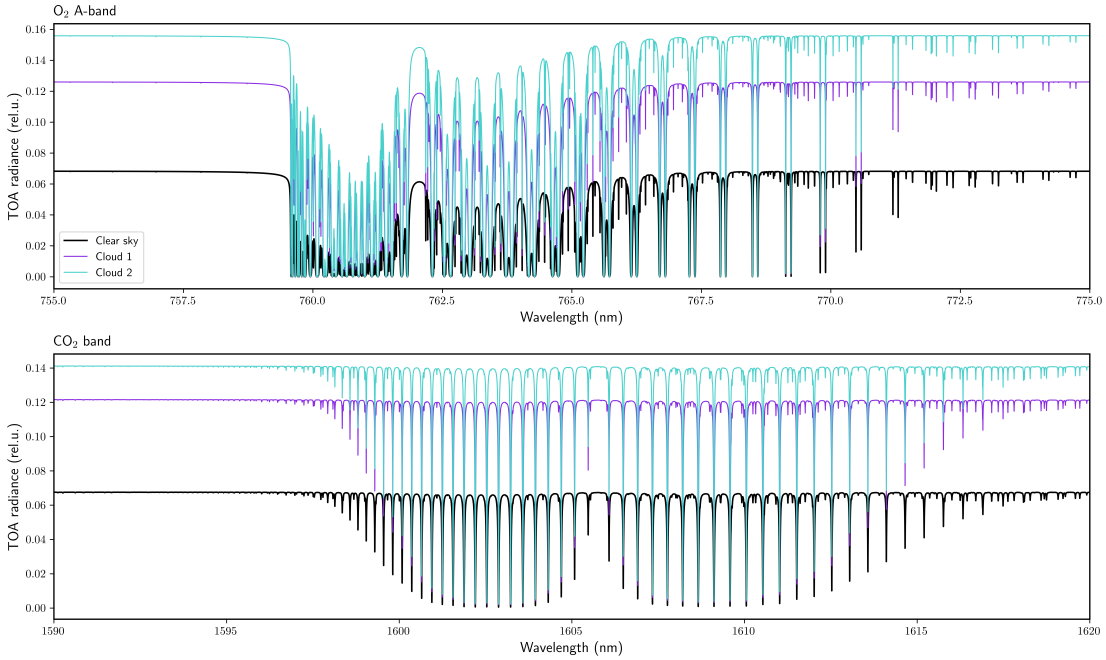


**Figure 5.** Radiance spectra computed by using the multi-stream RTM for three atmospheric scenarios: 'Clear sky' (black), 'Cloud 1' (purple) and 'Cloud 2' (blue). The upper panel corresponds to the $O_2$ A-band, while the bottom panel is for the $CO_2$ weak band.

Tables 2 and 3 provide the mean relative errors of the low-stream RTMs with respect to the multi-stream RTM for the $O_2$ A- and $CO_2$ absorption bands, respectively. Table 4 shows the computational performance for a different number of discrete ordinates $N_{\mathrm{do}}$. The results show that (i) the relative errors of the single-scattering model are higher than those of the two-stream model, (ii) the relative error increases as the optical thickness of the atmosphere increases, and (iii) the speedup factors of the single-scattering model are considerably higher than those of the two-stream model.

**Table 2.** Mean relative error $\varepsilon$ for the low-stream models (single-scattering and two-stream models) compared with the multi-stream model for the $O_2$ A-band.

| Scenario | Single-Scattering Model | Two-Stream Model |
|---|---|---|
| | $\varepsilon$ (%) | $\varepsilon$ (%) |
| Clear sky | 3.4 | 0.14 |
| Aerosol 1 | 36 | 1.0 |
| Aerosol 2 | 78 | 4.6 |
| Cloud 1 | 93 | 7.1 |
| Cloud 2 | 93 | 1.4 |

**Table 3.** Mean relative error $\varepsilon$ for the low-stream models (single-scattering and two-stream models) compared with the multi-stream model for the $CO_2$ weak band.

| Scenario | Single-Scattering Model | Two-Stream Model |
|---|---|---|
| | $\varepsilon$ (%) | $\varepsilon$ (%) |
| Clear sky | 0.203 | 0.004 |
| Aerosol 1 | 14.7 | 0.40 |
| Aerosol 2 | 55.5 | 2.1 |
| Cloud 1 | 97.56 | 6.13 |
| Cloud 2 | 97.91 | 2.00 |

**Table 4.** Computational time in seconds of the radiative transfer solution as a function of the number of streams, and speedup factors with respect to the case $N_{\mathrm{do}} = 32$.

| $N_{\mathrm{do}}$ | Time (s) | Speedup Factor |
|---|---|---|
| 0 | 0.4 | 5800 |
| 1 | 3.2 | 725 |
| 2 | 12.4 | 187.1 |
| 4 | 34.4 | 67.4 |
| 8 | 110 | 21 |
| 16 | 550 | 4.2 |
| 32 | 2320 | - |

### *3.2. Acceleration Techniques: Accuracy Evaluation*

#### 3.2.1. Spectral Residuals

The residuals (cf. Equation (8)) are computed for (i) different numbers of principal components in the case of the PCA-based RTM, and (ii) different numbers of regression points per cluster in the case of the CLSR method. Figure 6 shows these residuals for the 'Cloud 2' and the 'Aerosol 2' scenarios.

From Figure 6 the following conclusions can be drawn.

1. The residuals decrease when increasing the number of PCs and regression points.
2. The interquartile range for the CLSR method is substantially reduced when switching from 1–2 to 3 regression points for the $CO_2$ band, and from 3 to 4 regression points for the $O_2$ A-band.
3. In both spectral bands, the interquartile range for the PCA-based RTM decreases gradually with the number of principal components.
4. The residuals in the $O_2$ A-band are systematically higher than those in the $CO_2$ band. This behaviour is more pronounced when the gas optical thickness is large, thus resulting in larger discrepancies for the PCA-based RTM and almost negligible discrepancies for the CLSR approach.
5. In the $O_2$ A-band, the median values of the residuals Equation (8) are higher than 2%. The median values remain almost constant with the number of principal components and regression points. However, the median values as well as the interquartile range for the PCA method are generally higher than those of the CLSR method. This trend remains coherent using the single-scattering

RTM instead of the two-stream RTM, although the residuals are substantially higher for the PCA-based RTM.

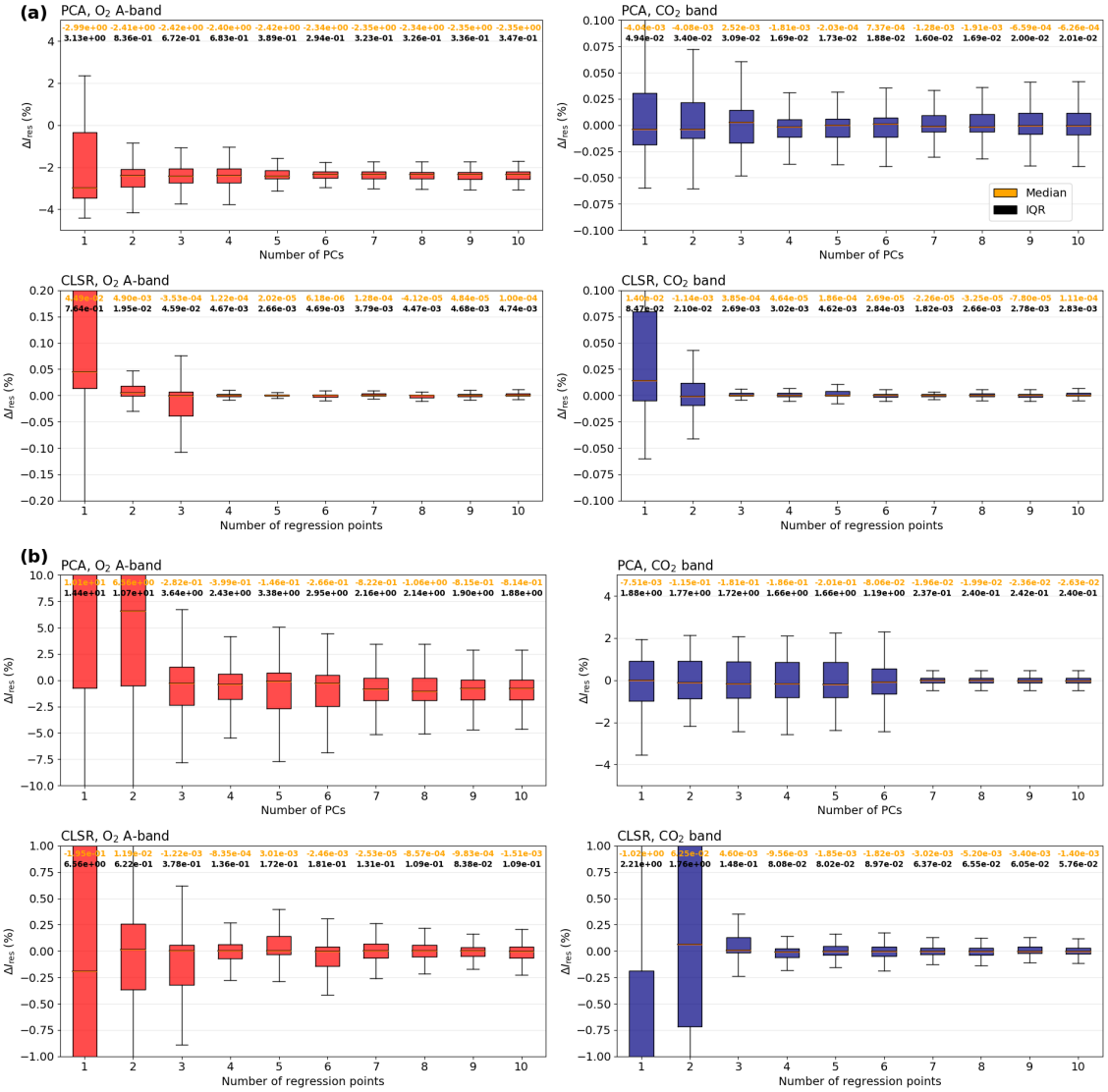


**Figure 6.** Box plots of the residuals with respect to the continuum in percentage for (upper panels) the PCA model and (bottom panels) the CLSR method, when the low-stream model is (**a**) the two-stream model for the 'Cloud 2' scenario or (**b**) the single-scattering model for the 'Aerosol 2' scenario. The red boxes indicate the $O_2$ A-band while the blue boxes indicate the $CO_2$ band. Box description: the upper and lower limits of the box represent the interquartile range (IQR), which is $(Q_3 - Q_1)$ being $Q_i$ the $i$-th quartile; the upper and lower whiskers indicate $(Q_3 + 1.5 \cdot \text{IQR})$ and $(Q_1 - 1.5 \cdot \text{IQR})$, respectively; the orange line inside the box represents the median; the orange values on top of each box indicate the median values and the black values correspond to the IQR value.

### 3.2.2. Estimation of the Required Parameters for the Acceleration Techniques

In order to select the optimal number of PCs, we compute the explained variance ratio of the optical data, which accounts for the variance associated with a given number of PCs. The results in Figure 7 show that in all scenarios almost 99% of the optical data variance can be explained within the first four PCs. Moreover, from Figure 6a we can infer that (i) the convergence of the two-stream

radiances is reached for 4-5 PCs, while (ii) the convergence of the single-scattering radiances requires 4-7 PCs (Figure 6b).

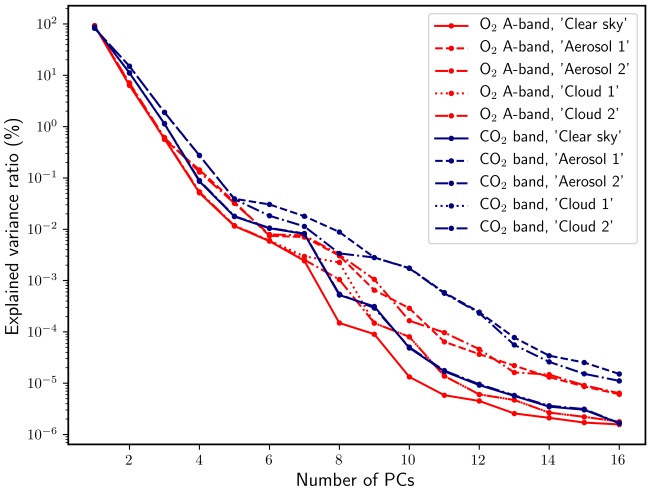


**Figure 7.** Explained variance ratio in percentage as a function of the number of PCs for all the atmospheric scenarios: 'Clear sky', 'Aerosol 1', 'Aerosol 2', 'Cloud 1' and 'Cloud 2'; and the two spectral bands ($O_2$ A- and $CO_2$ bands). The red color corresponds to the $O_2$ A-band and the blue color to the $CO_2$ weak band. The different dashing of the lines indicates the different atmospheric scenarios.

To select the number of clusters and regression points for the CLSR method, we estimate the mean error (cf. Equation (9)). As an example, we illustrate in Figure 8 the mean errors for the 'Aerosol 2' scenario. The results show that 4–5 clusters and 3–5 regression points guarantee a small mean error.

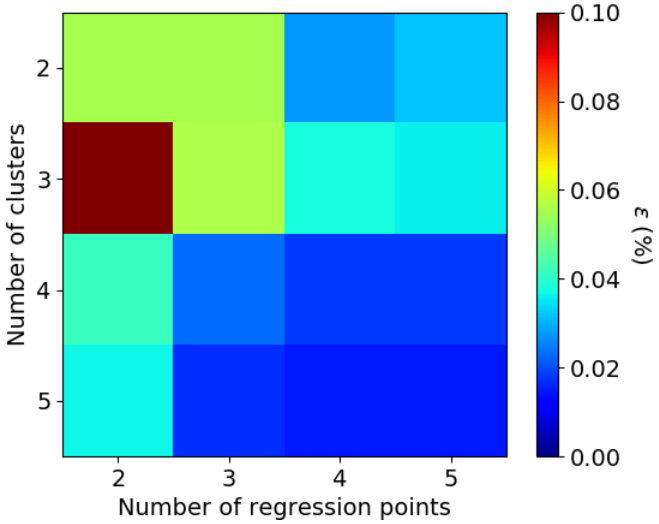


**Figure 8.** Dependence of the number of clusters and number of regression points with the mean error in percentage for the $CO_2$ band for the 'Aerosol 2' scenario. The low-stream model used is the two-stream model.

### 3.2.3. Accuracy of the CLSR and PCA-Based Methods

In this section, we assess the accuracy of the PCA-based and CLSR methods. The residuals (Equation (8)) and interquartile ranges corresponding to the PCA-based and CLSR methods, as well as for the two-stream and single-scattering models, are shown in Figure 9.

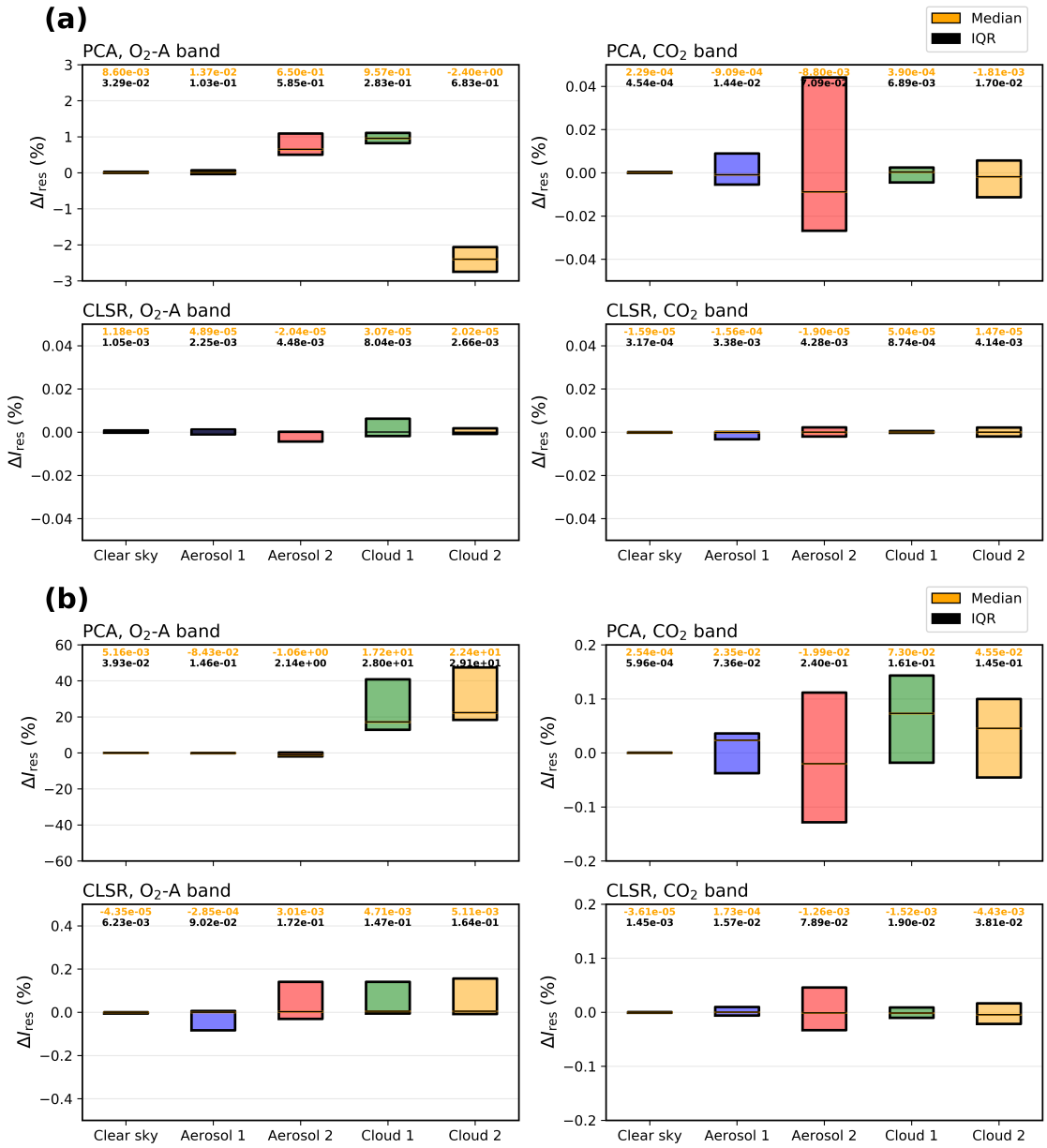


**Figure 9.** Comparison of the residuals for the methods PCA and CLSR for all the atmospheric scenarios (grey: 'Clear sky'; blue: 'Aerosol 1'; red: 'Aerosol 2'; green: 'Cloud 1'; yellow: 'Cloud 2') and gases ($O_2$ A- and $CO_2$ bands), when the low-stream model is (**a**) the two-stream model or (**b**) the single-scattering model. Note the differences in scales for the PCA technique for the $O_2$ A-band with the rest of cases. The orange values on top of each box indicate the median values and the black values correspond to the IQR value.

The results can be summarized as follows:

1. For both low-stream models and cloudy scenarios, (i) the residuals of the PCA-based method in the $O_2$ A-band are substantially larger than those of the CLSR method (on average, for the two-stream model in the $O_2$ A-band, the residuals are above 1% for the PCA-based method and below 0.01% for the CLSR method), while (ii) the PCA-based and CLSR methods yield comparable accuracies in the $CO_2$ band (on average, for the two-stream model in the $CO_2$ band, the residuals are below 0.1% for the PCA-based method and below 0.01% for the CLSR method). Please note similar results were established in previous studies, i.e., the errors of the PCA-based method generally increase with the optical depth [16]. The superiority of the CLSR method is also demonstrated by the interquartile ranges: these are much smaller for the CLSR method.
2. For both $O_2$ A- and $CO_2$ bands, (i) the residuals of the single-scattering model with the PCA-based method are higher than those corresponding to the two-stream model, while (ii) the residuals of the two-stream and single-scattering models with the CLSR method are comparable (on average, for the single-scattering model, the residuals of the CLSR method are below 0.2% in the $O_2$ A-band and below 0.1% in $CO_2$ band, while for the two-stream model the corresponding errors are below 0.01% in both spectral bands).

Thus, (i) the two-stream model with the PCA-based and CLSR methods yields accurate results, (ii) the efficiency of the PCA-based method decreases when increasing the optical thickness, and (iii) both the two-stream and the single-scattering models with the CLSR method provide reasonable accuracies.

Please note the considered atmospheric scenarios implicitly involve a set of assumptions which causes differences between RTM calculations and actual (physically realistic) atmospheric conditions. Furthermore, the discrete ordinate method itself is an approximate technique which in principle should be validated against rigorous analytical methods [46]. However, we use the same atmospheric definition inputs for comparing acceleration techniques.

### *3.3. Acceleration Techniques: Computational Performance*

In this section, we evaluate the computational efficiency of the CLSR method and compare it with the PCA-based method. The speedup factors computed by using Equations (10) and (11) for the PCA-based and CLSR methods, respectively, are given in Table 5. Since the single-scattering model with the PCA-based method provides large errors (see Figure 9b), the table includes the speedup factor $S_{\mathrm{PCA}}$ for the two-steam model with the PCA-based method, as well as for the two-stream and single-scattering models with the CLRS method, $S^{\mathrm{TS}}_{\mathrm{CLSR}}$ and $S^{\mathrm{SS}}_{\mathrm{CLSR}}$, respectively.

**Table 5.** Speedup factor of the PCA-based ($S_{\mathrm{PCA}}$) and CLSR methods with: the two-stream model ($S^{\mathrm{TS}}_{\mathrm{CLSR}}$) and single-scattering model ($S^{\mathrm{SS}}_{\mathrm{CLSR}}$).

| $S_{\mathrm{PCA}}$ | $S^{\mathrm{TS}}_{\mathrm{CLSR}}$ | $S^{\mathrm{SS}}_{\mathrm{CLSR}}$ |
|---|---|---|
| 534 | 505 | 1294 |

According to the optimal values of the PCs and regression points and clusters, the number of calls to the multi-stream model for the PCA-based method is 9, for the two-stream model with the CLSR method 12, and for the single-scattering model with the CLSR method 20. The computational time to perform the PCA calculation $t_{\mathrm{PCA}}$ is 3 orders of magnitude higher than that of the least-squares calculation $t_{\mathrm{LSM}}$. However, since the most time-consuming part is due to the number of calls to the multi-stream model, the efficiencies of the two-stream model with the CLRS and the PCA-based method are comparable. In contrast, the computational performance of the single-scattering model with the CLRS method is much higher than that with the PCA-based method due to the neglect of multiple scattering computations. Comparing the speedup factors of this study with those of other

authors (e.g., [21]), we find that our values are of the order of their speedup factors and one order of magnitude higher when considering the single-scattering model for the CLSR technique.

### *3.4. Computation of Convolved Spectra*

In this section, we examine the efficiency of the CLSR method for computing convolved spectra. The high-resolution spectra in the $O_2$ A-band are convolved with the slit functions corresponding to GOME-2 and TROPOMI instruments, while the radiance spectra in the $CO_2$ band are convolved with the GOSAT slit function. In this paper, slit functions are modeled with a Gaussian function. The corresponding full widths at half maximum (FWHM) are listed in Table 6. The FWHM considered for the $O_2$ A-band are based on pre-launch calibrations [47] and for the $CO_2$ band on [48].

