Technical University of Munich
TUM School of Engineering and Design

Retrieval of cloud properties from EPIC/DSCOVR

Víctor Molina García

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

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr.-Ing. habil. Florian Seitz

Prüfende der Dissertation:

1. Priv.-Doz. Dr. habil. Adrian Doicu
2. Prof. Dr.-Ing. habil. Richard Hans Georg Bamler
3. Prof. Dr. rer. nat. Bernhard Mayer, Ludwig Maximilian Universität München

Die Dissertation wurde am 28.06.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 19.10.2022 angenommen.

Acknowledgements

This dissertation would have not been possible without the support and help of many people.

First, I would like to express my gratitude to my academic supervisor, Dr. Adrian Doicu, who guided me throughout this long project, and from whom I learnt most of my current knowledge about radiative transfer and non-linear optimisation solvers. I want to show my appreciation to my mentor, Dr. Diego Loyola, for being present in the difficult moments of this project, and for trying to transfer me his knowledge about decision making and triage of priorities (with partial success, only my fault). I also want to thank Prof. Dr. Thomas Trautmann for all his support in the administrative duties and his willingness to help. And I would like to sincerely thank Prof. Dr. Richard Bamler and Prof. Dr. Bernhard Mayer for their generosity of dedicating their time to read and review this dissertation.

This endeavor would not have been possible without the financial support of the German Academic Exchange Service (DAAD) as well as the German Aerospace Center (DLR). I would always be grateful to both institutions for giving me the chance to develop this research.

The colleagues at the Atmospheric Processors department in DLR-IMF have made this journey worthwhile. This dissertation would not be in its current form without all the fruitful discussions with Dr. Dmitry Efremenko, ranging a broad scope of topics from radiative transfer itself, to dimensionality reduction techniques and automatic differentiation. I also want to sincerely thank Dr. Ronny Lutz for the discussions about the OCRA algorithm, and for being always there to help me with my never-ending doubts about German. I am very grateful to Fabian Romahn for all our conversations about neural network training. I want to thank the personal support by other colleagues during these years, including (but not limited to) Ana, Athina, Bernd, Franz, Jian, Sebastián, Song and Sruthy. Please forgive me if I forgot somebody.

I also do not want to finish these acknowledgements without a mention to my home university in Valladolid (Spain). Without the support of Dr. Abel Calle, Prof. Dr. Victoria Cachorro and Prof. Dr. Ángel de Frutos, my scientific career in the scope of atmospheric physics would have never started. And thanks to the Remote Sensing lectures of Prof. Dr. José Luis Casanova I decided to go into the world of atmospheric remote sensing, which now is a part of my life. Thank you for all your guidance during my studies.

Last but not least, I would like to thank my family and Cristina for supporting me during the last six years of my life (in person and virtually), and especially for *suffering* me in the last stages of this dissertation.

Done can also be better than perfect

Víctor Molina García

24 June 2022

Summary

The Earth Polychromatic Imaging Camera (EPIC) on board the Deep Space Climate Observatory (DSCOVR), launched in February 2015, is the first scientific mission offering the unique view of the Sun-illuminated half of the Earth from the vicinity of the Lagrangian point L_1, at about 1.5 million kilometers far away from Earth. The EPIC instrument, which provides measurements for 10 narrow channels ranging from ultraviolet to near-infrared, aims to offer global hourly information about the concentration of trace gases and the properties of clouds, aerosols, and the Earth's surface.

This dissertation describes a framework for the fast and accurate retrieval of liquid-water cloud macrophysical properties from radiance measurements acquired by the EPIC/DSCOVR instrument. The first point that the dissertation addresses is the solution of the registration issue discovered in the EPIC Level 1B Version 2 collection, which is essential for the usefulness of this collection in any scientific work. By developing an unsupervised registration correction algorithm, the overall registration accuracy of this Level 1B collection improves from 5 to about 1 ground pixel. In addition, the EPIC instrument degradation is studied for 4 years of mission data and correction factors are derived accordingly for the 10 EPIC channels.

The second part of this work analyses the combination of standard acceleration techniques to increase the time efficiency of the radiative transfer models describing how light is absorbed and scattered in the atmosphere and later measured by EPIC, and it shows that the computation times can be reduced at least two orders of magnitude without a significant impact on the model accuracy. For the computation of the derivatives of the physical model with respect to the retrieved parameters, the use of the forward-adjoint approach reduces the computation time by three orders of magnitude. The acceleration techniques reduce the time required for the generation of look-up tables from weeks into days.

The Optical Cloud Recognition Algorithm (OCRA), which estimates the cloud fraction by means of image analysis, is applied to four years of EPIC measurements and compared against the daily cloud fraction product from MODIS (Moderate-Resolution Imaging Spectroradiometer). Finally, the Retrieval Of Cloud Information Using Neural Networks (ROCINN), which estimates two cloud macrophysical parameters (cloud optical thickness and cloud-top height) from measurements in the oxygen A- und B-bands, is implemented for the EPIC instrument. From a sensitivity analysis performed with synthetic EPIC measurements under several sources of uncertainty, it can be concluded that only the cloud-top height can be retrieved with an acceptable accuracy. Finally, the combination of both OCRA and ROCINN is applied to one EPIC test dataset, and the retrieved results are compared with the operational cloud products of the EPIC/DSCOVR mission.

Zusammenfassung

Die „Earth Polychromatic Imaging Camera“ (EPIC) an Bord des „Deep Space Climate Observatory“ (DSCOVR) wurde im Februar 2015 gestartet und ist die erste wissenschaftliche Mission die einen einzigartigen Blick auf die von der Sonne erleuchtete Hälfte der Erdkugel aus der Nähe des Lagrange-Punktes L_1 ermöglicht, aus eine Entfernung von etwa 1.5 Millionen Kilometern. Das EPIC Instrument misst in zehn schmalen Spektralbereichen, welche vom Ultravioletten bis ins Nah-Infrarote reichen und zielt auf die Ableitung der Konzentration von Spurengasen und Eigenschaften von Wolken, Aerosolen sowie der Beschaffenheit der Erdoberfläche mit einer globalen räumlichen Abdeckung und stündlicher zeitlicher Auflösung.

Diese Dissertation beschreibt ein Verfahren für die schnelle und genaue Bestimmung von makrophysikalischen Parametern von Wasserwolken aus den Radianz-Messungen des EPIC Instruments. Als ersten Punkt behandelt die Dissertation eine Lösung des Problemes der Bildkorrektur, welche essentiell für eine wissenschaftliche Nutzbarkeit der EPIC Daten aus der Level 1B Version 2 Kollektion ist. Durch die Entwicklung eines automatisierten Korrekturalgorithmus wird die räumliche Genauigkeit der Level 1B Kollektion von 5 auf etwa 1 Boden-Pixel reduziert. Zusätzlich wird die Langzeitstabilität des EPIC Instruments über einen Zeitraum von 4 Jahren analysiert und Korrekturfaktoren für alle 10 EPIC Kanäle berechnet.

Der zweite Teil dieser Arbeit untersucht die Kombination von Standard-Beschleunigungstechniken um die Effizienz der verwendeten Strahlungstransportmodelle, welche die atmosphärischen Absorptions- und Streuprozesse der vom EPIC Instrument gemessenen Strahlung beschreiben, zu erhöhen. Es wird gezeigt, dass die Rechenzeit um mindestens zwei Größenordnungen reduziert werden kann ohne dabei einen signifikanten Effekt auf die Modellgenauigkeit zu haben. Für die Berechnung der Ableitungen des physikalischen Modells in Bezug auf die abgeleiteten Parameter reduziert der „forward-adjoint“ Ansatz die Rechenzeit um drei Größenordnungen. Diese Beschleunigungstechniken verkürzen damit die Dauer für das Erzeugen von Umsetzungstabellen von Wochen zu Tagen.

Der OCRA Algorithmus bestimmt den Wolkenbedeckungsgrad mittels Bildanalyse und wird auf 4 Jahre EPIC Daten angewendet und mit dem täglichen Wolkenprodukt des MODIS Instruments verglichen. Der ROCINN Algorithmus bestimmt die makrophysikalischen Wolkenparameter optische Dichte und Wolkenhöhe aus Messungen in den O_2 A- und B-Banden und wird ebenfalls an das EPIC Instrument angepasst. Aus Sensitivitätsstudien mit synthetischen EPIC Messungen unter der Berücksichtigung mehrerer Fehlerquellen, lässt sich schlussfolgern, dass nur die Wolkenhöhe mit genügender Genauigkeit bestimmt werden kann. Abschließend wird die Kombination OCRA/ROCINN auf einen EPIC Testdatensatz angewandt und die abgeleiteten Ergebnisse mit den operationellen EPIC Produkten verglichen.

Table of contents

Chapter 1
Introduction

The story of the Deep Space Climate Observatory (DSCOVR) dates back to 1998, when for the first time it was proposed to establish an Earth observation system orbiting around the Lagrangian point L_1, at about 1.5 million kilometers far away from Earth on the Sun-Earth path. Originally named as Triana, the initial goal of this mission of the National Aeronautics and Space Administration (NASA) was to provide an hourly view of the Sun-illuminated half of the Earth suitable for climate observations. Short after January 2001, the mission was put on hold. NASA renamed the spacecraft to Deep Space Climate Observatory (DSCOVR) in 2003. It was not until 2009 that the DSCOVR project was brought back to life. After several years of refurbishment, DSCOVR was finally launched in February 2015, and it reached its destination around the Lagrangian point L_1 in June 2015.

Among other instruments on board DSCOVR, the Earth Polychromatic Imaging Camera (EPIC) is responsible for the acquisition of measurements of the Sun-illuminated half of the Earth ranging from the ultraviolet to the near-infrared. Since the first image was acquired on 6 July 2015, a broad scope of scientific research products and analyses based on EPIC imagery have been developed [1]: trace gas estimation in the atmosphere, such as ozone [2–5] and sulfur dioxide [6] from volcanic eruptions; determination of aerosol properties [7–9]; studies on cloud dynamics [10, 11]; surface characterisation [12]; and vegetation monitoring [13–15].

Clouds are one crucial component of the Earth's atmosphere. They are an important component of the hydrological cycle, and since about 70% of the Earth is covered by clouds [16], the characterisation of the atmosphere necessarily needs to account for them. The retrieval of atmospheric trace gas components needs this cloud information as support data [17], and surface products need to be able to detect the presence of clouds in order to separate clear-sky scenes from cloud-contaminated scenes.

Atmospheric radiative transfer theory models how the incident Sun light is absorbed and scattered by the constituents of the Earth's atmosphere (gases, clouds and aerosols) and the Earth interfaces [18]. The solution of the radiative transfer equation allows to determine the radiance reflected at the top of the atmosphere (TOA) at a given spectral wavelength, which may be measured by instruments on board Earth observation satellites. While the estimation of trace gas constituents is performed on the basis of the Differential Optical Absorption Spectroscopy (DOAS) [19], the estimation of cloud properties is normally inferred from radiation measurements in regions of the electromagnetic spectrum where only the molecular oxygen (O_2) is responsible for the main absorption of the radiation. Since the molecular oxygen trace gas column concentration remains stable in the atmosphere, the impact of the O_2 presence to

the expected TOA radiation at the O_2 absorption spectral windows under clear-sky conditions is known in advanced. If this TOA radiation is higher than expected, it indicates that only a portion of the oxygen column is absorbing the incident light, while the remaining portion is shielded by an object in the atmosphere (e.g. a cloud). Two spectral windows of absorption by molecular oxygen in the visible and near-infrared are the oxygen A-band (759–770 nm) [20] and the oxygen B-band (685–695 nm) [21].

The retrieval of cloud properties based on TOA radiance measurements are optimisation problems in which the model function to fit is the solution of the radiative transfer equation for an atmosphere containing a cloud. The optimiser is a non-linear least squares solver [22, 23]. Examples of model functions are the libraries DISORT (DIScrete Ordinate Radiative Transfer) [24], DOME (Discrete Ordinate method with Matrix Exponential) [25], and LIDORT (LInearized Discrete Ordinate Radiative Transfer) [26]. The calls to these libraries are in general computationally expensive, especially when the number of streams, used to replace the multiple-scattering term of the radiative transfer equation with a quadrature sum, is high.

Since the EPIC instrument contains two pairs of absorption and reference channels in the oxygen A-band and B-band, information about the cloud properties could potentially be estimated from them. The discussion about the retrieval of cloud properties started years before the launch of DSCOVR. In 2013, Yang et al. published a first methodology that claimed that cloud-top height and cloud geometrical thickness could be retrieved from EPIC measurements in the oxygen bands simultaneously for fully-cloudy scenes over ocean surfaces [27]. Posterior analyses on cloud information content corrected these results, stating that only the cloud-top height could be retrieved with acceptable uncertainty in most of the cases [28, 29], and that only one piece of information is available.

1.1 Motivation

After this first description, at least four aspects in regard to the retrieval of cloud parameters from EPIC can summarise the motivation for this dissertation to be developed:

- First of all, there is a need to ensure the existence of a reliable source of measurements from EPIC/DSCOVR, i.e. well calibrated and geolocated measurements.
- There is a need to know which cloud macrophysical information is actually possible to retrieve from measurements of the EPIC/DSCOVR mission.
- The retrieval of cloud macrophysical properties applied to real data requires the cloud retrieval algorithms to be fast but at the same time accurate.
- Current standard radiative transfer models may be very accurate but at the cost of high computation times.

These four points will be addressed in the following chapters of the dissertation.

1.2 Overview of the dissertation

This dissertation is divided in six chapters. After the current Chapter 1, which acts as entry point to the dissertation topic, the next chapters address the different aspects required for a successful and fast implementation of the retrieval of cloud properties from EPIC/DSCOVR:

- The objective of Chapter 2 is to describe the pre-processing steps needed to sanitise the original EPIC measurements. After a short description of the EPIC instrument channels and its official calibration coefficients, the chapter presents an unsupervised registration correction algorithm that fixes the misregistration issues of the EPIC Level 1B collection version 2. After this correction is applied to four years of EPIC datasets, global trends of the EPIC measurements are analysed to estimate the degradation of the instrument and provide correction factors for the calibration coefficients of each of the 10 EPIC channels.
- In Chapter 3, the challenges associated to the radiative transfer modelling for EPIC are outlined. The goal of having a fast and accurate radiative transfer model for clouds is approached by using two variants: through model simplifications, and through the use of acceleration techniques based on dimensionality reduction of the optical properties. A similar investigation is performed for the computation of the radiative transfer model derivatives with respect to the cloud parameters, for which the forward-adjoint approach is compared to the standard linearisation of the radiative transfer problem.
- The Chapter 4 focuses on the application of the Optical Cloud Recognition Algorithm (OCRA) to retrieve the cloud fraction from EPIC/DSCOVR datasets. By taking advantage of the results from Chapter 2, clear-sky maps are created with high registration accuracy for the EPIC channels using 4 years of data from the EPIC Level 1B collection version 2. The OCRA scaling and offset parameters are estimated accordingly. The OCRA parameters and clear-sky maps are used to generate four years of EPIC radiometric cloud fraction datasets, which are validated with the daily MODIS cloud product.
- Chapter 5 describes the setup and application of the Retrieval Of Cloud Information using Neural Networks (ROCINN) to the EPIC Level 1B collection version 2. An extensive analysis of the ROCINN configuration is outlined, such as the proper description of the input space and the procedure to find out appropriate neural networks that emulate the time-expensive radiative transfer models accurately. These neural networks are used to perform a sensitivity analysis that determines the actual capabilities of the EPIC instrument to infer information about cloud macrophysical properties. The findings of this sensitivity analysis are evaluated with one EPIC test dataset, for which the OCRA/ROCINN combination is applied, and the resulting ROCINN cloud products are compared with the corresponding operational EPIC cloud products from NASA.

Finally, Chapter 6 recapitulates the findings of the former chapters from a global perspective, and it indicates in addition future lines of work.

References

[1] A. Marshak, J. Herman, A. Szabo, K. Blank, S. Carn, A. Cede, I. Geogdzhayev, D. Huang, L.-K. Huang, Y. Knyazikhin, M. Kowalewski, N. Krotkov, A. Lyapustin, R. McPeters, K. G. Meyer, O. Torres, Y. Yang, Earth observations from DSCOVR EPIC instrument, Bulletin of the American Meteorological Society 99 (9) (2018) 1829–1850. doi:10.1175/bams-d-17-0223.1.

[2] J. Herman, L. Huang, R. McPeters, J. Ziemke, A. Cede, K. Blank, Synoptic ozone, cloud reflectivity, and erythemal irradiance from sunrise to sunset for the whole earth as viewed by the DSCOVR spacecraft from the earth–sun lagrange 1 orbit, Atmospheric Measurement Techniques 11 (1) (2018) 177–194. doi:10.5194/amt-11-177-2018.

[3] J. Herman, A. Cede, L. Huang, J. Ziemke, O. Torres, N. Krotkov, M. Kowalewski, K. Blank, Global distribution and 14-year changes in erythemal irradiance, UV atmospheric transmission, and total column ozone for 2005–2018 estimated from OMI and EPIC observations, Atmospheric Chemistry and Physics 20 (14) (2020) 8351–8380. doi:10.5194/acp-20-8351-2020.

[4] N. A. Kramarova, J. R. Ziemke, L.-K. Huang, J. R. Herman, K. Wargan, C. J. Seftor, G. J. Labow, L. D. Oman, Evaluation of version 3 total and tropospheric ozone columns from Earth Polychromatic Imaging Camera on Deep Space Climate Observatory for studying regional scale ozone variations, Frontiers in Remote Sensing 2 (2021). doi:10.3389/frsen.2021.734071.

[5] K. Yang, X. Liu, Ozone profile climatology for remote sensing retrieval algorithms, Atmospheric Measurement Techniques 12 (9) (2019) 4745–4778. doi:10.5194/amt-12-4745-2019.

[6] S. A. Carn, On the detection and monitoring of effusive eruptions using satellite SO_2 measurements, Geological Society, London, Special Publications 426 (1) (2016) 277–292.

[7] C. Ahn, O. Torres, H. Jethva, R. Tiruchirapalli, L.-K. Huang, Evaluation of aerosol properties observed by DSCOVR/EPIC instrument from the Earth-Sun Lagrange 1 orbit, Journal of Geophysical Research: Atmospheres 126 (12) (2021) e2020JD033651. doi:https://doi.org/10.1029/2020JD033651.

[8] A. Lyapustin, S. Go, S. Korkin, Y. Wang, O. Torres, H. Jethva, A. Marshak, Retrievals of aerosol optical depth and spectral absorption from DSCOVR EPIC, Frontiers in Remote Sensing 2 (2021). doi:10.3389/frsen.2021.645794.

[9] Z. Lu, J. Wang, X. Xu, X. Chen, S. Kondragunta, O. Torres, E. M. Wilcox, J. Zeng, Hourly mapping of the layer height of thick smoke plumes over the western U.S. in 2020 severe fire season, Frontiers in Remote Sensing 2 (2021). doi:10.3389/frsen.2021.766628.

[10] A. Delgado-Bonal, A. Marshak, Y. Yang, L. Oreopoulos, Daytime variability of cloud fraction from DSCOVR/EPIC observations, Journal of Geophysical Research: Atmospheres 125 (10) (2020). doi:10.1029/2019jd031488.

[11] A. Delgado-Bonal, A. Marshak, Y. Yang, L. Oreopoulos, Cloud height daytime variability from DSCOVR/EPIC and GOES-R/ABI observations, Frontiers in Remote Sensing 3 (2022). doi:10.3389/frsen.2022.780243.

[12] A. Lyapustin, Y. Wang, S. Go, M. Choi, S. Korkin, D. Huang, Y. Knyazikhin, K. Blank, A. Marshak, Atmospheric correction of DSCOVR EPIC: Version 2 MAIAC algorithm, Frontiers in Remote Sensing 2 (2021). doi:10.3389/frsen.2021.748362.

[13] A. Marshak, Y. Knyazikhin, The spectral invariant approximation within canopy radiative transfer to support the use of the EPIC/DSCOVR oxygen B-band for monitoring vegetation, Journal of Quantitative Spectroscopy and Radiative Transfer 191 (2017) 7–12. doi:10.1016/j.jqsrt.2017.01.015.

[14] X. Ni, Y. Knyazikhin, Y. Sun, X. She, W. Guo, O. Panferov, R. B. Myneni, Vegetation angular signatures of equatorial forests from DSCOVR EPIC and Terra MISR observations, Frontiers in Remote Sensing 2 (2021). doi:10.3389/frsen.2021.766805.

[15] W. Song, X. Mu, T. R. McVicar, Y. Knyazikhin, X. Liu, L. Wang, Z. Niu, G. Yan, Global quasi-daily fractional vegetation cover estimated from the DSCOVR EPIC directional hotspot dataset, Remote Sensing of Environment 269 (2022) 112835. doi:10.1016/j.rse.2021.112835.

[16] C. J. Stubenrauch, W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. D. Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. C. Maddux, W. P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, G. Zhao, Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel, Bulletin of the American Meteorological Society 94 (7) (2013) 1031–1049. doi:10.1175/BAMS-D-12-00117.1.

[17] A. Doicu, D. S. Efremenko, D. Loyola, T. Trautmann, Discrete ordinate method with matrix exponential for stochastic radiative transfer in broken clouds, Journal of Quantitative Spectroscopy and Radiative Transfer 138 (2014) 1–16. doi:10.1016/j.jqsrt.2014.01.011.

[18] W. Zdunkowski, T. Trautmann, A. Bott, Radiation in the Atmosphere - A Course in Theoretical Meteorology, Cambridge University Press, 2007.

[19] U. Platt, J. Stutz, Differential Absorption Spectroscopy, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 135–174. doi:10.1007/978-3-540-75776-4_6.

[20] D. Q. Wark, D. M. Mercer, Absorption in the atmosphere by the oxygen A band, Applied Optics 4 (7) (1965) 839–845. doi:10.1364/AO.4.000839.

[21] F. R. Spiering, M. B. Kiseleva, N. N. Filippov, L. van Kesteren, W. J. van der Zande, Collision-induced absorption in the O_2 B-band region near 670 nm, Physical Chemistry Chemical Physics 13 (2011) 9616–9621. doi:10.1039/C1CP20403C.

[22] A. N. Tikhonov, V. I. Arsenin, Solutions of ill-posed problems, Scripta series in mathematics, Winston & Sons, Washington, D.C., 1977.

[23] A. Doicu, T. Trautmann, F. Schreier, Numerical Regularization for Atmospheric Inverse Problems, Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-05439-6.

[24] K. Stamnes, S.-C. Tsay, W. Wiscombe, K. Jayaweera, Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, Applied Optics 27 (12) (1988) 2502. doi:10.1364/ao.27.002502.

[25] A. Doicu, 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) (2009) 146–158. doi:10.1016/j.jqsrt.2008.09.014.

[26] R. Spurr, T. Kurosu, K. Chance, A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, Journal of Quantitative Spectroscopy and Radiative Transfer 68 (6) (2001) 689–735. doi:10.1016/s0022-4073(00)00055-8.

[27] Y. Yang, A. Marshak, J. Mao, A. Lyapustin, J. Herman, A method of retrieving cloud top height and cloud geometrical thickness with oxygen A and B bands for the Deep Space Climate Observatory (DSCOVR) mission: Radiative transfer simulations, Journal of Quantitative Spectroscopy and Radiative Transfer 122 (2013) 141–149. doi:10.1016/j.jqsrt.2012.09.017.

[28] A. B. Davis, G. Merlin, C. Cornet, L. C. Labonnote, J. Riédi, N. Ferlay, P. Dubuisson, Q. Min, Y. Yang, A. Marshak, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: An optimal estimation approach, Journal of Quantitative Spectroscopy and Radiative Transfer 216 (2018) 6–16. doi:10.1016/j.jqsrt.2018.05.007.

[29] A. B. Davis, N. Ferlay, Q. Libois, A. Marshak, Y. Yang, Q. Min, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: A physics-based approach, Journal of Quantitative Spectroscopy and Radiative Transfer 220 (2018) 84–96. doi:10.1016/j.jqsrt.2018.09.006.

Chapter 2
The EPIC instrument

In this chapter, we first provide a short overview of the EPIC instrument and the radiometric calibration and registration performed in the EPIC Level 1B product (version 2).