**Table 6.** Spectral ranges and FWHM of the Gaussian slit functions of the instruments used in the study: TROPOMI, GOME-2 and GOSAT.

| Instrument | Spectral Range | FWHM |
|---|---|---|
| TROPOMI | 710–775 nm | 0.183 nm |
| GOME-2 | 590–790 nm | 0.51 nm |
| GOSAT | 1.56–1.69 µm | 0.2 $cm^{-1}$ |

The examples of convolved spectra corresponding to the 'Cloud 1' scenario are shown in Figure 10. The computations are performed by using the PCA-based and the CLSR methods (in conjunction with the two-stream model) as well as the LBL approach. Tables 7 and 8 show the mean relative errors for the PCA-based ($\varepsilon_{PCA}$) and CLSR ($\varepsilon_{CLSR}$) methods for the $O_2$ A-band and the $CO_2$ band, respectively. In addition, the residuals for non-convolved spectra are shown for comparison. For the 'Clear sky' and aerosol scenarios the accuracies of both methods are comparable, while for cloud scenarios the CLSR method is more accurate. Please note the residuals estimated for the convolved spectra are very close to those for the non-convolved ones and the value of residuals according to Equation (8) are robust.

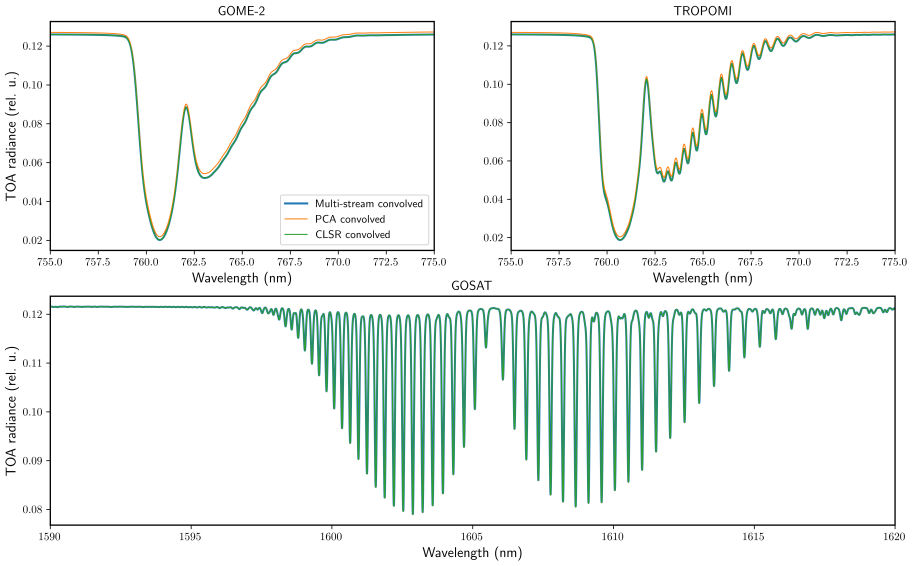


**Figure 10.** Convolved spectra for the multi-stream model and the two acceleration methods for the 'Cloud 1' scenario using PCA and CLSR methods for the sensors GOME-2, TROPOMI and GOSAT. For GOME-2 and TROPOMI the $O_2$ A-band spectra are convolved, while for GOSAT the $CO_2$ spectra are convolved.

**Table 7.** Mean relative error $\varepsilon$ for the convolved spectra compared with the multi-stream spectra for the PCA-based and CLSR methods and for the different atmospheric scenarios considered for the $O_2$ A-band. In all cases, the low-stream model considered is the two-stream model. The instruments analyzed are GOME-2 and TROPOMI, which are compared with the non-convolved values.

| Scenario | $O_2$ A-Band | | | | | |
|---|---|---|---|---|---|---|
| | GOME-2 | | TROPOMI | | Non-Convolved | |
| | $\varepsilon_{PCA}$(%) | $\varepsilon_{CLSR}$(%) | $\varepsilon_{PCA}$(%) | $\varepsilon_{CLSR}$(%) | $\varepsilon_{PCA}$(%) | $\varepsilon_{CLSR}$(%) |
| Clear sky | 0.021 | 0.004 | 0.021 | 0.004 | 0.022 | 0.006 |
| Aerosol 1 | 0.049 | 0.007 | 0.050 | 0.007 | 0.061 | 0.011 |
| Aerosol 2 | 0.837 | 0.019 | 0.837 | 0.019 | 0.856 | 0.026 |
| Cloud 1 | 1.23 | 0.011 | 1.23 | 0.011 | 1.24 | 0.017 |
| Cloud 2 | 2.92 | 0.006 | 2.92 | 0.006 | 2.93 | 0.009 |

**Table 8.** Mean relative error $\varepsilon$ for the convolved spectra compared with the multi-stream spectra for the PCA-based and CLSR methods and for the different atmospheric scenarios considered for the $CO_2$ band. In all cases, the low-stream model considered is the two-stream model. The instrument analyzed is GOSAT, which is compared with the non-convolved values.

| Scenario | $CO_2$ Band | | | |
|---|---|---|---|---|
| | GOSAT | | Non-Convolved | |
| | $\varepsilon_{PCA}$(%) | $\varepsilon_{CLSR}$(%) | $\varepsilon_{PCA}$(%) | $\varepsilon_{CLSR}$(%) |
| Clear sky | 0.0003 | 0.0003 | 0.0006 | 0.0006 |
| Aerosol 1 | 0.0075 | 0.0067 | 0.010 | 0.012 |
| Aerosol 2 | 0.035 | 0.012 | 0.044 | 0.017 |
| Cloud 1 | 0.011 | 0.007 | 0.013 | 0.008 |
| Cloud 2 | 0.011 | 0.005 | 0.016 | 0.006 |

## 4. Conclusions

In this study, we developed the Cluster Low-Streams Regression (CLSR) method for fast radiative transfer simulations of the $O_2$ A- and $CO_2$ absorption bands. The CLSR method exploits a strong close-to-linear relationship between the radiances computed with a low-stream model (which is either the two-stream or the single-scattering model) and the radiances computed with the multi-stream model. The spectral points are grouped in several clusters according to the values of the low-stream radiances. For each cluster, the regression model is established between the low-stream and multi-stream models, where the corresponding regression coefficients are found by using the least-squares method. This approach can be regarded as a variation of the low-streams interpolation method explained in [21], in which the binning is performed in the space of gas optical depths.

The CLRS method was compared with the PCA-based RTM in several atmospheric scenarios. For the 'Clear sky' scenario, the performance of both acceleration methods is comparable. For the cloud scenarios, the CLSR method shows more accurate results than the PCA-based method. However, to improve the performance of the PCA-based method, in [14–16] the optical depth binning technique was applied, i.e., the spectral points were grouped into bins according to the optical depth values and the PCA-based method was applied independently for each bin. Such an approach improves the accuracy of the PCA-based method, although at the cost of increasing the number of multi-stream computations. Comparing the results of [14–16] with the results obtained in our study, we can conclude that the PCA-based binned approaches and the CLSR method are comparable in terms of accuracy.

Our analysis has shown that, although the CLSR method requires more calls to the multi-stream model than the PCA-based model, the corresponding speedup factors are very similar, with slightly better results for the CLSR method. However, this difference is not significant. In addition, we note that the CLSR method does not require to deal with the eigenvalue problem as the PCA-based RTMs do. The CLSR method can be used in conjunction with either the two-stream or the single-scattering

model providing a performance enhancement of almost two orders of magnitude while keeping the maximum error below 0.1% (note that the use of the single-scattering model instead of the two-stream model within the framework of the PCA–based method drastically reduces the accuracy).

As future goals, following [18,49] we plan to extend the CLSR method for computing the Stokes parameters and analyze the possibility of hybrid use of the CLSR method, the PCA-based method and the correlated *k*-distribution technique. It is also planned to analyze different atmospheric scenarios such as cirrus clouds [50] as well as the efficiency of the CLRS method for improving the accuracy of other types of approximate models. In this case, one can expect reducing the number of regression points if instead of the single scattering model, the double scattering approximation is considered [51]. For optically thick media, as an approximate model, asymptotic radiative transfer models can be used [52].

**Author Contributions:** Methodology, A.d.Á. and D.S.E.; Software, A.d.Á. and V.M.G.; Validation, D.S.E. and M.Y.K.; Writing—Original Draft Preparation, A.d.Á. and D.S.E.; Writing—Review and Editing, V.M.G. and M.Y.K.; Visualization, A.d.Á.; Supervision, D.S.E.; Funding Acquisition, D.S.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Aerospace Center (DLR) and the German Academic Exchange Service (DAAD) through the programmes DLR/DAAD Research Fellowships 2018 (57424731) with reference number 91711709 and DLR/DAAD Research Fellowships 2015 (57186656) with reference number 91613528.

**Acknowledgments:** The authors would like to thank Adrian Doicu for his constructive criticism of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| CLSR | Cluster Low-Streams Regression |
| DOME | Discrete Ordinates with Matrix Exponential |
| EOF | Empirical Orthogonal Function |
| FWHM | Full Width at Half Maximum |
| GOME-2 | Global Ozone Monitoring Experiment–2 |
| GOSAT | Greenhouse gases Observing SATellite |
| HITRAN | HIgh-resolution TRANsmission molecular absorption database |
| IQR | InterQuartile Range |
| LBL | Line-By-Line |
| LS | Low-Streams |
| LSM | Least-Squares Method |
| MS | Multi-Streams |
| OPAC | Optical Properties of Aerosols and Clouds |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| Py4CAts | Python for Computational ATmospheric Spectroscopy |
| RTM | Radiative Transfer Model |
| TOA | Top Of the Atmosphere |
| TROPOMI | TROPOspheric Monitoring Instrument |

## References

1. Clough, S.A.; Rinsland, C.P.; Brown, P.D. Retrieval of tropospheric ozone from simulations of nadir spectral radiances as observed from space. *J. Geophys. Res.* **1995**, *100*, 16579. [CrossRef]
2. Ambartsumian, V. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars. *Publ. Astron. Obs. Leningr. State Univ.* **1936**, *6*, 7–18.
3. Goody, R.; West, R.; Chen, L.; Crisp, D. The correlated k-method for radiation calculations in nonhomogeneous atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **1989**, *42*, 539–550. [CrossRef]

4. Fu, Q.; Liou, K. On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *J. Atmos. Sci.* **1992**, *49*, 2139–2156. [CrossRef]
5. Fomin, B.A. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 1. FKDM, fast k-distribution model for the longwave. *J. Geophys. Res.* **2004**, *109*. [CrossRef]
6. Fomin, B.A. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. *J. Geophys. Res.* **2005**, *110*. [CrossRef]
7. Hunt, G.E.; Grant, I.P. Discrete space theory of radiative transfer and its application to problems in planetary atmospheres. *J. Atmos. Sci.* **1969**, *26*, 963–972. [CrossRef]
8. Wiscombe, W.; Evans, J. Exponential-sum fitting of radiative transmission functions. *J. Comput. Phys.* **1977**, *24*, 416–444. [CrossRef]
9. Moncet, J.L.; Uymin, G.; Lipton, A.E.; Snell, H.E. Infrared radiance modeling by optimal spectral sampling. *J. Atmos. Sci.* **2008**, *65*, 3917–3934. [CrossRef]
10. Liu, X.; Smith, W.L.; Zhou, D.K.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Appl. Opt.* **2006**, *45*, 201–208. [CrossRef]
11. Hollstein, A.; Lindstrot, R. Fast reconstruction of hyperspectral radiative transfer simulations by using small spectral subsets: application to the oxygen A band. *Atmos. Meas. Tech.* **2014**, *7*, 599–607. [CrossRef]
12. Natraj, V.; Jiang, X.; Shia, R.; Huang, X.; Margolis, J.; Yung, Y. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the $O_2$ A-band. *J. Quant. Spectrosc. Radiat. Transf.* **2005**, *95*, 539–556. [CrossRef]
13. Efremenko, D.; Doicu, A.; Loyola, D.; Trautmann, T. Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *133*, 128–135. [CrossRef]
14. Kopparla, P.; Natraj, V.; Spurr, R.; Shia, R.L.; Crisp, D.; Yung, Y.L. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *173*, 65–71. [CrossRef]
15. Kopparla, P.; Natraj, V.; Limpasuvan, D.; Spurr, R.; Crisp, D.; Shia, R.L.; Somkuti, P.; Yung, Y.L. PCA-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *198*, 104–111. [CrossRef]
16. Somkuti, P.; Boesch, H.; Natraj, V.; Kopparla, P. Application of a PCA-based fast radiative transfer model to $XCO_2$ retrievals in the shortwave infrared. *J. Geophys. Res. Atmos.* **2017**, *122*, 10477–10496. [CrossRef]
17. del Águila, A.; Efremenko, D.S.; Molina García, V.; Xu, J. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band. *Atmosphere* **2019**, *10*, 142. [CrossRef]
18. Molina García, V.; Sasi, S.; Efremenko, D.S.; Doicu, A.; Loyola, D. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *213*, 228–240. [CrossRef]
19. Efremenko, D.S.; Loyola, D.G.; Doicu, A.; Spurr, R.J.D. Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method. *Comput. Phys. Commun.* **2014**, *185*, 3079–3089. [CrossRef]
20. Amdahl, G.M. Validity of the single processor approach to achieving large scale computing capabilities. In Proceedings of the Spring Joint Computer Conference, Atlantic City, NJ, USA, 18–20 April 1967.
21. O'Dell, C.W. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. *J. Geophys. Res.* **2010**, *115*. [CrossRef]
22. Fischer, J.; Grassl, H. Detection of cloud-top height from backscattered radiances within the Oxygen A band. Part 1: Theoretical study. *J. Appl. Meteorol.* **1991**, *30*, 1245–1259. [CrossRef]
23. Kataev, M.Y.; Lukyanov, A.K. Empirical orthogonal functions and its modification in the task of retrieving of the total amount $CO_2$ and $CH_4$ with help of satellite Fourier transform spectrometer GOSAT (TANSO-FTS). In Proceedings of the 22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, Tomsk, Russia, 30 June–3 July 2016.
24. Munro, R.; Lang, R.; Klaes, D.; Poli, G.; Retscher, C.; Lindstrot, R.; Huckle, R.; Lacan, A.; Grzegorski, M.; Holdak, A.; et al. The GOME-2 instrument on the Metop series of satellites: Instrument design, calibration, and level 1 data processing—An overview. *Atmos. Meas. Tech.* **2016**, *9*, 1279–1301. [CrossRef]

25. Veefkind, J.; Aben, I.; McMullan, K.; Förster, H.; de Vries, J.; Otter, G.; Claas, J.; Eskes, H.; de Haan, J.; Kleipool, Q.; et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **2012**, *120*, 70–83. [CrossRef]
26. Butz, A.; Guerlet, S.; Hasekamp, O.; Schepers, D.; Galli, A.; Aben, I.; Frankenberg, C.; Hartmann, J.M.; Tran, H.; Kuze, A.; et al. Toward accurate $CO_2$ and $CH_4$ observations from GOSAT. *Geophys. Res. Lett.* **2011**, *38*. [CrossRef]
27. Doicu, A.; Trautmann, T. Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. *J. Quant. Spectrosc. Radiat. Transf.* **2009**, *110*, 146–158. [CrossRef]
28. Efremenko, D.; Doicu, A.; Loyola, D.; Trautmann, T. Acceleration techniques for the discrete ordinate method. *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *114*, 73–81. [CrossRef]
29. Efremenko, D.S.; Molina García, V.; Gimeno García, S.; Doicu, A. A review of the matrix-exponential formalism in radiative transfer. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *196*, 17–45. [CrossRef]
30. Korkin, S.; Lyapustin, A. Matrix exponential in C/C$^{++}$ version of vector radiative transfer code IPOL. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *227*, 106–110. [CrossRef]
31. Waterman, P.C. Matrix-exponential description of radiative transfer. *J. Opt. Soc. Am.* **1981**, *71*, 410. [CrossRef]
32. Nakajima, T.; Tanaka, M. Matrix formulations for the transfer of solar radiation in a plane-parallel scattering atmosphere. *J. Quant. Spectrosc. Radiat. Transf.* **1986**, *35*, 13–21. [CrossRef]
33. Afanas'ev, V.P.; Efremenko, D.S.; Lubenchenko, A.V. On the application of the invariant embedding method and the radiative transfer equation codes for surface state analysis. In *Light Scattering Reviews 8: Radiative Transfer and Light Scattering*; Kokhanovsky, A.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 363–423.
34. Schreier, F.; Gimeno García, S.; Hochstaffl, P.; Städt, S. Py4CAtS—PYthon for Computational ATmospheric Spectroscopy. *Atmosphere* **2019**, *10*, 262. [CrossRef]
35. Gordon, I.; Rothman, L.; Hill, C.; Kochanov, R.; Tan, Y.; Bernath, P.; Birk, M.; Boudon, V.; Campargue, A.; Chance, K.; et al. The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *203*, 3–69. [CrossRef]
36. Checa-Garcia, R.; Landgraf, J.; Galli, A.; Hase, F.; Velazco, V.A.; Tran, H.; Boudon, V.; Alkemade, F.; Butz, A. Mapping spectroscopic uncertainties into prospective methane retrieval errors from Sentinel-5 and its precursor. *Atmos. Meas. Tech.* **2015**, *8*, 3617–3629. [CrossRef]
37. Richard, C.; Gordon, I.; Rothman, L.; Abel, M.; Frommhold, L.; Gustafsson, M.; Hartmann, J.M.; Hermans, C.; Lafferty, W.; Orton, G.; et al. New section of the HITRAN database: Collision-induced absorption (CIA). *J. Quant. Spectrosc. Radiat. Transf.* **2012**, *113*, 1276–1285. [CrossRef]
38. Mlawer, E.J.; Payne, V.H.; Moncet, J.L.; Delamere, J.S.; Alvarado, M.J.; Tobin, D.C. Development and recent evaluation of the MT_CKD model of continuum absorption. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2012**, *370*, 2520–2556. [CrossRef]
39. Tran, H.; Boulet, C.; Hartmann, J.M. Line mixing and collision-induced absorption by oxygen in the A band: Laboratory measurements, model, and tools for atmospheric spectra computations. *J. Geophys. Res.* **2006**, *111*. [CrossRef]
40. Bodhaine, B.A.; Wood, N.B.; Dutton, E.G.; Slusser, J.R. On Rayleigh optical depth calculations. *J. Atmos. Ocean. Technol.* **1999**, *16*, 1854–1861. [CrossRef]
41. Anderson, G.; Clough, S.; Kneizys, F.; Chetwynd, J.; Shettle, E. *AFGL Atmospheric Constituent Profiles (0–120 km)*; Air Force Geophysics Laboratory, Hanscom Air Force Base: Bedford, MA, USA, 1986.
42. Hess, M.; Koepke, P.; Schult, I. Optical properties of aerosols and clouds: The software package OPAC. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 831–844. [CrossRef]
43. Deirmendjian, D. *Electromagnetic Scattering on Spherical Polydispersions*; Elsevier: Amsterdam, The Netherlands, 1969.
44. Natraj, V.; Shia, R.L.; Yung, Y.L. On the use of principal component analysis to speed up radiative transfer calculations. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 810–816. [CrossRef]
45. del Águila, A.; Efremenko, D.S.; Trautmann, T. A review of dimensionality reduction techniques for processing hyper-spectral optical signal. *Light Eng.* **2019**, *27*, 85–98, [CrossRef]
46. Rogovtsov, N.N.; Borovik, F. Application of general invariance relations reduction method to solution of radiation transfer problems. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *183*, 128–153. [CrossRef]

47. Beirle, S.; Lampel, J.; Lerot, C.; Sihler, H.; Wagner, T. Parameterizing the instrumental spectral response function and its changes by a super-Gaussian and its derivatives. *Atmos. Meas. Tech.* **2017**, *10*, 581–598. [CrossRef]
48. GOSAT Spectral Resolution. Available online: http://www.gosat-2.nies.go.jp/about/spacecraft_and_instruments/ (accessed on 17 March 2020).
49. Molina García, V.; Sasi, S.; Efremenko, D.S.; Doicu, A.; Loyola, D. Linearized radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *213*, 241–251. [CrossRef]
50. Butz, A.; Galli, A.; Hasekamp, O.; Landgraf, J.; Tol, P.; Aben, I. TROPOMI aboard Sentinel-5 Precursor: Prospective performance of CH4 retrievals for aerosol and cirrus loaded atmospheres. *Remote Sens. Environ.* **2012**, *120*, 267–276. [CrossRef]
51. Korkin, S.V.; Lyapustin, A.I.; Marshak, A.L. On the accuracy of double scattering approximation for atmospheric polarization computations. *J. Quant. Spectrosc. Radiat. Transf.* **2012**, *113*, 172–181. [CrossRef]
52. Kokhanovsky, A.A. Asymptotic radiative transfer. In *Light Scattering Reviews. Springer Praxis Books*; Kokhanovsky, A.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 253–289.

## A.4 Publication IV: The Cluster Low-Streams Regression Method for Fast Computations of Top-of-the-Atmosphere Radiances in Absorption Bands

### Reference

A. del Águila and D. S. Efremenko. The Cluster Low-Streams Regression Method for Fast Computations of Top-of-the-Atmosphere Radiances in Absorption Bands. Proceeding of the 30th International Conference on Computer Graphics and Machine Vision (Graphicon 2020). Part 2, pages short25–1–short25–9, 2020. doi: 10.51130/graphicon-2020-2-4-25

### Copyright



### Abstract

Atmospheric composition sensors provide a huge amount of data. A key component of trace gas retrieval algorithms are radiative transfer models (RTMs), which are used to simulate the spectral radiances in the absorption bands. Accurate RTMs based on line-by-line techniques are time-consuming. In this paper we analyze the efficiency of the cluster low-streams regression (CLSR) technique to accelerate computations in the absorption bands. The idea of the CLRS method is to use the fast two-stream RTM model in conjunction with the line-by-line model and then to refine the results by constructing the regression model between two- and multi-stream RTMs. The CLSR method is applied to the Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands for the clear sky and several aerosol types. The median error of the CLSR method is below 0.001 %, the interquartile range (IQR) is below 0.1 %, while the performance enhancement is two orders of magnitude.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko)

A.d.Á and D.S.E. designed the study together and A.d.Á performed the tests and created the figures and tables. The results were discussed between A.d.Á and D.S.E. Finally, D.S.E. supervised the study and the original draft was written by A.d.Á and D.S.E.

The overall own contribution of A.d.Á for the publication in Appendix A.4. is estimated at 90%, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.4).

**Table A.4:** Criteria and estimated own contribution for Appendix A.4

| Criteria | Estimated contribution |
|---|---|
| Conceptualization | 80% |
| Implementation | 90% |
| Analysis and discussion | 90% |
| Figures and tables | 100% |
| Manuscript writing | 90% |
| **Total** | **90** % |

# The Cluster Low-Streams Regression Method for Fast Computations of Top-of-the-Atmosphere Radiances in Absorption Bands *

Ana del Águila[1 [0000−0001−9006−9631]], Dmitry Efremenko[1 [0000−0002−7449−5072]]

[1]Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany
ana.delaguilaperez@dlr.de, dmitry.efremenko@dlr.de

**Abstract.** Atmospheric composition sensors provide a huge amount of data. A key component of trace gas retrieval algorithms are radiative transfer models (RTMs), which are used to simulate the spectral radiances in the absorption bands. Accurate RTMs based on line-by-line techniques are time-consuming. In this paper we analyze the efficiency of the cluster low-streams regression (CLSR) technique to accelerate computations in the absorption bands. The idea of the CLRS method is to use the fast two-stream RTM model in conjunction with the line-by-line model and then to refine the results by constructing the regression model between two- and multi-stream RTMs. The CLSR method is applied to the Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands for the clear sky and several aerosol types. The median error of the CLSR method is below 0.001 %, the interquartile range (IQR) is below 0.1 %, while the performance enhancement is two orders of magnitude.

**Keywords:** Radiative transfer model, Regression model, Line-by-line model

## 1 Introduction

The information about the atmospheric trace gases can be retrieved from the spectral radiances measured at the top of the atmosphere. The key component of atmospheric retrieval algorithms are the radiative transfer models (RTMs). Accurate simulations in the absorption bands are based on the so called line-by-line (LBL) model [1], which requires thousands of monochromatic RTM computations per absorption band due to strong spectral variability of the absorption coefficient. Alternatives to computationally expensive LBL models are the $k$-distribution method [2, 3] and the principal component analysis (PCA)-based RTMs [4–9], in which the redundancies in hyper-spectral data are eliminated and the spectrum can be computed by using a small number of RTM calls. These methods are reviewed in [10].



* Supported by the German Aerospace Center (DLR) and the German Academic Exchange Service (DAAD) through the programme DLR/DAAD Research Fellowship (57424731).

In our recent work [11], the Cluster Low-Streams Regression (CLSR) method was developed to accelerate hyper-spectral computations. The idea of the CLSR method is to perform LBL computations by using a fast two-stream RTM and then to refine results by using the correlation model for the two-stream and reference multi-stream RTMs. This approach was applied to the $O_2$A-band and the weak $CO_2$ band for different atmospheric scenarios. The results were compared with the PCA-based RTMs showing an improvement over the last in terms of accuracy. Note that the idea of improving accuracy of two-stream models was exploited in numerous theoretical studies (see e.g. [12, 13] and references therein).

In this study, the CLSR method is extended to ozone Hartley-Huggins band and the water vapour band in the ultra-violet and near infrared spectral ranges, respectively. Additionally, the CLSR method is applied to several atmospheric models containing different aerosol types.

## 2 Methodology

### 2.1 Data overview

We consider the computations of the reflected spectral radiances at the top of the atmosphere (TOA) in the Hartley-Huggins, $O_2$A-, water vapour and $CO_2$ bands. Table 1 summarizes the spectral bands examined with their corresponding spectral range, spectral resolution and number of spectral points to be simulated. As a reference RTM, we use the discrete ordinates with matrix exponential (DOME) method [14, 15]. The number of discrete ordinates (streams) in the polar hemisphere $N_{\text{do}}$ regulates the computational performance and accuracy. In the following, the model is called multi-stream (MS) when $N_{\text{do}} \geq 2$ and low-stream (LS) otherwise. Following previous analysis in [16], the multi-stream RTM with $N_{\text{do}} = 32$ discrete ordinates is used as a reference RTM.

The gaseous absorption coefficients for the $O_2$A-, water vapour and $CO_2$ bands are computed with the LBL model Py4CAtS [17], while the ozone absorption cross-sections in the Hartley-Huggins band are taken from the HITRAN 2016 database [18]. Rayleigh scattering is modeled as proposed in [19].

**Table 1.** Spectral ranges, spectral resolution and number of spectral points for the absorption bands used in this study.

| Band | Spectral range (nm) | Spectral resolution (nm) | Number of spectral points |
|---|---|---|---|
| Hartley-Huggins | 280 — 335 | 0.0188 | 2932 |
| $O_2$A | 755 — 775 | 0.0010 | 20000 |
| Water vapour | 770 — 1000 | 0.0058 | 40000 |
| $CO_2$ | 1590 — 1620 | 0.0015 | 20000 |

The atmosphere is discretized into 35 layers with a step of 1 km between 0 and 25 km, and a step of 2.5 km between 25 km and 50 km. For all the simulations, we

assume a Lambertian surface with an albedo of 0.3. The solar zenith angle, the viewing zenith angle and the relative azimuth angle are $45^\circ$, $35^\circ$ and $90^\circ$, respectively.

The atmosphere can contain one of the aerosols from the OPAC database [20], optical properties of which are summarized in Table 2 and Table 3 .

**Table 2.** Aerosol optical thickness (AOD) in the middle of the spectral range for the spectral bands and aerosol types considered.

| Aerosol type | Hartley-Huggins (315 nm) | $O_2A$ (760 nm) | Water vapour (885 nm) | $CO_2$ (1610 nm) |
|---|---|---|---|---|
| Tropospheric | 0.133 | 0.018 | 0.015 | 0.010 |
| Continental clean | 0.42 | 0.20 | 0.17 | 0.08 |
| Urban | 4.28 | 0.46 | 0.35 | 0.16 |
| Desert | 0.71 | 0.20 | 0.20 | 0.19 |
| Continental polluted | 2.4 | 1.2 | 0.9 | 0.4 |

**Table 3.** Single scattering albedo (SSA) for each spectral band and aerosol type.

| Aerosol type | Hartley-Huggins | $O_2A$ | Water vapour | $CO_2$ |
|---|---|---|---|---|
| Tropospheric | 0.950 | 0.947 | 0.942 | 0.936 |
| Continental clean | 0.959 | 0.962 | 0.959 | 0.958 |
| Urban | 0.940 | 0.935 | 0.929 | 0.913 |
| Desert | 0.932 | 0.953 | 0.950 | 0.945 |
| Continental polluted | 0.951 | 0.960 | 0.957 | 0.950 |

Figure 1 shows the radiance spectra computed by using the multi-stream and the low-stream (i.e. two-stream) RTMs for the different absorption bands. The computations are perfomed using the LBL-framework. Note that both spectra have a similar spectral behavior. Hence, it is possible to establish a regression model between the low- and multi-stream radiances, which is a subject to the CLSR method considered further.

### 2.2 Cluster Low-Streams Regression (CLSR) method

The Cluster Low-Streams Regression (CLSR) method is described in detail in [11] and can be summarized as follows:

First, let us consider a LBL spectrum $\{I_{\mathrm{LS}}(\lambda_i)\}_{i=1}^N$ computed at $N$ spectral points $\{\lambda_i\}_{i=1}^N$ by means of a low-stream RTM. We sort the radiances in ascending order and consider $C$ clusters in the sorted radiance set (with $N_C = N/C$ radiance points). Secondly, we select $n$ equidistant radiance points in each cluster and we compute the multi-stream radiances for the corresponding wavelengths $\{\overline{I}^c_{\mathrm{MS},q}\}_{q=1}^n$. Assuming that in each cluster $c$ there is a linear relationship between low- and multi-stream radiances, we get

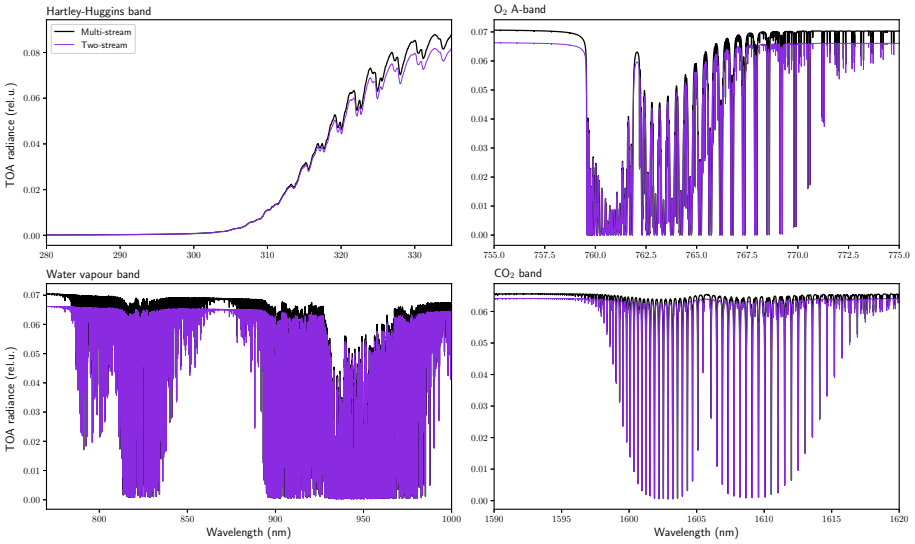


**Fig. 1.** Top of the atmosphere radiances computed for the absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands for one aerosol case. Black lines correspond to the multi-stream RTM while purple lines correspond to the two-stream RTM.

$$\hat{I}^c_{\mathrm{MS},i} = \alpha^c \hat{T}^c_i + \beta^c \hat{I}^c_{\mathrm{LS},i} + \gamma^c,\ i = 1, ..., N_C, \tag{1}$$

where $\alpha^c$, $\beta^c$ and $\gamma^c$ are the regression coefficients of the $c$-th cluster and $\hat{T}$ is the corresponding direct transmittance. Equation (1) can be also written as follows:

$$Y = A \cdot X \tag{2}$$

with $Y = \left[\hat{I}^c_{\mathrm{MS},i}\right]$, $A = [\alpha^c, \beta^c, \gamma^c]$ and $X = \left[\hat{T}^c_i, \hat{I}^c_{\mathrm{LS},i}, 1\right]$. Finally, we find the regression coefficients as a solution to the following least square problem:

$$A = \arg\min_A \sum_{q=1}^{n} \left[\overline{I}^c_{\mathrm{MS},q} - Y\right]^2. \tag{3}$$

In this way, we can restore the spectra of the multi-stream radiances $\{\tilde{I}_{\mathrm{MS},i}\}_{i=1}^N$. Here, the "hat" notation $\hat{I}$ refers to the sorted radiances, the "bar" notation $\overline{I}$ refers to the equidistant radiances entering the regression model and the "tilde" notation $\tilde{I}$ refers to the predicted radiances. The total number of regression points, and thus the number of calls to the multi-stream RTM, is $nC$. Note that unlike the $k$-distribution method, the CLSR method provides a spectrum at the same spectral resolution as the LBL approach.

# 3 Simulations

## 3.1 Simulations of the absorption bands by using the CLSR method for the clear sky atmosphere

In this section we apply the CLSR method to simulate absorption bands. In addition to $O_2$A- and $CO_2$ bands analyzed in [11], we consider Hartley-Huggins and water vapour bands.

To estimate the accuracy of the CLSR method, we compute the residuals, the median and interquartile range (IQR). The residual for the radiance is computed at each spectral point $\lambda_i$ as follows:

$$\Delta I_{\mathrm{res},i} = \frac{\tilde{I}_{\mathrm{MS},i} - I_{\mathrm{MS},i}}{I_{\mathrm{MS}}^{\mathrm{cont}}} \cdot 100, \tag{4}$$

where $\tilde{I}_{\mathrm{MS},i}$ is the radiance calculated with the CLSR method (cf. Eq. (1)), while $I_{\mathrm{MS}}^{\mathrm{cont}}$ is the radiance without absorption (i.e. the continuum radiance, which is used to avoid radiance values close to zero in the denominator of Eq. (4), when strong gas absorption is present [11, 7]).

Figure 2 shows the residuals of the CLSR method for different number of regression points per cluster for the four bands considered.

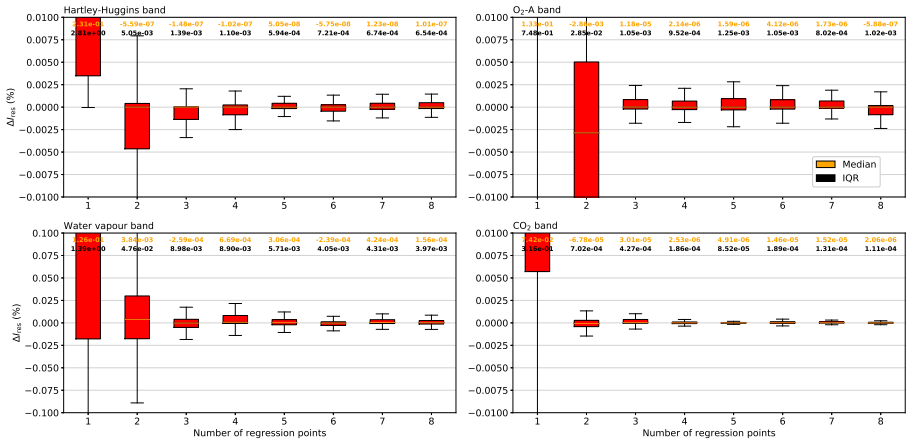


**Fig. 2.** Box plots of the residuals for the CLSR method for the clear sky scenario in the following absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The orange and black values on top of each box indicate the median values and the IQR values, respectively.

The residuals gradually decrease with the number of spectral points. In fact, they are significantly reduced when switching from 1—2 to 3 regression points. Therefore, the median values remain almost constant from 3 regression points. This trend is identical to the one found in [11] for different atmospheric scenarios. Note that the scale of residuals for the water vapour band is one order of magnitude higher than for the Hartley-Huggins, $O_2$A- and $CO_2$ bands.

### 3.2 Application of the CLSR method in the case of aerosols: accuracy results

In our previous work, we applied the CLSR method to the $O_2$A- and $CO_2$ bands for several atmospheric scenarios like aerosols and clouds at different heights and thicknesses. In this paper we examine the application of the CLSR method for several aerosol types and we extend the analysis to the Hartley-Huggins and water vapour bands. The computations are performed for the aerosol types outlined in Section 2.1.

Figure 3 shows the residuals for four bands and 5 aerosol models.

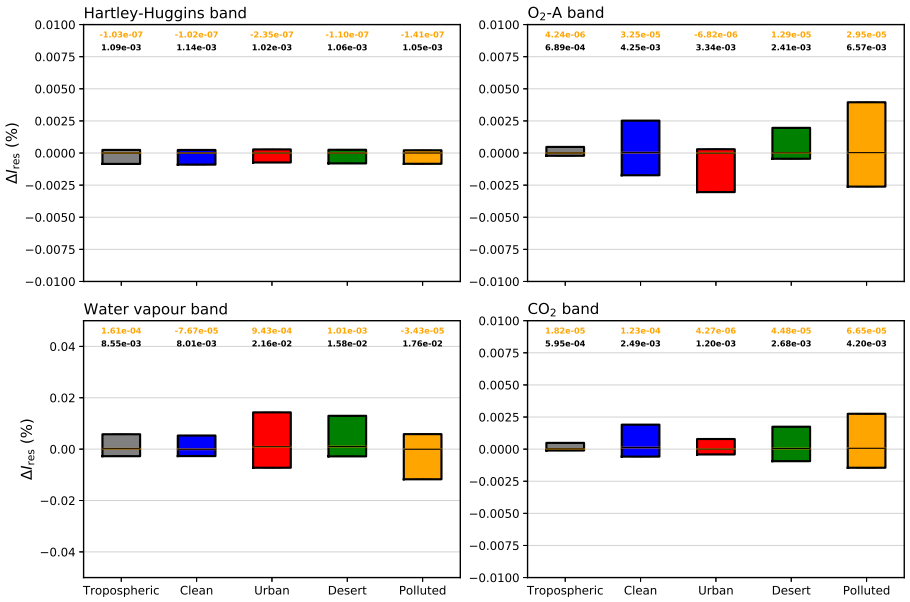


**Fig. 3.** Residuals for the CLSR method for the following absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The color of the boxes represents the type of aerosol: (grey) tropospheric; (blue) continental clean; (red): urban; (green): desert; (yellow): polluted. Note the different scale for the water vapour band compared with the other absorption bands.

Note that the residuals for the Hartley-Huggins, $O_2$A- and $CO_2$ bands are substantially smaller than those for the water vapour band. However, the median residuals are below 0.001 % and the results of the CLSR model are not biased.

In general, we conclude that the efficiency of the CLSR method is comparable to that of alternative methods like PCA-based RTMs (e.g. [7]) and our previous studies ([11]).

### 3.3 Assessment of the CLSR computational efficiency

In this section we analyse the computational performance of the CLSR method. Table 4 shows the number of calls to two- and multi-stream RTMs, the computational time

and the corresponding speedup factor with respect to the multi-stream LBL simulations for the $O_2$A-band. The computational time for monochromatic computations for the TS RTM is $t_{\mathrm{TS}}$ =1.6e-4 s, while for the MS RTM with $N_{\mathrm{do}} = 32$ discrete ordinates per hemisphere is $t_{\mathrm{MS}}$ =0.12 s, i.e., around three order of magnitude larger. However, most of the computational burden is still due to the MS RTM, and the computations of the approximate spectrum with a high spectral resolution by using the TS RTM is not a performance bottleneck in the whole CLSR processing chain. The results show that using the matrix of coefficients $X$ with 5 clusters and 4 regression points for the CLSR method is 420 faster than using the LBL model. The speedup factor is of the same order of magnitude as in [21].

**Table 4.** Summary of number of calls, computational time and speedup factors for the $O_2$A-band with LBL and CLSR methods.

| | LBL | CLSR |
|---|---|---|
| Number of calls to MS RTM | 20000 | 20 |
| Number of calls to TS RTM | 0 | 20000 |
| Time for MS RTM (s) | 2320 | 2.32 |
| Time for TS RTM (s) | 0 | 3.20 |
| Total computational time (s) | 2320 | 5.52 |
| Speedup factor | — | 420 |

## 4 Conclusions

In this study, we have analysed the efficiency of the Cluster Low-Streams Regression (CLSR) method to accelerate spectral computations in several absorption bands. The CLSR method exploits the linear relationship between the low-stream and multi-stream models, where the corresponding regression coefficients are found by using the least-squares method. In our simulations several OPAC aerosol models have been considered. We reproduced the spectra with a median error below 0.001 % as compared to the reference multi-stream line-by-line model and IQR values below 0.1%. Thus, the errors present low variation and stability.

The number of calls to the multi-stream model was reduced by 3 orders of magnitude (e.g. from 20000 to 20 calls in the case of $O_2$A-band). The resulting performance enhancement is about 400 times. Note, that since the CLSR method is two orders of magnitude faster than the LBL model, it can be used for computations of the aerosol spectra in near-real-time applications.

In our future work, we plan to extend the CLSR method by using the asymptotic radiative transfer theory [22] and the diffuse approximation [23] instead of the two-stream RTM. Also it is of high interest to apply the CLSR method for modelling of the Stokes parameters.

## References


1. Clough, S.A., Rinsland, C.P., Brown, P.D.: Retrieval of tropospheric ozone from simulations of nadir spectral radiances as observed from space. Journal of Geophysical Research **100**(D8), 16579 (1995). https://doi.org/10.1029/95jd01388
2. Fomin, B.A.: A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. Journal of Geophysical Research **110**(D2) (2005). https://doi.org/10.1029/2004jd005163
3. Fu, Q., Liou, K.: On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. Journal of the Atmospheric Sciences **49**(22), 2139–2156 (1992). https://doi.org/10.1175/1520-0469(1992)049¡2139:OTCDMF¿2.0.CO;2
4. del Águila, A., Efremenko, D.S., Molina García, V., Xu, J.: Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the hartley-huggins band. Atmosphere **10**(3), 142 (3 2019). https://doi.org/10.3390/atmos10030142
5. Efremenko, D.S., Loyola, D.G., Doicu, A., Spurr, R.J.D.: Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method. Computer Physics Communications **185**(12), 3079–3089 (2014). https://doi.org/10.1016/j.cpc.2014.07.018
6. Efremenko, D., Doicu, A., Loyola, D., Trautmann, T.: 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 (2014). https://doi.org/10.1016/j.jqsrt.2013.07.023
7. Kopparla, P., Natraj, V., Limpasuvan, D., Spurr, R., Crisp, D., Shia, R.L., Somkuti, P., Yung, Y.L.: Pca-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning. Journal of Quantitative Spectroscopy and Radiative Transfer **198**, 104–111 (2017). https://doi.org/https://doi.org/10.1016/j.jqsrt.2017.05.005
8. Liu, X., Smith, W.L., Zhou, D.K., Larar, A.: Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. Applied Optics **45**(1), 201–208 (2006). https://doi.org/10.1364/ao.45.000201
9. Natraj, V., Jiang, X., Shia, R., Huang, X., Margolis, J., Yung, Y.: Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the $O_2$ A-band. Journal of Quantitative Spectroscopy and Radiative Transfer **95**(4), 539–556 (2005). https://doi.org/10.1016/j.jqsrt.2004.12.024
10. del Águila, A., Efremenko, D.S., Trautmann, T.: A review of dimensionality reduction techniques for processing hyper-spectral optical signal. Light & Engineering pp. 85–98 (2019). https://doi.org/10.33383/2019-017
11. del Águila, A., Efremenko, D.S., Molina García, V., Kataev, M.Y.: Cluster low-streams regression method for hyperspectral radiative transfer computations: Cases of O2 A- and CO2 bands. Remote Sensing **12**(8), 1250 (Apr 2020). https://doi.org/10.3390/rs12081250
12. Afanas'ev, V., Basov, A.Y., Budak, V., Efremenko, D., Kokhanovsky, A.: Analysis of the discrete theory of radiative transfer in the coupled "ocean–atmosphere" system: Current status, problems and development prospects. Journal of Marine Science and Engineering **8**(3), 202 (Mar 2020). https://doi.org/10.3390/jmse8030202
13. Budak, V., Efremenko, D., Shagalov, O.: Efficiency of algorithm for solution of vector radiative transfer equation in turbid medium slab. Journal of Physics: Conference Series **369**, 012021 (Jun 2012). https://doi.org/10.1088/1742-6596/369/1/012021
14. Doicu, A., Trautmann, T.: Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. Journal of Quantitative Spectroscopy and Radiative Transfer **110**(1-2), 146–158 (2009). https://doi.org/10.1016/j.jqsrt.2008.09.014

15. Efremenko, D.S., Molina García, V., Gimeno García, S., Doicu, A.: A review of the matrix-exponential formalism in radiative transfer. Journal of Quantitative Spectroscopy and Radiative Transfer **196**, 17–45 (Jul 2017). https://doi.org/10.1016/j.jqsrt.2017.02.015
16. Molina García, V., Sasi, S., Efremenko, D., Doicu, A., Loyola, D.: Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. Journal of Quantitative Spectroscopy and Radiative Transfer **213**, 228–240 (2018). https://doi.org/10.1016/j.jqsrt.2018.03.014
17. Schreier, F., Gimeno García, S., Hochstaffl, P., Städt, S.: Py4cats—PYthon for computational ATmospheric spectroscopy. Atmosphere **10**(5), 262 (2019). https://doi.org/10.3390/atmos10050262
18. Gordon, I., Rothman, L., Hill, C., Kochanov, R., Tan, Y., Bernath, P., Birk, M., Boudon, V., Campargue, A., Chance, K., Drouin, B., Flaud, J.M., Gamache, R., Hodges, J., Jacquemart, D., Perevalov, V., Perrin, A., Shine, K., Smith, M.A., Tennyson, J., Toon, G., Tran, H., Tyuterev, V., Barbe, A., Császár, A., Devi, V., Furtenbacher, T., Harrison, J., Hartmann, J.M., Jolly, A., Johnson, T., Karman, T., Kleiner, I., Kyuberis, A., Loos, J., Lyulin, O., Massie, S., Mikhailenko, S., Moazzen-Ahmadi, N., Müller, H., Naumenko, O., Nikitin, A., Polyansky, O., Rey, M., Rotger, M., Sharpe, S., Sung, K., Starikova, E., Tashkun, S., Auwera, J.V., Wagner, G., Wilzewski, J., Wcisło, P., Yu, S., Zak, E.: The HITRAN2016 molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer **203**, 3–69 (2017). https://doi.org/10.1016/j.jqsrt.2017.06.038
19. Bodhaine, B., Wood, N., Dutton, E., Slusser, J.: On Rayleigh optical depth calculations. Journal of Atmospheric and Oceanic Technology **16**(11), 1854–1861 (1999). https://doi.org/10.1175/1520-0426(1999)016¡1854:orodc¿2.0.co;2
20. Hess, M., Koepke, P., Schult, I.: Optical properties of aerosols and clouds: The software package OPAC. Bulletin of the American Meteorological Society **79**(5), 831–844 (1998). https://doi.org/10.1175/1520-0477(1998)079¡0831:opoaac¿2.0.co;2
21. O'Dell, C.W.: Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. Journal of Geophysical Research **115**(D10) (2010). https://doi.org/10.1029/2009jd012803
22. Kokhanovsky, A.: Cloud Optics. Springer Netherlands (2006). https://doi.org/10.1007/1-4020-4020-2
23. Budak, V.P., Zheltov, V.S., Lubenchenko, A.V., Freidlin, K.S., Shagalov, O.V.: A fast and accurate synthetic iteration-based algorithm for numerical simulation of radiative transfer in a turbid medium. Atmospheric and Oceanic Optics **30**(1), 70–78 (Jan 2017). https://doi.org/10.1134/s1024856017010031

## A.5 Publication V: Accuracy Enhancement of the Two-Stream Radiative Transfer Model for Computing Absorption Bands at the Presence of Aerosols

### Reference

A. del Águila and D. S. Efremenko. Accuracy Enhancement of the Two-Stream Radiative Transfer Model for Computing Absorption Bands at the Presence of Aerosols. Light and Engineering, 29, 2, 79-86, 2021. doi: 10.33383/2020-078

### Copyright



### Abstract

The two-stream model is the fastest radiative transfer model among those based on the discrete ordinate method. Although its accuracy is not high enough to be used in applications, the two-stream model gets more attention in computationally demanding tasks such as line-by-line simulations in the gaseous absorption bands. For this reason, we designed the cluster low-streams regression (CLSR) technique, in which a spectrum computed with a two-stream model, is refined by using statistical dependencies between two- and multi-stream radiative transfer models. In this paper, we examine the efficiency of this approach for computing Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands at the presence of aerosols. The numerical results evidence that the errors of the CLSR method is not biased and around 0.05 %, while the performance enhancement is two orders of magnitude.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko)

A.d.Á and D.S.E. designed the study together and A.d.Á performed the tests and created the figures and tables. The results were discussed between A.d.Á and D.S.E. Finally, D.S.E. supervised the study and the original draft was written by A.d.Á and D.S.E.