The EPIC operational processor is in charge of converting the Level 0 raw EPIC measurements, i.e. the number of photons reaching the detector, into *count rates* by dividing over the exposure times [30]. In this conversion, several instrument effects are also corrected, e.g. the dark offset and dark rate signals; the nonlinearity of EPIC electronics response; the variation of EPIC sensitivity with the temperature; the differences in pixel-to-pixel sensitivity (*flat fielding*); the vignetting and etaloning effects; and the stray-light effect. These corrections together with the evaluation of the EPIC geolocation algorithm [31] for every channel produce the EPIC Level 1A product. This product is used as input to generate the Level 1B product, in which the EPIC images within every dataset are collocated to share common geolocation arrays. After an analysis of the EPIC Level 1B dataset collection, we concluded that at least two additional corrections need to be performed before we can use these data for the retrieval of cloud properties. In summary, we developed two algorithms that solve the following issues:

1. The EPIC geolocation for the Level 1B datasets (version 2) has a misregistration of about 5 pixels between the arrays of measurements and the arrays of Earth coordinates. We developed an automatic algorithm that improves the geolocation by calculating a motion (i.e. rotation and translation) plus a radial distortion correction. The optimal registration parameters are obtained by comparison of expected and actual locations of coastline features found using computer vision techniques. With our approach, the misregistration is reduced from 5 to about 1 ground pixel [32].

2. The EPIC calibration factors are not constant in time: the instrument degradation effects need to be considered. The degradation factors f_m^{degr} are modelled as a linear polynomial $f_m^{\text{degr}}(t) = 1 + c_m^{\text{degr}}(t - t_0)$, being t_0 a reference point in time t and c_m^{degr} the degradation ratio, and they must be applied to fix the original calibration factors. After this correction, the calibration factors for the ultraviolet channels increase yearly by about 0.6%, and for the visible and near-infrared channels they decrease yearly from 0.3% to 0.8%.

Scientific contributions

V. Molina García, S. Sasi, D. S. Efremenko, D. Loyola, Improvement of EPIC/DSCOVR image registration by means of automatic coastline detection, Remote Sensing 11 (15) (2019) 1747.
doi:10.3390/rs11151747

2.1 Instrument description and radiometric calibration

The instrument EPIC (Earth Polychromatic Imaging Camera) is a spectroradiometer with ten spectral channels distributed in the ultraviolet, visible and near infrared, ranging from 317 to 780 nm. The instrument consists of a CCD (Charge-Coupled Device) detector together with a 30-cm aperture Cassegrain telescope. Its maximum signal-to-noise ratio (SNR) is 290 [33].

A description of the EPIC channels is shown in Table 2.1. The mechanism allowing EPIC to measure these ten spectral channels consists of two wheels, with six holes per wheel in which ten spectral filters are placed (two of the wheel holes remain empty). The transmission functions of the filters, which determine the instrument spectral response function (ISRF), is shown in Figure 2.1. For all EPIC channels the spectral width is found below 3 nm. Among these channels, EPIC has two pairs of reference and absorption channels in the oxygen A-band (780 and 764 nm) and B-band (680 and 688 nm). These oxygen bands have been used successfully with other sensors to retrieve macrophysical cloud properties [34, 35].

Table 2.1: Description of the EPIC channels, adapted from EPIC's main website [36].

Channel	Central wavelength / nm	Full width / nm	Primary application
1	317.5 ± 0.1	1.0 ± 0.2	Ozone, sulfur dioxide
2	325.0 ± 0.1	2.0 ± 0.2	Ozone
3	340.0 ± 0.3	3.0 ± 0.6	Ozone, aerosols
4	388.0 ± 0.3	3.0 ± 0.6	Aerosols, clouds
5	443.0 ± 1.0	3.0 ± 0.6	Aerosols, clouds
6	551.0 ± 1.0	3.0 ± 0.6	Aerosols
7	680.0 ± 0.2	3.0 ± 0.6	Aerosols, clouds, vegetation
8	687.75 ± 0.20	0.80 ± 0.20	Aerosols, clouds, vegetation
9	764.0 ± 0.2	1.0 ± 0.2	Cloud height
10	779.5 ± 0.3	2.0 ± 0.4	Clouds, vegetation

Every image for every EPIC channel has a shape of 2048 × 2048 detector pixels, with a temporal resolution within one or two hours depending on the season, and a varying ground pixel size of about 12 × 12 km^2 at the center of the images. The effective spatial resolution is coarser, however, because the EPIC images are averaged on board the DSCOVR spacecraft to reduce the amount of data that needs to be transmitted.

The EPIC instrument lacks of an onboard facility for radiometric calibration. To obtain physical magnitudes from the calibrated counts per second (cps) measured by the EPIC instrument, it is mandatory to perform a vicarious calibration in which the EPIC observations are compared to other measurements that can be considered as reference. In this regard, Geogdzhayev and Marshak [37] calibrated four EPIC visible and near-infrared channels (443, 551, 680 and 780 nm) by comparison with measurements from the MODIS sensor on board the Terra and Aqua satellites, and two oxygen-absorbing bands (688 and 764 nm) by means

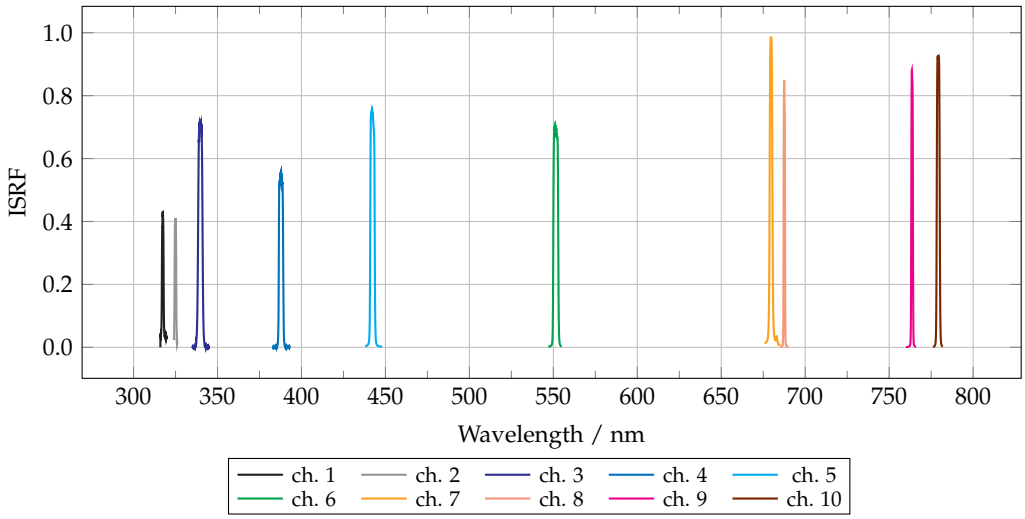


Figure 2.1: ISRF for the EPIC channels. The last four EPIC channels constitute the absorption-reference pairs in the oxygen A- and B- bands.

of EPIC lunar observations, and it was found that for these EPIC channels it is enough to define a constant calibration factor to convert from engineering units into top-of-atmosphere reflectance. Herman et al. calibrated the four ultraviolet channels by comparison with measurements from OMI on board Aura and OMPS on board Suomi-NPP [33], although for this case the calibration coefficients are not constant but they increase about 1.6% per year on average. In Table 2.2, the original calibration factors $\{k_{m0}\}_{m=1}^{10}$ for the ten EPIC channels are summarised. Their accuracy is about 1 to 3%. The calibration factors k_m can be used to convert the EPIC measurements C_m (in cps) into top-of-atmosphere reflectances $[\rho_m \cos\theta_0]$ by the relation

$$\rho_m \cos\theta_0 = k_m C_m = \frac{\pi L_m}{\overline{E}_{0m}/R_{\mathrm{ES}}^2}, \tag{2.1}$$

where, for every EPIC channel m, we denote with L_m the radiance, $\overline{E}_{0m}$ the exoatmospheric mean solar irradiance, ρ_m the true reflectance factor of the Earth, $\cos\theta_0$ the solar zenith-directional cosine on the Earth view scene, and R_{ES} is the ratio of the Earth-Sun mean distance r_0 and the Earth-Sun distance r for the given day number n_{day}:

$$R_{\mathrm{ES}} = \frac{r}{r_0} \simeq 1 - 0.016729 \cos\left[\frac{2\pi}{360}\, 0.9856(n_{\mathrm{day}} - 4)\right]. \tag{2.2}$$

Table 2.2: Calibration factors k_{m0} for the EPIC channels [38].

Channel m	1	2	3	4	5
k_{m0} **/ cps**$^{-1}$	1.216×10^{-4}	1.111×10^{-4}	1.975×10^{-5}	2.685×10^{-5}	8.340×10^{-6}
Channel m	6	7	8	9	10
k_{m0} **/ cps**$^{-1}$	6.660×10^{-6}	9.300×10^{-6}	2.020×10^{-5}	2.360×10^{-5}	1.435×10^{-5}

2.2 Geolocation

The operational EPIC geolocation algorithm computes the Earth latitudes and longitudes corresponding to every EPIC pixel location. These arrays of Earth coordinates are computed by means of a navigation algorithm that determines the DSCOVR spacecraft location and orientation and, taking into account the optical distortions of the EPIC telescope, it maps the 3D Earth coordinates into the 2D detector grid [31]. In the EPIC Level 1A products, every channel stores its own array of Earth coordinates, as the images within one acquisition are taken asynchronously and they are not collocated. In the EPIC Level 1B products, the images from every Level 1A product are collocated into a common array of Earth coordinates.

With few exceptions over the entire EPIC mission, the channel collocation is performed successfully. However, and although the operational navigation algorithm provides a good initial estimation for the geolocation of the EPIC Level 1B images, they still present a misregistration that is not neglectable and it is especially noticeable when observing the expected and actual locations of the land bodies (Figure 2.2). The misregistration issue was already reported and briefly described for the version 1 of the EPIC Level 1B products by Haney et al. [39], and it persists in the EPIC Level 1B version 2.

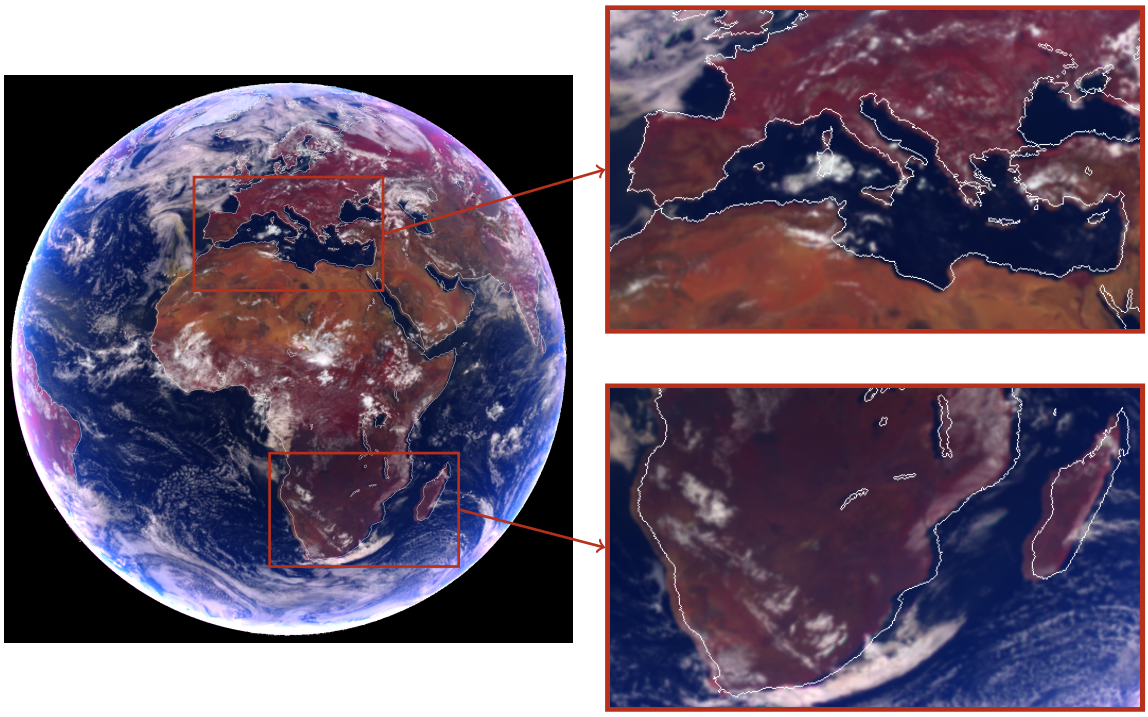


Figure 2.2: Reflectances from EPIC Level 1B version 2 on 2018-08-04 10:44:25 UTC using the false-RGB channels at 780, 551 and 388 nm. The coastlines are drawn in white line. On the left side, the complete Earth image is shown. On the right side, we zoom into two regions in which the misregistration is visible.

The scientific algorithms for the retrieval of atmospheric properties either make use of the EPIC geolocation information directly (e.g. the Optical Cloud Recognition Algorithm [40]) or rely on ancillary information that depends on the Earth coordinates, e.g. the surface altitude and albedo. Therefore, the misregistration of the EPIC Level 1B version 2 imagery difficults the use of the EPIC measurements to generate EPIC Level 2 atmospheric products.

To solve the misregistration, we developed an automatic EPIC registration algorithm [32] that is based on the following assumptions:

1. The original geolocation error is smaller than 10 detector pixels.
2. It is possible to find pairs of EPIC Level 1B pixels corresponding to expected and actual ground locations.
3. It is possible to model the registration correction as a function that transforms wrong measurement pixel indices into correct measurement pixel indices, leaving the original geolocation arrays untouched.

This registration algorithm was applied on false-RGB images consisting of the EPIC true reflectance values at 780, 551 and 388 nm, with the northern polar region of the Earth located at the top of the image (as in Figure 2.2). The true reflectance ρ can be obtained from the EPIC Level 1B images by converting the engineering units with the calibration factors described in Section 2.1 and dividing by the solar zenith-directional cosine $\cos\theta_0$. The choice of the false-RGB image channels is arbitrary, i.e. other channels can be used as long as oceans and land bodies can be clearly distinguished. The registration obtained for such false-RGB images can be used for the remaining EPIC channels because they are collocated.

2.2.1 Identification of expected and actual feature locations

The first step to fix the misregistration of an EPIC Level 1B image is to identify singular locations (features) for which it is possible to know where they should be found and where they are actually found. The coastlines are a good source of these features, since the expected (correct) coastlines can be inferred from the arrays of Earth coordinates in combination with a land-ocean mask, and the actual (incorrect) coastlines can be found in the EPIC measurement images by means of edge detection algorithms (e.g. the Canny algorithm [41]).

We show one example of such theoretical (expected) and radiometric (actual) coastlines in Figure 2.3. From this figure it can be observed that the radiometric coastline is polluted in general with other detected edges, e.g. cloud borders or interfaces between two different land cover types. However, they do not need to be filtered as long as we restrict the identification of features to the neighbourhood of the theoretical coastline.

Once the expected and actual coastline masks are computed, we must find features that are common to both theoretical and radiometric coastlines, as long as the misregistration of the radiometric coastline is not severe. This is an image matching problem that can be addressed

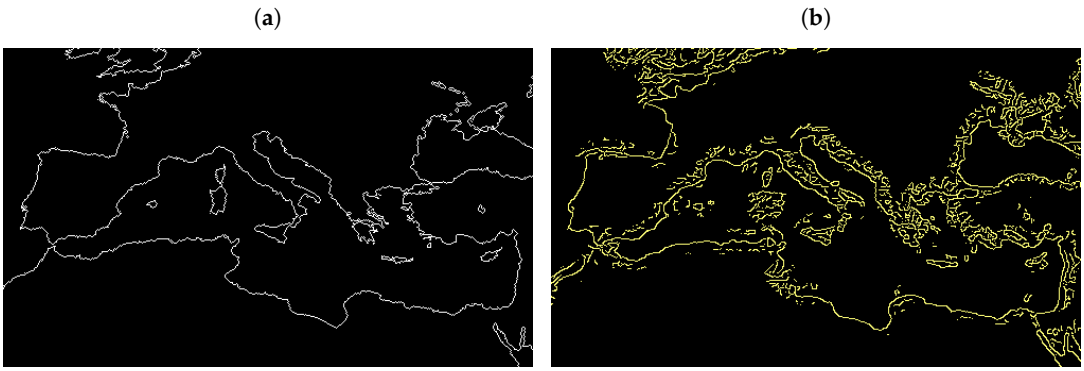


Figure 2.3: Comparison of the (**a**) theoretical and (**b**) radiometric coastlines for one of the EPIC regions shown in Figure 2.2. The radiometric coastline is polluted with other type of borders detected by the Canny algorithm (e.g. clouds edges and land cover interfaces).

with the conventional pattern recognition techniques of common use in computer vision, and whose procedure is usually divided in three tasks:

1. Detect keypoints (e.g. edges, corners or regions of interest) in the compared images.
2. Describe every keypoint by a descriptor vector that uses neighbourhood information.
3. Match pairs of keypoints from both images using the similarity of their descriptors.

We choose the detector and descriptor ORB (Oriented FAST and Rotated BRIEF) [42] through its OpenCV [43] interface because this algorithm is not patented and the amount of detected keypoints in our coastline binary images is enough for this application. Because the descriptor vectors that ORB computes are binary (i.e. their components are either 0 or 1), the keypoints detected and described with ORB for both the theoretical and radiometric coastlines are matched by using a brute-force matcher with the Hamming distance as measurement, i.e. the number of different bits in the descriptor vectors of the compared keypoints.

Given a set of n_1 theoretical keypoints $\{\boldsymbol{z}_{1k_1}\}_{k_1=1}^{n_1}$ with descriptor vectors $\{\boldsymbol{d}_{1k_1}\}_{k_1=1}^{n_1}$ and a set of n_2 radiometric keypoints $\{\boldsymbol{z}_{2k_2}\}_{k_2=1}^{n_2}$ with descriptor vectors $\{\boldsymbol{d}_{2k_2}\}_{k_2=1}^{n_2}$, the brute-force matcher states that a pair of keypoints $(\boldsymbol{z}_{1i}, \boldsymbol{z}_{2j})$ is a valid matching pair if $\boldsymbol{z}_{2j}$ is the keypoint of the radiometric coastline with the smallest descriptor distance to the keypoint $\boldsymbol{z}_{1i}$ in the theoretical coastline and the reciprocal statement also holds. After finding matching keypoint pairs with the minimum Hamming distance criterion, which only considers the similarity of descriptor vectors, we reject clear outliers based on the spatial constraints of our problem:

1. The spatial distance between a radiometric keypoint $\boldsymbol{z}_{2j}$ and the theoretical coastline.
2. The spatial distance between the keypoints of a matching pair $(\boldsymbol{z}_{1i}, \boldsymbol{z}_{2j})$.

If any of these spatial distances is greater than a threshold (e.g. 10 pixels), the matching pair is discarded, otherwise it is preserved (Figure 2.4).

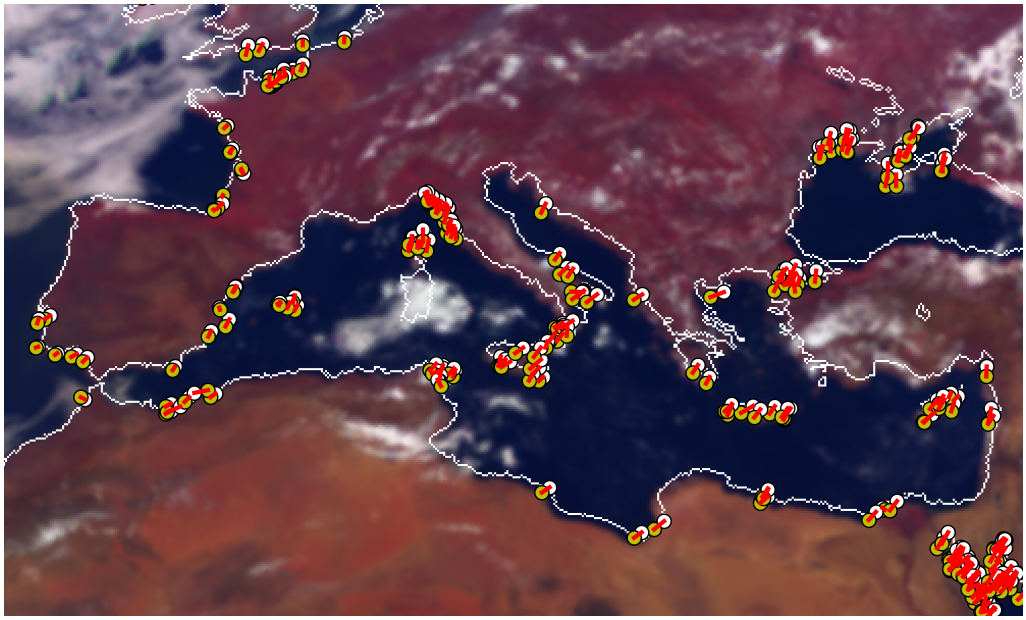


Figure 2.4: Pairs of matching keypoints found in one of the EPIC regions shown in Figure 2.2 after applying the outlier rejection step based on spatial criteria. The keypoints detected in the theoretical coastline (white lines) are shown as white points, while the corresponding keypoints detected in the EPIC radiometric coastline are shown as yellow points. Every matching pair is connected with a red segment.

2.2.2 Description of the registration function

The registration function f is the mathematical expression that describes how a "distorted" pixel located at indices $[x_d, y_d]^T$ (hereinafter the superscript ^{T} stands for "transpose") can be converted into its registered counterpart located at indices $[x_r, y_r]^T$. We model this transformation f using an empirical approach that describes common registration issues, such as optical distortions or incorrect instrument orientations, and whose expression is the following:

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_u - x_0 \\ y_u - y_0 \end{bmatrix}, \tag{2.3}$$

where $[x_s, y_s]^T$ is the shift vector, θ is the rotation angle around the center of rotation $[x_0, y_0]^T$, and $[x_u, y_u]^T$ are the pixel indices after applying a radial distortion correction by means of the single-parameter Fitzgibbon division model [44, 45]:

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + g(\lambda, r) \begin{bmatrix} x_d - x_c \\ y_d - y_c \end{bmatrix}, \qquad g(\lambda, r) = \frac{1}{1 + \lambda r^2}, \tag{2.4}$$

where λ is the first radial distortion coefficient, $[x_c, y_c]^T$ is the center of distortion and r is the distance from the distorted point to the center of distortion:

$$r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}. \tag{2.5}$$

2.2.3 Optimisation of the registration parameters

The registration function in Eq. 2.3 depends on eight unknown parameters which we need to estimate. Based on empirical observations, we apply two restrictions to simplify the problem:

1. The center of distortion and the center of rotation are identical, i.e. $x_0 = x_c$ and $y_0 = y_c$.
2. The center of distortion/rotation is assumed known and located at the image center.

With these assumptions, the transformation from $[x_d, y_d]^T$ to $[x_r, y_r]^T$ (Eq. 2.3) is simplified to:

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} g(\lambda, r)\,(x_d - x_c) \\ g(\lambda, r)\,(y_d - y_c) \end{bmatrix}, \tag{2.6}$$

which depends on four parameters (i.e. x_s, y_s, θ, λ), as the point $[x_c, y_c]^T$ is now fixed.

In order to compute the optimal transformation parameters, we adapt Eq. 2.6 for $n > 2$ known pairs of distorted and registered points $\{([x_{di}, y_{di}]^T, [x_{ri}, y_{ri}]^T)\}_{i=1}^{n}$, i.e.

$$\begin{bmatrix} x_{r1} \\ y_{r1} \\ \vdots \\ x_{rn} \\ y_{rn} \end{bmatrix} = \begin{bmatrix} x_s + x_c \\ y_s + y_c \\ \vdots \\ x_s + x_c \\ y_s + y_c \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta & \cdots & 0 & 0 \\ -\sin\theta & \cos\theta & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cos\theta & \sin\theta \\ 0 & 0 & \cdots & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} g(\lambda, r_1)\,(x_{d1} - x_c) \\ g(\lambda, r_1)\,(y_{d1} - y_c) \\ \vdots \\ g(\lambda, r_n)\,(x_{dn} - x_c) \\ g(\lambda, r_n)\,(y_{dn} - y_c) \end{bmatrix}, \tag{2.7}$$

which can be rewritten in compact form as

$$\boldsymbol{z}_r = \boldsymbol{f}(\boldsymbol{z}_d, \boldsymbol{z}_c, \boldsymbol{p}), \tag{2.8}$$

where $\boldsymbol{f}$ denotes the transformation of the distorted data vector $\boldsymbol{z}_d = [x_{d1}, y_{d1}, \ldots, x_{dn}, y_{dn}]^T$ into the registered data vector $\boldsymbol{z}_r = [x_{r1}, y_{r1}, \ldots, x_{rn}, y_{rn}]^T$ by means of a known center of distortion/rotation $\boldsymbol{z}_c = [x_c, y_c]^T$ and an unknown state vector $\boldsymbol{p} = [x_s, y_s, \theta, \lambda]^T$.