The overall own contribution of A.d.Á for the publication in Appendix A.5. is estimated at 90%, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.5).

**Table A.5:** Criteria and estimated own contribution for Appendix A.5

| Criteria | Estimated contribution |
|---|---|
| Conceptualization | 80% |
| Implementation | 90% |
| Analysis and discussion | 90% |
| Figures and tables | 100% |
| Manuscript writing | 90% |
| **Total** | **90** % |

*Light & Engineering*
*Vol. 29, No. 2, pp. 79–86, 2021*

*https://doi.org/10.33383/2020-078*
*Svetotekhnika # 2, 2021, pp. 44–49*

# ACCURACY ENHANCEMENT OF THE TWO-STREAM RADIATIVE TRANSFER MODEL FOR COMPUTING ABSORPTION BANDS AT THE PRESENCE OF AEROSOLS

Ana del Águila and Dmitry S. Efremenko

*Remote Sensing Technology Institute, German Aerospace Centre (DLR), Oberpfaffenhofen, Germany*
*E-mails: ana.delaguilaperez@dlr.de, dmitry.efremenko@dlr.de*

## ABSTRACT

The two-stream model is the fastest radiative transfer model among those based on the discrete ordinate method. Although its accuracy is not high enough to be used in applications, the two-stream model gets more attention in computationally demanding tasks such as line-by-line simulations in the gaseous absorption bands. For this reason, we designed the cluster low-streams regression (CLSR) technique, in which a spectrum computed with a two-stream model, is refined by using statistical dependencies between two- and multi-stream radiative transfer models. In this paper, we examine the efficiency of this approach for computing Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands at the presence of aerosols. The numerical results evidence that the errors of the CLSR method is not biased and around 0.05 %, while the performance enhancement is two orders of magnitude.

**Keywords:** cluster low-streams regression method, hyperspectral data, regression model, radiative transfer, aerosol models, line-by-line model

## 1. INTRODUCTION

Optical sensors with high spectral resolution are attracting increasing attention in atmospheric science. The measurements of the spectral radiances in the absorption bands carry information about trace gas concentrations, and often, about their profiles. As a matter of fact, due to the high spectral resolution, it is possible not only to retrieve this information, but also to filter out the influence of aerosols, clouds, and ground surface on the retrieval result (e.g. [1, 2, 3]).

The radiance scattered by the terrestrial atmosphere can be computed by means of the radiative transfer models (RTMs). The fastest RTM among those based on the discrete ordinate method is the two-stream model, in which the radiances averaged over the polar hemispheres are considered [4]. However, the error of the two-stream model may reach tens of percent [5] which exceed the acceptable error level in atmospheric remote sensing applications [6]. Note that several techniques were developed to improve the accuracy of two-stream model (see e.g. [7, 8] and references therein).

Accurate simulations in the absorption bands are based on the so called line-by-line (LBL) model [9], which requires thousands of monochromatic RTM computations per absorption band due to strong spectral variability of the absorption coefficient. Alternatives to computationally expensive LBL models are the *k*-distribution method [10, 11] originally proposed by Ambartsumyan [12], the k-binning approach [13] and the principal component analysis (PCA)-based RTMs [14, 15, 16, 17, 18]. In the latter case, the two-stream model is used to compute a spectrum in a LBL manner and then a correction function is estimated in a reduced basis of optical parameters. These methods are reviewed in [19, 20]. Note that the *k*-distribu-

**Table 1. Spectral Ranges, Spectral Resolution and Number of Spectral Points for the Absorption Bands Used in this Study**

| Band | Spectral range, nm | Spectral resolution, nm | Number of spectral points |
|---|---|---|---|
| Hartley-Huggins | 280–335 | 0.1800 | 300 |
| $O_2$ A | 755–775 | 0.0010 | 20000 |
| Water vapour | 770–1000 | 0.0058 | 40000 |
| $CO_2$ | 1590–1620 | 0.0015 | 20000 |

**Table 2. AOD in the Middle of the Spectral Range for the Spectral Bands and Aerosol Types Considered**

| Aerosol type | Hartley-Huggins (315 nm) | $O_2$ A (760 nm) | Water vapour (885 nm) | $CO_2$ (1610 nm) |
|---|---|---|---|---|
| Tropospheric | 0.133 | 0.018 | 0.015 | 0.010 |
| Continental clean | 0.42 | 0.20 | 0.17 | 0.08 |
| Urban | 4.28 | 0.46 | 0.35 | 0.16 |
| Desert | 0.71 | 0.20 | 0.20 | 0.19 |
| Continental polluted | 2.4 | 1.2 | 0.9 | 0.4 |

tion method is suitable for computing a spectrum convolved with an instrument slit function.

In our recent work [21], the Cluster Low-Streams Regression (CLSR) method was developed to accelerate hyper-spectral computations. The idea of the CLSR method is to perform LBL computations by using a fast two-stream RTM and then to refine results by using the correlation model for the two-stream and reference multi-stream RTMs. This approach was applied to the $O_2$ A-band and the weak $CO_2$ band.

In this paper, the CLSR method is extended to ozone Hartley-Huggins band and the water vapour band in the ultra-violet and near infrared spectral ranges, respectively. Additionally, the CLSR method is applied to several atmospheric models containing different aerosol models.

## 2. METHODOLOGY

### 2.1 Data Overview

The reflected spectral radiances at the top of the atmosphere (TOA) are considered for this study. We have focused our study in the ultra-violet region with the Hartley-Huggins band, and in the infrared region with the $O_2$ A-, water vapour and $CO_2$ bands. Table 1 summarizes the spectral bands examined with their corresponding spectral range, spectral resolution and number of spectral points to be simulated. As a reference RTM, we use the discrete ordinates with matrix exponential (DOME) method [22, 23]. The number of discrete ordinates (streams) in the polar hemisphere $N_{do}$ regulates the computational performance and accuracy. In the following, the model is called multi-stream (MS) when $N_{do} \geq 2$ and low-stream (LS) otherwise. In particular, the case $N_{do} = 1$ corresponds to the two-stream (TS) model. Also note that the computational time increases with $N_{do}$ increase, and hence, the two-stream model presents the lowest computational time. Following previous analysis in [24], the multi-stream RTM with $N_{do} = 32$ discrete ordinates is used as a reference RTM.

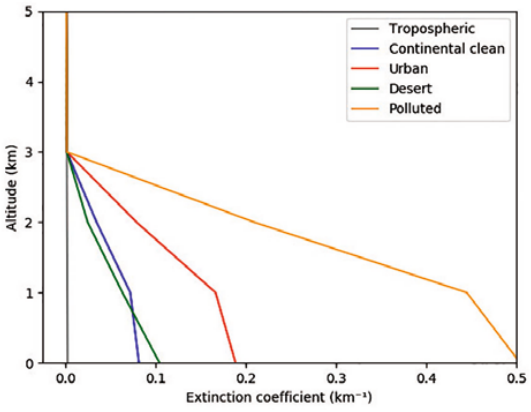


Fig. 1. Extinction coefficient profile for the $O_2$ A-band and all aerosol models

**Table 3. SSA for Each Spectral Band and Aerosol Type Considered**

| Aerosol type | Hartley-Huggins | $O_2$ A | Water vapour | $CO_2$ |
|---|---|---|---|---|
| Tropospheric | 0.950 | 0.947 | 0.942 | 0.936 |
| Continental clean | 0.959 | 0.962 | 0.959 | 0.958 |
| Urban | 0.940 | 0.935 | 0.929 | 0.913 |
| Desert | 0.932 | 0.953 | 0.950 | 0.945 |
| Continental polluted | 0.951 | 0.960 | 0.957 | 0.950 |

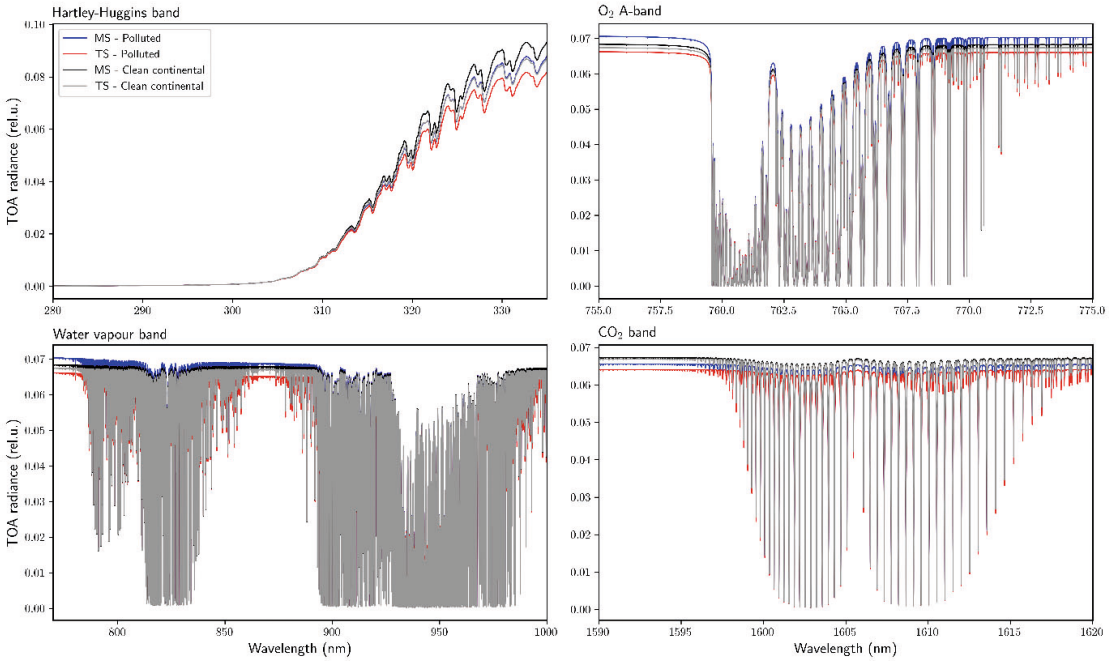


Fig. 2. Top-of-the-atmosphere (TOA) radiances computed for the absorption bands: Hartley- Huggins, $O_2$ A-, water vapour and $CO_2$ bands for two aerosol cases: (blue) polluted aerosol and (red) clean continental aerosol; solid lines correspond to the multi-stream (MS) RTM while dashed lines correspond to the two-stream (TS) RTM

The gaseous absorption coefficients for the $O_2$ A-, water vapour and $CO_2$ bands are computed with the LBL model Py4CAtS [25], while the ozone absorption cross-sections in the Hartley-Huggins band are taken from the HITRAN2016 database [26]. Rayleigh scattering is treated as in [27].

The atmosphere is discretized into 35 layers with a step of 1 km between 0 and 25 km, and a step of 2.5 km between 25 km and 50 km. For all the simulations, we assume a Lambertian surface with an albedo of 0.3. The solar zenith angle, the viewing zenith angle and the relative azimuth angle are 45°, 35°, and 9°, respectively.

For modelling the aerosol properties, the OPAC database [28] is used. The aerosol optical thickness (AOD) and the single scattering albedo (SSA) and are summarized in Table 2 and Table 3, respectively.

Fig. 1 provides an example of the extinction coefficients profile for the $O_2$ A-band and all aerosol models considered in this study. The chosen extinction profiles are in agreement with recent data provided by TROPOMI [29].

Fig. 2 shows the radiance spectra of two aerosol models computed by using the multi-stream (MS, solid lines) and the two-stream (TS, dashed lines) RTMs for the different absorption bands. The computations are performed using the LBL-framework. Note that both pairs of spectra have a similar spectral behaviour. It allows us to establish a regression model between the low- and multi-stream radiances, which is a subject to the CLSR method considered further.

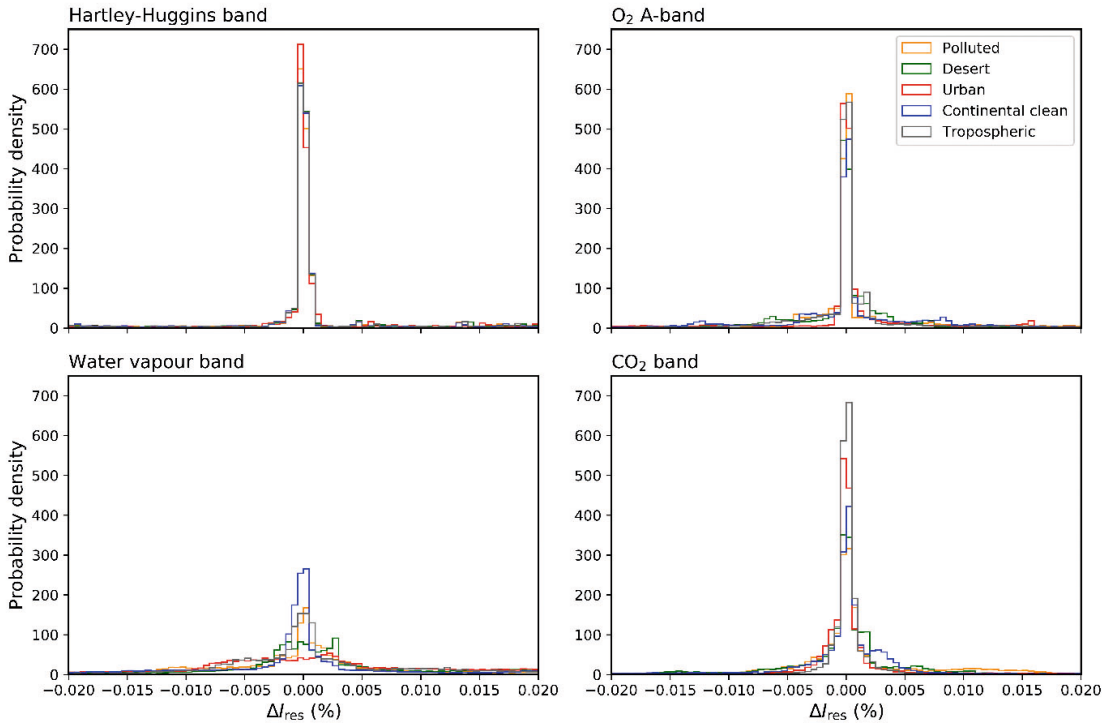


Fig. 3. Probability density of the residuals for the CLSR method for the following absorption bands: Hartley- Huggins, $O_2$ A-, water vapour and $CO_2$ bands (colour of the lines represents the type of aerosol: grey – tropospheric; blue – continental clean; red – urban; green – desert; yellow – polluted)

### 2.2. Cluster Low-Streams Regression (CLSR) Method

The theory of the Cluster Low-Streams Regression (CLSR) method can be found in [21]. The method can be described as follows. Let us consider a low-stream RTM spectrum $I_{LS}(\lambda_i)$ for $i = 1,...,N$ $\lambda_i$ spectral points. First, we sort the radiances in ascending order and consider $C$ clusters in the sorted radiance set (with $N_C = N / C$ radiance points). Secondly, we select $n$ equidistant radiance points in each cluster in the sorted space and we compute the multi-stream radiances for the corresponding wavelengths $\overline{I}^c_{MS,q}$ for $q = 1,...,n$. Therefore, assuming that there is a linear relationship between the low- and multi-stream radiances in each $c$ cluster, we obtain the following regression equation:

$$\hat{I}^c_{MS,i} = \alpha^c \hat{T}^c_i + \beta^c \hat{I}^c_{LS,i} + \gamma^c, i = 1,...,N_C, \qquad (1)$$

where $\alpha^c$, $\beta^c$, and $\gamma^c$ are the regression coefficients of the $c$-th cluster, $\hat{T}^c_i$ is the direct transmittance for the corresponding cluster, which is defined as $\hat{T} = \exp(-\tau)$, where $\tau$ is the total optical thickness. Equation (1) can be written in a matrix form:

$$Y = A \cdot X, \qquad (2)$$

with $Y = \left[\hat{I}^c_{MS,i}\right]$, $A = \left[\alpha^c, \beta^c, \gamma^c\right]$, and $X = [\hat{T}^c_i, \hat{I}^c_{LS,i}, 1]$. We find the regression coefficients as a solution of the least-squares problem:

$$A = \arg\min \sum_{q=1}^{n} \left[\overline{I}^c_{MS,q} - Y\right]^2. \qquad (3)$$

Hence, we can restore the spectra of the multi-stream radiances $\tilde{I}_{MS,i}$ for $i = 1,...,N$. The "hat" notation $\hat{I}$ refers to the sorted radiances, the "bar" notation $\overline{I}$ refers to the equidistant radiances entering the regression model and the "tilde" notation $\tilde{I}$ refers to the predicted radiances. The total number of regression points, and thus the number of calls to the multi-stream RTM, is $nC$. Note that unlike the $k$-distribution method, the CLSR method provides a spectrum at the same spectral resolution as the LBL approach.

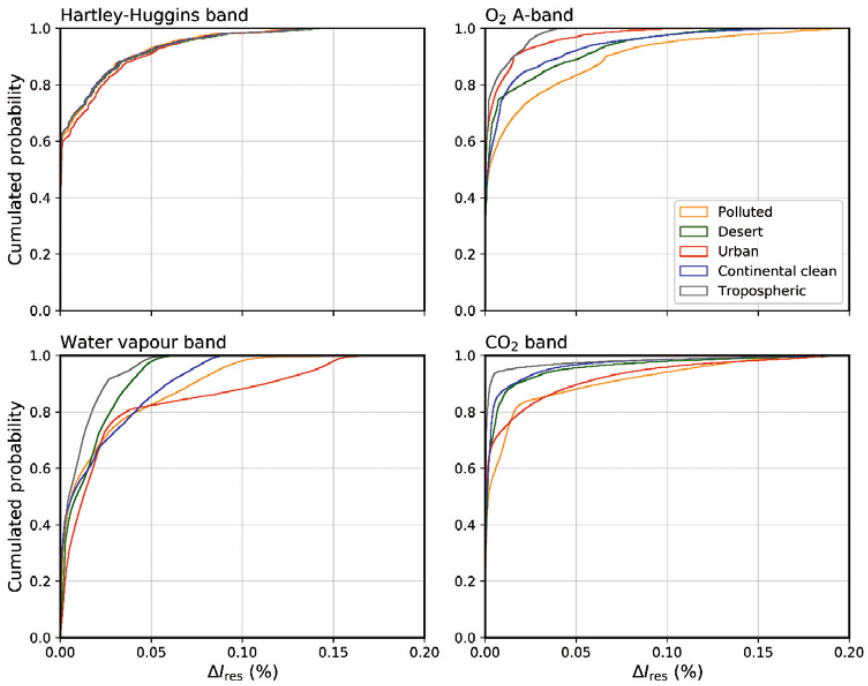


Fig. 4. Cumulative probability distribution of the residuals for the CLSR method for the following absorption bands: Hartley-Huggins, $O_2$ A-, water vapour and $CO_2$ bands (colour of the lines represents the type of aerosol: grey – tropospheric; blue continental clean; red – urban; green – desert; yellow – polluted)

## 3. SIMULATIONS

### 3.1. Accuracy of the Two-Stream Model Equipped with the CLSR Method

In this section we apply the two-stream model equipped with the CLSR method to simulate the spectra in Hartley-Huggins, $O_2$ A- and $CO_2$, and water vapour bands at the presence of aerosols. To estimate the accuracy of the method, the residuals, the probability density and cumulative probability distribution are computed. The residual for the radiance is estimated at each spectral point $\lambda_i$ as follows:

$$\Delta I_{res} = \frac{I_{CLSR} - I_{LBL}}{I_{LBL}^{cont}} \cdot 100, \qquad (4)$$

where $I_{CLSR}$ is the radiance calculated with the CLSR method (Eq. (1)) for every spectral point, while $I_{LBL}$ and $I_{LBL}^{cont}$ are, respectively, the LBL radiance with and without absorption (i.e. the continuum radiance, which is used to avoid radiance values close to zero in the denominator of Eq. (4), when strong gas absorption is present [18, 21]).

Fig. 3 shows the probability density of residuals for four bands and 5 aerosol models. Note that most of the highest probabilities of the residuals are found below 0.005 % for the Hartley-Huggins, $O_2$ A- and $CO_2$ bands. However, the probability density of the water vapour band shows a wider distribution than for the other bands. Therefore, with the information obtained from the probability density functions, we conclude that the residuals of the CLSR method are not biased.

Fig. 4 shows the cumulative probability distribution for the four bands and 5 aerosol models. The main conclusions that can be drawn from the figure are presented below.

- For the Hartley-Huggins band and for all aerosol types, the 90 % of the residuals are below 0.05 %.
- For the $O_2$ A-band, the 90 % of the probability for the urban and tropospheric aerosol are below 0.02 % while for the other aerosols, the residuals are higher.
- For the $CO_2$ band and the tropospheric aerosol, most of their residuals are below 0.01 %. Regarding the continental clean and desert aerosols, 90 % of the residuals are below 0.03 % and for urban and polluted aerosols are below 0.05 %. The reason these latter aerosols present higher errors might be that they have higher AODs and this influences their values. This also occurs for the water vapour band and those specific aerosols.
- Similarly to the other spectral bands, more than 80 % of the residuals in the water vapour band

**Table 4. Number of Calls to the MS and TS RTMs and Speedup Factors for the Hartley- Huggins, $O_2$ A-, water vapour, and $CO_2$ Bands with LBL and CLSR Methods**

| | Hartley-Huggins | | $O_2$ A- and $CO_2$ | | Water vapour | |
|---|---|---|---|---|---|---|
| | LBL | CLSR | LBL | CLSR | LBL | CLSR |
| Calls to MS RTM | 300 | 20 | 20000 | 20 | 40000 | 20 |
| Calls to TS RTM | — | 300 | — | 20000 | — | 40000 |
| Speedup factor | — | 15 | — | 1000 | — | 2000 |

present the accuracy below 0.05 % for all aerosols. However, their influence on the water vapour band is more pronounced.