The state vector $\boldsymbol{p}$ that optimally transforms $\boldsymbol{z}_d$ into $\boldsymbol{z}_r$ can be computed by minimisation of a selected objective function $\Phi(\boldsymbol{p})$. Our fitting problem is much more sensitive to the parameters $[\theta, \lambda]^T$, i.e. small perturbations added to them cause large transformations for points far from the center of distortion/rotation $[x_c, y_c]^T$. This will have a negative impact on those images for which the number of matching features is very low, since the remaining outliers in the matching pairs will perturbate the solutions for $[\theta, \lambda]^T$ and the optimal state vector will not have enough quality. For this reason, the minimisation problem needs to include some constraints to succeed globally (i.e. in as many EPIC Level 1B datasets as possible).

In [32], we included this constraint by selecting the Tikhonov function $\mathcal{T}(\boldsymbol{p})$ as the objective function to be minimised [46]:

$$\mathcal{T}(\boldsymbol{p}) = \frac{1}{2}\left(||\boldsymbol{f}(\boldsymbol{z}_d, \boldsymbol{z}_c, \boldsymbol{p}) - \boldsymbol{z}_r||^2 + \alpha||\boldsymbol{L}(\boldsymbol{p} - \boldsymbol{p}_a)||^2\right), \tag{2.9}$$

where α is the regularisation parameter, $\boldsymbol{L}$ is the regularisation matrix, and $\boldsymbol{p}_a$ is the *a priori* state vector. Such objective function penalises the solutions that deviate in excess from the

a priori state vector $\boldsymbol{p}_\mathrm{a}$, i.e. from the external knowledge about the solution. One advantage of this approach is that the retrieval of the optimal state vector $\boldsymbol{p}$ is done individually for every EPIC Level 1B dataset. However, it also requires the selection of the regularisation parameter α and the regularisation matrix $\boldsymbol{L}$, which can turn into solutions that barely deviate from the *a priori* state vector $\boldsymbol{p}_\mathrm{a}$ if they overconstrain the trust region for the solution.

An alternative procedure to the Tikhonov approach may combine simple Gauss-Newton optimisations (i.e. the objective function is the sum of squared residuals) with constraints that can be inferred from at least three different sources:

1. Quality constraints, i.e. the quality of the matching pairs used to create $\boldsymbol{z}_\mathrm{d}$ and $\boldsymbol{z}_\mathrm{r}$.
2. Bound constraints for the state vector components p_j for $j = 0, \ldots, 3$.
3. Temporal constraints for some of the state vector components.

This alternative procedure consists of two steps, in which the first step is used to compute a first estimation of the optimal solutions and infer two of the state vector components based on the constraint analysis, and the second step repeats the retrieval of the remaining state vector components using the knowledge from the first step. It proceeds as follows:

1. Compute the unsupervised estimation of the registration vector $\boldsymbol{p}^\mathrm{u}$ for each EPIC dataset, given as the Gauss-Newton optimal solution of the registration problem with $[x_\mathrm{s}, y_\mathrm{s}, \theta, \lambda]^\mathsf{T}$ as retrieved parameters and $[0,\ 0,\ 0.5,\ -5 \times 10^{-9}]^\mathsf{T}$ as initial guess.
   (a) For every unsupervised registration vector, discard it if the quality and bound constraints are not met. The following empirical constraints were chosen:
      i. An unsupervised registration vector is discarded when there are less than 100 matching pairs, or the descriptor distance of the best pair is more than 8, or the mean descriptor distance of all matching pairs is more than 40 (Figure 2.5).
      ii. An unsupervised registration vector is discarded if any of the following bounding conditions is true:

$$x_\mathrm{s} \notin [-20, 20], \quad y_\mathrm{s} \notin [-20, 20], \quad \theta \notin [0, 2], \quad \lambda \notin [-1 \times 10^{-6}, 0]. \tag{2.10}$$

      From empirical analysis, θ is non-negative because the rotation must be counterclockwise, and λ is non-positive because the images have barrel distortion.
   (b) With the unsupervised registration vectors that were not discarded, estimate the Lomb-Scargle temporal signals of the daily-averaged state vector components.
2. Compute the supervised estimation of the registration vector $\boldsymbol{p}^\mathrm{s}$ for every EPIC dataset, given as the Gauss-Newton optimal solution of the registration problem, but using the Lomb-Scargle values from Step 1b for every EPIC dataset timestamp as initial guesses, and with only $[x_\mathrm{s}, y_\mathrm{s}]^\mathsf{T}$ as retrieved parameters, leaving $[\theta, \lambda]^\mathsf{T}$ fixed.

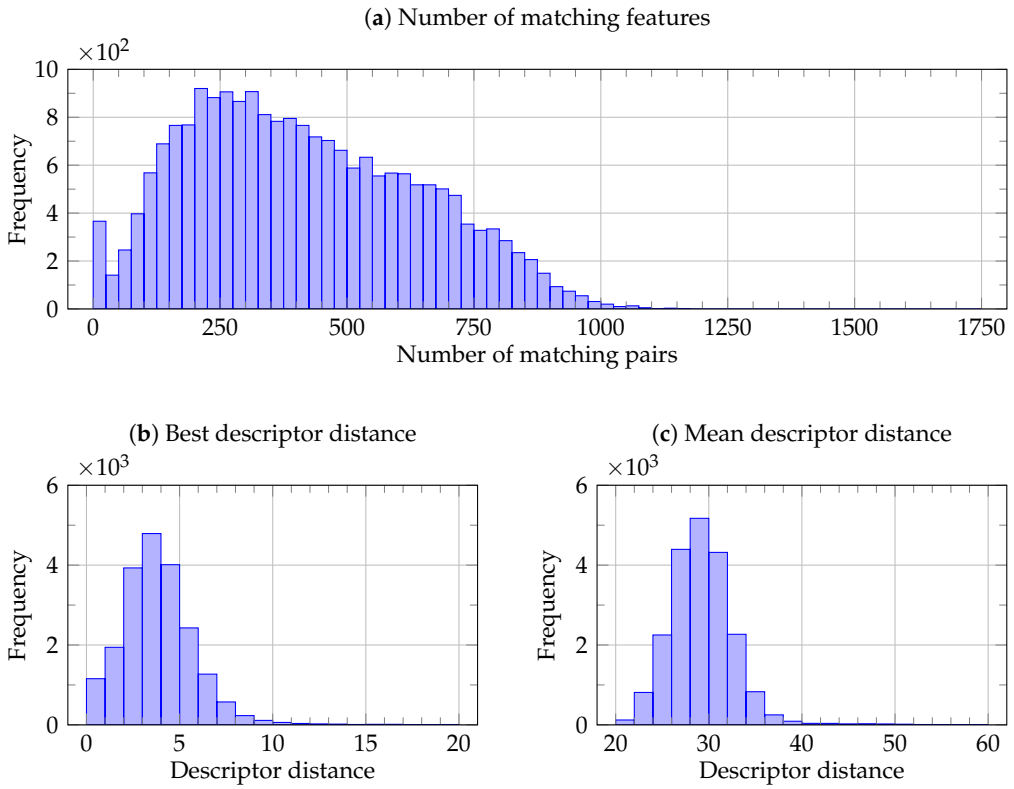


Figure 2.5: Histograms of quality constraint parameters for the EPIC mission from 2015-06-12 until 2019-06-27 containing more than 20000 datasets. The histograms show (**a**) the number of matching features found per image, (**b**) the descriptor distance for the best matching pair on each, and (**c**) the mean descriptor distance for the matching features found per image.

2.2.4 Results

We estimated the optimal transformation parameters for all the available datasets from EPIC Level 1B version 2 within the time period from 12 June 2015 to 27 June 2019. The parameters were computed with an Intel(R) Xeon(R) Gold 6152 CPU @ 2.10 GHz processor, distributed in 48 parallel processes with 2 threads per process. The clock time required for the complete period of study, which includes 20775 datasets, remained below 12 hours.

From the complete period of study, 19841 datasets (i.e. 95.5%) returned optimal unsupervised registration vectors satisfying the bounding conditions from Step 1a; this number was increased to 20711 (i.e. 99.7%) after the supervised registration, from which we infer that the constraints applied to the registration parameters θ and λ allow to improve the registration of the EPIC images that fail without supervision. The remaining 64 EPIC datasets (i.e. 0.3%) were not registered because no matching features were found, either because any of the chosen false-RGB channels was missing or the measurements were faulty. Figure 2.6 shows two examples in which the registration before and after applying our procedure can be observed.

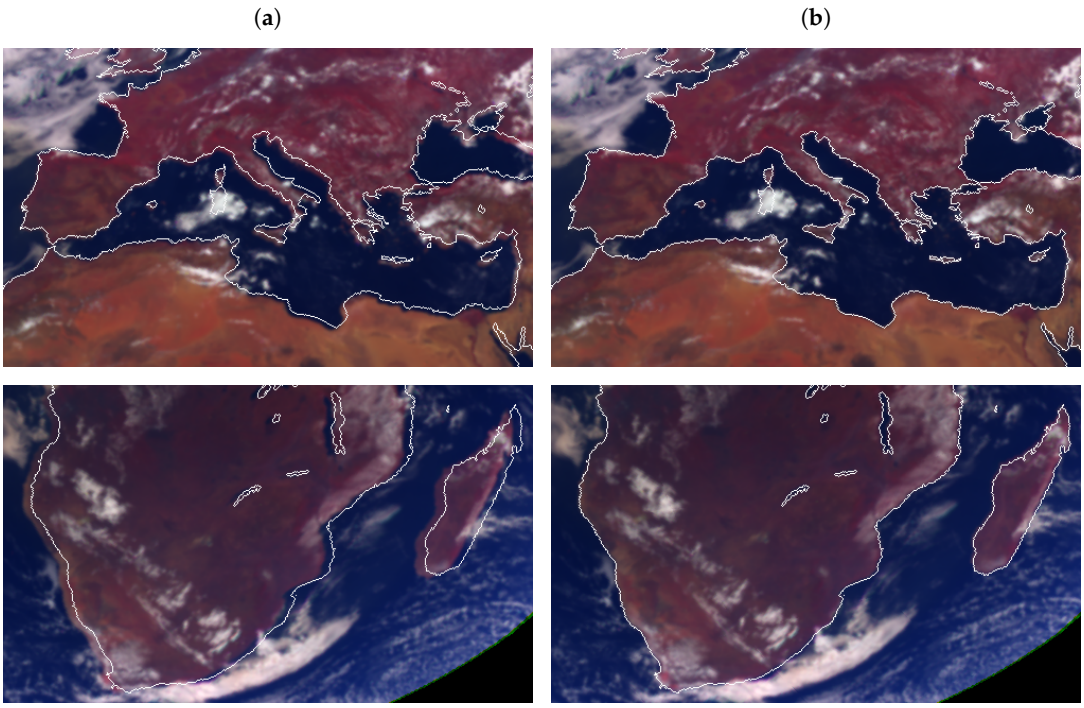


Figure 2.6: Image registration for the zoomed EPIC regions in Figure 2.2 (**a**) before and (**b**) after applying the correction. The theoretical coastlines are shown in white lines.

In Figure 2.7, we present the histograms of spatial distances between the points of every matching pair within the time series in study before and after applying the registration procedure from Section 2.2.3. In comparison to the results from [32], the time series now contains one additional year of EPIC datasets and the optimisation process is changed to the two-step Gauss-Newton approach, but the conclusions from [32] remain.

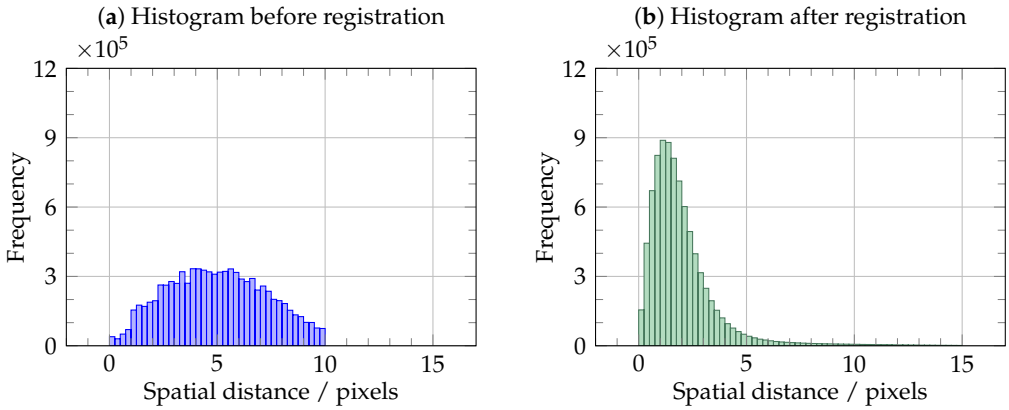


Figure 2.7: Histograms of spatial distances between the points of matching pairs (**a**) before applying registration, i.e. where they are originally located, and (**b**) after applying registration, i.e. after converting the distorted points into registered points using our registration model, for the time period under study.

In Figure 2.7**a**, the distribution of spatial distances before registration presents the higher density of occurrences in the interval of $[2.50, 7.75)$ pixels, and about 50% of the values are found in the range of $[0, 5)$ pixels. In Figure 2.7**b**, after registration, this distribution presents higher density of occurrences in the range of $[0, 2.5)$ pixels, and about 50% of the values occur in the range of $[0, 1.75)$ pixels. The mode is found in the bin of $[1.00, 1.25)$ pixels.

The existence of the right tail in the histogram of Figure 2.7**b** indicates that some matching pairs do not reduce their spatial distance after the registration, or they may even increase it. This fact may indicate us that some invalid matching pairs can pass the outlier rejection step during the matching of the theoretical and radiometric coastlines. However, it also shows that the amount of wrong matching pairs is not substantial, as they do not significantly influence in the retrieval of the registration parameters.

The histogram mode in Figure 2.7**b** is found above one pixel because of the necessary discretisation of the coastline and the actual EPIC ground pixel size. We computed the theoretical coastlines as the last contour of land pixels from a reference land-ocean mask, while the radiometric coastlines were obtained by finding intensity borders in the generated false-RGB images, which usually coincide with the actual land borders (Figure 2.8), but for certain cases the radiometric coastline can be found one pixel inside land (when the pixel is a mixture of land and ocean information) or one pixel outside land (due to water turbidity).

In Figure 2.9 we present the time evolution of the unsupervised registration parameters for the EPIC datasets that passed the constraint conditions. After an analysis of dominant frequencies based on the Lomb-Scargle method [47–49], we observe that all the parameters have an oscillatory behaviour, which can be described for every parameter as follows:

1. The horizontal shift x_s oscillates around its mean of 2.0 pixels, with one main frequency of period 174 days.

2. The vertical shift y_s oscillates around its mean value of -0.4 pixels, with two main frequencies of periods 362 and 184 days. The signal has a damping behaviour which the standard Lomb-Scargle method does not cover.

3. The rotation angle θ has a mean value of 0.4 deg, and we can observe up to three main frequencies of periods 347, 182 and 120 days. When compared to the initial results in [32], the mean value is now slightly smaller and the signal amplitude is higher.

4. The distortion parameter λ has a mean value of -5.4×10^{-9}, and its two main frequencies have periods of 175 and 90 days. As it was described for θ, the signal amplitude for λ is higher when compared to the results in [32].

It can be observed that the periods of all these main frequencies stay close to proper fractions of the Earth's orbit period around the Sun. In particular, all the registration parameters share one common frequency of period in the interval $[174, 184]$ days, which probably corresponds to the 180-day orbit period of DSCOVR's spacecraft around the Lagrangian point L_1 [50].

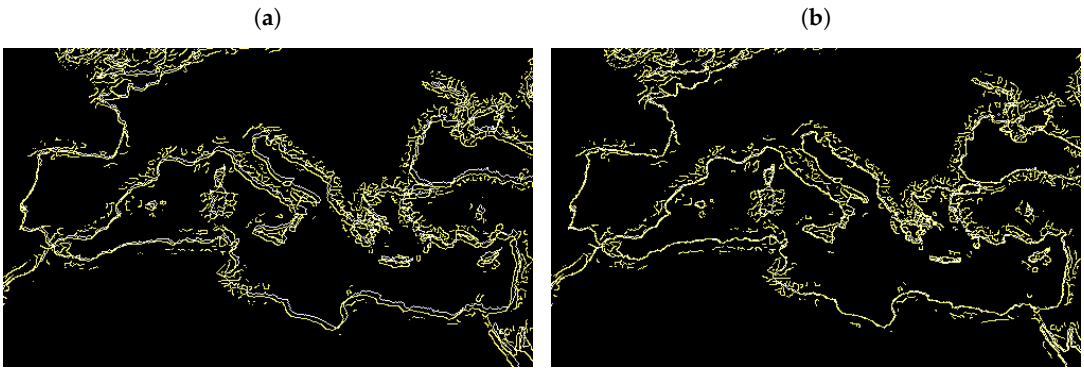


Figure 2.8: Theoretical (white lines) and the radiometric (yellow lines) coastlines for one of the EPIC regions in Figure 2.2 (**a**) before and (**b**) after applying the registration algorithm.

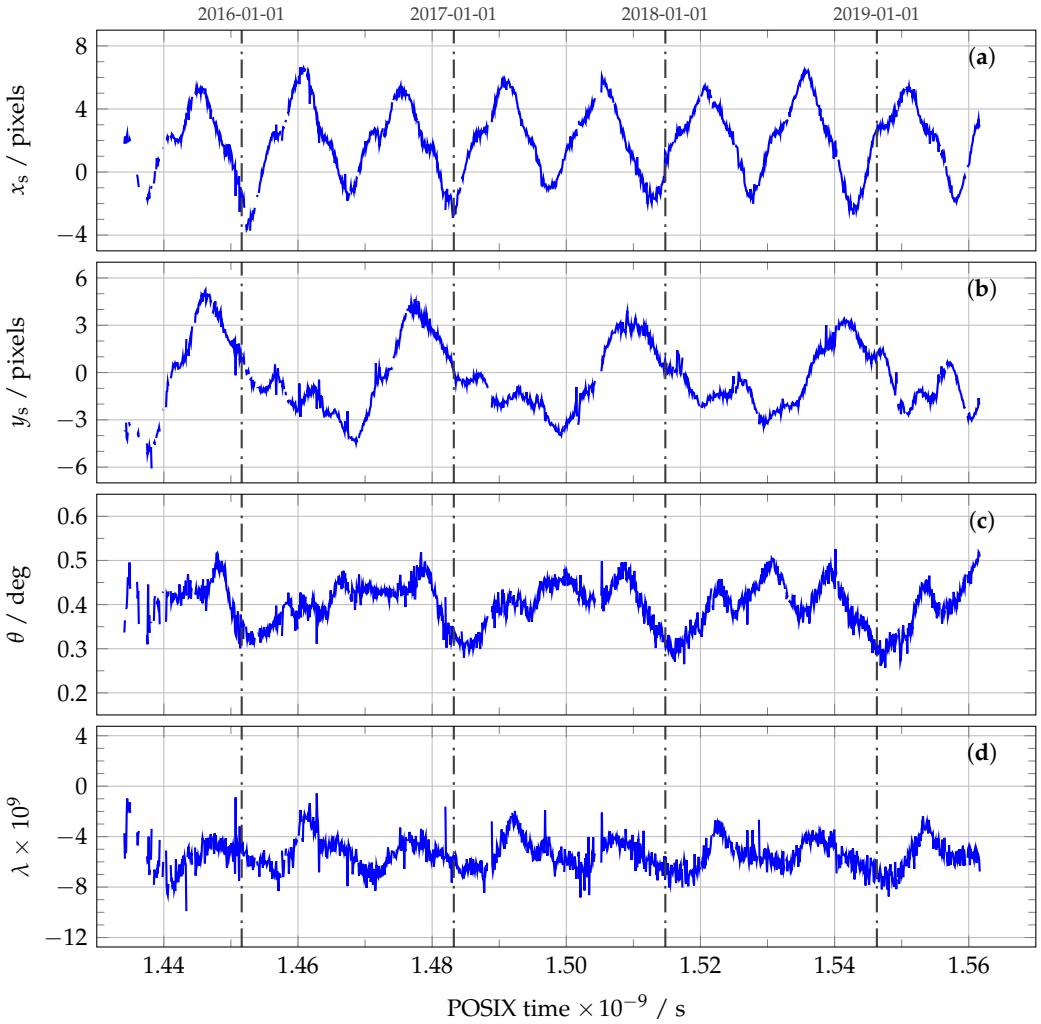


Figure 2.9: Time series plot of daily means for every registration parameter: (**a**) the horizontal shift x_s; (**b**) the vertical shift y_s; (**c**) the rotation angle θ; and (**d**) the radial distortion parameter λ. The analysis corresponds to the unsupervised step.

2.3 Degradation

The sensors on board satellites usually suffer from in-flight instrument degradation, i.e. a variation in the sensitivity and calibration parameters caused by the severe conditions outside the Earth's atmosphere. To maintain the stability of the sensor measurements during the mission lifetime, the degradation of the instrument must be estimated and removed from the measurements. This correction is normally implemented by comparison with a reference time in which the calibration parameters are known. Examples of this degradation correction can be found in the literature for low-Earth orbit sensors, such as GOME on board ERS-2 [51], SCIAMACHY on board ENVISAT [52], and GOME-2 on board the MetOp satellites [40].

For the EPIC instrument, Herman et al. reported the existence of degradation at least for the ultraviolet channels [33], which translates into a yearly increase of about 1.6% in their corresponding calibration factors. Geogdzhayev and Marshak did not report such degradation for the visible and near-infrared channels in their vicarious calibration [37].

We performed a global degradation analysis over the 10 EPIC channels for the time period from 12 June 2015 to 27 June 2019. For every date within the time period of study, we computed the global daily mean true reflectances $\{\bar{\rho}_m(t)\}_{m=1}^{10}$ using all the EPIC datasets available for that date after applying the registration correction explained in Section 2.2 and after converting from engineering units C_m into true reflectance ρ_m using the calibration coefficients $\{k_{m0}\}_{m=1}^{10}$ that are summarised in Section 2.1 and dividing by the zenith-directional cosine $\cos\theta_0$.

We define the degradation trend $\bar{\rho}_m^{\text{degr}}$ for the m-th channel as the fitting line of the global daily mean true reflectance $\bar{\rho}_m$ with the time t:

$$\bar{\rho}_m^{\text{degr}}(t) = a_m^{\text{degr}} + b_m^{\text{degr}}(t - t_0), \tag{2.11}$$

where $a_m^{\text{degr}} \neq 0$ is the value of $\bar{\rho}_m^{\text{degr}}$ at the reference time point t_0, and b_m^{degr} is the degradation slope. For the EPIC instrument, we set the reference point on 1 January 2016 12:00 UTC (with POSIX time $t_0 = 1451649600$ s), because this is the reference time in which the calibration factors for the ultraviolet channels were computed. It is common to reformulate Eq. 2.11 dividing by a_m^{degr} and so define the degradation factor f_m^{degr} as

$$f_m^{\text{degr}}(t) = \frac{\bar{\rho}_m^{\text{degr}}}{a_m^{\text{degr}}} = 1 + c_m^{\text{degr}}(t - t_0), \tag{2.12}$$

where $c_m^{\text{degr}} = b_m^{\text{degr}}/a_m^{\text{degr}}$ is the degradation ratio. Once c_m^{degr} is determined, the degradation effect can be removed from the EPIC measurements dividing the uncorrected true reflectance ρ_m by the degradation factor f_m^{degr}:

$$\rho_m^{\text{corr}}(t) = \frac{\rho_m(t)}{f_m^{\text{degr}}(t)} = \frac{k_{m0}}{f_m^{\text{degr}}(t)} \frac{C_m(t)}{\cos\theta_0} = k_m(t) \frac{C_m(t)}{\cos\theta_0}, \tag{2.13}$$

where ρ_m^{corr} is the degradation-corrected true reflectance, C_m is the EPIC measurement in cps, $\cos\theta_0$ is the solar zenith-directional cosine, and k_m is the degradation-corrected calibration factor for the m-th EPIC channel:

$$k_m(t) = \frac{k_{m0}}{f_m^{\text{degr}}(t)} = \frac{k_{m0}}{1 + c_m^{\text{degr}}(t - t_0)}, \tag{2.14}$$

being k_{m0} the original m-th calibration factor from Table 2.2. If $c_m^{\text{degr}} < 0$, then the calibration factor k_m increases with time; conversely, if $c_m^{\text{degr}} > 0$, then k_m decreases with time.

The results obtained for the degradation ratios c_m^{degr} are presented in Table 2.3. The time evolution of the global daily mean reflectances and their trend lines are shown in Figure 2.10 for the ultraviolet channels and in Figure 2.11 for the visible and near-infrared channels.

Table 2.3: Degradation ratios c_m^{degr} for the EPIC channels.

Channel m	1	2	3	4	5
c_m^{degr} / s^{-1}	-2.064×10^{-10}	-1.926×10^{-10}	-1.745×10^{-10}	-1.621×10^{-11}	9.293×10^{-11}
Channel m	6	7	8	9	10
c_m^{degr} / s^{-1}	1.214×10^{-10}	2.430×10^{-10}	2.300×10^{-10}	1.826×10^{-10}	1.750×10^{-10}

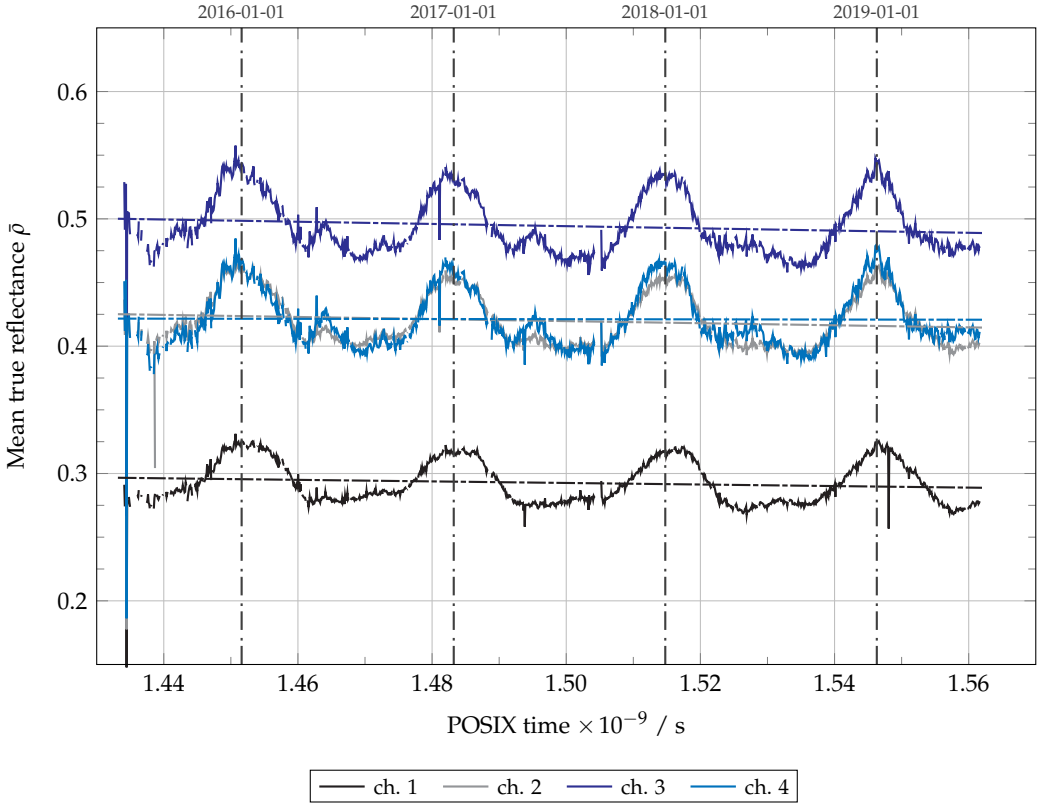


Figure 2.10: Global daily mean true reflectance $\bar{\rho}$ as a function of time t for the EPIC ultraviolet channels. The trend lines are plotted for every channel as dashed lines with the same colours.

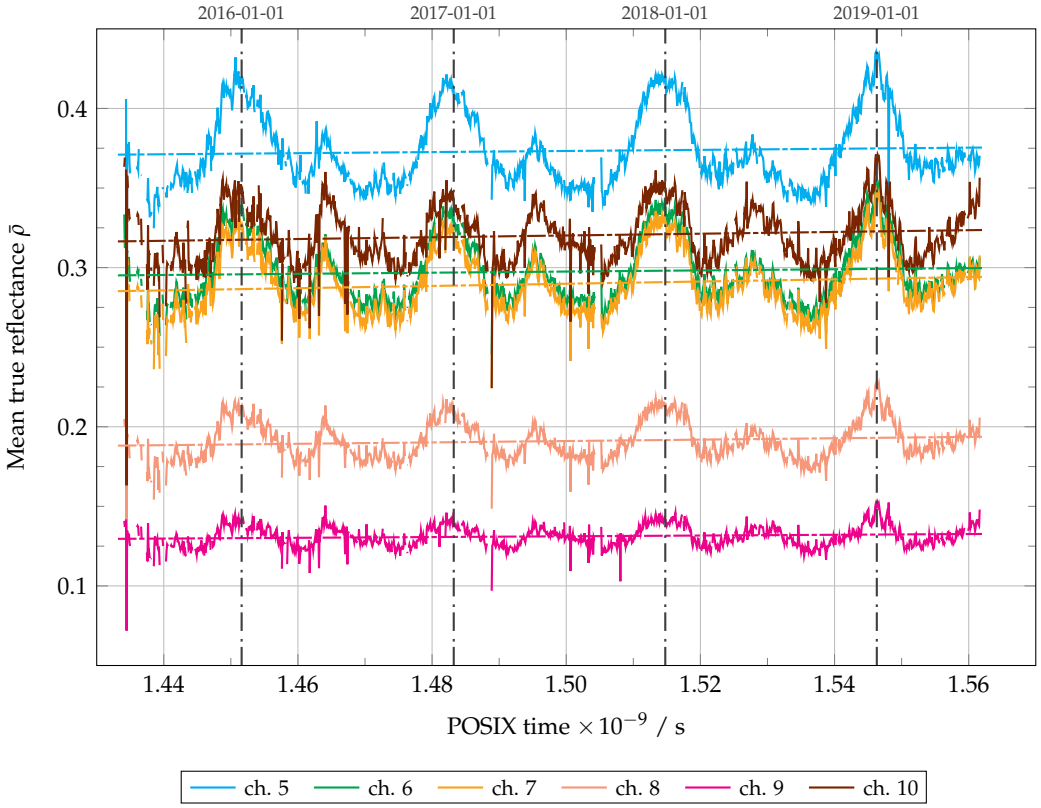


Figure 2.11: Global daily mean true reflectance $\bar{\rho}$ as a function of time t for the EPIC visible and near-infrared channels. The trend lines are plotted for every channel as dashed lines with the same colours.

From Figure 2.10, we observe decreasing time trends for the global daily mean reflectance of the EPIC ultraviolet channels. This implies, as seen in Table 2.3, that the degradation ratios c_m^{degr} are smaller than 0 and, thus, the calibration factors k_m for the ultraviolet channels increase with time. This conclusion points into the same direction as in [33], although our yearly gain is smaller. For the first three channels, f_m^{degr} is about 0.994 one year after the reference time t_0, which is equivalent to a yearly increase of 0.6% for the original calibration factors k_{m0}. The degradation of channel 4 is one order of magnitude smaller when compared to the others.

From Figure 2.11, we can see that the trends of the visible and near-infrared channels are slightly positive, and thus their degradation ratios c_m^{degr} are greater than 0. One year after the reference time t_0, the degradation ratio f_m^{degr} is roughly between 1.003 and 1.008, which means a yearly decrease from 0.3 to 0.8% in the original calibration factors k_{m0}.

References

[30] A. Marshak, J. Herman, A. Szabo, K. Blank, S. Carn, A. Cede, I. Geogdzhayev, D. Huang, L.-K. Huang, Y. Knyazikhin, M. Kowalewski, N. Krotkov, A. Lyapustin, R. McPeters, K. G. Meyer, O. Torres, Y. Yang, Earth observations from DSCOVR EPIC instrument, Bulletin of the American Meteorological Society 99 (9) (2018) 1829–1850. doi:10.1175/bams-d-17-0223.1.

[31] K. Blank, EPIC geolocation and color imagery algorithm revision 5, Tech. rep. (2017). URL https://eosweb.larc.nasa.gov/project/dscovr/DSCOVR_EPIC_Geolocation_V02.pdf

[32] V. Molina García, S. Sasi, D. S. Efremenko, D. Loyola, Improvement of EPIC/DSCOVR image registration by means of automatic coastline detection, Remote Sensing 11 (15) (2019) 1747. doi:10.3390/rs11151747.

[33] J. Herman, L. Huang, R. McPeters, J. Ziemke, A. Cede, K. Blank, Synoptic ozone, cloud reflectivity, and erythemal irradiance from sunrise to sunset for the whole earth as viewed by the DSCOVR spacecraft from the earth–sun lagrange 1 orbit, Atmospheric Measurement Techniques 11 (1) (2018) 177–194. doi:10.5194/amt-11-177-2018.

[34] D. Loyola, W. Thomas, Y. Livschitz, T. Ruppert, P. Albert, R. Hollmann, Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, IEEE Transactions on Geoscience and Remote Sensing 45 (9) (2007) 2747–2758. doi:10.1109/tgrs.2007.901043.

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

[36] Available online, https://epic.gsfc.nasa.gov/epic, (Accessed on 2018-03-17).

[37] I. V. Geogdzhayev, A. Marshak, Calibration of the DSCOVR EPIC visible and NIR channels using MODIS Terra and Aqua data and EPIC lunar observations, Atmospheric Measurement Techniques 11 (1) (2018) 359–368. doi:10.5194/amt-11-359-2018.

[38] Available online, https://eosweb.larc.nasa.gov/project/dscovr/DSCOVR_EPIC_Calibration_Factors_V02.pdf, (Accessed on 2018-03-17).

[39] C. Haney, D. Doelling, P. Minnis, R. Bhatt, B. Scarino, A. Gopalan, The calibration of the DSCOVR EPIC multiple visible channel instrument using MODIS and VIIRS as a reference, in: J. J. Butler, X. J. Xiong, X. Gu (Eds.), Earth Observing Systems XXI, SPIE, 2016. doi:10.1117/12.2238010.

[40] R. Lutz, D. Loyola, S. Gimeno García, F. Romahn, OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B, Atmospheric Measurement Techniques 9 (5) (2016) 2357–2379. doi:10.5194/amt-9-2357-2016.

[41] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (6) (1986) 679–698. doi:10.1109/tpami.1986.4767851.

[42] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision, Institute of Electrical and Electronics Engineers (IEEE), 2011, pp. 2564–2571. doi:10.1109/iccv.2011.6126544.

[43] A. Kaehler, G. Bradski, Learning OpenCV 3: Computer vision in C++ with the OpenCV library, O'Reilly Media, 2016.

[44] A. W. Fitzgibbon, Simultaneous linear estimation of multiple view geometry and lens distortion, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Institute of Electrical and Electronics Engineers (IEEE), 2001, pp. 125–132. doi:10.1109/cvpr.2001.990465.

[45] A. Wang, T. Qiu, L. Shao, A simple method of radial distortion correction with centre of distortion estimation, Journal of Mathematical Imaging and Vision 35 (3) (2009) 165–172. doi:10.1007/s10851-009-0162-1.

[46] A. N. Tikhonov, V. I. Arsenin, Solutions of ill-posed problems, Scripta series in mathematics, Winston & Sons, Washington, D.C., 1977.

[47] N. Lomb, Least-squares frequency analysis of unequally spaced data, Astrophysics and Space Science 39 (2) (1976) 447–462. doi:10.1007/bf00648343.

[48] J. Scargle, Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data, The Astrophysical Journal 263 (1982) 835–853. doi:10.1086/160554.

[49] K. Hocke, Phase estimation with the Lomb-Scargle periodogram method, Annales Geophysicae 16 (1998) 356–358.

[50] J. H. Jiang, A. J. Zhai, J. Herman, C. Zhai, R. Hu, H. Su, V. Natraj, J. Li, F. Xu, Y. L. Yung, Using Deep Space Climate Observatory measurements to study the Earth as an exoplanet, The Astronomical Journal 156 (1) (2018) 26. doi:10.3847/1538-3881/aac6e2.

[51] M. Coldewey-Egbers, S. Slijkhuis, B. Aberle, D. Loyola, Long-term analysis of GOME in-flight calibration parameters and instrument degradation, Applied Optics 47 (26) (2008) 4749–4761. doi:10.1364/ao.47.004749.

[52] L. G. Tilstra, M. de Graaf, I. Aben, P. Stammes, In-flight degradation correction of SCIAMACHY UV reflectances and absorbing aerosol index, Journal of Geophysical Research: Atmospheres 117 (D06209) (2012). doi:10.1029/2011jd016957.

Chapter 3
Radiative transfer model

In this chapter, we provide an analysis of accuracy and efficiency of several radiative transfer models that describe the dependency between the cloud macrophysical parameters and the measurements as observed by the EPIC instrument. The analysis includes two *quasi-exact* models—the discrete ordinate and matrix operator methods with matrix exponential (DOME and MOME, respectively)—and two *approximate* models—thick-cloud asymptotic model and equivalent-Lambertian cloud model. For the *quasi-exact* models, two approaches to compute the Jacobian of the measurements with respect to the cloud parameters are also studied: the linearised and the forward-adjoint approaches. The main findings are summarised as follows:

1. The approximate cloud models show worse accuracy than the *quasi-exact* models when simulating measurements in the EPIC oxygen A-band absorption channel. When compared to reference spectra, the asymptotic model errors are large for cloud optical thicknesses smaller than 10, and the equivalent-Lambertian cloud model errors are large for all cloud optical thicknesses and have a bias of 10% for optically-thick clouds.

2. The *quasi-exact* models provide measurements with a relative error less than 1.7% when using 32 discrete ordinates, and the time efficiency can be improved two orders of magnitude by combining two acceleration techniques (correlated k-distribution and principal component analysis) to 18 s per measurement without significant error increase.

3. The forward-adjoint and the linearised approaches provide similar results when evaluating radiance derivatives with respect to cloud optical thickness and cloud-top height; the computation time is decreased three orders of magnitude combining the forward-adjoint approach and PCA on correlated k-distribution to 13 s for each derivative pair.

Scientific contributions

D. S. Efremenko, V. Molina García, S. Gimeno García, A. Doicu, A review of the matrix-exponential formalism in radiative transfer, JQSRT 196 (2017) 17–45.
doi:10.1016/j.jqsrt.2017.02.015

V. Molina García, S. Sasi, D. S. Efremenko, A. Doicu, D. Loyola, Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, JQSRT 123 (2018) 228–240.
doi:10.1016/j.jqsrt.2018.03.014

V. Molina García, S. Sasi, D. S. Efremenko, A. Doicu, D. Loyola, Linearized radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, JQSRT 123 (2018) 241–251.
doi:10.1016/j.jqsrt.2018.03.008

3.1 Introduction

The estimation of atmospheric properties from remote sensing measurements (e.g. the layer height and the optical depth of clouds and aerosols) aims to find the set of atmospheric parameters $\boldsymbol{x}$ (i.e. state vector) for which a certain radiative transfer model $\boldsymbol{f}$ reproduces a set of radiances $\boldsymbol{y}^{\delta}$ (i.e. measurement vector):

$$\boldsymbol{y}^{\delta} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{b}) + \boldsymbol{\delta}, \tag{3.1}$$

where $\boldsymbol{b}$ is a set of auxiliary parameters that are assumed known (e.g. the surface height and albedo) and $\boldsymbol{\delta}$ denotes the error data vector. Such ill-posed problems are solved on the basis of non-linear optimisation techniques [53] combined with Tikhonov regularisation [54], which involves the evaluation of the radiative transfer model $\boldsymbol{f}$ and the derivatives of the model with respect to the state vector components at every iteration of the optimisation procedure.

For operational scientific applications, there is the need to find a radiative transfer model with high accuracy and time efficiency, since the lack of accuracy would produce wrong estimates of the set of radiances $\boldsymbol{y}^{\delta}$, and the lack of time efficiency would not allow to process the huge amount of measurements acquired by the instrument. An accurate method to model the radiative transfer in a molecular atmosphere generally requires a line-by-line (LBL) calculation procedure to compute the gas absorption coefficients, which may be computationally expensive if all the gas absorption lines over the spectral region of interest are considered. The computation time of a *quasi-exact* model can be reduced at the cost of the accuracy using two approaches: (1) *approximate models*, i.e. simplifications on the underlying physical description that reduce the complexity of the radiative transfer problem; and (2) *acceleration techniques*, i.e. methods that exploit the redundancy present in the optical data to reduce the number of wavelengths in which the *quasi-exact* radiative transfer problem is solved.

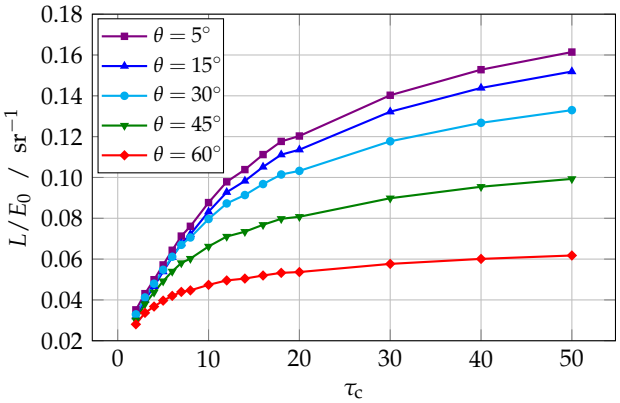


Figure 3.1: Simulated DOME radiances L at EPIC absorption oxygen A-band of the test scenarios for several viewing zenith angles θ, as functions of the cloud optical thickness τ_c. The reference number of discrete ordinates is $M = 128$. The radiances L are normalised by the solar top-of-atmosphere irradiance E_0.

3.2 Quasi-exact models

By *quasi-exact* radiative transfer model, we refer to the models that provide solutions of the radiative transfer equation under a minimum set of assumptions (i.e. the discretisation of the inhomogeneous atmosphere in N levels separating $N-1$ homogeneous layers), and which are used as starting point to design approximate models after imposing additional assumptions. The *quasi-exact* models analysed for its applicability in retrieval of cloud properties from EPIC/DSCOVR are the scalar discrete ordinate method with matrix exponential (DOME) and the scalar matrix operator method with matrix exponential (MOME). A complete description of DOME is found in [55, 56], and the specific details of the DOME and MOME implementations for the retrieval of cloud properties from the EPIC mission are found in [57].

The results in this chapter show radiance comparisons with respect to a set of reference cloud radiance simulations in the EPIC absorption oxygen A-band. The test scenarios utilise clouds modelled as a homogeneous layer of water droplets following a Gamma size distribution with mean radius of 8 µm, the cloud-top height is fixed to 4 km, the cloud geometrical thickness is set as a function of the cloud optical thickness τ_c, and the surface is assumed as Lambertian with fixed albedo of 0.06. We assume that the solar and viewing zenith angles are similar, i.e. $\theta_0 = \theta + \varepsilon_\theta$ with $0 < \varepsilon_\theta \ll 1$ deg, and the relative azimuth angle is $\Delta\varphi = 176$ deg. The number of discrete ordinates is fixed to $M = 128$. The viewing zenith angle and the cloud optical thickness are varied to generate a total number of 90 radiance simulations. The full description is found in [57]. The computation times in this chapter are user times, i.e. they are the sum of all the computation times from the parallel threads involved in the computations of the 90 test scenarios. The reference simulated radiances are presented in Figure 3.1.

We concentrated on the effects of the number of discrete ordinates M on the accuracy of the *quasi-exact* models. The number of discrete ordinates is the parameter used to discretise the viewing zenith angle in a fixed number of values and so convert the multiple-scattering integral term from the radiative transfer equation into a Gauss-Legendre quadrature sum. The number of discrete ordinates has an impact on the accuracy of the simulated radiances, as it modulates the error in the quadrature sum, but it also has an impact on the time efficiency, since more quadrature points require more computations and more time. In addition to this, the number of discrete ordinates M limits the maximum number of iterations in which the radiance L is expanded as a Fourier series in $\Delta\varphi$, which is critical due to the viewing geometry of EPIC, close to the glory region in the scattering phase function of the Mie water droplets [58], and which requires a long series expansion for the single-scattering contribution to describe the glory effect properly. In Figure 3.2, we show the relative errors in the EPIC simulated radiances for the reference scenarios when the number of discrete ordinates M is decreased from the original value 128 into 8, 16, 32 and 64. Their corresponding computation times and the speed-up factors with respect to the reference simulations are shown in Table 3.1.

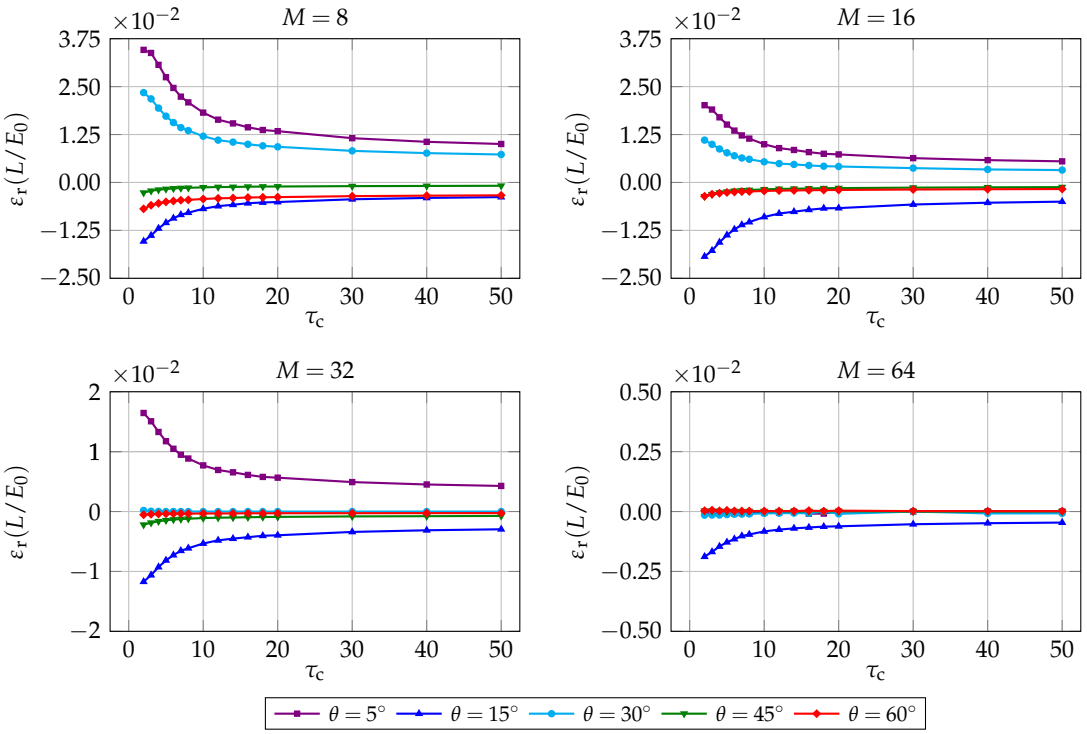


Figure 3.2: Relative errors ε_r in the DOME radiances L/E_0 of the test scenarios for different cloud optical thicknesses τ_c and viewing zenith angles θ when the number of discrete ordinates M is reduced.