In general, polluted and urban aerosols for the $O_2$A, water vapour and $CO_2$ bands are the ones providing higher residuals, and this is due to the higher AOD for those aerosol types. To sum up, the 90 % of the cumulated probability for all spectral bands is below the accuracy of 0.05 %, which is in agreement with recent results for the same region (e.g. [30]). The accuracy of the two-stream model equipped with the CLSR method is comparable to that of alternative methods like PCA-based RTMs (e.g. [18]) and our previous studies, [21].

### 3.2 Assessment of the CLSR Computational Efficiency

In this section, we analyse the computational performance of the CLSR method in comparison with the LBL model. Table 4 shows the number of calls to TS and MS RTMs and the speedup factors with respect to the multi-stream LBL simulations for the Hartley-Huggins, $O_2$ A-, water vapour, and $CO_2$ bands.

We have used the vector of coefficients $X$ with 5 clusters and 4 regression points for the CLSR method for all the spectral bands. The same number of CLSR method computations, applied to the different absorption bands with different number of spectral points, provides different speedup factors. For instance, the speedup factor for the Hartley-Huggins band is around 15, while for the $O_2$ A- and $CO_2$ bands the speedup is of 1000. Presumably, the water vapour band has more spectral points in the LBL RTM so that with 20 calls to the MS RTM, we obtain a much superior speed of 2000 compared with the LBL. The speedup factors obtained in this study are of the same order of magnitude for other studies as for the Hartley-Huggins band (e.g. [15]), the $O_2$ A- and $CO_2$ band (e.g. [30, 31]).

## 4. CONCLUSIONS

It has been shown that in the case of line-by-line (LBL) simulations, the accuracy of the two-stream model can be enhanced by using the CLSR technique, which exploits the linear relationship between the two-stream (TS) and multi-stream (MS) models, while the corresponding regression coefficients are found by using the least-squares method. The efficiency of this approach has been tested for computing the top-of-the-atmosphere radiances in Hartley-Huggins, $O_2$ A-, water vapour, and $CO_2$ bands at the presence of aerosols.

Five OPAC aerosol models with different optical properties have been considered in simulations. The error of the computations is generally below 0.05 % and robust with varying aerosol properties.

The number of calls to the time-consuming multi-stream model has been reduced by 1–3 orders of magnitude, depending on the spectral band. For example, for the $O_2$ A- and $CO_2$ bands, the number of MS RTM calls is reduced from 20000 calls to 20 calls with respect to the LBL model. The resulting performance enhancement is about two-orders of magnitude, and the two-stream model in conjunction with the CLSR method can be used for computations of the aerosol spectra in near-real-time applications or in aerosol retrieval algorithms, which take into account the uncertainty in aerosol model selection [32].

Finally, we note that the similar idea based on the CLSR technique can be applied to other approximate RTMs (e.g. based on the asymptotic radiative transfer theory [33] and the diffuse approximation [34]), which will be the topic of our future research.

## ACKNOWLEDGMENTS

This research was funded by the German Aerospace Centre (DLR) and the German Academic Exchange Service (DAAD) through the programme

DLR/DAAD Research Fellowship 2018 (57424731) granted to Ana del Águila with reference number 91711709.

## REFERENCES

1. Dubovik O., Li Z., Mishchenko M.I., Tanrй D., Karol Y., Bojkov B., Cairns B., Diner D.J., Espinosa W.R., Goloub P., Gu X., Hasekamp O., Hong J., Hou W., Knobelspiesse K.D., Landgraf J., Li L., Litvinov P., Liu Y., Lopatin A., Marbach T., Maring H., Martins V., Meijer Y., Milinevsky G., Mukai S., Parol F., Qiao Y., Remer L., Rietjens J., Sano I., Stammes P., Stamnes S., Sun X., Tabary P., Travis L.D., Waquet F., Xu F., Yan C., Yin D. Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives// Journal of Quantitative Spectroscopy and Radiative Transfer, 2019, # 224, pp. 474–511. doi:10.1016/j.jqsrt.2018.11.024, https://doi.org/10.1016/j.jqsrt.2018.11.024.

2. Chan K.L., Valks P., Slijkhuis S., Kuhler C., Loyola D. Total column water vapour retrieval for global ozone monitoring experience-2 (GOME-2) visible blue observations//Atmospheric Measurement Techniques, 2020, 13(8), pp. 4169–4193. doi:10.5194/amt-13–4169–2020, https://doi.org/10.5194/amt-13–4169–2020

3. Zeng Z.C., Chen S., Natraj V., Le T., Xu F., Merrelli A., Crisp D., Sander S.P., Yung Y.L. Constraining the vertical distribution of coastal dust aerosol using $CO_2$ and $O_2$ A-band measurements// Remote Sensing of Environment, 2020, 236, 111494. doi:10.1016/j.rse.2019.111494, https://doi.org/10.1016/j. rse.2019.111494.

4. Schuster A. Radiation through a foggy atmosphere// The Astrophysical Journal 1905, 21, 1. doi:10.1086/141186, https://doi.org/10.1086/141186

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

6. Barker H.W., Cole J.N.S., Li J., Yi B., Yang P. Estimation of errors in two-stream approximations of the solar radiative transfer equation for cloudy-sky conditions// Journal of the Atmospheric Sciences, 2015, Vol. 72, #11, pp. 4053–4074. doi:10.1175/jas-d-15–0033.1, https://doi.org/10.1175/jas-d-15–0033.1.

7. Afanas'ev V., Basov A.Y., Budak V., Efremenko D., Kokhanovsky A. Analysis of the discrete theory of radiative transfer in the coupled "ocean–atmosphere" system: Current status, problems and development prospects// Journal of Marine Science and Engineering, 2020, Vol. 8, #3, 202. doi:10.3390/jmse8030202.

8. Budak V.P., Zheltov V.S., Lubenchenko A.V., Freidlin K.S., Shagalov O.V. A fast and accurate synthetic iteration-based algorithm for numerical simulation of radiative transfer in a turbid medium// Atmospheric and Oceanic Optics, 2017, Vol. 30, #1, pp. 70–78. doi:10.1134/s1024856017010031.

9. Clough S.A., Rinsland C.P., Brown P.D. Retrieval of tropospheric ozone from simulations of nadir spectral radiances as observed from space// Journal of Geophysical Research, 1995, 100 (D8), 16579. doi:10.1029/95jd01388

10. Fomin B.A. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave// Journal of Geophysical Research, 2005, 110 (D2). doi:10.1029/2004jd005163.

11. Fu Q., Liou K. On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres// Journal of the Atmospheric Sciences, 1992, Vol. 49, # 22, pp. 2139–2156. doi:10.1175/1520–0469(1992)049<2139: OTCDMF>2.0.CO;2

12. Ambartzumyan V. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars// Publ. Obs. Astron. Univ. Leningrad 1936, 6, pp. 7–18.

13. Boesche E., Stammes P., Preusker R., Bennartz R., Knap W., Fischer J. Polarization of skylight in the $O_2$ A band: effects of aerosol properties// Applied Optics, 2008, 47 (19), 3467. doi:10.1364/ao.47.003467.

14. Natraj V., Jiang X., Shia R., Huang X., Margolis J., Yung Y. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the $O_2$ A-band// Journal of Quantitative Spectroscopy and Radiative Transfer, 2005, Vol. 95, #4, pp. 539–556. doi:10.1016/j.jqsrt.2004.12.024

15. del Águila A., Efremenko D.S., Molina Garcia V., Xu J. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band// Atmosphere, 2019, Vol. 10, #3, 142. doi:10.3390/atmos10030142.

16. Efremenko D.S., Loyola D.G., Doicu A., Spurr R.J.D. Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method// Computer Physics Communications, 2014, Vol. 185, #12, pp. 3079–3089. doi:10.1016/j.cpc.2014.07.018

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

18. Kopparla P., Natraj V., Limpasuvan D., Spurr R., Crisp D., Shia R.L., Somkuti P., Yung Y.L. PCA-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning// Journal of Quantitative Spectroscopy and Radiative Transfer, 2017, Vol. 198, pp. 104–111. doi: https://doi.org/10.1016/j.jqsrt.2017.05.005.

19. Natraj V. A review of fast radiative transfer techniques// Light Scattering Reviews, 2013, Vol. 8, pp. 475–504. Springer Berlin Heidelberg. doi: https://doi.org/10.1007/978–3–642–32106–1_10.

20. del Águila A., Efremenko D.S., Trautmann T. A review of dimensionality reduction techniques for processing hyperspectral optical signal// Light & Engineering, 2019, Vol. 27, #3, pp. 85–98. doi:10.33383/2019–017

21. del Águila A., Efremenko D.S., Molina Garcia V., Kataev M.Y. Cluster low-streams regression method for hyperspectral radiative transfer computations: Cases of $O_2$ A- and $CO_2$ bands// Remote Sensing, 2020, 12(8), 1250. doi:10.3390/rs12081250.

22. Doicu A., Trautmann T. Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. Journal of Quantitative Spectroscopy and Radiative Transfer, 2009, Vol. 110, #1–2, pp. 146–158. doi:10.1016/j.jqsrt.2008.09.014.

23. Efremenko D.S., Molina Garcia V., Gimeno Garcia S., Doicu A. A review of the matrix-exponential formalism in radiative transfer// Journal of Quantitative Spectroscopy and Radiative Transfer, 2017, 196, pp. 17–45. doi:10.1016/j.jqsrt.2017.02.015.

24. Molina Garcia V., Sasi S., Efremenko D., Doicu A., Loyola D. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements// Journal of Quantitative Spectroscopy and Radiative Transfer, 2018, 213, pp. 228–240. doi:10.1016/j.jqsrt.2018.03.014.

25. Schreier F., Gimeno Garcia S., Hochstaffl P., Städt S. Py4cats – PYthon for computational atmospheric spectroscopy// Atmosphere, 2019, Vol. 10, #5, 262. doi:10.3390/atmos10050262.

26. Gordon I., Rothman L., Hill C., Kochanov R., Tan Y., Bernath P., Birk M., Boudon V., Campargue A., Chance K., Drouin B., Flaud J.M., Gamache R., Hodges J., Jacquemart D., Perevalov V., Perrin A., Shine K., Smith M.A., Tennyson J., Toon G., Tran H., Tyuterev V., Barbe A., Császár, A., Devi V., Furtenbacher T., Harrison J., Hartmann J.M., Jolly A., Johnson T., Karman T., Kleiner I., Kyuberis A., Loos J., Lyulin O., Massie S., Mikhailenko S., Moazzen-Ahmadi N., Müller H., Naumenko O., Nikitin A., Polyansky O., Rey M., Rotger M., Sharpe S., Sung K., Starikova E., Tashkun S., Auwera J.V., Wagner G., Wilzewski J., Wcisło P., Zak E. The HITRAN2016 molecular spectroscopic database// Journal of Quantitative Spectroscopy and Radiative Transfer, 2017, 203, pp. 3–69. doi:10.1016/j.jqsrt.2017.06.038

27. Bodhaine B., Wood N., Dutton E., Slusser J. On Rayleigh optical depth calculations// Journal of Atmospheric and Oceanic Technology, 1999, Vol. 16, #11, pp. 1854–1861. doi:10.1175/1520–0426(1999)016<1854: orodc>2.0.co;2.

28. Hess M., Koepke P., Schult I. Optical properties of aerosols and clouds: The software package OPAC// Bulletin of the American Meteorological Society, 1998, Vol. 79**,** # 5, pp. 831–844 (1998). doi:10.1175/1520–0477(1998)079<0831: opoaac>2.0.co;2.

29. Nanda S., de Graaf M., Veefkind J.P., Sneep M., ter Linden M., Sun J., Levelt P.F. A first comparison of TROPOMI aerosol layer height (ALH) to CALIOP data// Atmospheric Measurement Techniques, Jun. 2020, Vol. 13, # 6, pp. 3043–3059. doi:10.5194/amt-13–3043–2020, https://doi.org/10.5194/amt-13–3043–2020.

30. Liu C., Yao B., Natraj V., Kopparla P., Weng F., Le T., Shia R.L., Yung Y.L. A spectral data compression (SDCOMP) radiative transfer model for high-spectral-resolution radiation simulations// Journal of the Atmospheric Sciences, 2020, Vol. 77**,** # 6, pp. 2055–2066. doi:10.1175/jas-d-19–0238.1, https://doi.org/10. 1175/jas-d-19–0238.1

31. O'Dell C.W. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation// Journal of Geophysical Research, 2010, 115 (D10). doi:10.1029/2009jd012803.

32. Sasi S., Natraj V., Garcia V.M., Efremenko D.S., Loyola D., Doicu A. Model selection in atmospheric remote sensing with application to aerosol retrieval from DSCOVR/EPIC, part 2: Numerical analysis// Remote Sensing, 2020, 12 (21), 3656. doi:10.3390/rs12213656, https://doi.org/10.3390/rs12213656.

33. Kokhanovsky A. Cloud Optics// Springer Netherlands, 2006. doi:10.1007/1–4020–4020–2.

34. Budak V.P., Zheltov V.S., Lubenchenko A.V., Freidlin K.S., Shagalov O.V. A fast and accurate synthetic iteration-based algorithm for numerical simulation of radiative transfer in a turbid medium// Atmospheric and Oceanic Optics, 2017, Vol. 30**,** # 1, pp. 70–78. doi:10.1134/s1024856017010031.

***Ana del Águila***
graduated in Physics from the Granada University (UGR) in 2015. From 2016–2018 she worked as an early-stage researcher at the National Institute for Aerospace Technology (INTA) in Spain.
At present, she is doing the Ph.D. in the German Aerospace Centre (DLR) with a DAAD/DLR scholarship. Her scientific interests are in-situ atmospheric aerosols, lidar systems, remote sensing, radiate transfer and Big Data analysis

***Dmitry S. Efremenko***
graduated from the Moscow Power Engineering institute (MPEI) in 2009. He received his Ph.D. degree from the Moscow State University in 2011 and the habilitation degree from MPEI in 2017.
Since 2011 he works as a research scientist at the German Aerospace Centre (DLR). His scientific interests include radiate transfer, remote sensing, and Big Data analysis

## A.6 Publication VI: Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method

### Reference

A. del Águila and D. S. Efremenko. Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method. Remote Sensing, 13(3):434, 2021. doi: 10.3390/rs13030434

### Copyright



### Abstract

Fast radiative transfer models (RTMs) are required to process a great amount of satellite-based atmospheric composition data. Specifically designed acceleration techniques can be incorporated in RTMs to simulate the reflected radiances with a fine spectral resolution, avoiding time-consuming computations on a fine resolution grid. In particular, in the cluster low-streams regression (CLSR) method, the computations on a fine resolution grid are performed by using the fast two-stream RTM, and then the spectra are corrected by using regression models between the two-stream and multi-stream RTMs. The performance enhancement due to such a scheme can be of about two orders of magnitude. In this paper, we consider a modification of the CLSR method (which is referred to as the double CLSR method), in which the single-scattering approximation is used for the computations on a fine resolution grid, while the two-stream spectra are computed by using the regression model between the two-stream RTM and the single-scattering approximation. Once the two-stream spectra are known, the CLSR method is applied the second time to restore the multi-stream spectra. Through a numerical analysis, it is shown that the double CLSR method yields an acceleration factor of about three orders of magnitude as compared to the reference multi-stream fine-resolution computations. The error of such an approach is below 0.05%. In addition, it is analysed how the CLSR method can be adopted for efficient computations for atmospheric scenarios containing aerosols. In particular, it is discussed how the precomputed data for clear sky conditions can be reused for computing the aerosol spectra in the framework of the CLSR method. The simulations are performed for the Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ weak absorption bands and five aerosol models from the optical properties of aerosols and clouds (OPAC) database.

### Contribution

(A.d.Á: Ana del Águila; D.S.E: Dmitry S. Efremenko)
As stated in the publication, the author contributions are the following: the conceptualization, methodology, investigation, implementation of the tests and data visualization were done by A.d.Á. The analysis and supervision of the results were done by D.S.E. as well as the funding acquisition. The original manuscript was written by A.d.Á with the review and editing of D.S.E.

The overall own contribution of A.d.Á for the publication in Appendix A.6. is estimated at 94%, which is the average value of the percentage values estimated for the criteria listed in the table below (Table A.6).

**Table A.6:** Criteria and estimated own contribution for Appendix A.5

| **Criteria** | **Estimated contribution** |
|---|---|
| Conceptualization | 95% |
| Implementation | 95% |
| Analysis and discussion | 90% |
| Figures and tables | 100% |
| Manuscript writing | 90% |
| **Total** | **94** % |

*remote sensing*

MDPI

*Article*

# Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method

**Ana del Águila * and Dmitry S. Efremenko**

Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany; dmitry.efremenko@dlr.de

* Correspondence: Ana.delAguilaPerez@dlr.de; Tel.: +49-8153-28-1983

**Abstract:** Fast radiative transfer models (RTMs) are required to process a great amount of satellite-based atmospheric composition data. Specifically designed acceleration techniques can be incorporated in RTMs to simulate the reflected radiances with a fine spectral resolution, avoiding time-consuming computations on a fine resolution grid. In particular, in the cluster low-streams regression (CLSR) method, the computations on a fine resolution grid are performed by using the fast two-stream RTM, and then the spectra are corrected by using regression models between the two-stream and multi-stream RTMs. The performance enhancement due to such a scheme can be of about two orders of magnitude. In this paper, we consider a modification of the CLSR method (which is referred to as the double CLSR method), in which the single-scattering approximation is used for the computations on a fine resolution grid, while the two-stream spectra are computed by using the regression model between the two-stream RTM and the single-scattering approximation. Once the two-stream spectra are known, the CLSR method is applied the second time to restore the multi-stream spectra. Through a numerical analysis, it is shown that the double CLSR method yields an acceleration factor of about three orders of magnitude as compared to the reference multi-stream fine-resolution computations. The error of such an approach is below 0.05%. In addition, it is analysed how the CLSR method can be adopted for efficient computations for atmospheric scenarios containing aerosols. In particular, it is discussed how the precomputed data for clear sky conditions can be reused for computing the aerosol spectra in the framework of the CLSR method. The simulations are performed for the Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ weak absorption bands and five aerosol models from the optical properties of aerosols and clouds (OPAC) database.

**Keywords:** hyper-spectral resolution; fast radiative transfer model; gaseous absorption; low-streams regression; aerosols



**Citation:** del Águila, A.; Efremenko, D.S. Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method. *Remote Sens.* **2021**, *13*, 434. https://doi.org/10.3390/rs13030434

Academic Editor: Thomas Ruhtz
Received: 28 December 2020
Accepted: 23 January 2021
Published: 27 January 2021



## 1. Introduction

The operational processing of remote sensing data requires fast radiative transfer models (RTMs), which simulate the radiance field scattered in the atmosphere. The high spectral resolution simulation in the gaseous absorption bands is a demanding task. As the gaseous absorption coefficient changes rapidly with wavelength, the accurate computations based on a fine resolution grid may require thousands of calls to monochromatic RTMs. To accelerate these computations, several techniques have been developed over the years. For instance, the correlated-*k* models [1–4] take into account that the mean radiance across a spectral range depends more on the distribution of the absorption coefficient than on its variation in the spectral range. The state-of-the-art acceleration techniques are based on predicting the spectrum by using a fast two-stream RTM (instead of relatively more time consuming multi-stream RTMs) and then refining the result by introducing a correction function. The latter can be estimated in the original basis of optical parameters, as in the low-stream interpolation (LSI) method [5], or in the reduced basis, as in some principal component analysis (PCA)-based RTMs [6–13]. Recently, we introduced the cluster low-streams regression (CLSR) method, in which such a correction function is found in the

space of radiances computed with a two-stream RTM [14]. This approach was applied to the simulations in the $O_2$A and the weak $CO_2$ bands for different atmospheric scenarios, including aerosols and clouds. It was shown that the error of the CLSR method did not exceed 0.1%, while providing the speedup factor of about two orders of magnitude compared to the line-by-line (LBL) model. Here, we refer to LBL computations as the computations of the radiance field in a fine resolution grid. Therefore, to alleviate the computational burden of the fine resolution radiances, a reduction in the number of radiative transfer simulations is performed. Significant time can be saved by using the LBL model to precompute hyperspectral radiances in different atmospheric conditions, to store them in look-up-tables (LUT) for further use in the development of retrieval algorithms. In this sense, LUT are implemented for monochromatic radiances while other applications precompute the monochromatic transmittances or absorption cross sections (e.g., [15]).

Additional performance can be achieved by combining several acceleration techniques or by using an acceleration method twice. For example, in the double *k*-approach, the integration is performed across the total absorption optical depth and the absorption optical depth from the top of the atmosphere to the scattering layer [16]. The dimensionality reduction scheme based on PCA can be utilized twice, first, to the atmospheric optical properties and then to the radiance datasets. Such a technique has been applied for simulations in the UV range [12] and for the solar spectral range (the spectral data compression (SDCOMP) method [17]). In Molina García et al. [11], the correlated-*k* method was used in conjunction with the PCA-based RTM. A similar approach but for three dimensional computations was applied in Doicu et al. [18]. In Kopparla et al. [19], the PCA-based RTM was combined with a sort of spectral binning. Since the performance bottleneck of the CLSR method is the two-stream RTM used for the LBL computations, our intention is to examine the effect of the double application of the CLSR method.

In this paper, two modifications of the CLSR method are proposed. In the first modification, the two-stream spectra are computed by using the CLSR method and the single-scattering RTM as an approximate RTM. Thus, the CLSR method is applied twice. Such a scheme is referred to as double CLSR. In the second modification, the influence of the aerosols is modelled as a perturbation of the clear sky spectra. In this regard, a spectrum for actual aerosol conditions is estimated from a spectrum computed for clear sky conditions in the framework of the CLSR method.

The rest of the paper is organized as follows. In Section 2, we briefly outline the CLSR method and describe the double CLSR method, as well as an improvement for the aerosol scenarios. In Section 3, we present the results of the simulations and comparisons between the single CLSR and double CLSR methods in terms of accuracy and computation time. The obtained acceleration factors are compared with the state-of-the-art acceleration techniques. In addition, the efficiency of the improved aerosol treatment in the single and double CLSR method is analysed.

## 2. Methodology

### *2.1. Data Overview*

To check the efficiency of the proposed modifications of the CLSR method, we consider high spectral resolution computations of the top-of-the-atmosphere (TOA) radiances in the Hartley–Huggins (315 nm), $O_2$ A-band (760 nm), water vapour (885 nm) and $CO_2$ band (1610 nm), as in our previous work [12,14]. The information about absorption bands is summarized in Table 1. The corresponding absorption coefficients in the $O_2$ A-, water vapour and $CO_2$ bands are computed with the LBL model Py4CAtS [20], while the ozone absorption cross-sections in the Hartley–Huggins band are taken from the HITRAN 2016 database [21]. Hence, the absorption coefficients are pre-computed and stored. The Rayleigh scattering coefficients are modelled as proposed in [22]. The atmosphere assumed for the tests is the mid-latitude summer reference atmospheric model profile [23]. In our simulations, the atmosphere is discretized with a step of 1 km between 0 and 25 km, and a step of 2.5 km between 25 km and 50 km, resulting in 35 layers. The chosen altitude

grid is used for the purpose of comparing the computational speed of different models. The boundary conditions at the bottom are defined by the Lambertian surface with an albedo of 0.3. The solar zenith angle, the viewing zenith angle and the relative azimuth angle are 45°, 35° and 90°, respectively.