As expected, from Figure 3.2 and Table 3.1 we see that reducing the number of discrete ordinates has a positive impact in the time efficiency, but it has a negative impact in the accuracy, especially for small cloud optical thicknesses, i.e. when the single-scattering term contributes more than the multiple-scattering term. The relative errors increase up to about 3.7% if $M = 8$. A compromise between accuracy and time efficiency is found when $M = 32$. In this case, the radiance relative errors remain below 1% when the cloud optical thickness is greater than 10, and below 1.7% when it is lower than 10, and the speed-up factor with respect to the reference simulations is 93.6. This is the number of discrete ordinates that we will use in the following.

Finally, in [57] we also analysed the discrepancies between the *quasi-exact* models DOME and MOME for the case in which the number of discrete ordinates is equal to 32. It was found that the radiances from both models show negligible differences for the set of reference scenarios. For this reason, in the next sections we will refer to DOME as our *quasi-exact* model.

Table 3.1: Computation times and speed-up factors for the radiance simulations in Figure 3.2 when compared to the reference simulated radiances in Figure 3.1.

M	128	64	32	16	8
Time / min	160116	14548	1711	284	67
Speed-up factor	–	11.0	93.6	563.8	2389.8

3.3 Approximate models

By *approximate* radiative transfer model, we refer to the models that are based on a *quasi-exact* model (such as DOME) and make additional assumptions in order to reduce the complexity of the radiative transfer problem. The first assumption in the approximate models is to replace the atmosphere below the cloud and the surface with an equivalent-Lambertian surface at the cloud-bottom height. From this assumption, we inspect two approximate models:

1. the asymptotic models, which, based on the assumption of a cloud optical thickness τ_c sufficiently large, derive simple analytical expressions for the reflection and transmission matrices from the asymptotic theory of optically-thick layers [59]; and
2. the equivalent-Lambertian cloud model, in which the atmosphere below the cloud-top height is replaced by an equivalent-Lambertian surface with a ground albedo equal to the spherical albedo of the atmosphere.

Figure 3.3 presents the relative errors in the simulated radiances at the EPIC absorption oxygen A-band for the test scenarios when utilising the 0-order asymptotic model and the equivalent-Lambertian cloud model. In the asymptotic model, we can observe that the model has acceptable results when the cloud optical thickness is very large, which is expected since the model starts from the assumption $\tau_c \to \infty$. However, the relative errors increase up to 24% for small values of the cloud optical thickness. The maximum relative error for $\tau_c < 10$ can be reduced to 4% if asymptotic models of higher order are used [57]. The equivalent-Lambertian cloud model appears inaccurate in the complete range of cloud optical thicknesses of study. When $\tau_c < 10$, the relative errors in the simulated radiances increase up to 40%, while for $\tau_c \to \infty$ the model tends to an underestimation bias of about 10%.

Because neither the asymptotic model nor the equivalent-Lambertian cloud model offer the accuracy of DOME with $M = 32$ discrete ordinates, these models are discarded.

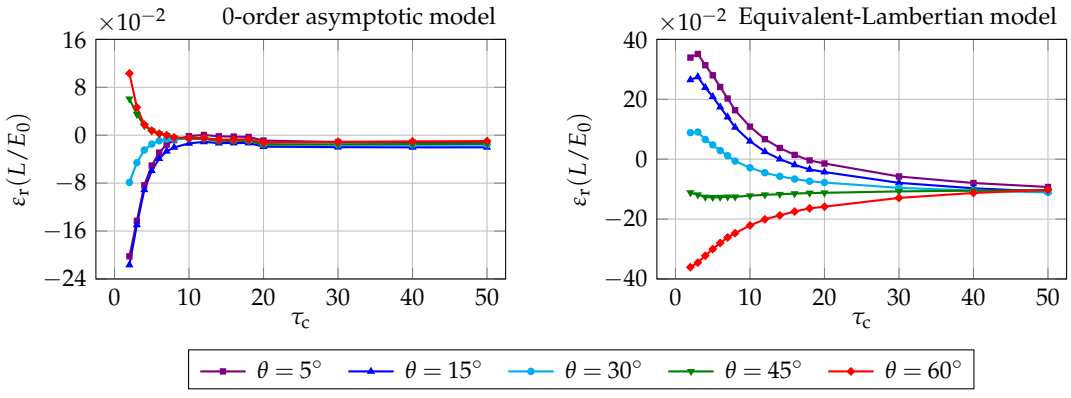


Figure 3.3: Relative errors ε_r in the radiances L/E_0 of the tests scenarios for several cloud optical thicknesses τ_c and viewing zenith angles θ when approximating with the 0-order asymptotic model and the equivalent-Lambertian cloud model. These simulations used a fixed number of discrete ordinates $M = 32$.

3.4 Acceleration techniques

The computation of a spectrum by means of a *quasi-exact* model is time-consuming in general because it implies that we need to solve the radiative transfer equation for a large number of wavelengths in order to capture all the features of the spectral signal. For a given instrument channel of central wavelength λ_c and slit function $g(\lambda_c - \lambda)$ of width s, the signal received by the instrument is the convolution of the slit function and the monochromatic radiance L_λ, i.e.

$$L(\lambda_c) = \int_{\lambda_c - s/2}^{\lambda_c + s/2} g(\lambda_c - \lambda)\, L_\lambda(\lambda)\, \mathrm{d}\lambda. \tag{3.2}$$

If we assume that the slit function $g(\lambda_c - \lambda)$ varies smoothly with the wavelength λ, then the convolution may be approximated by discretising the wavelength space into a set $\{\lambda_i\}_{i=1}^{N_\lambda}$ of N_λ equally-spaced wavelengths in the spectral interval with a discretisation step $\Delta\lambda$ so that

$$\lambda_1 - \frac{\triangle\lambda}{2} = \lambda_c - \frac{s}{2} \quad \text{and} \quad \lambda_{N_\lambda} + \frac{\triangle\lambda}{2} = \lambda_c + \frac{s}{2},$$

and then Eq. 3.2 may be rewritten as

$$L(\lambda_c) \approx \sum_{i=1}^{N_\lambda} g(\lambda_c - \lambda_i) \int_{\lambda_i - \triangle\lambda/2}^{\lambda_i + \triangle\lambda/2} L_\lambda(\lambda)\, \mathrm{d}\lambda. \tag{3.3}$$

Note that the assumption of smoothness done for the slit function g is in general not applicable to L_λ, since we are normally interested in spectral regions in which strong gas absorption takes place, and so L_λ may show big variations within a discretisation step $\Delta\lambda$ (even of several orders of magnitude).

The aim of the acceleration techniques for radiative transfer modelling is to find an accurate approximation of the integral term in Eq. 3.3, since a naive numerical integration would require, for each $i = 1, \ldots, N_\lambda$, a fine gridding of the spectral space $[\lambda_i - \triangle\lambda/2, \lambda_i + \triangle\lambda/2]$ and, thus, a huge amount of radiative transfer calls to estimate $L_\lambda(\lambda)$. Two main approches are presented here:

1. The correlated *k*-distribution method, which simplifies the integral term with an appropriate change of variable that permits us to evaluate the integral using a Gauss-Legendre quadrature with few quadrature points.

2. The principal component analysis (PCA) method, which simplifies the computation of the integrand $L_\lambda(\lambda)$ by using a prediction-correction approach, in which the prediction is performed using a simpler model (i.e. the predictor) and the correction is estimated with a correction factor evaluated in the space of optical data.

Both methods exploit the redundancy of information in the gas optical properties. We claim that these methods may also be combined to speed up the radiative transfer computations.

3.4.1 Correlated *k*-distribution method

The correlated *k*-distribution method is based on the grouping of spectral intervals according to the strength of the absorption coefficients [60]. First, since the absorption coefficient $k_{\mathtt{abs}}^{\mathtt{gas}}$ of the main absorbing gas has a greater spectral variability than the molecular and particulate scattering, we may redefine L_λ as:

$$L_\lambda(\lambda) = L_\lambda(k_{\mathtt{abs}}^{\mathtt{gas}}(\lambda)).$$

For a homogeneous atmospheric layer, the transmission within a spectral interval is independent of the line-by-line variability of $k_{\mathtt{abs}}^{\mathtt{gas}}$ with respect to the wavelength λ, but it only depends on the distribution of $k_{\mathtt{abs}}^{\mathtt{gas}}$ within that interval [61]. Let be $F_i = F_i(k_{\mathtt{abs}}^{\mathtt{gas}})$ the cumulative distribution function of $k_{\mathtt{abs}}^{\mathtt{gas}}(\lambda)$ in the spectral interval $[\lambda_i - \triangle\lambda/2, \lambda_i + \triangle\lambda/2]$, and $k_{\mathtt{abs},i}^{\mathtt{gas}}(F)$ the quantile function (i.e. the inverse of the cumulative distribution function). The correlated *k*-distribution method rewrites Eq. 3.3 after a change of variable from λ to F as:

$$L(\lambda_\mathrm{c}) = \triangle\lambda \sum_{i=1}^{N_\lambda} g(\lambda_\mathrm{c} - \lambda_i) \int_0^1 L_\lambda(k_{\mathtt{abs},i}^{\mathtt{gas}}(F))\,\mathrm{d}F, \tag{3.4}$$

and, selecting a Gauss-Legendre quadrature of pairs $\{(F_j, w_j)\}_{j=1}^{N_\mathrm{q}}$, Eq. 3.4 becomes:

$$L(\lambda_\mathrm{c}) = \triangle\lambda \sum_{i=1}^{N_\lambda} g(\lambda_\mathrm{c} - \lambda_i) \sum_{j=1}^{N_\mathrm{q}} w_j\, L_\lambda(k_{\mathtt{abs},i}^{\mathtt{gas}}(F_j)). \tag{3.5}$$

The values $k_{\mathtt{abs},i}^{\mathtt{gas}}(F_j)$ can be computed by inverting the cumulative distribution functions F_i of the line-by-line gas absorption coefficients or, when using the exponential-sum fitting of transmittance method [62], by solving a nonlinear least squares problem. Eq. 3.5 can be rewritten as a quadrature rule of the original Eq. 3.2:

$$L(\lambda_\mathrm{c}) = \sum_{p=1}^{\overline{N}_\lambda} \overline{w}_p\, g(\lambda_\mathrm{c} - \overline{\lambda}_p)\, L_\lambda(k_{\mathtt{abs}}^{\mathtt{gas}}(\overline{\lambda}_p)) \tag{3.6}$$

after defining a new set of $\overline{N}_\lambda = N_\lambda\, N_\mathrm{q}$ wavelengths $\{\overline{\lambda}_p\}_{p=1}^{\overline{N}_\lambda}$ and weights $\{\overline{w}_p\}_{p=1}^{\overline{N}_\lambda}$, through the following relations:

$$\overline{\lambda}_p = \lambda_i, \qquad \overline{w}_p = \triangle\lambda\, w_j, \tag{3.7}$$

with index p given by

$$p = (i-1)\, N_\mathrm{q} + j$$

for each $i = 1, \ldots, N_\lambda$ and $j = 1, \ldots, N_\mathrm{q}$. Note that $\{\overline{\lambda}_p\}_{p=1}^{\overline{N}_\lambda}$ consists of N_λ groups of N_q identical wavelengths, and that $k_{\mathtt{abs}}^{\mathtt{gas}}(\overline{\lambda}_p) = k_{\mathtt{abs},i}^{\mathtt{gas}}(F_j)$.

3.4.2 Principal component analysis method

The principal component analysis (PCA) method is used in the context of radiative transfer forward modelling as a dimensionality reduction technique of the optical properties [63–65]. This method assumes that the spectral radiance $L_\lambda(\lambda)$ in Eq. 3.3 can be written as the product of a predictor and a corrector, i.e.

$$L_\lambda(\lambda) = L_{\lambda,\mathrm{a}}(\lambda)\exp[f_L(\lambda)], \tag{3.8}$$

where $L_{\lambda,\mathrm{a}}(\lambda)$ is a radiance provided by a predictor radiative transfer model, and $f_L(\lambda)$ is a correction factor. Setting a number of wavelengths N_λ large enough to capture the gas absorption features, we can convert Eq. 3.3 into the classical naive numerical integration, i.e.

$$L(\lambda_\mathrm{c}) = \sum_{i=1}^{N_\lambda} g(\lambda_\mathrm{c}-\lambda_i)L_\lambda(\lambda_i)\,\Delta\lambda = \sum_{i=1}^{N_\lambda} g(\lambda_\mathrm{c}-\lambda_i)L_{\lambda,\mathrm{a}}(\lambda_i)\exp[f_L(\lambda_i)]\,\Delta\lambda,$$

requiring N_λ evaluations of the predictor radiative transfer model, which is faster than the *quasi-exact* radiative transfer model, and N_λ evaluations of the correction factor f_L.

An efficient and accurate method for computing $f_L(\lambda)$ is given by Natraj et al. [63, 64]. If the atmosphere is discretised in N levels, for every wavelength in the set $\{\lambda_i\}_{i=1}^{N_\lambda}$ we can define the $(2N-2)$-dimensional optical vector

$$\boldsymbol{x}_i \coloneqq \boldsymbol{x}(\lambda_i) = \left[\log \bar{k}^{\mathtt{gas}}_{\mathtt{abs},n}(\lambda_i),\ \log \bar{k}^{\mathtt{mol}}_{\mathtt{sct},n}(\lambda_i)\right]^\mathsf{T},$$

where $\bar{k}^{\mathtt{gas}}_{\mathtt{abs},n}(\lambda_i)$ and $\bar{k}^{\mathtt{mol}}_{\mathtt{sct},n}(\lambda_i)$ are the mean gas absorption coefficient and the mean molecular scattering coefficient, respectively, for the wavelength λ_i and the n-th atmospheric layer, with $n = 1,\ldots,N-1$. This vector $\boldsymbol{x}_i$ encapsulates the variability of the optical parameters for a given wavelength. Principal component analysis is then performed over the set of optical vectors $\{\boldsymbol{x}_i\}_{i=1}^{N_\lambda}$, i.e. we apply a change into an orthogonal basis $\{\boldsymbol{u}_l\}_{l=1}^{N_\mathrm{u}}$ in an alternative N_u-dimensional space such that

$$\boldsymbol{x}_i \approx \bar{\boldsymbol{x}} + \sum_{l=1}^{N_\mathrm{u}} q_{il}\,\boldsymbol{u}_l \coloneqq \bar{\boldsymbol{x}} + \Delta\boldsymbol{x}_i,$$

where $\bar{\boldsymbol{x}}$ is the mean vector of $\{\boldsymbol{x}_i\}_{i=1}^{N_\lambda}$, and $\{q_{il}\}_{l=1}^{N_\mathrm{u}}$ is the set of components of $\boldsymbol{x}_i$ under the new basis. Finally, the correction factor f_L is expanded as a second-order Taylor expansion:

$$f_L(\lambda_i) \equiv f_L(\boldsymbol{x}_i) = f_L(\bar{\boldsymbol{x}}) + \nabla f_L(\bar{\boldsymbol{x}})\,\Delta\boldsymbol{x}_i + \frac{1}{2}\Delta\boldsymbol{x}_i^\mathsf{T}\,\nabla^2 f_L(\bar{\boldsymbol{x}})\,\Delta\boldsymbol{x}_i.$$

The partial derivatives in $\nabla f_L(\bar{\boldsymbol{x}})$ and $\nabla^2 f_L(\bar{\boldsymbol{x}})$, which are computed with first- and second-order central finite differences, and the reference value $f_L(\bar{\boldsymbol{x}})$ require the evaluation of f_L for a small number of optical vectors $\boldsymbol{x}$ as the logarithm of the quotient of L_λ and $L_{\lambda,\mathrm{a}}$ using the optical properties encapsulated in $\boldsymbol{x}$. These are the unique calls to the *quasi-exact* model, while the predictor model needs to be called N_λ additional times.

3.4.3 Principal component analysis on correlated *k*-distribution method

In the previous subsections, we described the formalism of the correlated *k*-distribution and the principal component analysis methods in the context of radiative transfer modelling. The PCA formalism is initially applied to a fine wavelength grid $\{\lambda_i\}_{i=1}^{N_\lambda}$, such that it is possible to generate convolved spectra from numerical integration of Eq. 3.3. However, the PCA formalism remains valid if instead of using the set of wavelengths $\{\lambda_i\}_{i=1}^{N_\lambda}$ we consider the set of wavelengths $\{\overline{\lambda}_p\}_{p=1}^{\overline{N}_\lambda}$ of the correlated *k*-distribution method, in which case the wavelength variability is encapsulated in the optical vectors $\boldsymbol{x}(\overline{\lambda}_p)$.

Figure 3.4 shows a comparison of radiative transfer model errors for the same scenarios as presented in Figure 3.1 when using different acceleration techniques and a number of discrete ordinates $M = 32$. On the top left, the original LBL calculation is shown, while the other plots show the simulations when using the correlated *k*-distribution method, the PCA method and the PCA on correlated *k*-distribution method. For the correlated *k*-distribution method, we used $N_\lambda = 60$ and $N_\mathrm{q} = 4$. For the simple PCA method, $N_\lambda = 20000$ was selected, the PCA orthogonal basis had size $N_\mathrm{u} = 4$, and the predictor model was the two-stream DOME. For the PCA on correlated *k*-distribution method, we used $N_\mathrm{u} = 4$ principal components on the set of $\overline{N}_\lambda = 240$ wavelengths of the correlated *k*-distribution method.

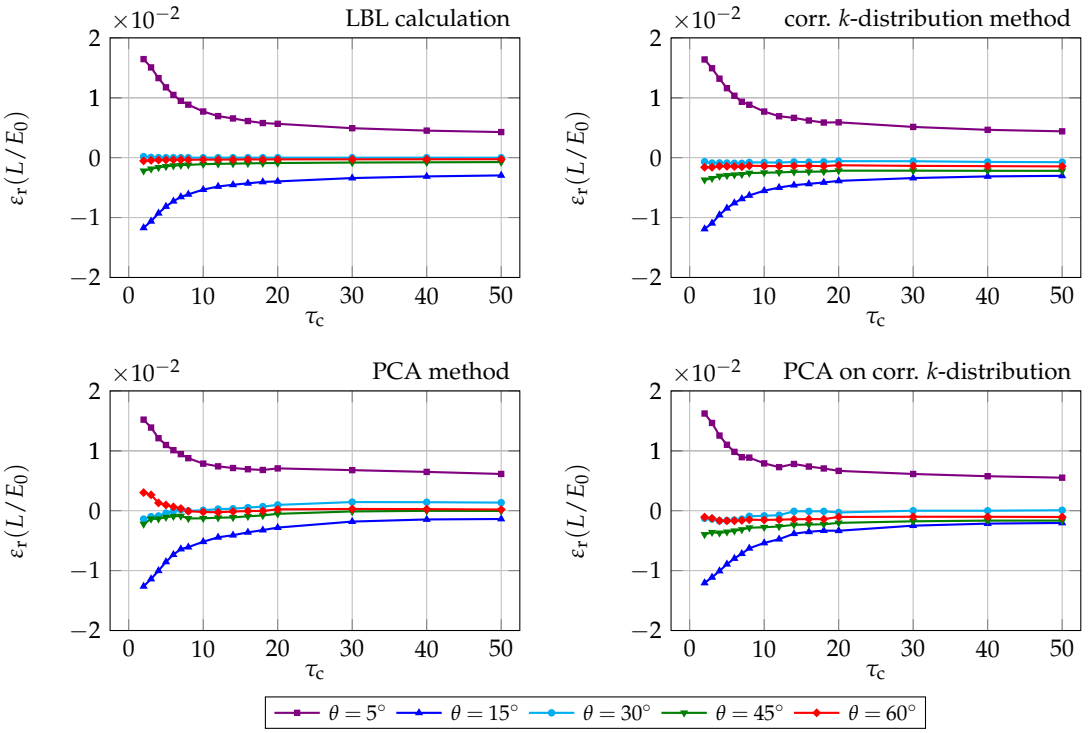


Figure 3.4: Relative errors ε_r in the simulated DOME radiances L/E_0 of the reference example scenario (Figure 3.1) for different cloud optical thicknesses τ_c and viewing zenith angles θ when applying acceleration techniques such as correlated *k*-distribution, PCA and a combination of both.

Table 3.2: Computation times for the simulated spectra in Figure 3.4, and speed-up factors with respect to the LBL case.

	LBL calculation	Corr. k-distribution	PCA	PCA on corr. k-distribution
Time / min	1711	159	22	5
Speed-up factor	–	10.8	77.8	342.2

From Figure 3.4 we observe that it is possible to simulate EPIC spectral measurements in the oxygen A-band of absorption with any of these acceleration techniques, with neglectable variation in the relative errors when compared to the LBL computation for $M = 32$. However, the time efficiency of the acceleration techniques is different, and it is connected to the number of necessary calls to the *quasi-exact* model. This computation time is shown in Table 3.2. It may be observed that the technique of applying PCA on the correlated k-distribution method is the one providing the highest speed-up with respect to the LBL computation, and the computation time is reduced two orders of magnitude with respect to the LBL case.

3.5 Linearisation of the radiative transfer model

The retrieval of the atmospheric constituents from satellite measurements is an optimisation problem and thus requires not only an efficient and accurate radiative transfer forward model, but also the derivatives of this model with respect to the atmospheric parameters to be retrieved. The process of obtaining this set of partial derivatives to construct the Jacobian of the optimisation problem is commonly referred to as linearisation analysis. Two common linearisation approaches may be considered:

1. In the *linearised forward* approach, the partial derivatives of the radiance L with respect to the retrieved atmospheric parameters are computed analytically. This approach has been developed for the conventional discrete ordinates method and for matrix operator method by Spurr [66–68].

2. In the *forward-adjoint* approach, the radiance L is expressed as the scalar product of the solution of the adjoint problem and the source function of the forward problem. By performing the linearisation of the forward and adjoint problems, it is possible to determine analytical expressions for the derivatives of the radiance with respect to the atmospheric parameters [69, 70].

In the context of the retrieval of cloud macrophysical parameters from measurements of the EPIC/DSCOVR instrument, we have implemented the computation of the partial derivatives with respect to the cloud optical thickness and the cloud-top height for the DOME and MOME radiative transfer models by means of both the linearised forward and the forward-adjoint methods [71], which also benefit from the acceleration techniques developed for the forward

models. In particular, it may be noted that the principal component analysis in the linearised problem may be implemented in two different manners:

1. The first method follows the same procedure as described for the forward model, i.e. the derivative of the radiance L with respect to each retrieved parameter ς_m is computed by means of a predictor-corrector approach:

$$\frac{\partial L_\lambda(\lambda)}{\partial \varsigma_m} = \frac{\partial L_{\lambda,\mathrm{a}}(\lambda)}{\partial \varsigma_m} \exp[f_{\partial L/\partial \varsigma_m}(\lambda)],$$

   where $L_{\lambda,\mathrm{a}}(\lambda)$ is a radiance from a predictor radiative transfer model, and $f_{\partial L/\partial \varsigma_m}(\lambda)$ is the correction factor for the derivatives of $L_\lambda(\lambda)$ with respect to ς_m.

2. The second method performs the linearisation directly on the restoration formula from Eq. 3.8, which involves the computation of the partial derivatives of the components q_{il} with respect to ς_m.

For the derivative analysis, we chose a test of scenarios utilising the same cloud model, EPIC absorption oxygen A-band and configuration as in the radiance analysis, with the following modifications: the viewing zenith angle is fixed to $\theta = 30$ deg, the surface albedo is set to 0.2, the number of discrete ordinates is set to $M = 32$ according to the conclusions in Section 3.2, and the derivatives with respect to the cloud optical thickness τ_c are evaluated for a fixed cloud-top height $h_\mathrm{t} = 4$ km, while the derivatives with respect to the cloud-top height h_t are computed for a fixed cloud optical thickness $\tau_\mathrm{c} = 5$.

Figure 3.5 shows the derivatives of the radiance for the test of scenarios when using the linearised DOME model (LDOME). On the left, we observe that $\partial(L/E_0)/\partial\tau_\mathrm{c}$ is always non-negative, so the increases in the cloud optical thickness for a fixed cloud-top height imply increases in the observed radiances, although this increase is less evident as τ_c becomes higher. On the right, we see that $\partial(L/E_0)/\partial h_\mathrm{t}$ is also non-negative and more significant as h_t increases, which is a clear consequence of the shielding effect of the molecular oxygen below the cloud.