**Table 1.** Spectral ranges, resolutions and number of spectral points for the absorption bands used in this study.

| Band | Spectral Range (nm) | Spectral Resolution (nm) | Number of Spectral Points |
|---|---|---|---|
| Hartley–Huggins | 280–335 | 0.18 | 300 |
| $O_2$ A | 755–775 | 0.0010 | 20,000 |
| Water vapour | 770–1000 | 0.0058 | 40,000 |
| $CO_2$ | 1590–1620 | 0.0015 | 20,000 |

The radiative transfer solver used in the study is based on the discrete ordinates with matrix exponential (DOME) method [24,25]. The number of discrete ordinates per hemisphere ($N_{\mathrm{do}}$), often referred to as streams, controls the accuracy and performance of the computations. The RTM is called multi-stream when $N_{\mathrm{do}} \geq 2$. For the specific case of $N_{\mathrm{do}} = 1$, the model is called two-stream. In the single-scattering approximation, the multiple scattering term of the radiative transfer equation is neglected and the solution can be derived analytically without using the discrete ordinate method. In calculations involving aerosol single-scattering phase functions, the delta-M scaling [26] and the TMS-correction [27] procedures are applied. The boundary value problem for the multilayer atmosphere is solved by using the matrix operator method [28], which merges layers into a single layer. The radiance along a viewing direction is computed by using the false discrete ordinate method [8,29].

The optical properties of aerosols are computed by using the optical properties of aerosols and clouds (OPAC) database [30]. The following aerosol types are considered: tropospheric, continental clean, urban, desert and continental polluted. For this study, clouds are not taken into account. The values of the aerosol optical depth (AOD), single scattering albedo (SSA) and asymmetry factor (g) are summarized in Figure 1. For this study, the atmospheric composition affects the accuracy results for the different aerosol types. Therefore, we have included several types of aerosols with different optical properties in order to test as many cases as possible.

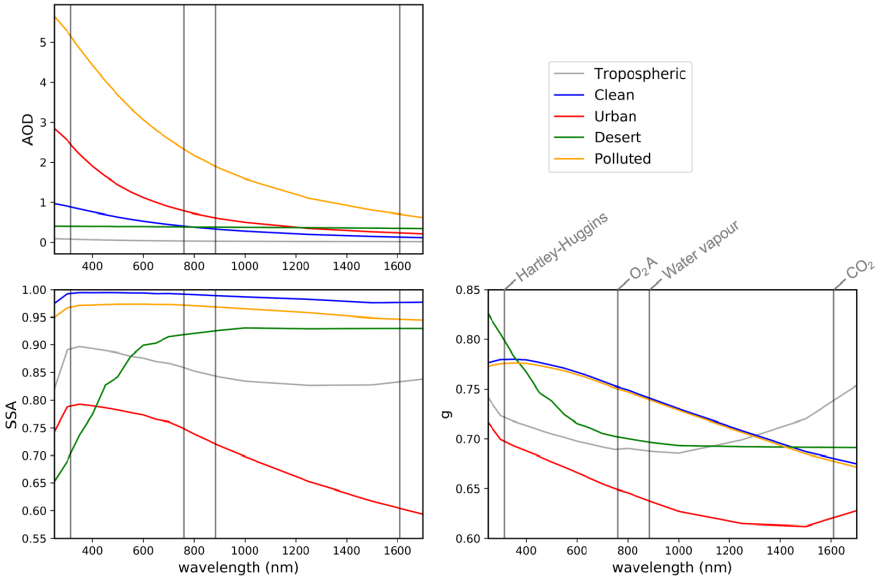


**Figure 1.** Overview of the optical properties (AOD, SSA and g) obtained from the OPAC database for the aerosol types: tropospheric, clean continental, urban, desert and polluted. Each vertical line corresponds with the middle wavelength of the spectral bands in order: Hartley–Huggins, $O_2$ A, water vapour and $CO_2$ bands.

### *2.2. Acceleration Techniques*

#### 2.2.1. Summary of the Cluster Low-Streams Regression (CLSR) Method

The idea of the CLSR method is to obtain the spectrum corresponding to the reference RTM by using the approximate RTM and the regression model between approximate and reference RTMs [14]. The method can be summarized as follows:

We consider a LBL spectrum $\{I_{\mathrm{TS}}(\lambda_i)\}_{i=1}^{N}$ computed at $N$ spectral points $\{\lambda_i\}_{i=1}^{N}$ by means of the two-stream (TS) RTM. Then, the radiances are sorted in ascending order, obtaining the set $\{\hat{I}_{\mathrm{TS}}(\lambda_i)\}_{i=1}^{N}$, which is split into $C$ clusters. In each cluster containing $N_C = N/C$ radiance points, we select $n$ equidistant radiance points, and for the corresponding wavelengths, we compute the radiances $\{\hat{I}^c_{\mathrm{MS},q}\}_{q=1}^{n}$ by using the multi-stream (MS) RTM. Assuming a regression model between TS and MS radiances within each cluster $c$, we obtain

$$\hat{I}^c_{\mathrm{MS},i} = \alpha^c \hat{T}^c_i + \beta^c \hat{I}^c_{\mathrm{TS},i} + \gamma^c, \; i = 1, ..., N_C, \tag{1}$$

where $\alpha^c$, $\beta^c$ and $\gamma^c$ are the regression coefficients of the $c$-th cluster and $\hat{T}^c_i$ is the corresponding direct transmittance ($T = \exp(-\tau^{\mathrm{aer}})$ with $\tau^{\mathrm{aer}}$ being the total AOD). Equation (1) can also be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{X}, \tag{2}$$

where

$$\begin{cases} \mathbf{Y} = \left[\hat{I}^c_{\mathrm{MS},i}\right], \\ \mathbf{A} = [\alpha^c, \beta^c, \gamma^c], \\ \mathbf{X} = \left[\hat{T}^c_i, \hat{I}^c_{\mathrm{TS},i}, 1\right]. \end{cases} \tag{3}$$

The regression coefficients are found as a solution of the following least squares problem:

$$\mathbf{A} = \arg\min_{\mathbf{A}} \sum_{q=1}^{n} \left[ \bar{I}^{c}_{\mathrm{MS},q} - \mathbf{Y} \right]^2. \quad (4)$$

Finally, the MS radiances $\{\tilde{I}_{\mathrm{MS},i}\}_{i=1}^{N}$ are restored at full spectral resolution using the TS radiances and the regression coefficients. The number of calls to the MS RTM equals to $nC$. Note that the TS spectra are computed in a LBL framework, imposing a performance bottleneck.

2.2.2. Double Cluster Low-Streams Regression Method

In the double CLSR method, the TS spectra are computed also by applying the CLSR technique. In this case, as an approximate model, we use the single-scattering (SS) RTM. Thus, the algorithm can be described as follows:

**Step 1**: We compute the LBL spectra $\{I_{\mathrm{SS}}(\lambda_i)\}_{i=1}^{N}$ by using the SS RTM and apply sorting and clustering to the space of SS radiances. Assuming a regression model between SS and TS radiances within each cluster $z$, we obtain

$$\mathrm{Step\,1}: \hat{I}^{z}_{\mathrm{TS},i} = a^z \hat{T}^{z}_{i} + b^z \hat{I}^{z}_{\mathrm{SS},i} + d^z,\ i = 1, ..., N_Z, \quad (5)$$

with number of radiance points $N_Z = N/Z$ and cluster index $z$. The regression coefficients $[a^z, b^z, d^z]$ are found as a solution of the following least squares problem

$$\mathbf{A} = \arg\min_{\mathbf{A}} \sum_{q=1}^{n} \left[ \bar{I}^{z}_{\mathrm{TS},q} - \mathbf{Y} \right]^2, \quad (6)$$

where $\mathbf{A} = [a^z, b^z, d^z]$ and $\mathbf{Y} = \left[ \hat{I}^{z}_{\mathrm{TS},i} \right]$. By knowing the regression coefficients, the TS spectra $\{\tilde{I}_{\mathrm{TS},i}\}_{i=1}^{N}$ can be restored from $\{I_{\mathrm{SS}}(\lambda_i)\}_{i=1}^{N}$ at high spectral resolution.

**Step 2**: We apply the CLSR method as described in the previous section (Section 2.2.1) using the TS spectra computed in Step 1.

Figure 2 shows a schematic representation of the CLSR and double CLSR methods. Note that the TS spectra derived at Step 1 by using the CLSR method differ from those computed by the TS RTM in a LBL manner. However, the possible bias obtained in the double CLSR method at Step 1 is removed by the regression model at Step 2.

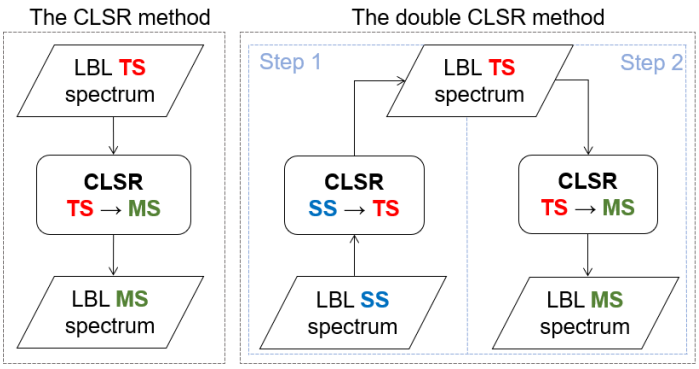


**Figure 2.** Scheme of the CLSR method vs. the double CLSR method.

*2.3. CLSR Method: Improvement to Aerosol Scheme*

It is important to include aerosols in the simulations, using a numerically efficient RTM to quantify the impact of aerosol scattering [31]. To compute the spectrum in the case of the atmosphere with aerosol, the CLSR and double CLSR methods can be applied. For this case, the regression model (Equation (2)) is used, in which

$$\mathbf{Y} = [I_{\mathrm{MS}}^{\mathrm{aer}}] \tag{7}$$

and the original **X**- matrix is substituted by

$$\mathbf{X}_0 = [T, I_{\mathrm{TS}}^{\mathrm{aer}}, 1], \tag{8}$$

where the upper index 'aer' explicitly says that the computations are performed for the aerosol case. However, as shown in Figure 3, the spectra for the clear sky atmosphere (i.e., Rayleigh atmosphere) with and without aerosols have a similar spectral behaviour, i.e., the aerosol spectra depend almost linearly on the clear sky spectra. In this regard, alternative formulations of the CLSR method can be considered. For instance, taking the **X**- matrix as

$$\mathbf{X}_1 = \left[T, I_{\mathrm{MS}}^{\mathrm{clear}}, 1\right], \tag{9}$$

we obtain a method, which converts the MS clear sky spectra into spectra corresponding to the aerosol conditions. The upper index 'clear' indicates that the computations are performed for the clear sky case. The possible benefit of such a scheme is that the clear sky spectra can be precomputed and stored in LUTs and perform the computations offline, while the computations for actual aerosol properties can be performed online.

Alternatively, we consider the **X**- matrix in the following form:

$$\mathbf{X}_2 = \left[T, I_{\mathrm{TS}}^{\mathrm{clear}} - I_{\mathrm{TS}}^{\mathrm{aer}}, I_{\mathrm{MS}}^{\mathrm{clear}}, 1\right]. \tag{10}$$

In this case, the regression model is supplied with the first order perturbation computed by using the TS RTM. As a matter of fact, in this case, we do not expect performance enhancement compared to the $\mathbf{X}_0$-scheme. The question is if the error can be reduced by involving precomputed LUTs for clear sky cases as compared to the original $\mathbf{X}_0$-scheme. Note that for all these cases, the MS RTM for the aerosol scenarios is called for a few spectral points.

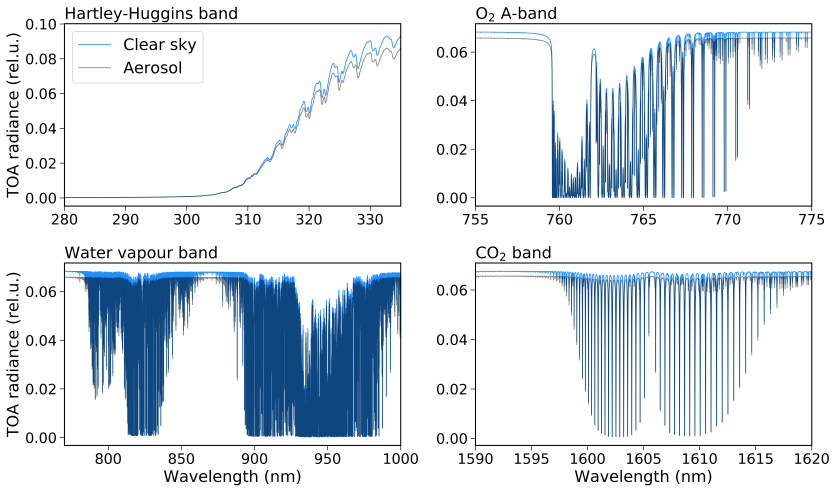


**Figure 3.** TOA radiances computed by using the MS RTM for the absorption bands: Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands. Blue lines correspond to the case without aerosol, while the black lines correspond to the specific case of desert aerosol.

## 3. Results and Discussion

### *3.1. Single CLSR vs. Double CLSR: Accuracy Results*

In this section, we compare the single and double application of the CLSR method in terms of accuracy. To assess it, we compute the relative error (residual) with respect to the continuum at each spectral point in an LBL manner,

$$\Delta I_{\text{res}} = \frac{I_{\text{CLSR}} - I_{\text{MS}}^{\text{ref}}}{I_{\text{MS}}^{\text{cont}}} \cdot 100\,(\%), \tag{11}$$

where $I_{\text{CLSR}}$ is the radiance calculated with either the CLSR method or the double CLSR method, $I_{\text{MS}}^{\text{ref}}$ is the reference radiance with absorption, while $I_{\text{MS}}^{\text{cont}}$ is the radiance in the continuum, i.e., without absorption, which is used to avoid values close to zero in the denominator, as in [19]. Other metrics used to estimate the accuracy of the simulations are the mean absolute relative error with respect to the continuum, the probability density functions and the cumulated probability.

Figure 4 shows the probability density function of the spectral residuals $\Delta I_{\text{res}}$ for both methods and the four spectral bands. The main conclusions that can be drawn from the figure are the following:

- More than 70% and 60% of the residuals are below 0.01% for the single and double CLSR methods, respectively, for all bands, with the exception of the water vapour band.
- The residuals of the water vapour band present a wider distribution in comparison with the other spectral bands.
- The probability densities are almost indistinguishable for both acceleration methods, demonstrating that both techniques provide accurate results among the different spectral bands.

It can be seen that the accuracy of the CLSR method is slightly higher than that of the double CLSR method. This result can be expected, since the TS spectra used in the double CLSR method are approximate and obtained from the SS spectrum (see Figure 2).

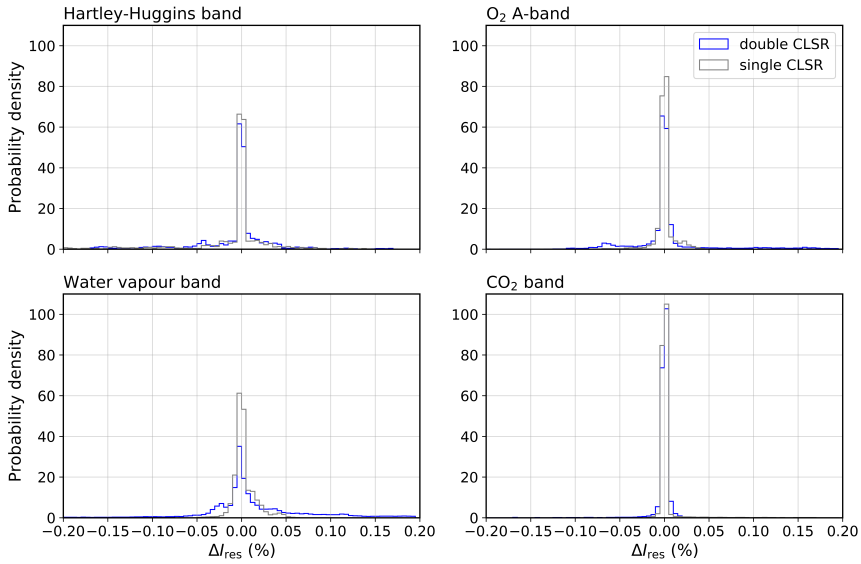


**Figure 4.** Probability density function of the residuals for the single CLSR (grey) and the double CLSR (blue) methods for the Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands and for the tropospheric aerosol case.

Figure 5 shows the cumulated probability functions of the CLSR and double CLSR methods for all spectral bands. Over 90% residuals are less than 0.05% in the case of the Hartley–Huggins band and 0.01% in the case of the $CO_2$ band. Higher differences can be seen between the CLSR and double CLSR methods for the $O_2$A- and water vapour bands. For these bands, over 90% of the CLSR residuals are less than 0.025%. Meanwhile, the double CLSR provides slightly larger errors: over 60% and 80% of the residuals are less than 0.05% for the $O_2$ A- and water vapour band, respectively. However, these residual values are still low and of the same order as those obtained by using the PCA-based RTMs. For instance, in Liu et al. [17] PCA was applied to optical parameters and spectral radiances yielding an error lower than 0.2% in the solar region (775–920 nm). Kopparla et al. [19] combined the PCA technique for optical parameters and the spectral binning for accurate computations in the case of aerosols. The residuals were below 0.01% for the $O_2$ A-band, which are of the same order as our results.

In our previous paper on the CLSR method [14], the convolved spectra were computed, showing errors of the same order as the non-convolved spectra. Here, the same conclusions apply for the double CLSR method.

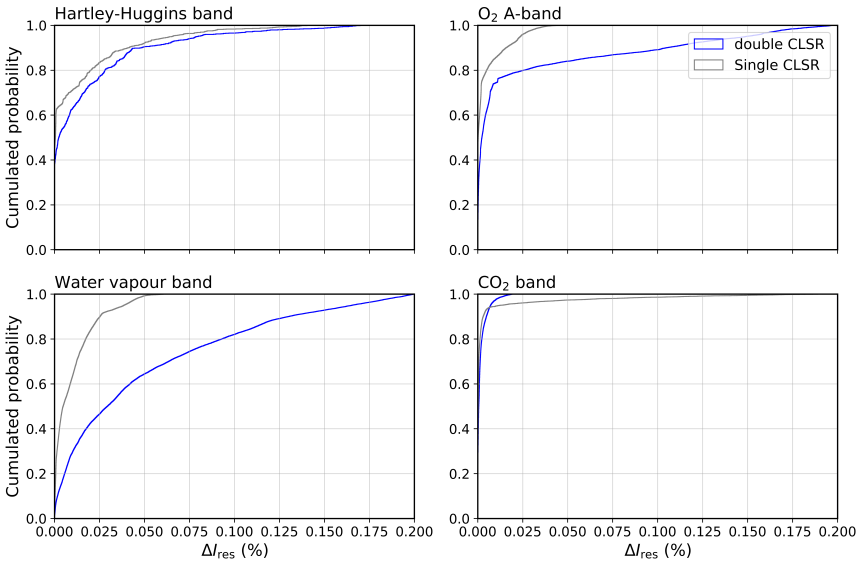


**Figure 5.** Same as Figure 4, but for the cumulated probability of the residuals.

### *3.2. Single CLSR vs. Double CLSR: Computational Performance*

In this section, we analyse the computational performance of the single and double CLSR methods. Tables 2–4 show the number of calls to the SS, TS and MS RTMs, computational times and corresponding acceleration factors with respect to the MS LBL simulations. We recall that, as the reference model, the MS LBL RTM with 32 streams is used.

For the single CLSR method, 5 clusters and 4 regression points per cluster are used. Hence, computations for each absorption band involve 20 MS RTM calls, while the number of TS RTM calls is equal to the number of spectral points in the high-resolution LBL computations (i.e., 300, 20,000 and 40,000 calls to the TS RTM for the Hartley–Huggins, $O_2$ A- and $CO_2$ and water vapour band, respectively). As can be seen in Tables 2–4, the TS RTM imposes the computational burden of the single CLSR method for the $O_2$ A-, $CO_2$ and water vapour bands, consuming up to 70% of the whole computation time.

**Table 2.** Summary of number of calls, computational time and acceleration factors for the Hartley–Huggins band. The computational times marked in red indicate the computational burden.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 300 | 20 | 20 |
| | TS | 0 | 300 | 32 |
| | SS | 0 | 0 | 300 |
| Computation time (s) | MS | 35 | 2.32 | 2.32 |
| | TS | 0 | 0.048 | 0.005 |
| | SS | 0 | 0 | 0.006 |
| Total computational time (s) | | 35 | 2.37 | 2.33 |
| Acceleration factor | | – | 14.8 | 15.0 |

**Table 3.** Same as for Table 2 but for the $O_2$A- and $CO_2$ bands.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 20,000 | 20 | 20 |
| | TS | 0 | 20,000 | 32 |
| | SS | 0 | 0 | 20,000 |
| Computation time (s) | MS | 2320 | 2.32 | 2.32 |
| | TS | 0 | 3.2 | 0.005 |
| | SS | 0 | 0 | 0.4 |
| Total computational time (s) | | 2320 | 5.52 | 2.725 |
| Acceleration factor | | – | 420 | 850 |

**Table 4.** Same as for Table 2 but for the water vapour band.

| | RTM | LBL | Single CLSR | Double CLSR |
|---|---|---|---|---|
| Number of calls | MS | 40,000 | 20 | 20 |
| | TS | 0 | 40,000 | 32 |
| | SS | 0 | 0 | 40,000 |
| Computation time (s) | MS | 4640 | 2.32 | 2.32 |
| | TS | 0 | 6.4 | 0.005 |
| | SS | 0 | 0 | 0.8 |
| Total computational time (s) | | 4640 | 8.72 | 3.13 |
| Acceleration factor | | – | 532 | 1482 |

At the first step of the double CLSR method, 8 clusters and 4 regression points are used. Thus, the TS RTM is called for 32 spectral points. In the case of the double CLSR, the SS RTM is utilized for the LBL computations, resulting in an additional performance enhancement by 2 times for the $O_2$A- and $CO_2$ bands and 3 times for the water vapour band. We note that in the double CLSR, the computational burden corresponds to the MS RTM, while the computation times related to the TS and SS RTMs are three and two orders of magnitude lower than those of the MS RTM, respectively. In the case of the Hartley–Huggins band, the computational burden is still due to the MS RTM [12] and the double CLSR does not further improve the performance.

As a final remark, the accuracy is crucial to determine the number of calls needed for the CLSR methods, and we could improve it by increasing the number of calls to the RTM models. However, this would add a computational burden to the simulations, while providing little improvement in the accuracy. Several tests have been performed by increasing the number of calls to the SS RTM or TS RTM but the errors are of the same order of magnitude as the actual values.

### *3.3. Computational Performance: State-of-the-Art Acceleration Techniques*

In this section, we compare the CLSR and double CLSR methods against other state-of-the-art acceleration techniques. Table 5 summarizes the spectral bands/regions and the acceleration factors for the selected acceleration techniques, including this study. In this analysis, we consider the studies covering absorption bands in the 280–3000 nm spectral range.