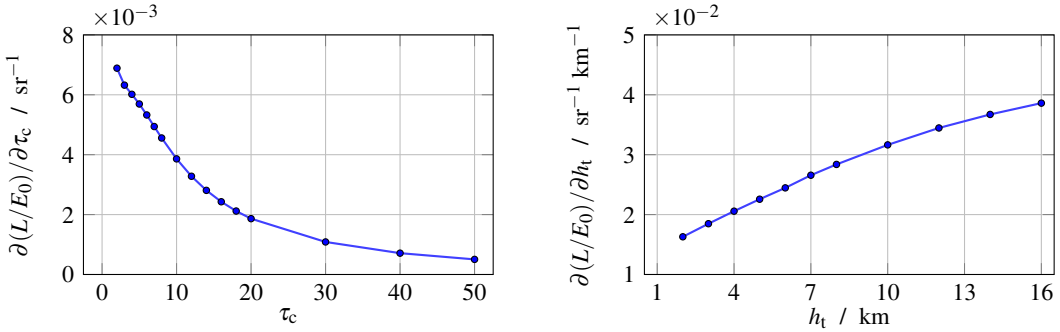


Figure 3.5: Simulated derivatives of the DOME radiances with respect to the cloud optical thickness τ_c and the cloud-top hegiht h_t at EPIC absorption oxygen A-band for the test scenarios of viewing zenith angle $\theta = 30$ deg. The reference number of discrete ordinates is $M = 32$.

Since it was observed that the behaviour of LDOME, the linearised MOME (LMOME) and the forward-adjoint DOME (FADOME) is practically identical [71], we focused the analysis on the impact of the acceleration techniques in the accuracy and time efficiency of the derivative computation. Figure 3.6 shows the relative errors in the computation of radiance derivatives for the test scenarios when compared to the simulations in Figure 3.5. We may observe that the relative errors remain below 0.5% for $\partial(L/E_0)/\partial\tau_c$ and below 1% for $\partial(L/E_0)/\partial h_t$.

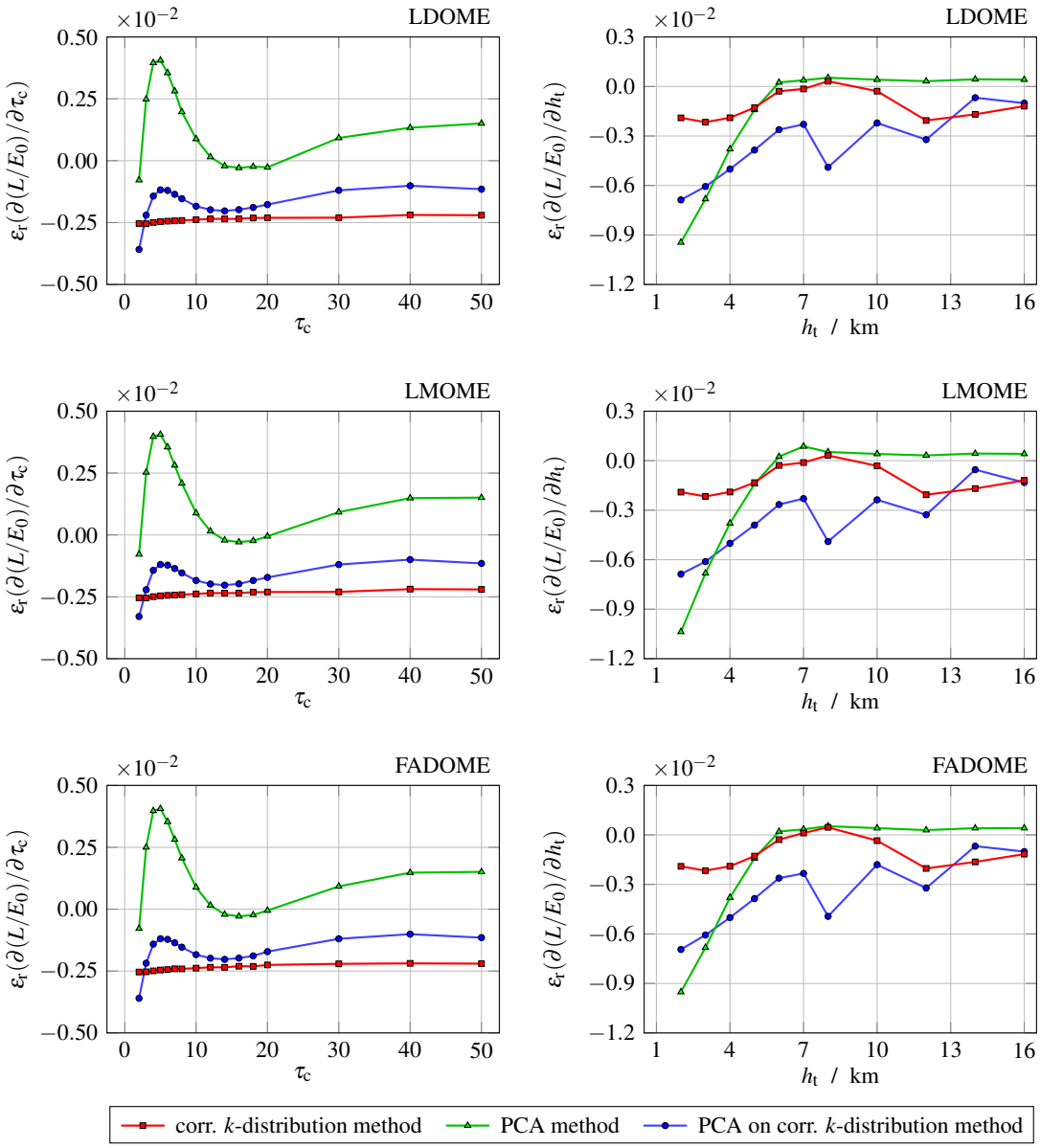


Figure 3.6: Relative errors ε_r in the simulated derivatives of L/E_0 when compared to the reference values in Figure 3.5 for the models LDOME, LMOME and FADOME using the three acceleration techniques (correlated k-distribution, PCA, and PCA on correlated k-distribution). The number of discrete ordinates is $M = 32$.

Table 3.3: Computation times in minutes to simulate the derivatives of L/E_0 with respect to τ_c for the different models and acceleration techniques. The speed-up factor corresponds to the ratio between LDOME and FADOME.

Acceleration techniques	Linearised models			Speed-up factor
	LDOME	LMOME	FADOME	
LBL calculation	3625	4602	1804	2.0
Correlated k-distribution method	678	869	335	2.0
PCA method	8	9.5	6	1.3
PCA on corr. k-distribution method	6	7.5	3.5	1.7

Table 3.4: Computation times in minutes to simulate the derivatives L/E_0 with respect to h_t for the different models and acceleration techniques. The speed-up factor corresponds to the ratio between LDOME and FADOME.

Acceleration techniques	Linearised models			Speed-up factor
	LDOME	LMOME	FADOME	
LBL calculation	2451	3079	1240	2.0
Correlated k-distribution method	456	574	224	2.0
PCA method	5.5	6.5	4.5	1.2
PCA on corr. k-distribution method	4	5	2.5	1.6

Finally, in Table 3.3 and Table 3.4 we may observe the computation times to simulate the derivatives shown in Figure 3.4 for the three linearised models and the three possible acceleration techniques as well as the simple LBL calculation. Each case (i.e. model and acceleration technique) consists of 16 simulations of $\partial(L/E_0)/\partial\tau_c$ and 11 simulations of $\partial(L/E_0)/\partial h_t$. It is observable that FADOME is in general the fastest method, followed by LDOME, with independence of the acceleration technique being used. The best configuration is FADOME combined with the PCA on correlated k-distribution method, which provides a global speed-up factor of about 1000 when compared to the equivalent LDOME LBL calculation.

References

[53] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, Journal of the Society for Industrial and Applied Mathematics 11 (2) (1963) 431–441. doi:10.1137/0111030.

[54] A. Doicu, T. Trautmann, F. Schreier, Numerical regularization for atmospheric inverse problems, Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-05439-6.

[55] A. Doicu, 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) (2009) 146–158. doi:10.1016/j.jqsrt.2008.09.014.

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

[57] V. Molina García, S. Sasi, D. Efremenko, A. Doicu, D. Loyola, Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, Journal of Quantitative Spectroscopy and Radiative Transfer 123 (2018) 228–240. doi:10.1016/j.jqsrt.2018.03.014.

[58] B. Mayer, M. Schröder, R. Preusker, L. Schüller, Remote sensing of water cloud droplet size distributions using the backscatter glory: A case study, Atmospheric Chemistry and Physics 4 (5) (2004) 1255–1263. doi:10.5194/acp-4-1255-2004.

[59] E. Yanovitskij, Light scattering in inhomogeneous atmospheres, Springer-Verlag, 1997.

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

[61] 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 (1936) 7–18.

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

[63] V. Natraj, X. Jiang, R. Shia, X. Huang, J. Margolis, Y. Yung, 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) (2005) 539–556. doi:10.1016/j.jqsrt.2004.12.024.

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

[65] D. Efremenko, A. Doicu, D. Loyola, 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 (2014) 128–135. doi:10.1016/j.jqsrt.2013.07.023.

[66] R. Spurr, T. Kurosu, K. Chance, A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, Journal of Quantitative Spectroscopy and Radiative Transfer 68 (6) (2001) 689–735. doi:10.1016/s0022-4073(00)00055-8.

[67] R. Spurr, Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, Journal of Quantitative Spectroscopy and Radiative Transfer 75 (2) (2002) 129–175. doi:10.1016/s0022-4073(01)00245-x.

[68] 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. Kokhanovsky (Ed.), Light scattering reviews, Vol. 3, 2008, pp. 229–275. doi:10.1007/978-3-540-48546-9_7.

[69] J. Landgraf, O. Hasekamp, M. Box, T. Trautmann, A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory approach, Journal of Geophysical Research: Atmospheres 106 (D21) (2001) 27291–27305. doi:10.1029/2001jd000636.

[70] A. Doicu, T. Trautmann, Two linearization methods for atmospheric remote sensing, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (8) (2009) 477–490. doi:10.1016/j.jqsrt.2009.02.001.

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

Chapter 4

EPIC-OCRA: Estimation of cloud fraction from EPIC/DSCOVR

In this chapter, we present the Optical Cloud Recognition Algorithm (OCRA) applied to the estimation of the radiometric cloud fraction from EPIC/DSCOVR measurements, which is crucial in atmospheric remote sensing, as it is one key cloud macrophysical parameter needed for trace gas retrieval [72]. Our main contributions are summarised as follows:

1. We give an estimation of the viewing angle dependency in the EPIC top-of-atmosphere (TOA) reflectances ρ_m, $m = 1, \ldots, 10$. This dependency relates to the BRDF effect and it is modelled separately for water and land locations for the 10 EPIC channels by means of polynomial factors $f_m^{\text{view}}(\theta)$ evaluated for the scene viewing zenith angle θ. The TOA reflectance divided by $f_m^{\text{view}}(\theta)$ is used to define the OCRA radiometric quantity ρ_m^{ocra}.
2. We created reference clear-sky maps with a spatial grid resolution of 0.2 deg $\times$ 0.2 deg for the OCRA radiometric quantities ρ_m^{ocra}, $m = 1, \ldots, 10$, for the 10 EPIC channels, using a temporal kernel reduction method performed independently for every EPIC channel, and which is based on the assumption that, for every Earth location, the likelihood of finding a clear-sky scene increases when the OCRA radiometric quantities decrease.
3. We determine OCRA scaling and offset parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$) for the EPIC channels at 388, 551 and 780 nm, which have a time-oscillating behaviour related to the yearly evolution of the global TOA reflectance. We show that a proper selection of the OCRA threshold parameter p is crucial to ensure that the upper bound in the OCRA radiometric cloud fraction is not lower than 1 over bright surfaces.
4. We compute the EPIC OCRA radiometric cloud fraction, which is the geophysical parameter needed for trace gas retrieval and for quantitative cloud studies. Our OCRA radiometric cloud product for EPIC provides more information than the operational EPIC cloud product version 1 [73], which only includes a simple cloud mask [74].
5. We compare the EPIC radiometric cloud fractions with the MODIS/Terra geometric cloud fractions, and a median discrepancy of about -0.186 is found, which decreases to -0.040 if we discard water locations with clouds of optical thickness lower than 10. EPIC UV–VIS channels are almost insensitive to this kind of clouds.

Conference contributions

V. Molina García, R. Lutz, D. Loyola, Applying the OCRA algorithm to the retrieval of the cloud fraction from EPIC/DSCOVR, oral presentation at 21st EGU General Assembly, 7–12 April 2019, Vienna (Austria), European Geosciences Union, 2019.

4.1 Introduction

The Optical Cloud Recognition Algorithm (OCRA) is a method for determining the cloud coverage [75], i.e. the fraction of a remote sensing scene covered by clouds, from UV–VIS–NIR measurements under the assumption that the sensor measurements may be divided into two main components: a background term that is cloud-free and a remainder term that describes the influence of the clouds.

Let be $S_{\text{ocra}} \subseteq \{m\}_{m=1}^{10}$ a subset of size n_{ocra} from the sensor channel indices, and let be ρ_j^{ocra}, $j \in S_{\text{ocra}}$, a physical quantity derived from the sensor measurements for a number n_{ocra} of sensor channels. Given a certain remote sensing scene, we define the vector of differences $\boldsymbol{\Delta} = [\Delta_j]^{\mathsf{T}}$ as a vector of n_{ocra} components in which each component provides the difference between the OCRA quantity value $\rho_{j,0}^{\text{ocra}}$ expected in the absence of clouds and the value ρ_j^{ocra} actually measured by the instrument, i.e.

$$\Delta_j = \rho_j^{\text{ocra}} - \rho_{j,0}^{\text{ocra}}. \tag{4.1}$$

For a given remote sensing scene, OCRA defines the *radiometric cloud fraction* χ_{CF} in terms of the vector of differences $\boldsymbol{\Delta}$ as:

$$\chi_{\text{CF}} = \min\left\{1, \sum_{j \in S_{\text{ocra}}} \alpha_j \max\left\{0, (\Delta_j - \beta_j)\right\}^2\right\}^{1/2}, \tag{4.2}$$

which can be interpreted as the norm of the positive components from the difference vector between $\boldsymbol{\Delta}$ and a quantity bias vector $\boldsymbol{\beta} = [\beta_j]^{\mathsf{T}}$, with $\beta_j \ll 1$, weighted by the components of a scaling vector $\boldsymbol{\alpha} = [\alpha_j]^{\mathsf{T}}$, with $\alpha_j \geq 1$. In particular:

- The scaling terms $\{\alpha_j\}_{j \in S_{\text{ocra}}}$ weight the components of the vector of differences $\boldsymbol{\Delta}$ based on an effective valid range for each Δ_j, $j \in S_{\text{ocra}}$. They are responsible of setting the upper limits of $\{\Delta_j\}_{j \in S_{\text{ocra}}}$ for which the scene is considered as fully cloudy.
- The offset terms $\{\beta_j\}_{j \in S_{\text{ocra}}}$ correct the clear-sky OCRA quantities from biases of the actual values $\{\rho_j^{\text{ocra}}\}_{j \in S_{\text{ocra}}}$. They are responsible of setting the lower limits of $\{\Delta_j\}_{j \in S_{\text{ocra}}}$ for which the scene is considered as cloud-free.

The norm-based definition of the radiometric cloud fraction is clipped to ensure that the valid range of χ_{CF} lies within the interval $[0, 1]$.

The description of the proper OCRA quantity ρ^{ocra} extracted from the EPIC measurements is described in Section 4.2. The estimation of the OCRA quantities $\{\rho_{j,0}^{\text{ocra}}\}_{j \in S_{\text{ocra}}}$ under cloud-free conditions, i.e. the generation of the OCRA clear-sky maps, is detailed in Section 4.3. The determination of $\{\alpha_j\}_{j \in S_{\text{ocra}}}$ and $\{\beta_j\}_{j \in S_{\text{ocra}}}$ is outlined in Section 4.4. The application of OCRA to the EPIC/DSCOVR mission, as well as a global comparison between the EPIC radiometric cloud fraction and the MODIS geometric cloud fraction products, is shown in Section 4.5.

4.2 Nadir-effective TOA reflectance

The computation of the radiometric cloud fraction by means of OCRA requires the use of an input quantity that is statistically independent from the time t and the viewing geometry (i.e. the viewing zenith angle θ). The time dependency can be handled for a time-dependent quantity by generating clear-sky maps on a periodic basis that can capture the variations of the input quantity over the seasons. The viewing geometry dependency, however, imposes restrictions on the quantity that can be used by the OCRA algorithm. In particular, the TOA reflectance ρ is not suitable as OCRA input, as ρ may depend strongly on the viewing zenith angle θ (Figure 4.1**a**) [76]. The use of the TOA reflectance ρ would translate into two main consequences within the OCRA framework:

1. As the OCRA clear-sky map generation relies on a kernel-reduction approach in which the clear-sky scenario of every map cell is the farthest measurement from the white in a given time period, OCRA clear-sky maps using ρ as input quantity will be mainly built with values from small viewing zenith angles, whose measurements tend to be smaller.
2. The evaluation of the radiometric cloud fraction by means of the differences in the TOA reflectance ρ with respect to a clear-sky scenario computed as in the previous point will identify the positive deviations from the clear-sky scenario due to high viewing zenith angles as if they were due to the presence of clouds.

To avoid this problem, we define the TOA reflectance $\rho_m(t,\theta)$ for the EPIC m-th channel as the product of the *nadir-effective* TOA reflectance $\rho^*_m(t)$, which is defined as independent from the viewing geometry [77], and a viewing correction factor $f_m^{\text{view}}(\theta)$:

$$\rho_m(t,\theta) = \rho^*_m(t)\, f_m^{\text{view}}(\theta), \tag{4.3}$$

and such nadir-effective TOA reflectance $\rho^*_m(t)$ can be used as a valid OCRA input quantity ρ_m^{ocra} (see Figure 4.1**b**).

(**a**) (**b**)

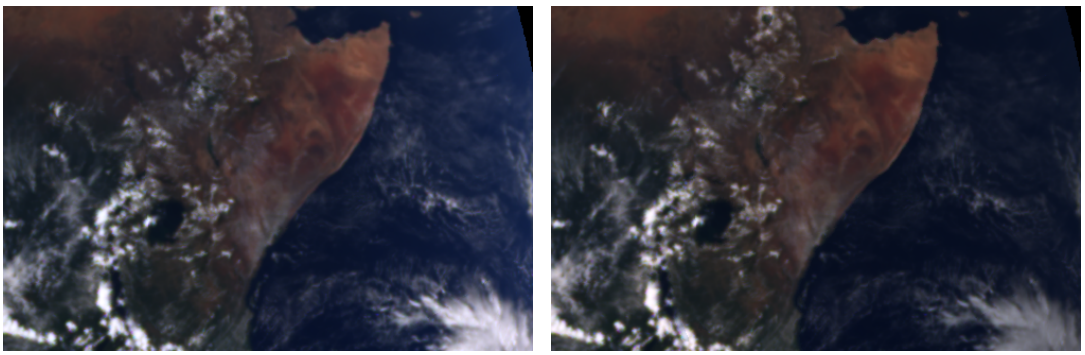


Figure 4.1: Horn of Africa as seen from EPIC Level 1B version 2 RGB on 2019-03-20 11:15:31 UTC: (**a**) TOA reflectance and (**b**) nadir-effective TOA reflectance.

The viewing correction factors $\{f_m^{\text{view}}(\theta)\}_{m=1}^{10}$ encapsulate the reflectance variations mainly due to the non-Lambertian behaviour of the Earth's land and ocean surfaces (i.e. the BRDF effects), as well as the Rayleigh scattering in the atmosphere. These correction factors can be parametrised in terms of an n-th degree polynomial in the viewing zenith angle θ (in degrees):

$$f_m^{\text{view}}(\theta) = f_m^{\text{glint}}(\theta) \sum_{\substack{i=0 \\ i\neq 1}}^{n} a_{m,i}^{\text{view}} \theta^i, \tag{4.4}$$

where $f_m^{\text{glint}}(\theta)$ is a factor describing the glint effect for small values of θ, given by the following expression in the viewing zenith angle θ (in degrees):

$$f_m^{\text{glint}}(\theta) = \begin{cases} 1 & \text{over land,} \\ 1 + \exp(b_{m,0}^{\text{glint}} + b_{m,2}^{\text{glint}} \theta^2) & \text{over ocean,} \end{cases} \tag{4.5}$$

where $b_{m,0}^{\text{glint}}$ determines the glint enhancement at the nadir view and $b_{m,2}^{\text{glint}} < 0$ modulates the exponential decay of the glint enhancement.

In order to determine the viewing correction factors $\{f_m^{\text{view}}(\theta)\}_{m=1}^{10}$ for the ten EPIC channels, we selected all the available datasets from EPIC Level 1B version 2 within the time period from 12 June 2015 to 27 June 2019 that could be geocorrected by means of the registration algorithm in Section 2.2 [78]. For every geocorrected dataset and m-th channel, $m = 1, \ldots, 10$, we applied the following steps if the channel was present:

1. Convert the EPIC engineering units into TOA reflectance ρ_m^{corr} (Eq. 2.13) using the undegraded calibration factor defined in Eq. 2.14 and the degradation ratios from Table 2.3.
2. Separate the TOA reflectances by their associated land-ocean mask value.
3. For every land-ocean case:
   (a) Group the measurements in 1-deg bins for the viewing zenith angle $\theta \in [0, 60)$ deg.
   (b) Recompute the mean TOA reflectance $\bar{\rho}_m^{\text{corr}}(\theta)$ for each viewing zenith angle bin. The mean is recalculated from the mean value stored from the previous datasets and that from the current dataset, weighted by the number of measurements used for the stored result and the number coming from the current dataset.

Once these corrections were applied to all the datasets, the resulting mean TOA reflectances over land were least-squares fitted to a fifth degree polynomial in θ as defined in Eq. 4.4, with $f_m^{\text{glint}}(\theta) = 1$. For the resulting values over ocean, the values $\bar{\rho}_m^{\text{corr}}(\theta < 15 \text{ deg})$ were used first to determine the glint coefficients $[b_{m,0}^{\text{glint}}, b_{m,2}^{\text{glint}}]^\intercal$, and the viewing correction coefficients were fitted as for the land case after removing the glint effect.

Figure 4.2 shows the mean TOA reflectance $\bar{\rho}_m^{\text{corr}}$ for every EPIC channel as a function of the viewing zenith angle θ, together with the fitting polynomials $f_m^{\text{view}}(\theta)$ computed as described in the previous paragraphs, and normalised to the mean TOA reflectance value at the first viewing zenith angle bin.

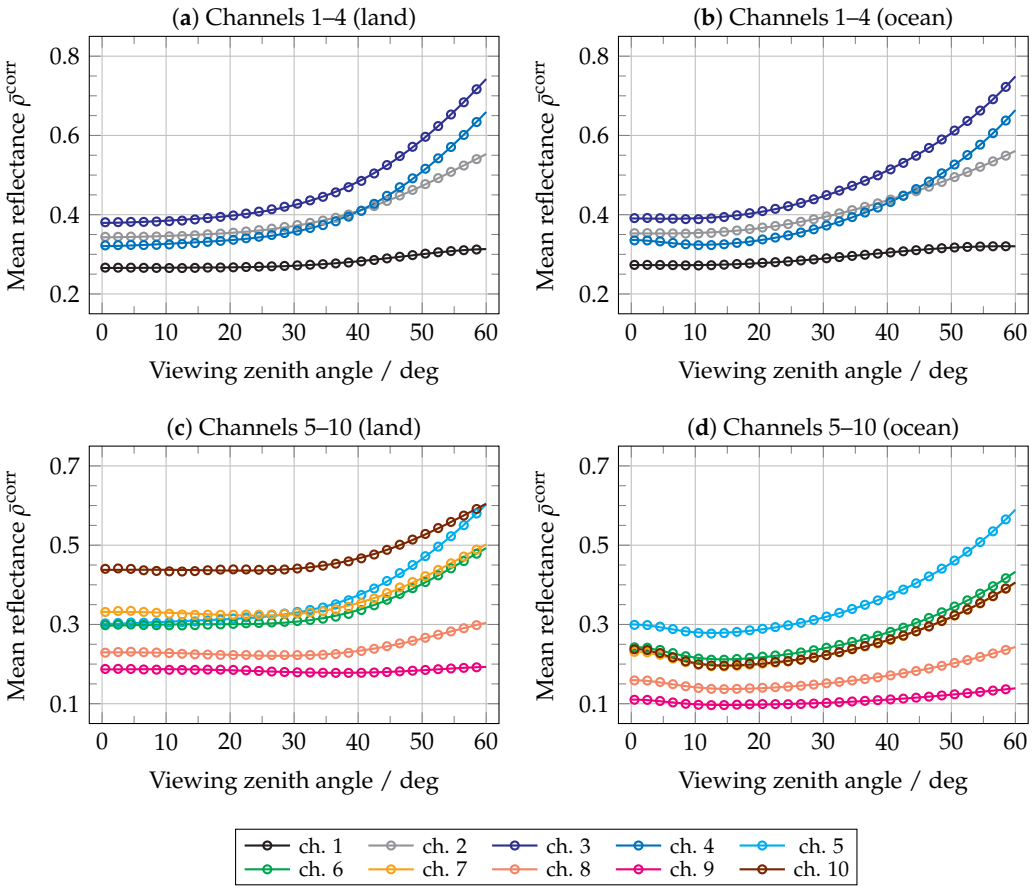


Figure 4.2: Dependency between the EPIC global mean reflectances and the viewing zenith angle for the UV channels (**a**) over land and (**b**) over ocean, and for the VNIR channels (**c**) over land and (**d**) over ocean. The circles represent the mean values obtained from the EPIC datasets, and the solid lines correspond to the fitting polynomials $f_m^{\text{view}}(\theta)$ normalised to nadir view. The glint effect is observable over ocean for the VNIR channels and small viewing zenith angles.