Note that the acceleration factors depend not only on the method used, but also on other aspects, such as the number of discrete ordinates $N_{do}$ used for the reference RTM. In turn, the required number of $N_{do}$ depends on aerosol properties, surface properties, geometry and the required accuracy. Therefore, the acceleration factors are ambiguous, as different numbers of streams or atmospheric parameters are used for the reference RTM in the different studies. Nevertheless, the following conclusions can be made:

**Table 5.** Selected acceleration techniques with the corresponding spectral region or band, acceleration factor (computed with respect to the LBL model) and their reference. Note that the acceleration factors are sometimes given in orders of magnitude compared to the LBL approach. One or two order of magnitude are indicated as 10× and 100×, respectively. The references are ordered chronologically.

| Acceleration Technique | Band/Spectral Region | Acceleration Factor | Reference |
|---|---|---|---|
| *k*-distribution | $H_2O$, $CO_2$ , $O_3$, and $O_2$ | 10× * | Fomin [32] |
| double-*k* approach | $O_2$ A | 1000 | Duan [16] |
| LSI | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 45 [b], 210 [c] | O'Dell [5] |
| PCA | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 50 | Natraj et al. [33] |
| PCA | 290–340 nm | 10 | Spurr et al. [34] |
| LEM | 325–335 nm | 10 | Efremenko et al. [9] |
| PCA | 325–335 nm | 2 | Efremenko et al. [35] |
| PCA | 300–3000 nm | 10× | Kopparla et al. [19] |
| PCA | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 100× | Somkuti et al. [36] |
| *k*-distribution + PCA | $O_2$ A | 342 | Molina García et al. [11] |
| PCA [a] | Hartley-Huggins | 18 | del Águila et al. [12] |
| NN | $O_2$ A, $CO_2$ weak, $CO_2$ strong | 250 [d] | Le et al. [37] |
| LEM | $NO_2$ (425–450 nm) | 12 [e] | Doicu et al. [18] |
| CLSR | $O_2$ A, $CO_2$ weak | 505 | del Águila et al. [14] |
| SDCOMP [a] | 750–920 nm | 1000 [d] | Liu et al. [17] |
| double CLSR | Hartley-Huggins | 15 | This study |
| | $O_2$ A, $CO_2$ weak | 850 | |
| | Water vapour | 1500 | |

[a] PCA is applied twice, (1) to the optical properties and (2) the radiance data set. [b] Nadir observations. [c] Glint observations. [d] Computation times are not considered, only the relative computational efficiency with respect to the accurate simulations. [e] Relative to the *k*-distribution method. * It is estimated from the information found in the reference, but the value does not appear explicitly.

1. For simulations in the Hartley–Huggins band, PCA techniques, linear embedding methods (LEM) and double CLSR have been applied. The double CLSR does not further improve the performance, since the computational burden is due to the MS RTM computations (see Section 3.2). The highest acceleration factor is provided by the method described in [12], in which PCA is applied to both optical parameters and spectral radiances. The performance enhancement in this case is up to 18 times.
2. There are several studies in which fast RTMs for the $O_2$ A- and $CO_2$ bands (either weak or strong) have been designed. In general, all considered techniques provide acceleration factors of about 2–3 orders of magnitude, including those based on artificial neural networks (NN) [37].
3. The water vapour band represents a challenge for acceleration techniques due to its complicated spectral structure. Therefore, the accuracy of the acceleration techniques is lower than for the $O_2$ A-band. For this band, the double CLSR method provides an acceleration factor of about 3 orders of magnitude, while the *k*-distribution [32] and PCA-based RTMs [19] achieve lower acceleration factors, of one order of magnitude.

### *3.4. Further Improvements to Aerosol Schemes*

In this section, the efficiency of the CLSR method for various configurations outlined in Section 2.3 is examined for several OPAC aerosol models. The results obtained using $\mathbf{X}_1$ and $\mathbf{X}_2$ (corresponding to Equations (9) and (10), respectively) are compared against the original CLSR method in which the matrix $\mathbf{X}_0$ (Equation (8)) is used. The mean absolute relative errors are shown in Figure 6. In general, mean relative absolute errors are below 0.05% for all aerosol types and bands when using the original $\mathbf{X}_0$-configuration for the CLSR method. However, these errors are slightly higher for the water vapour band due to its higher spectral complexity.

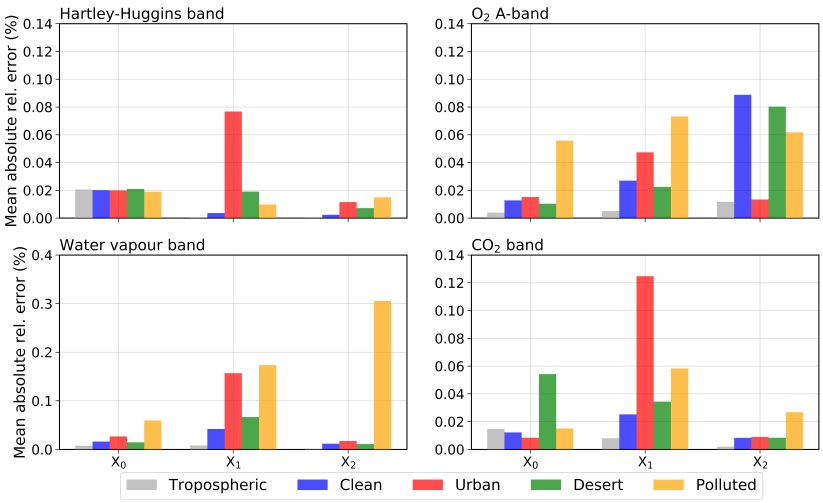


**Figure 6.** Mean absolute relative error in % for the four spectral bands, as well as for all aerosol types. Each bar colour corresponds to an aerosol type, while each group of bars corresponds to one of the $\mathbf{X}_i$-matrix ($i = 0, 1, 2$) (Equations (8)–(10)) used in the CLSR method. Note that the scale of the water vapour band is different from the rest of the spectral bands.

Almost in all cases, the $\mathbf{X}_1$ configuration is less accurate than the original CLSR method. This result can be expected, as $\mathbf{X}_1$ carries no information about the radiances of aerosols. However, for low aerosol load (tropospheric and clean aerosols), the errors are of the same order or below 0.05% compared to the ones from $\mathbf{X}_0$. The advantage of the $\mathbf{X}_1$ configuration is that it is based on spectra computed for the clear sky atmosphere. The corresponding LUT, therefore, is independent on aerosol properties. By using the CLSR method, the spectrum for actual aerosol conditions can be computed by calling MS RTM at a few spectral points (in our case, 20 spectral points).

The $\mathbf{X}_2$ configuration comprises $\mathbf{X}_0$, which contains the spectra corresponding to the TS RTM, and $\mathbf{X}_2$, which corresponds to TS and MS RTMs for clear sky scenarios. An improvement with respect to the original configuration is provided by the $\mathbf{X}_2$-configuration for the Hartley–Huggins and $CO_2$ bands, where absolute relative errors are of the same order as for $\mathbf{X}_0$-configuration or below 0.01%. For the $O_2$ A-band, there is almost no enhancement compared to the original configuration. We have excluded the transmittance in the numerical simulations for the $CO_2$ band in order to obtain slightly more accurate results.

In sum, we present two alternative configurations to the original $\mathbf{X}_0$, in order to obtain more accurate results for: (1) low aerosol loading conditions with the $\mathbf{X}_1$ configuration for all bands; and (2) the Hartley–Huggins, water vapour and $CO_2$ bands with the $\mathbf{X}_2$ configuration.

### *3.5. Combined Application of the Single CLSR vs. Double CLSR Method for Aerosol Scenarios*

Finally, the single and double CLSR methods are tested for the full set of aerosol models and $\mathbf{X}_i$-configurations. The probability density functions of the residuals for the tropospheric aerosol model are shown in Figures 7 and 8.

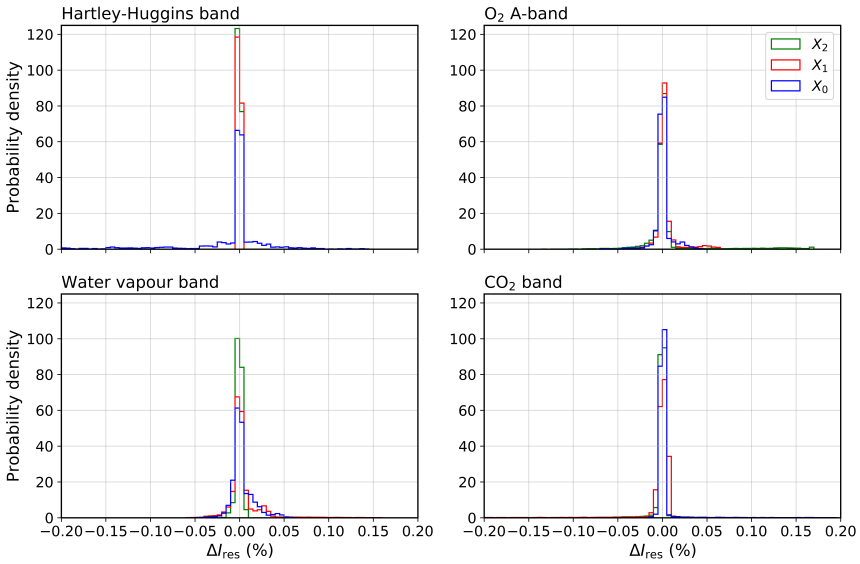


**Figure 7.** Probability density of the residuals for the single CLSR method and for several matrix configurations: $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$. Each plot represents the absorption bands: Hartley–Huggins, $O_2$ A-, water vapour and $CO_2$ bands. The case presented corresponds to the tropospheric aerosol.

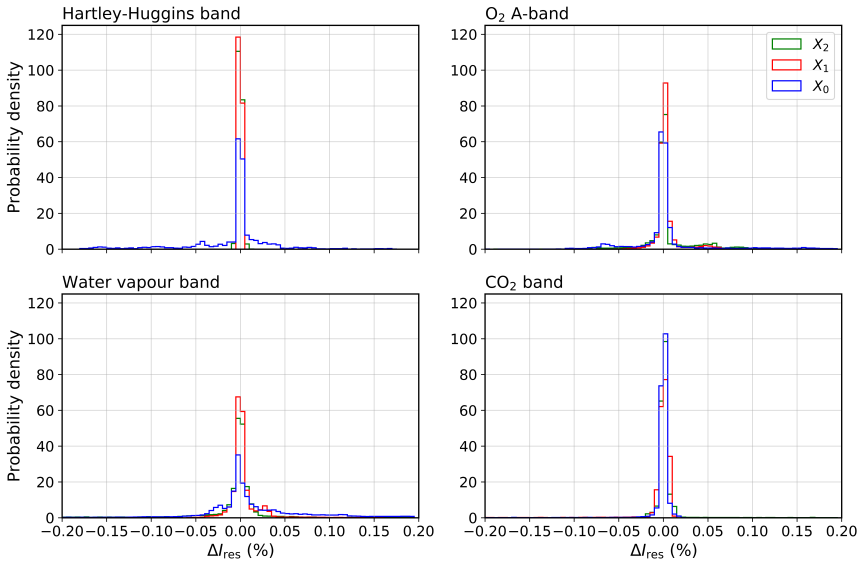


**Figure 8.** Same as Figure 7, but for the double CLSR method.

The conclusions drawn from the comparison of the two figures can be summarized as follows:

- The residual distributions of the single CLSR method are narrower than those of the double CLSR method, meaning that the single CLSR method is more accurate. However, in general, the residuals are below 0.01% for both methods and all spectral bands, except for the water vapour band, where the residual distributions are slightly wider and still below 0.05%. The distributions are not biased.

- For the Hartley–Huggins, $O_2$ A- and $CO_2$ bands, the residuals are below 0.05% for both single and double CLSR methods. Regarding the water vapour band, the residuals are below 0.05% and 0.1% for the single and double CLSR method, respectively. Similar accuracies were achieved in Kopparla et al. [19] for the water vapour band using the PCA-based RTM.
- In the case of the low aerosol load, the probability density functions are similar for all $\mathbf{X}_i$-configurations. However, as the aerosol load increases, the residual distributions for the $\mathbf{X}_1$ configuration provided by the single and double CLSR methods sometimes become biased for the water vapour band.

## 4. Conclusions

In this study, we have proposed two modifications to the CLSR method for fast computations of radiance spectra in absorption bands. In the first modification, the CLSR method is used twice, the first time for assessing the spectra corresponding to the TS RTM, and the second time for computing the spectra corresponding to the MS RTM. Our tests reveal that the double CLSR method further improves the computational performance of the original CLSR method by two ($CO_2$ and $O_2$A- bands) and three times (water vapour band), yet keeping the error below 0.05% for all spectral points on a high resolution grid. The designed approach has been compared with the state-of-the-art techniques found in the literature. Although the acceleration factors are ambiguous, in our simulations, the double CLSR method seems to provide a slightly higher performance than the PCA-based RTMs.

The second modification of the CLSR method is proposed for the atmospheric scenarios involving aerosols. The LBL computations performed for clear sky scenarios can be reused in the CLSR method as an approximate spectrum instead of the TS RTM, which is then corrected by calling the RTM for the atmospheric model with aerosol, but in a few spectral points. Therefore, by using the configuration of the CLSR method only with the information of the clear sky spectra, we can reproduce the LBL spectra with low aerosol load. We showed that in the case of low aerosol load, such a configuration provides absolute relative errors below 0.05% for all spectral bands, while enhancing the computational performance by three orders of magnitude. Low aerosol loading conditions occur when the atmosphere is in practice mainly clean. As the aerosol load increases (e.g., for polluted aerosol, with AOD equal to 5), the error also increases and reaches $\sim$0.2% for water vapour band and $\sim$0.1% in the $O_2$ A- and $CO_2$ band. The new configurations have been tested for both CLSR and double CLSR methods, revealing similar probability density functions. These results have the potential to be applied in near-real-time applications of aerosol computations or in aerosol retrieval algorithms considering an appropriate aerosol model selection [38,39]. It also seems beneficial to implement the CLSR method in conjunction with gradient-based LUT generators [40,41], to reduce the size of LUTs and interpolation errors. Besides, the CLSR method provides an accurate alternative to the LUT interpolation, which combines the regression and online RTM simulations. Therefore, the single and double CLSR methods could potentially be incorporated into the generation of LUT radiance fields in a fine resolution grid.

Our future work will be focused on coupling the CLSR method and machine learning approaches. Recent studies have shown that the accuracy of hyperspectral radiative transfer simulations can be improved by applying machine learning techniques (e.g., see [37]). Essentially, NNs are used as universal approximators which replace time-consuming RTMs. Further studies will be undertaken to replace the regression model with an NN, within the framework of the CLSR method. By doing this, we expect to reduce the number of clusters and spectral points for which the MS RTM is called.

**Author Contributions:** Conceptualization, A.d.Á.; methodology, A.d.Á.; software, A.d.Á.; formal analysis, D.S.E.; investigation, A.d.Á.; writing—original draft preparation, A.d.Á.; writing—review and editing, D.S.E.; visualization, A.d.Á.; supervision, D.S.E.; funding acquisition, D.S.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Aerospace Center (DLR) and the German Academic Exchange Service (DAAD) through the programme DLR/DAAD Research Fellowships 2018 (57424731) to A. del Águila with reference number 91711709.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AOD | Aerosol Optical Depth |
| CLSR | Cluster Low-Streams Interpolation |
| DOME | Discrete Ordinates with Matrix Exponential |
| HITRAN | High-Resolution Transmission Molecular Absorption Database |
| LBL | Line-By-Line |
| LEM | Linear Embedding Methods |
| LSI | Low-Streams Interpolation |
| LUT | LookUp Table |
| MS | Multi-Stream |
| NN | Neural Network |
| OPAC | Optical Properties of Aerosols and Clouds |
| PCA | Principal Component Analysis |
| Py4CAts | Python for Computational Atmospheric Spectroscopy |
| RTM | Radiative Transfer Model |
| SDCOMP | Spectral Data Compression |
| SS | Single-Scattering |
| SSA | Single Scattering Albedo |
| TOA | Top-of-the-Atmosphere |
| TS | Two-Stream |
| UV | Ultraviolet |

## References

1. Goody, R.; West, R.; Chen, L.; Crisp, D. The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **1989**, *42*, 539–550. [CrossRef]
2. Fu, Q.; Liou, K. On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *J. Atmos. Sci.* **1992**, *49*, 2139–2156. [CrossRef]
3. Kato, S.; Ackerman, T.P.; Mather, J.H.; Clothiaux, E.E. The k-distribution method and correlated-k approximation for a shortwave radiative transfer model. *J. Quant. Spectrosc. Radiat. Transf.* **1999**, *62*, 109–121. [CrossRef]
4. Zhang, F.; Zhu, M.; Li, J.; Li, W.; Di, D.; Shi, Y.N.; Wu, K. Alternate Mapping Correlated k-Distribution Method for Infrared Radiative Transfer Forward Simulation. *Remote Sens.* **2019**, *11*, 994. [CrossRef]
5. O'Dell, C.W. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. *J. Geophys. Res.* **2010**, *115*. [CrossRef]
6. Natraj, V.; Jiang, X.; Shia, R.; Huang, X.; Margolis, J.; Yung, Y. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the $O_2$ A-band. *J. Quant. Spectrosc. Radiat. Transf.* **2005**, *95*, 539–556. [CrossRef]
7. Kopparla, P.; Natraj, V.; Spurr, R.; Shia, R.L.; Crisp, D.; Yung, Y.L. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *173*, 65–71. [CrossRef]
8. Efremenko, D.; Doicu, A.; Loyola, D.; Trautmann, T. Acceleration techniques for the discrete ordinate method. *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *114*, 73–81. [CrossRef]
9. Efremenko, D.; Doicu, A.; Loyola, D.; Trautmann, T. Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *133*, 128–135. [CrossRef]
10. Liu, X.; Smith, W.L.; Zhou, D.K.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Appl. Opt.* **2006**, *45*, 201–208. [CrossRef]
11. Molina García, V.; Sasi, S.; Efremenko, D.; Doicu, A.; Loyola, D. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *213*, 228–240. [CrossRef]

12. del Águila, A.; Efremenko, D.S.; Molina García, V.; Xu, J. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band. *Atmosphere* **2019**, *10*, 142. [CrossRef]
13. del Águila, A.; Efremenko, D.S.; Trautmann, T. A review of dimensionality reduction techniques for processing hyper-spectral optical signal. *Light Eng.* **2019**, *27*, 85–98. [CrossRef]
14. del Águila, A.; Efremenko, D.S.; Molina García, V.; Kataev, M.Y. Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of O2 A- and CO2 Bands. *Remote Sens.* **2020**, *12*, 1250. [CrossRef]
15. Vincent, R.A.; Dudhia, A. Fast radiative transfer using monochromatic look-up tables. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *186*, 254–264. [CrossRef]
16. Duan, M.; Min, Q.; Li, J. A fast radiative transfer model for simulating high-resolution absorption bands. *J. Geophys. Res. Atmos.* **2005**, *110*. [CrossRef]
17. Liu, C.; Yao, B.; Natraj, V.; Kopparla, P.; Weng, F.; Le, T.; Shia, R.L.; Yung, Y.L. A Spectral Data Compression (SDCOMP) Radiative Transfer Model for High-Spectral-Resolution Radiation Simulations. *J. Atmos. Sci.* **2020**, *77*, 2055–2066. [CrossRef]
18. Doicu, A.; Efremenko, D.; Trautmann, T. A Spectral Acceleration Approach for the Spherical Harmonics Discrete Ordinate Method. *Remote Sens.* **2020**, *12*, 3703. [CrossRef]
19. Kopparla, P.; Natraj, V.; Limpasuvan, D.; Spurr, R.; Crisp, D.; Shia, R.L.; Somkuti, P.; Yung, Y.L. PCA-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *198*, 104–111. [CrossRef]
20. Schreier, F.; Gimeno García, S.; Hochstaffl, P.; Städt, S. Py4CAtS—PYthon for Computational Atmospheric Spectroscopy. *Atmosphere* **2019**, *10*, 262. [CrossRef]
21. Gordon, I.; Rothman, L.; Hill, C.; Kochanov, R.; Tan, Y.; Bernath, P.; Birk, M.; Boudon, V.; Campargue, A.; Chance, K.; et al. The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *203*, 3–69. [CrossRef]
22. Bodhaine, B.; Wood, N.; Dutton, E.; Slusser, J. On Rayleigh optical depth calculations. *J. Atmos. Ocean. Technol.* **1999**, *16*, 1854–1861. [CrossRef]
23. Anderson, G.; Clough, S.; Kneizys, F.; Chetwynd, J.; Shettle, E. *AFGL Atmospheric Constituent Profiles (0.120 km)*; Air ForceGeophysics Lab.: Hanscom AFB, MA, USA, 1986; p. 46.
24. Doicu, A.; Trautmann, T. Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case. *J. Quant. Spectrosc. Radiat. Transf.* **2009**, *110*, 146–158. [CrossRef]
25. Efremenko, D.S.; Molina García, V.; Gimeno García, S.; Doicu, A. A review of the matrix-exponential formalism in radiative transfer. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *196*, 17–45. [CrossRef]
26. Wiscombe, W.J. The Delta-M Method: Rapid Yet Accurate Radiative Flux Calculations for Strongly Asymmetric Phase Functions. *J. Atmos. Sci.* **1977**, *34*, 1408–1422. [CrossRef]
27. Nakajima, T.; Tanaka, M. Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation. *J. Quant. Spectrosc. Radiat. Transf.* **1988**, *40*, 51–69. [CrossRef]
28. Fischer, J.; Grassl, H. Radiative transfer in an atmosphere-ocean system: An azimuthally dependent matrix-operator approach. *Appl. Opt.* **1984**, *23*, 1032. [CrossRef]
29. Chalhoub, E.; Garcia, R. The equivalence between two techniques of angular interpolation for the discrete-ordinates method. *J. Quant. Spectrosc. Radiat. Transf.* **2000**, *64*, 517–535. [CrossRef]
30. Hess, M.; Koepke, P.; Schult, I. Optical properties of aerosols and clouds: The software package OPAC. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 831–844. [CrossRef]
31. Huang, Y.; Natraj, V.; Zeng, Z.C.; Kopparla, P.; Yung, Y.L. Quantifying the impact of aerosol scattering on the retrieval of methane from airborne remote sensing measurements. *Atmos. Meas. Tech.* **2020**, *13*, 6755–6769. [CrossRef]
32. Fomin, B. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. *J. Geophys. Res.* **2005**, *110*. [CrossRef]
33. Natraj, V.; Shia, R.L.; Yung, Y.L. On the use of principal component analysis to speed up radiative transfer calculations. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 810–816. [CrossRef]
34. Spurr, R.; Natraj, V.; Lerot, C.; Roozendael, M.V.; Loyola, D. Linearization of the Principal Component Analysis method for radiative transfer acceleration: Application to retrieval algorithms and sensitivity studies. *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *125*, 1–17. [CrossRef]
35. Efremenko, D.S.; Loyola, D.G.; Spurr, R.J.; Doicu, A. Acceleration of radiative transfer model calculations for the retrieval of trace gases under cloudy conditions. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *135*, 58–65. [CrossRef]
36. Somkuti, P.; Boesch, H.; Natraj, V.; Kopparla, P. Application of a PCA-Based Fast Radiative Transfer Model to XCO2 Retrievals in the Shortwave Infrared. *J. Geophys. Res. Atmos.* **2017**, *122*, 10,477–10,496. [CrossRef]
37. Le, T.; Liu, C.; Yao, B.; Natraj, V.; Yung, Y.L. Application of machine learning to hyperspectral radiative transfer simulations. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *246*, 106928. [CrossRef]
38. Sasi, S.; Natraj, V.; Molina García, V.; Efremenko, D.; Loyola, D.; Doicu, A. Model Selection in Atmospheric Remote Sensing with an Application to Aerosol Retrieval from DSCOVR/EPIC, Part 1: Theory. *Remote Sens.* **2020**, *12*, 3724. [CrossRef]
39. Sasi, S.; Natraj, V.; Molina García, V.; Efremenko, D.; Loyola, D.; Doicu, A. Model Selection in Atmospheric Remote Sensing with Application to Aerosol Retrieval from DSCOVR/EPIC. Part 2: Numerical Analysis. *Remote Sens.* **2020**, *12*, 3656. [CrossRef]

40. Vicent, J.; Verrelst, J.; Rivera-Caicedo, J.P.; Sabater, N.; Muñoz-Marí, J.; Camps-Valls, G.; Moreno, J. Emulation as an Accurate Alternative to Interpolation in Sampling Radiative Transfer Codes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4918–4931. [CrossRef]
41. Vicent Servera, J.; Alonso, L.; Martino, L.; Sabater, N.; Verrelst, J.; Camps-Valls, G.; Moreno, J. Gradient-Based Automatic Lookup Table Generator for Radiative Transfer Models. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1040–1048. [CrossRef]

# B Related Publications

The journals as well as the conference contributions included in this Appendix correspond with all my contributions up to date, i.e., the ones obtained during my thesis period and my previous experience in which I am co-author.