Table 4.1 and Table 4.2 present the viewing correction polynomial coefficients $\{a_{m,i}^{\text{view}}\}$ over land and ocean, respectively, for each of the ten EPIC channels. Table 4.3 provides the glint correction polynomial coefficients $\{b_{m,i}^{\text{glint}}\}$, which are applicable only to the EPIC measurements over ocean for channels 4 to 10.

From this procedure, we define the OCRA input quantity ρ_m^{ocra} for every pixel acquired at time t as the nadir-effective TOA reflectance, i.e. from Eq. 4.3 and Eq. 2.13:

$$\rho_m^{\text{ocra}}(t) := (\rho_m^{\text{corr}}(t))^* = \frac{\rho_m^{\text{corr}}(t,\theta)}{f_m^{\text{view}}(\theta)} = k_m(t)\,\frac{C_m(t,\theta)}{f_m^{\text{view}}(\theta)\cos\theta_0}, \tag{4.6}$$

where $k_m(t)$ is the degradation-corrected calibration factor, $C_m(t,\theta)$ is the EPIC measurement in cps, $f_m^{\text{view}}(\theta)$ is the viewing correction factor, θ is the viewing zenith angle, and θ_0 is the solar zenith angle.

Table 4.1: Polynomial coefficients for EPIC viewing correction over land.

Channel	$a_{m,0}^{\text{view}}$	$a_{m,2}^{\text{view}}$	$a_{m,3}^{\text{view}}$	$a_{m,4}^{\text{view}}$	$a_{m,5}^{\text{view}}$
1	0.997	$+6.429 \times 10^{-5}$	-5.881×10^{-6}	$+2.066 \times 10^{-7}$	-1.878×10^{-9}
2	0.996	$+1.628 \times 10^{-4}$	-9.032×10^{-6}	$+2.932 \times 10^{-7}$	-2.347×10^{-9}
3	0.996	$+2.246 \times 10^{-4}$	-1.150×10^{-5}	$+3.557 \times 10^{-7}$	-2.548×10^{-9}
4	0.995	$+3.083 \times 10^{-4}$	-1.873×10^{-5}	$+5.296 \times 10^{-7}$	-3.702×10^{-9}
5	0.996	$+3.565 \times 10^{-4}$	-2.411×10^{-5}	$+6.597 \times 10^{-7}$	-4.667×10^{-9}
6	0.996	$+2.625 \times 10^{-4}$	-2.269×10^{-5}	$+6.521 \times 10^{-7}$	-4.937×10^{-9}
7	1.001	-5.263×10^{-7}	-9.780×10^{-6}	$+4.080 \times 10^{-7}$	-3.420×10^{-9}
8	1.004	$+1.898 \times 10^{-5}$	-1.134×10^{-5}	$+4.151 \times 10^{-7}$	-3.437×10^{-9}
9	0.994	$+1.444 \times 10^{-4}$	-1.539×10^{-5}	$+3.940 \times 10^{-7}$	-2.918×10^{-9}
10	0.994	$+6.284 \times 10^{-5}$	-8.509×10^{-6}	$+2.897 \times 10^{-7}$	-2.268×10^{-9}

Table 4.2: Polynomial coefficients for EPIC viewing correction over ocean.

Channel	$a_{m,0}^{\text{view}}$	$a_{m,2}^{\text{view}}$	$a_{m,3}^{\text{view}}$	$a_{m,4}^{\text{view}}$	$a_{m,5}^{\text{view}}$
1	0.999	-7.680×10^{-5}	$+8.907 \times 10^{-6}$	-1.623×10^{-7}	$+8.084 \times 10^{-10}$
2	0.998	-6.541×10^{-5}	$+1.137 \times 10^{-5}$	-2.015×10^{-7}	$+1.259 \times 10^{-9}$
3	0.998	-1.368×10^{-4}	$+1.759 \times 10^{-5}$	-3.318×10^{-7}	$+2.455 \times 10^{-9}$
4	1.003	-1.336×10^{-4}	$+1.719 \times 10^{-5}$	-3.138×10^{-7}	$+2.444 \times 10^{-9}$
5	1.004	-1.096×10^{-4}	$+1.601 \times 10^{-5}$	-2.857×10^{-7}	$+2.307 \times 10^{-9}$
6	1.004	-4.907×10^{-5}	$+1.190 \times 10^{-5}$	-1.833×10^{-7}	$+1.400 \times 10^{-9}$
7	1.005	-5.436×10^{-5}	$+1.300 \times 10^{-5}$	-2.061×10^{-7}	$+1.564 \times 10^{-9}$
8	1.004	$+5.740 \times 10^{-6}$	$+6.370 \times 10^{-6}$	-7.013×10^{-8}	$+4.414 \times 10^{-10}$
9	1.004	$+1.335 \times 10^{-4}$	-5.188×10^{-6}	$+1.627 \times 10^{-7}$	-1.283×10^{-9}
10	1.006	-4.857×10^{-5}	$+1.129 \times 10^{-5}$	-1.561×10^{-7}	$+1.146 \times 10^{-9}$

Table 4.3: Polynomial coefficients for EPIC glint correction over ocean.

Channel	$b_{m,0}^{\text{glint}}$	$b_{m,2}^{\text{glint}}$
1	–	–
2	–	–
3	–	–
4	−3.126	−0.019
5	−2.361	−0.016
6	−1.713	−0.014
7	−1.465	−0.013
8	−1.615	−0.012
9	−1.779	−0.016
10	−1.389	−0.015

4.3 Clear-sky map generation

Once the OCRA quantity ρ^{ocra} is defined, the next preparation step is the creation of the OCRA clear-sky maps. These maps provide the OCRA quantity values $\{\rho^{\text{ocra}}_{j,0}\}_{j\in S_{\text{ocra}}}$ expected under cloud-free conditions for the OCRA channels $j \in S_{\text{ocra}}$ at every Earth's location at a given time t, and they are utilised to estimate the differences in ρ^{ocra} for every ground-pixel measurement.

Given a cell in a clear-sky map, the cell value is obtained by means of a kernel-reduction method over the OCRA quantity values $\{\rho^{\text{ocra}}_j(x,y)\}_{j\in S_{\text{ocra}}}$ computed from the sensor measurements that lie within the map cell for a given period of time. The kernel function must be chosen so that it guarantees the selection of the cell measurements that correspond most likely to a clear-sky scene.

Depending on the channel interdependency during the generation of the clear-sky maps, we can distinguish two different approaches:

1. In the *channel-independent* approach, the kernel-reduction method is applied to the OCRA quantities $\{\rho^{\text{ocra}}_j\}_{j\in S_{\text{ocra}}}$ directly. The generation of the clear-sky maps can be performed for every OCRA channel independently, and the kernel function follows a minimum-based criterion: for every map cell centered at Earth's coordinates (x,y) and for every OCRA channel $j \in S_{\text{ocra}}$, the clear-sky scene value $\rho^{\text{ocra}}_{j,0}(x,y)$ is determined as the smallest $\rho^{\text{ocra}}_j(x,y)$ value within the time series of study.

2. In the *channel-dependent* approach, the kernel-reduction method is applied to the OCRA quantities $\{\rho^{\text{ocra}}_j\}_{j\in S_{\text{ocra}}}$ after their conversion to the normalised *rgb* color space. For this approach, n_{ocra} is set to 3, and the color vectors $\boldsymbol{v}_{\text{rgb}} = [r,g,b]^{\mathsf{T}}$ are defined by means of the following components:

$$r = \frac{\rho^{\text{ocra}}_1(x,y)}{\sum\limits_{j\in S_{\text{ocra}}} \rho^{\text{ocra}}_j(x,y)}, \qquad g = \frac{\rho^{\text{ocra}}_2(x,y)}{\sum\limits_{j\in S_{\text{ocra}}} \rho^{\text{ocra}}_j(x,y)}, \qquad b = 1 - r - g.$$

   The generation of the clear-sky maps using the color vector $\boldsymbol{v}_{\text{rgb}}$ adds interdependency accross the OCRA quantities ρ^{ocra}_j from different channels, and the kernel function uses a maximum-based criterion: for every map cell centered at Earth's coordinates (x,y), the clear-sky scene is determined as the set of values $\{\rho^{\text{ocra}}_j(x,y)\}_{j\in S_{\text{ocra}}}$ within the time series of study for which the distance of $\boldsymbol{v}_{\text{gb}} = [g,b]^{\mathsf{T}}$ to the white point $\boldsymbol{w}_{\text{gb}} = [1/3, 1/3]^{\mathsf{T}}$ is maximum. This approach has been applied to the estimation of the radiometric cloud fraction for the sensor GOME on board ERS-2 [79] and GOME-2 on board MetOp-A and MetOp-B [80]. The TROPOMI mission implements a variant of the channel-dependent approach in which two OCRA channels are used instead of three [81], and for which the white point corresponds to $\boldsymbol{w}_{\text{gb}} = [1/2, 1/2]^{\mathsf{T}}$.

We created latitude-longitude clear-sky maps for the 10 EPIC channels on a daily basis with spatial grid of 0.2 deg $\times$ 0.2 deg using the channel-independent approach over the time period from 12 June 2015 to 27 June 2019. We chose the channel-independent approach because it allows to calculate the clear-sky maps channel by channel, and such clear-sky maps can be used for any combination of EPIC channels that is chosen as OCRA channels. This is not possible under the channel-dependent approach, as the clear-sky values $\{\rho_{j,0}^{\text{ocra}}(x,y)\}_{j\in S_{\text{ocra}}}$ for each map cell are coupled, and a hypothetical redefinition of the OCRA channels would require to recompute the clear-sky maps. The kernel reduction is done in three steps:

1. For every EPIC dataset in the time period and every EPIC channel $m = 1, \ldots, 10$, project the OCRA quantities ρ_m^{ocra} onto an equidistant cylindrical projection map with a spatial grid of 0.2 deg $\times$ 0.2 deg using nearest neighbours. If multiple values lie on the same cell, apply an average kernel. We refer to these projected datasets as *individual maps*.
2. For every date within the time period, let S_{I} be the set of all the available individual maps belonging to that given date. We calculate the *daily map* as the result of applying a day-kernel on the maps in S_{I} along the time dimension. The day-kernel used to generate the EPIC daily maps was the median function.
3. For every day d of the year, let S_{D} be the set of daily maps whose day number d_i verifies $|d_i - d \mod 365| \leq 14$. We compute the *clear-sky daily map* as the percentile $q \ll 100$ of the maps in S_{D} along the time dimension. For the EPIC clear-sky maps, we chose $q = 1$.

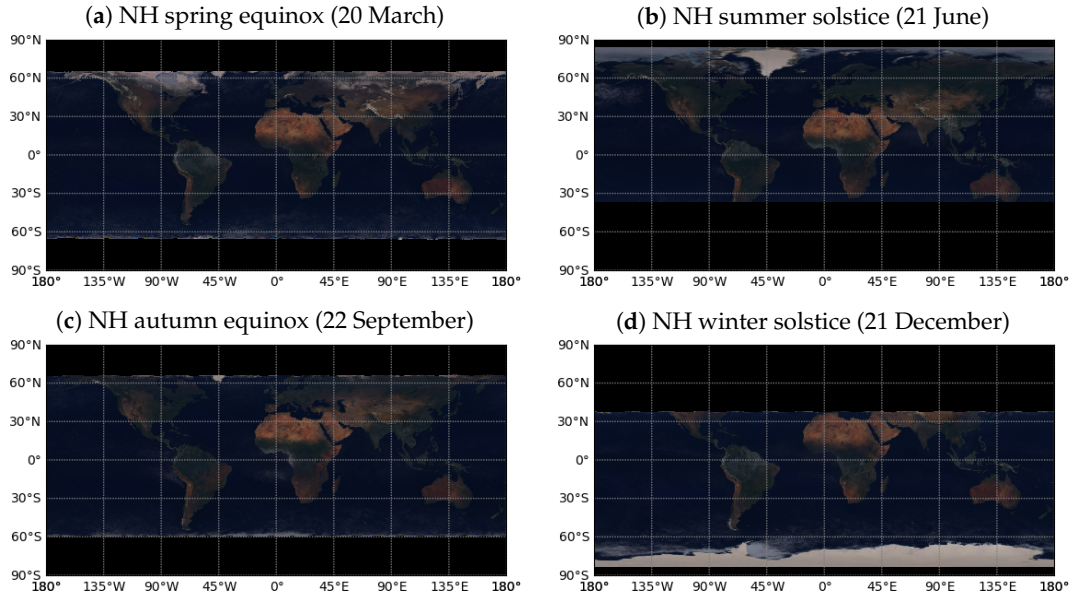


Figure 4.3: EPIC clear-sky maps generated with the channel-independent approach using the EPIC channels at 680, 551 and 433 nm. These maps correspond to the configuration RGB (Red-Green-Blue). Cells in which no measurements verified the conditions $\theta_0 < 60$ deg and $\theta < 60$ deg are shown in black; these regions change with the season: Northern Hemisphere spring (**a**), summer (**b**), autumn (**c**) and winter (**d**).

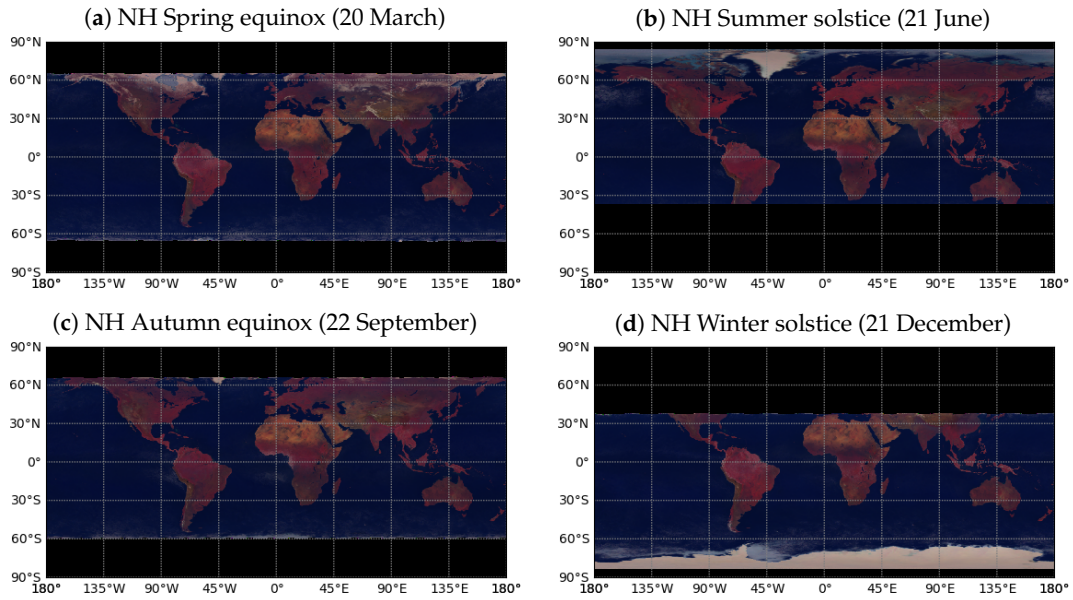


Figure 4.4: EPIC clear-sky maps generated with the channel-independent approach using the EPIC channels at 780, 443 and 388 nm. We will refer to this configuration as NGU (Near infrared-Green-Ultraviolet).

Figure 4.3 shows the resulting EPIC clear-sky maps at the solstices and equinoxes for the EPIC RGB channels (Red-Green-Blue) at 680, 551 and 433 nm. We can observe that the kernel reduction method based on the channel-independent approach is capable of producing maps in which the basic land/ocean cover classifications are visible (water, barren land, vegetated land, permanent snow/ice).

In the same way, Figure 4.4 shows the EPIC clear-sky maps at the solstices and equinoxes for the EPIC NGU channels (Near infrared-Green-Ultraviolet) at 780, 551 and 388 nm. These clear-sky maps use EPIC channels that are similar to those from the OCRA implementation for GOME-2 in [80].

From the results presented in Figure 4.3 and Figure 4.4, we can also observe the existence of regions in which the clouds are still not completely removed. This happens especially over specific areas over ocean, such as the Northern Pacific Ocean (Figure 4.4**b**) and the Antarctic Ocean (Figure 4.4**d**). Two main reasons may describe this issue:

1. There is a limitation in the geocorrection algorithm from Section 2.2. The geocorrection depends on the detection of coastline features, therefore the likelihood of not finding a solution for the transformation vector is higher for the EPIC datasets in which the Pacific Ocean covers most of the sunlit half of the Earth because the coastlines are limited and concentrated in the outer part of the EPIC image, which has worse spatial resolution. As we do not use the EPIC datasets that cannot be geocorrected well enough, the number of measurements available over the Pacific Ocean is smaller (see Figure 4.5).

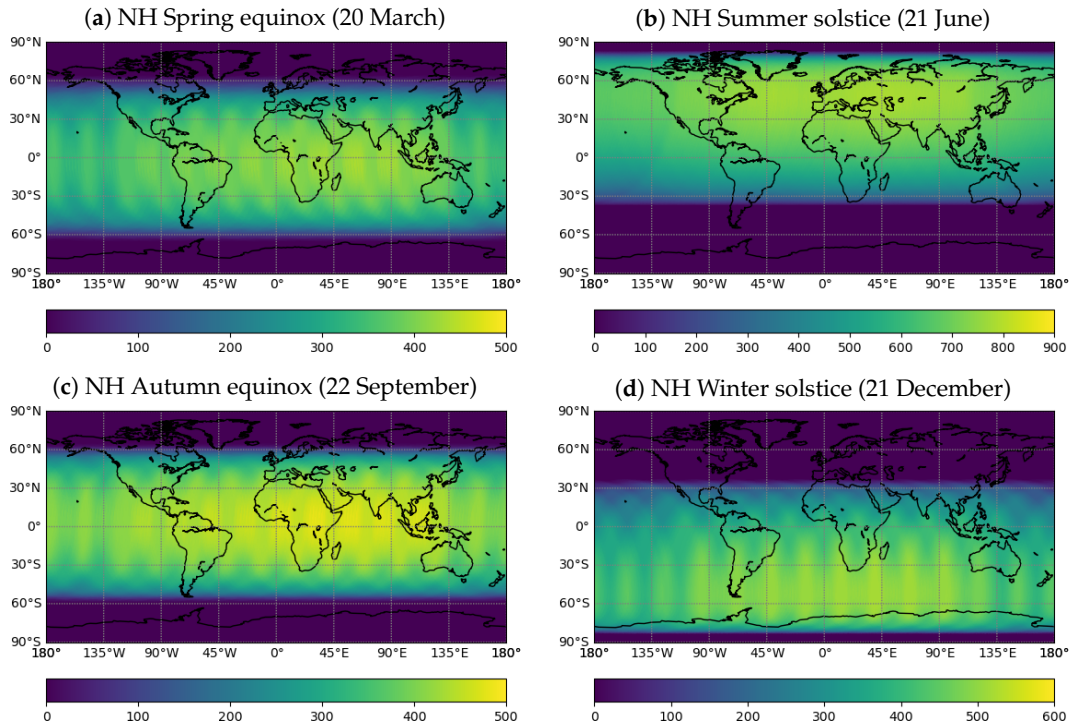


Figure 4.5: Number of EPIC measurements used to create the EPIC NGU clear-sky maps from Figure 4.4. For every example case, the number of available measurements over the Pacific Ocean is in general smaller than in other regions.

2. The selection of the median function as kernel in the Step 2 of the kernel reduction procedure may play a negative role over regions that are mostly cloudy. When generating a daily map, every map cell will store a clear-sky measurement only if at least one half of the measurements within that location and day correspond to scenes without clouds, a condition that might not be found in certain regions for any of the days within the time period of analysis. We also studied the use of a small percentile as kernel function for Step 2. While this improves the cloud cleaning over ocean map cells, it has a severe negative impact over the land map cells, because there is a tendency to select measurements with elevated viewing zenith angles as optimal clear-sky land scenes, and these measurements have much worse spatial resolution and also worse registration (even after the geocorrection of Section 2.2). A compromise solution could be the use of different day-kernels for ocean and land map cells.

The Step 3 of the kernel reduction procedure differs on the temporal gridding with respect to the OCRA implementations for the GOME, GOME-2 and TROPOMI missions. While the latter ones construct the clear-sky maps on a monthly basis and use interpolation for dates in between maps, the OCRA implementation for EPIC uses a moving kernel that allows to generate global clear-sky maps for every day of the year. This will have a positive impact on the description of events that occur at small temporal scales, e.g. ice/snow melting.

4.4 Scaling and offset parameters

For each instrument channel used in the OCRA algorithm, the scaling parameter α_j and the offset parameter β_j can be obtained from the statistical analysis of the histograms of differences Δ_j available within a representative period of time (e.g. for each date). Let $P_{\Delta_j}(x)$ and $F_{\Delta_j}(x)$ be the probability and cumulative distribution functions, respectively, derived from the histograms of Δ_j at a certain date:

- We define the *scaling parameter* α_j by means of the value x_p of the variable x for which the cumulative distribution function $F_{\Delta_j}(x)$ reaches a predefined threshold value p, with $0 \ll p \leq 1$ (Figure 4.6**a**):

$$\alpha_j = \frac{1}{x_p^2} \quad | \quad F_{\Delta_j}(x_p) = p. \tag{4.7}$$

  Typical values of p are 0.99 in the GOME and GOME-2 implementations [79, 80], which use three OCRA channels, and 0.90 in the TROPOMI implementation [81], which uses two OCRA channels.