## B.1 Journals

- **A. del Águila**, D.S. Efremenko. Accuracy enhancement of the two-stream model for computing absorption bands at the presence of aerosols by means of the cluster low-streams regression. Light and Engineering, 29, 2, 79-86, 2021. doi: 10.33383/-2020-078

- **A. del Águila**, D.S. Efremenko. Fast Hyper-Spectral Radiative Transfer Model Based on the Double Cluster Low-Streams Regression Method. Remote Sensing, 13(3):434, 2021. doi: 10.3390/rs13030434

- **A. del Águila**, D. S. Efremenko, V. Molina García, M.Y. Kataev. Cluster Low-Streams Regression Method for Hyperspectral Radiative Transfer Computations: Cases of O2 A- and CO2 Bands. Remote Sensing, 12(8):1250, 2020.
doi: 10.3390/rs12081250

- C. Córdoba-Jabonero, L. Gómez-Martin, **A. del Águila**, J.M. Vilaplana, M.-A. López-Cayuela, M.-P. Zorzano. Cirrus-induced shortwave radiative effects depending on their optical and physical properties: Case studies using simulations and measurements. Atmospheric Research, 246, 105095, 2020.
doi: 10.1016/j.atmosres.2020.105095

- M. Sicard, M.J. Granados-Muñoz, L. Alados-Arboledas, R. Barragán, A.E. Bedoya-Velásquez, J.A. Benavent-Oltra, D. Bortoli, A. Comerón, C. Córdoba-Jabonero, M.J. Costa, **A. del Águila**, A.J. Fernández, J.L. Guerrero-Rascado, O. Jorba, F. Molero, C. Muñoz-Porcar, P. Ortiz-Amezcua, N. Papagiannopoulos, M. Potes, M. Pujadas, F. Rocadenbosch, A. Rodríguez-Gómez, R. Román, R. Salgado, V. Salgueiro, Y. Sola, M. Yela. Ground/space, passive/active remote sensing observations coupled with particle dispersion modelling to understand the inter-continental transport of wildfire smoke plumes. Remote Sensing of Environment, 232, 111294, 2019. doi: 10.1016/j.rse.2019.111294

- **A. del Águila**, D. S. Efremenko, T. Trautmann. A review of dimensionality reduction techniques for processing hyper-spectral optical signal. Light and Engineering, 27(3):85-98, 2019. doi: 10.33383/2019-017

- C. Córdoba-Jabonero, M. Sicard, **A. del Águila**, M. Jimenez, M.-P. Zorzano. Performance of a dust model to predict the vertical mass concentration of an extreme Saharan dust event in the Iberian Peninsula: comparison with continuous, elastic, polarization-sensitive lidars. Atmospheric Environment, 214, 116828, 2019. doi: 10.1016/j.atmosenv.2019.116828

- **A. del Águila**, D. S. Efremenko, V. Molina Garcia, J. Xu. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band. Atmosphere, 10(3):142, 2019. doi: 10.3390/atmos10030142

- M. Sorribas, E. Andrews, J.A. Ogren, **A. del Águila**, R. Fraile, P. Sheridan, M. Yela. Climatological study for understanding the aerosol radiative effects at southwest Atlantic coast of Europe. Atmospheric Environment, 205, 52-66, 2019. doi: 10.1016/j.atmosenv.2019.02.017

- J.A. Adame, L. Lope, P.J. Hidalgo, M. Sorribas, I. Gutiérrez-Álvarez, **A. del Águila**, A. Saiz-Lopez, M. Yela. Study of the exceptional meteorological conditions, trace gases and particulate matter measured during the 2017 forest fire in Doñana Natural Park, Spain. Science of the Total Environment, 645, 710-720, 2018. doi: 10.1016/j.scitotenv.2018.07.181

- C. Córdoba-Jabonero, M. Sicard, A. Ansmann, **A. del Águila**, H. Baars. Separation of the optical and mass features of particle components in different aerosol mixtures by using POLIPHON retrievals in synergy with continuous polarized Micro-Pulse Lidar (P-MPL) measurements. Atmos. Meas. Tech., 11, 4775-4795, 2018. doi: 10.5194/amt-11-4775-2018

- **A. del Águila**, M. Sorribas, H. Lyamani, F.J. Olmo, G. Arruda-Moreira, M. Yela and L. Alados-Arboledas. Sources and physicochemical characteristics of submicron aerosols during three intensive campaigns in Granada (Spain). Atmospheric Research, 213, 398-410, 2018. doi: 10.1016/j.atmosres.2018.06.004

- E. Alonso-Blanco, F.J. Gómez-Moreno, B. Artíñano, S. Iglesias-Samitier, V. Juncal-Bello, M. Piñeiro-Iglesias, P. López-Mahía, N. Pérez, M. Brines, A. Alastuey, M.I. García, S. Rodríguez, M. Sorribas, **A. del Águila**, G. Titos, H. Lyamani and L. Alados-Arboledas. Temporal and spatial variability of atmospheric particle number size distributions across Spain. Atmospheric Environment, 190, 146-160, 2018. doi: 10.1016/j.atmosenv.2018.06.046

- G. Titos, **A. del Águila**, A. Cazorla, H. Lyamani, J.A. Casquero-Vera, C. Colombi, E. Cuccia, V. Gianelle, G. Mocnik, A. Alastuey, F.J. Olmo, L. Alados-Arboledas. Spatial and temporal variability of carbonaceous aerosols: assesing the impact of biomass burning in the urban environment, Science of the Total Environment, 578, 613-625, 2017. doi: 10.1016/j.scitotenv.2016.11.007

## B.2 Conferences

- **A. del Águila**, D. S. Efremenko. The Cluster Low-Streams Regression Method for Fast Computations of Top-of-the-Atmosphere Radiances in Absorption Bands. Oral presentation. Part 2, pages short25–1–short25–9, 2020. doi: 10.51130/graphicon-2020-2-4-25. doi: 10.51130/graphicon-2020-2-4-25
  30th International Conference on Computer Graphics and Machine Vision (Graphicon 2020), Saint Petersburg, 22.09.2020-25.09.2020

- **A. del Águila**, D. S. Efremenko, D. Loyola, V. Molina García, T. Trautmann, Jian Xu. AI4RTM: Artificial Intelligence for Radiative Transfer Modeling. Oral presentation.
  Phi-Week 2019, ESA-ESRIN, Frascati, 09.09.2019-13.09.2019

- **A. del Águila**, D. S. Efremenko, V. Molina García, Jian Xu. Dimensionality reduction techniques for simulations of the spectral radiance in the Hartley-Huggins band. Poster presentation. Vol. 21, EGU2019-507, 2019
  EGU General Assembly 2019, Vienna, 03.05.2019-08.05.2019

- **A. del Águila**, V. Molina García, D. S. Efremenko. Dimensionality reduction of optical data: application to total ozone column retrieval. Poster presentation. doi:10.2760/848593
  Big Data from Space 2019 (BiDS'19), Munich, 19.02.2019-21.02.2019

- Víctor Molina García, Dmitry S. Efremenko, **Ana del Águila Pérez**. Automatic differentiation for Jacobian computations in radiative transfer problems. Oral presentation.
  21st European Workshop on Automatic Differentiation, Jena (Germany), 19.11.2018-20.11.2018

- C. Córdoba-Jabonero, M. Sicard, S. Basart, **A. del Águila**, M. Jiménez, M.-P. Zorzano. Comparison of the vertical distribution of dust mass concentration between a Dust forecast model and continuous polarized-lidar observations. Poster presentation.
  ACTRIS-2 WP2 Workshop, Hatfield, 19.11.2018-23.11.2018

- C. Córdoba-Jabonero, **A. del Águila**, M. Yela. Aerosol/cloud vertical monitoring at the "El Arenosillo" Atmospheric Observatory (SW Iberian Peninsula) with a polarized Micro-Pulse Lidar (P-MPL). Poster presentation.
  European Lidar Conference (ELC 2018), Thessaloniki, 03.07.2018-05.07.2018

- C. Córdoba-Jabonero, S. Basart, **A. del Águila**, M. Jiménez. Dust mass concentration profiles derived from polarized Micro-Pulse Lidar (P-MPL) measurements and NMMB/BSC-Dust modelling simulations: A case study of an intense dusty event unusually occurred in wintertime over Madrid (Central Spain). Poster presentation.
  European Lidar Conference (ELC 2018), Thessaloniki, 03.07.2018-05.07.2018

- M. Sicard, M.-J. Granados-Muñoz, N. Papagiannopoulos, L. Alados-Arboledas, R. Barragán, A.E. Bedoya-Velásquez, J.A. Benavent-Oltra, D. Bortoli, A. Comerón, C. Córdoba-Jabonero, A. M. J. Costa, **A. del Águila**, A. J. Fernández, J.L. Gómez Amo, J.L. Guerrero-Rascado, O. Jorba, J.A. Martínez-Lozano, F. Molero, C. Muñoz-Porcar, P. Ortiz-Amezcua, M. Potes, M. Pujadas, F. Rocadenbosch, A. Rodríguez-Gómez, R. Román, R. Salgado, V. Salgueiro, M.P. Utrillas, M. Yela. Tropospheric and stratospheric impact of Canadian smoke plumes over the Iberian Peninsula in August-September 2017. Poster presentation.
  European Lidar Conference (ELC 2018), Thessaloniki, 03.07.2018-05.07.2018

- **A. del Águila**, M. Sorribas, H. Lyamani, G. Titos, F.J. Olmo, G. Arruda-Moreira, M. Yela and L. Alados-Arboledas. Sources and physicochemical characteristics of submicron aerosols during three intensive campaigns at Granada (Spain). Poster presentation.
  4th ACTRIS-2 General Meeting 2018, Nafplio, Greece, 17.04.2018-19.04.2018

- **A. del Águila**, L. Gómez, J. M. Vilaplana and C. Córdoba-Jabonero. Assessment of the radiative effect of Cirrus clouds on surface: Radiative transfer simulations and solar radiation measurements at 'El Arenosillo' Atmospheric Observatory (SW

Iberian Peninsula). Poster presentation.
4th ACTRIS-2 General Meeting 2018, Nafplio, Greece, 17.04.2018-19.04.2018

- C. Córdoba-Jabonero, **A. del Águila** and M. Yela. Polarized Micro-Pulse Lidar (P-MPL) observations at 'El Arenosillo' Atmospheric Observatory (SW Iberian Peninsula): Case studies of Saharan dust, Smoke plumes and Cirrus clouds. Poster presentation.
  4th ACTRIS-2 General Meeting 2018, Nafplio, Greece, 17.04.2018-19.04.2018

- M. Sicard, M.-J. Granados-Muñoz, N. Papagiannopoulos, L. Alados-Arboledas, R. Barragán, A.E. Bedoya-Velásquez, J.A. Benavent-Oltra, D. Bortoli, A. Comerón, C. Córdoba-Jabonero, A. M. J. Costa, **A. del Águila**, A. J. Fernández, J.L. Gómez Amo, J.L. Guerrero-Rascado, O. Jorba, J.A. Martínez-Lozano, F. Molero, C. Muñoz-Porcar, P. Ortiz-Amezcua, M. Potes, M. Pujadas, F. Rocadenbosch, A. Rodríguez-Gómez, R. Román, R. Salgado, V. Salgueiro, M.P. Utrillas, M. Yela. Tropospheric and stratospheric impact of Canadian smoke plumes over the Iberian Peninsula in August-September 2017. Oral presentation.
  4th ACTRIS-2 General Meeting 2018, Nafplio, Greece, 17.04.2018-19.04.2018

- L. Gómez, C. Córdoba-Jabonero, O. Puentedura, M. Navarro-Comas, **A. del Águila**, L. Lope and M. Yela. Validation of MAX-DOAS aerosol profiles retrieval by comparison with LIDAR measurements. Poster presentation.
  4th ACTRIS-2 General Meeting 2018, Nafplio, Greece, 17.04.2018-19.04.2018

- **A. del Águila**, M. Sorribas, J. A. Adame, J. M. Vilaplana, J. A. Bogeat, C. Córdoba-Jabonero, M. Yela. Long-term study of new particle formation and shrinkage events in a coastal environment: meteorology, gas phase and solar radiation implications. Poster presentation.
  European Aerosol Conference (EAC 2017), Zurich , 27.08.2017-01.09.2017

- **A. del Águila**, L. Gómez, J. M. Vilaplana, J. A. Adame, M. Sorribas, C. Córdoba-Jabonero. Cirrus clouds observations at the Atmospheric Observatory 'El Arenosillo' (SW Iberian Peninsula) and their effects on the surface solar radiation. Poster presentation.
  European Aerosol Conference (EAC 2017), Zurich , 27.08.2017-01.09.2017

- M. Sorribas, E. Andrews, J. A. Adame, **A. del Águila**, M. Yela, P. Sheridan, J.A. Ogren. 11 years continuous monitoring of the in-situ aerosol optical properties at El Arenosillo observatory: emphasis on an extreme Saharan desert dust episode in February 2016. Poster presentation.
  European Aerosol Conference (EAC 2017), Zurich , 27.08.2017-01.09.2017

- F. J. Gómez-Moreno, E. Alonso-Blanco, B. Artíñano, J. Rodríguez-Maroto, E. Rojas, M. Piñeiro-Iglesias, P. López, M. Mahía, M. Sorribas, **A. del Águila**, J. A. Casquero-Vera, H. Lyamani, L. Alados-Arboledas, E. Borrás, A. Muñoz, T. Tritschen, A. Koched, E. Latorre. The REDMAAS 2017 intercomparison campaign: CPC and SMPS results. Poster presentation.
  5th Iberian Meeting on Aerosol Science and Technology (RICTA2017), Barcelona, 03.07.2017-06.07.2017

- **A. del Águila**, L. Gómez, J. M. Vilaplana, M. Sorribas, C. Córdoba-Jabonero. Cirrus clouds properties derived from Polarized Micro-Pulse Lidar (P-MPL) observations at the atmospheric observatory 'El Arenosillo' (SW Iberian Peninsula): a

case study for radiative implications. Poster presentation. The European Physical Journal Conferences 176(6):05042. DOI10.1051/epjconf/201817605042
ILRC 28 – 28th International Laser-Radar Conference, Bucharest, 25.06.2017 – 30.06.2017

- C. Córdoba-Jabonero, M. Sicard, A. Ansmann, **A. del Águila**, H. Baars. Vertical separation of the atmospheric aerosol components by using POLIPHON retrieval in Polarized Micro-Pulse Lidar (P-MPL) measurements: case studies of specific climate-relevant aerosol types. Poster presentation. The European Physical Journal Conferences 176(4):05041. DOI10.1051/epjconf/201817605041
ILRC 28 – 28th International Laser-Radar Conference, Bucharest, 25.06.2017 – 30.06.2017

- **A. del Águila**, M. Sorribas, J.A. Adame, J.M. Vilaplana, J.A. Bogeat, C., Córdoba-Jabonero and M. Yela. Long-term study of new particle and shrinkage episodes in a coastal environment: meteorology, gas phase and solar radiation implications. Oral presentation.
3rd ACTRIS-2 General and MTR Meeting, Granada, 30.01.2017-03.02.2017

- **A. del Águila**, L. Gómez, J.M. Vilaplana, M. Sorribas and C. Córdoba-Jabonero. Cirrus clouds properties observations at the Atmospheric Observatory "El Arenosillo" (SW Iberian Peninsula): A case study for radiative implications. Poster presentation.
3rd ACTRIS-2 General and MTR Meeting, Granada, 30.01.2017-03.02.2017

- M. Sorribas, E. Andrews, J.A. Adame, **A. del Águila**, M. Yela, P. Sheridan and J.A. Ogren. Multi-year aerosol study based on aerosol radiative properties at El Arenosillo Observatory: emphasis on an extreme Saharan desert dust episode in February 2016. Poster presentation.
3rd ACTRIS-2 General and MTR Meeting, Granada, 30.01.2017-03.02.2017

- C. Córdoba-Jabonero, M. Sicard, A. Ansmann, **A. del Águila** and H. Baars. Vertical separation of the atmospheric aerosol components by using POLIPHON retrieval in polarized Micro Pulse Lidar (P-MPL) measurements: Case studies of specific climate-relevant aerosol types. Poster presentation.
3rd ACTRIS-2 General and MTR Meeting, Granada, 30.01.2017-03.02.2017

# C List of Abbreviations

| Abbreviation | Description |
|---|---|
| ACS | Atmospheric Composition Sensors |
| AOD | Aerosol Optical Depth |
| AVHRR | Advanced Very High Resolution Radiometer |
| CIA | Collision-Induced Absorption |
| CLSR | Cluster Low-Streams Regression |
| DLR | Deutsches zentrum für Luft- und Raumfahrt (German Aerospace Center) |
| DOME | Discrete Ordinates with Matrix Exponential |
| EM | ElectroMagnetic |
| EO | Earth Observation |
| EOF | Empirical Orthogonal Function |
| ESFT | Exponential Sum Fitting of Transmittances |
| FWHM | Full Widths at Half Maximum |
| GOME | Global Ozone Monitoring Experiment |
| GOSAT | Greenhose gases Observing SATellite |
| HITRAN | HIgh-resolution TRANsmission molecular absorption database |
| HPC | High Performance Computing |
| IR | InfraRed |
| IQR | InterQuartile Range |
| LBL | Line-By-Line |
| LEM | Linear Embedding Methods |
| LS | Low Streams |
| LSI | Low-Streams Interpolation |
| LUT | Look-Up Table |
| L1B | Level-1B |
| ML | Machine Learning |
| MODTRAN | MODerate resolution atmospheric TRANsmission |
| MS | Multi-Stream |
| NIR | Near InfraRed |
| NN | Neural Network |
| NRT | Near Real Time |
| OCO | Orbiting Carbon Observatory |
| OMI | Ozone Monitoring Instrument |
| OPAC | Optical Properties of Aerosols and Clouds |
| OSS | Optimal Spectral Sampling |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PCRTM | Principal Component-based Radiative Transfer Model |
| PRISM | Picosatellite for Remote sensing and Innovative Space Missions |
| PYDOME | PYthon Discrete Ordinates with Matrix Exponential |

| Abbreviation | Description |
|---|---|
| Py4CAts | Python for Computational Atmospheric Spectroscopy |
| RT | Radiative Transfer |
| RTE | Radiative Transfer Equation |
| RTM | Radiative Transfer Model |
| RTTOV | Radiative Transfer for TOVS |
| SDCOMP | Spectral Data COMPression |
| SMART | Spectral Mapping Atmospheric Radiative Transfer |
| SRF | Spectral Response Function |
| SS | Single-Scattering |
| SSA | Single Scattering Albedo |
| S5P | Sentinel 5 Precursor |
| TOA | Top-of-the-Atmosphere |
| TROPOMI | TROPospheric Ozone Monitoring Instrument |
| TS | Two-Stream |
| UV | UltraViolet |
| VIS | VISible |

# D List of Symbols

| Symbol | Description |
|---|---|
| $\mathcal{B}$ | absolute temperature |
| $k_{\text{abs}}$ | absorption coefficient |
| $k$ | absorption coefficient |
| $g$ | analytical function |
| $\varphi$ | azimuthal angle |
| $I_m$ | azimuthal component of the radiance |
| $m$ | azimuthal expansion term |
| $\mathbf{f}$ | basis vector |
| $k_{\mathbf{B}}$ | Boltzmann constant |
| $c$ | cluster for the CLSR method |
| $z$ | cluster for the double CLSR method |
| $t$ | computational time |
| $I^{\text{cont}}$ | continuum radiance |
| $f(\mathbf{x}_w)$ | correction function for PCA |
| $\mu$ | cosine of the polar angle |
| $G$ | cumulative frequency distribution |
| $M$ | dimension of subspace using PCA |
| $\mathbf{\Omega}$ | direction of the light |
| $T$ | direct transmittance |
| $f$ | distribution function |
| $\Lambda$ | eigenvalue matrix |
| $\mathbf{V}$ | eigenvector matrix |
| $\mathbf{q}$ | empirical orthogonal functions |
| $\overline{I}$ | equidistant radiances for the regression model |
| $\xi$ | error of bilinear interpolation |
| $k_{\text{ext}}$ | extinction coefficient |
| $\tau_{\text{gas}}$ | gas absorption optical depth |
| $l$ | layer |
| $\mathbf{A}$ | layer matrix |
| $I_{\text{LS}}$ | low streams component of the radiance |
| $\mathbf{Y}$ | matrix of atmospheric radiances |
| $\mathbf{T}$ | matrix of atmospheric radiances in the new basis |
| $\mathbf{A}$ | matrix of regression coefficients in the CLSR method |
| $\mathbf{X}$ | matrix of the CLSR method |
| $\mathbf{F}$ | matrix of the new basis |
| $\epsilon$ | mean relative error |
| $I_{\text{MS}}$ | multiple-scattering component of the radiance |
| $\mathbf{n}$ | normalized vertical distribution |

| Symbol | Description |
|---|---|
| $K$ | number of coordinates in the new basis |
| $N_{\mathrm{do}}$ | number of discrete ordinates |
| $C$ | number of clusters for the CLSR method |
| $Z$ | number of cluster for the double CLSR method |
| $n$ | number of equidistant radiance points on each cluster |
| $N$ | number of homogeneous layers |
| $L$ | number of layers |
| $N_c$ | number of radiance points on each cluster |
| $J$ | number of spectral points in the new basis |
| $W$ | number of wavelengths |
| $\tau$ | optical depth |
| $r$ | path |
| $\theta$ | polar angle |
| $\tilde{I}$ | predicted radiances |
| $y$ | principal component |
| $u$ | quantity of absorber |
| $\mathbf{i}_m$ | radiance vector |
| $\mathbf{R}$ | reflectivity matrix |
| $\alpha$ | regression coefficient |
| $\beta$ | regression coefficient |
| $\gamma$ | regression coefficient |
| $a$ | regression coefficient |
| $b$ | regression coefficient |
| $d$ | regression coefficient |
| $I_{\mathrm{res}}$ | residual error of the radiance |
| $W_{\mathrm{rot}}$ | rotational energy |
| $I$ | scalar radiance |
| $k_{\mathrm{sca}}$ | scattering coefficient |
| $S$ | set of spectra |
| $\omega$ | single scattering albedo |
| $I_{\mathrm{SS}}$ | single-scattering component of the radiance |
| $P$ | single scattering phase function |
| $\hat{I}$ | sorted radiance in ascending order |
| $\mathbf{P}$ | source matrix |
| $\mathbf{b}$ | source vector |
| $S_{\mathrm{PCA}}$ | speedup factor for the PCA-based RTM |
| $S_{\mathrm{CLSR}}$ | speedup factor for the CLSR method |
| $s$ | surface boundary condition |
| $\tau_{abs}$ | total absorption optical thickness |
| $W_{\mathrm{tot}}$ | total energy |
| $\mathbf{T}$ | transmission matrix |
| $W_{\mathrm{tra}}$ | translational energy |
| $\mathbf{t}$ | vector in the new basis |
| $\mathbf{y}$ | vector of atmospheric radiances or spectra |
| $\mathbf{x}$ | vector of optical parameters |
| $W_{\mathrm{vib}}$ | vibrational energy |
| $\lambda$ | wavelength |
| $w$ | wavelength index |
| $w_i$ | weights |