- We define the *offset parameter* β_j as the value of the variable x in which the probability distribution function $P_{\Delta_j}(x)$ finds its maximum value (Figure 4.6**b**):

$$\beta_j = \arg\max P_{\Delta_j}(x). \tag{4.8}$$

Because the cumulative distribution function $F_{\Delta_j}(x)$ is positive monotonic, an increase in the selected threshold value p translates into an increase in the value x_p and, thus, a decrease in the scaling parameter α_j; conversely, a decrease in the threshold value p causes a decrease in the value x_p and an increase in the scaling parameter α_j. The proper selection of p is a crucial aspect of the scaling parameter estimation, as the upper limit for the radiometric cloud fraction χ_{CF} may depend on the scaling parameters $\{\alpha_j\}_{j\in S_{\text{ocra}}}$. Let consider a scenario in which the offset terms $\{\beta_j\}_{j\in S_{\text{ocra}}}$ are small enough, and let assume that the OCRA quantities $\{\rho_j^{\text{ocra}}\}_{j\in S_{\text{ocra}}}$ are greater than the clear-sky quantities $\{\rho_{j,0}^{\text{ocra}}\}_{j\in S_{\text{ocra}}}$ but smaller than 1:

$$\beta_j \approx 0, \qquad 0 < \rho_{j,0}^{\text{ocra}} < \rho_j^{\text{ocra}} < 1, \qquad \forall\, j \in S_{\text{ocra}},$$

then an upper limit $\chi_{\text{CF}}^{\text{upper}}$ for the radiometric cloud fraction χ_{CF} can be inferred from Eq. 4.2 as follows. First, let the offset terms $\{\beta_j\}_{j\in S_{\text{ocra}}}$ be neglected and, as $\rho_j^{\text{ocra}} > \rho_{j,0}^{\text{ocra}}$:

$$\chi_{\text{CF}} \approx \sqrt{\sum_{j\in S_{\text{ocra}}} \alpha_j \max\left\{0, (\Delta_j)^2\right\}} = \sqrt{\sum_{j\in S_{\text{ocra}}} \alpha_j (\rho_j^{\text{ocra}} - \rho_{j,0}^{\text{ocra}})^2}.$$

This expression can be converted into a lower-equal inequality by replacing each ρ_j^{ocra} with its maximum value 1, and by replacing $\rho_{j,0}^{\text{ocra}}$ with the minimum clear-sky OCRA quantity from the set $\{\rho_{j,0}^{\text{ocra}}\}_{j\in S_{\text{ocra}}}$, i.e.

$$\chi_{\mathrm{CF}} \lesssim \sqrt{\sum_{j\in S_{\mathrm{ocra}}} \alpha_j (1-\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0})^2} = (1-\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0}) \sqrt{\sum_{j\in S_{\mathrm{ocra}}} \alpha_j} = (1-\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0}) \sqrt{\|\boldsymbol{\alpha}\|_1},$$

where $j_{\mathrm{min}} = \arg\min\{\rho^{\mathrm{ocra}}_{j,0}\}_{j\in S_{\mathrm{ocra}}}$ and $\|\cdot\|_1$ denotes the Manhattan norm, and we can conclude that the upper limit $\chi^{\mathrm{upper}}_{\mathrm{CF}}$ for the radiometric cloud fraction χ_{CF} verifies:

$$\chi^{\mathrm{upper}}_{\mathrm{CF}} \lesssim \min\left\{1, (1-\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0}) \sqrt{\|\boldsymbol{\alpha}\|_1}\right\}.$$

In particular, $\chi^{\mathrm{upper}}_{\mathrm{CF}}$ will be lower than 1 if the following criterion holds:

$$(1-\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0})^2 < \frac{1}{\|\boldsymbol{\alpha}\|_1}. \tag{4.9}$$

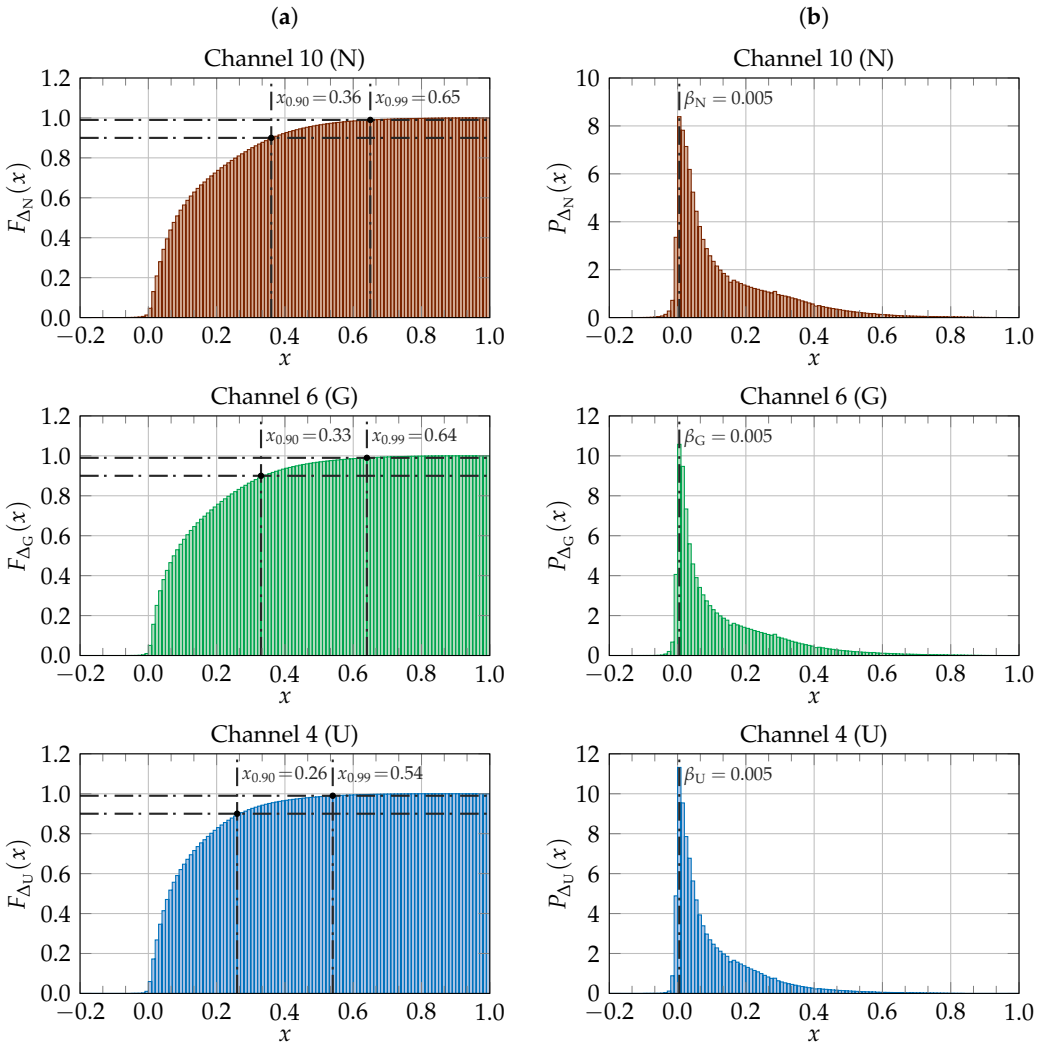


Figure 4.6: Estimation of OCRA scaling and offset parameters for NGU channels of EPIC on 26 September 2017. (**a**) Cumulative distribution functions $F_{\Delta_j}(x)$. (**b**) Probability distribution functions $P_{\Delta_j}(x)$.

From the criterion derived in Eq. 4.9, we infer that χ_{CF} may have an upper-bound below 1 for scenes with high surface reflectivity (i.e. high value of $\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0}$) if the Manhattan norm of $\boldsymbol{\alpha}$ is too small, i.e. if the threshold value p used to determine $\{\alpha_j\}_{j\in S_{\mathrm{ocra}}}$ is too high. For example, a scene in which $\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0} \approx 0.6$ (e.g. permanent ice) would require $\|\boldsymbol{\alpha}\|_1 \gtrsim 6.25$ to ensure that the upper bound for the radiometric cloud fraction definition χ_{CF} is 1. Considering the scaling parameters for the three channels in Figure 4.6:

$$\boldsymbol{\alpha}^{p=0.90}_{\mathrm{NGU}} = [7.72, 6.06, 14.79]^{\mathsf{T}}, \qquad \boldsymbol{\alpha}^{p=0.99}_{\mathrm{NGU}} = [2.37, 2.44, 3.43]^{\mathsf{T}},$$

we observe that the inequality Eq. 4.9 does not hold neither for $\boldsymbol{\alpha}^{p=0.90}_{\mathrm{NGU}}$ nor for $\boldsymbol{\alpha}^{p=0.99}_{\mathrm{NGU}}$, so the radiometric cloud fraction range for scenes with $\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0} \approx 0.6$ will be $[0, 1]$ when using three OCRA channels, as desired. If we restrict OCRA to only use the last two channels in Figure 4.6, i.e. if the scaling vectors are

$$\boldsymbol{\alpha}^{p=0.90}_{\mathrm{GU}} = [6.06, 14.79]^{\mathsf{T}}, \qquad \boldsymbol{\alpha}^{p=0.99}_{\mathrm{GU}} = [2.44, 3.43]^{\mathsf{T}},$$

then the criterion Eq. 4.9 allows us to conclude that, for the scenes in which $\rho^{\mathrm{ocra}}_{j_{\mathrm{min}},0} \approx 0.6$, it is not possible to reach the maximum radiometric cloud fraction of 1 if $p = 0.99$ is combined with the OCRA two-channel approach, while $p = 0.90$ allows to keep the radiometric cloud fraction upper limit still equal to 1.

Since the EPIC global mean TOA reflectances are periodic functions of time t (Figure 2.10 and Figure 2.11) due to the changes in the illumination conditions and the Lissajous orbit of DSCOVR, we also analysed the temporal evolution of the scaling and offset parameters α_j and β_j for two different threshold values $p = 0.99$ and $p = 0.90$. While the offset parameters β_j were found to be stable and can be assumed constant for the EPIC mission, the scaling parameters α_j show an oscillating behaviour (Figure 4.7 and Figure 4.8) which can be modelled in terms of a series of $N = 2$ Lomb-Scargle signals [82–84]:

$$\alpha_j(t) = \bar{\alpha}_j + \sum_{n=1}^{N} \left[A_{j,n} \cos(\omega_{j,n}(t - t_0) + \phi_{j,n})\right], \qquad \omega_{j,n} = \frac{2\pi}{T_{j,n}}, \tag{4.10}$$

where t_0 is the reference time on 1 January 2016 12:00 UTC (with POSIX time 1451649600 s), $\bar{\alpha}_j$ is the mean scaling parameter, and $A_{j,n}$, $\omega_{j,n}$, $T_{j,n}$ and $\phi_{j,n}$ denote the amplitude, pulse, period and initial phase of the n-th Lomb-Scargle signal.

Table 4.4 and Table 4.5 present $\{\beta_j\}_{j\in S_{\mathrm{ocra}}}$ and the Lomb-Scargle signal parameters used to model $\{\alpha_j(t)\}_{j\in S_{\mathrm{ocra}}}$ for the thresholds $p = 0.99$ and $p = 0.90$, respectively, for the EPIC NGU channels. The three scaling parameters are modelled with signals of similar frequency, with mean periods $\overline{T}_{j,1} \approx 363.5$ days and $\overline{T}_{j,2} \approx 183.9$ days, which are similar to the periodicities in the time evolution of the EPIC global mean TOA reflectances in Figure 2.10 and Figure 2.11. Furthermore, for each channel, the location of the signal maxima and minima in Figure 4.7 and Figure 4.8 correspond to the signal minima and maxima in Figure 2.10 and Figure 2.11, which is expected since α_j is inversely correlated to ρ^{ocra}_j, and thus to the EPIC reflectances ρ_j.

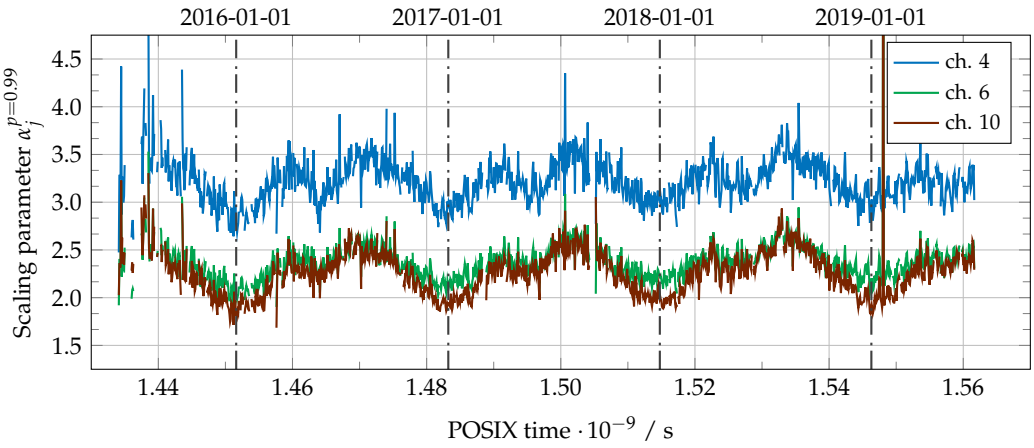


Figure 4.7: Time evolution of α_j for the EPIC NGU channels with $p = 0.99$.

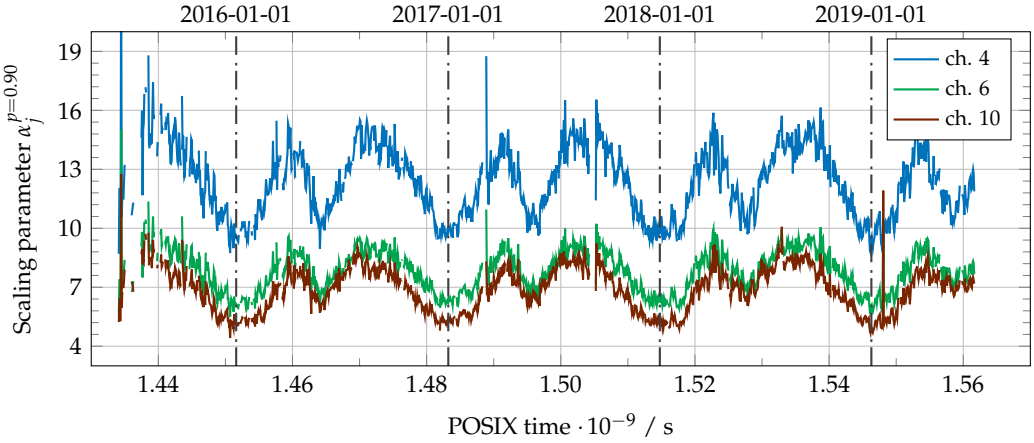


Figure 4.8: Time evolution of α_j for the EPIC NGU channels with $p = 0.90$.

Table 4.4: OCRA offset/scaling parameters for the EPIC NGU channels ($p = 0.99$).

Channel	β_j	$\bar{\alpha}_j$	First signal			Second signal		
			$A_{j,1}$	$T_{j,1}$ / day^{-1}	$\phi_{j,1}$ / rad	$A_{j,2}$	$T_{j,2}$ / day^{-1}	$\phi_{j,2}$ / rad
4	0.007	3.213	0.169	364.9	+2.731	0.136	182.0	−2.420
6	0.006	2.370	0.152	363.2	+2.901	0.074	180.7	−2.358
10	0.005	2.263	0.240	365.3	+2.938	0.093	186.0	−2.044

Table 4.5: OCRA offset/scaling parameters for the EPIC NGU channels ($p = 0.90$).

Channel	β_j	$\bar{\alpha}_j$	First signal			Second signal		
			$A_{j,1}$	$T_{j,1}$ / day^{-1}	$\phi_{j,1}$ / rad	$A_{j,2}$	$T_{j,2}$ / day^{-1}	$\phi_{j,2}$ / rad
4	0.007	12.316	1.397	363.6	+2.576	1.626	184.1	+3.837
6	0.006	7.815	0.897	361.7	+2.673	0.884	184.6	+3.768
10	0.005	6.883	1.183	362.1	+2.821	0.726	185.8	−2.346

4.5 Results

In this section, we show the OCRA radiometric cloud fractions generated for the EPIC mission within the time period from 12 June 2015 to 27 June 2019, and its comparison with the geometric cloud fractions from the MODIS/Terra Level 3 daily product MOD08_D3 [85]. As OCRA quantities ρ^{ocra}, we used the nadir-effective TOA reflectance outlined in Section 4.2 for the EPIC channels at 551 and 388 nm (GU, i.e. $n_{\text{ocra}} = 2$), and the reference clear-sky maps were generated by using the channel-independent approach in Section 4.3. The OCRA scaling and offset parameters are taken from Table 4.5 for the EPIC channels 6 and 4 (i.e. $p = 0.90$).

In contrast to the 3-channel OCRA implementation utilised for the GOME and GOME-2 missions, and following the 2-channel OCRA implementation for TROPOMI, it was decided not to use the EPIC channel 10 at 780 nm, since this is the EPIC reference channel in the oxygen A-band, which will be employed in the retrieval of other cloud macrophysical properties (cloud-top height and cloud optical thickness), and the cloud information content in the oxygen A-band is limited [86–89]. Due to the use of two channels as OCRA quantities instead of three channels, we set $p = 0.90$ as threshold and not $p = 0.99$, in order to avoid the upper-bound limitations in the radiometric cloud fraction definition described in Section 4.4.

Figure 4.9 shows an example of the OCRA procedure. In Figure 4.9**a**, we observe an EPIC NGU image subset converted into nadir-effective reflectances (channel N is included only for visualisation), and Figure 4.9**b** shows the corresponding expected clear-sky scenario. The 2-channel differences of Figure 4.9**a** with respect to Figure 4.9**b** are used to evaluate the OCRA formula in Eq. 4.2 and generate the radiometric cloud fraction in Figure 4.9**c**. We see that the cloud features in Figure 4.9**a** are well captured by the radiometric cloud fraction in Figure 4.9**c**.

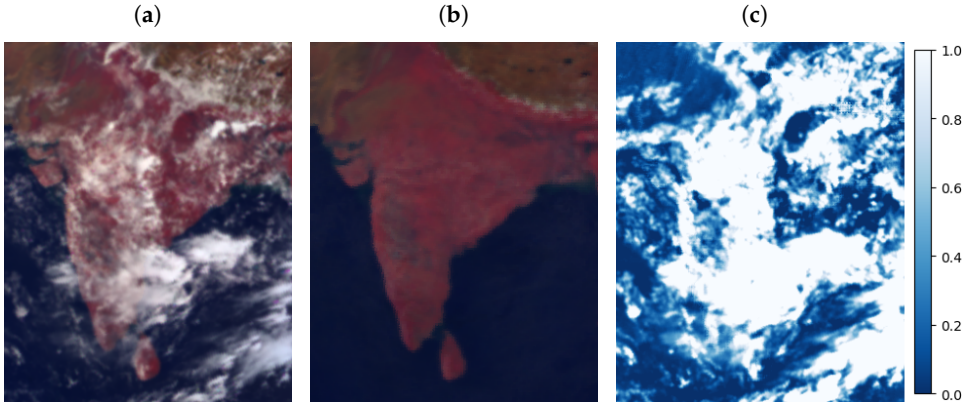


Figure 4.9: Example of OCRA application on an EPIC Level 1B image subset (India and Sri Lanka) on 2017-08-24 05:07:46 UTC: (**a**) nadir-effective reflectance for the EPIC NGU channels, (**b**) clear-sky scene reprojected to the EPIC Level 1B image Earth coordinates, and (**c**) radiometric cloud fraction χ_{CF} using the 2-channel approach (GU) and threshold $p = 0.90$. The first EPIC channel N is included in (**a**) and (**b**) only for visualisation purposes.

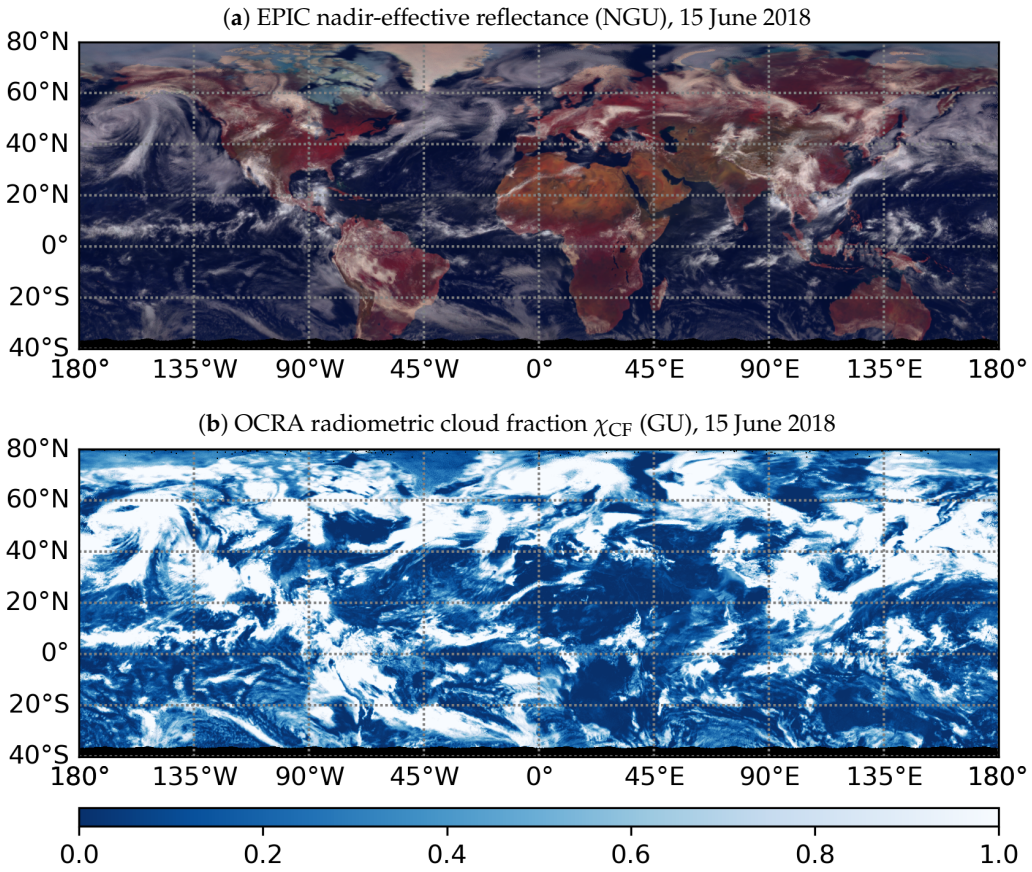


Figure 4.10: Example of daily median-aggregated OCRA radiometric cloud fraction on 15 June 2018, regridded onto an equidistant cylindrical projection map with a spatial grid resolution of 0.2 deg $\times$ 0.2 deg: (**a**) nadir-effective reflectance for EPIC NGU channels, and (**b**) OCRA radiometric cloud fraction χ_{CF} using the 2-channel approach (GU) and $p = 0.90$. The first EPIC channel N is included in (**a**) only for visualisation purposes.

Figure 4.10 presents the median-aggregated OCRA radiometric cloud fraction for all the EPIC datasets available on 15 June 2018. Similarly to Figure 4.9, we can observe that the main cloud features that are visible in the nadir-effective EPIC Level 1B product (Figure 4.9**a**) are well captured in the OCRA radiometric cloud fraction (Figure 4.9**b**).

To perform the comparison between the EPIC OCRA radiometric cloud fraction and the MODIS/Terra geometric cloud fraction, we computed collocated EPIC OCRA daily mean radiometric cloud fraction maps with the same projection and spatial resolution as the MODIS Level 3 daily aggregated products (i.e. equidistant cylindrical projection and spatial grid of 1 deg $\times$ 1 deg). For every MODIS map cell, the selection of EPIC cloud fraction values lying onto the cell which are time-collocated is done using the mean solar time expected for the MODIS cell, i.e. the EPIC values are time-collocated with the MODIS cell if the absolute difference between their hour angles and the cell hour angle is smaller than a given threshold.

The time collocation of MODIS and EPIC values requires, thus, two filtering steps:

1. From the MODIS daily geometric cloud fraction maps, we discard the cells for which the standard deviation of the solar azimuth angle φ_0 is greater than an empirical threshold of 10 deg. For these map cells, the daily geometric cloud fraction uses measurements from different orbits (i.e. measurements acquired with a significant time difference), so they are not suitable for the comparison. As a consequence, measurements from higher latitudes are in general discarded. For the remaining MODIS cells, we estimate the mean local hour angle ω_h using the MODIS/Terra Level 3 mean solar angles:

$$\omega_\mathrm{h} = \arcsin\left[-\frac{\sin\varphi_0 \, \sin\theta_0}{\cos\delta(d_\mathrm{J})}\right], \tag{4.11}$$

   where φ_0 is the solar azimuth angle (defined clockwise from due North), θ_0 is the solar zenith angle, and $\delta(d_\mathrm{J})$ is the Sun's declination approximated by the formula:

$$\delta(d_\mathrm{J}) = (0.410152 \text{ rad}) \, \sin\left[\frac{2\pi}{365}(284 + d_\mathrm{J})\right], \tag{4.12}$$

   in which d_J denotes the Julian day number.

2. For every EPIC dataset used to generate an EPIC daily radiometric cloud fraction map, we determine the expected local hour angle $\omega_\mathrm{h}^\mathrm{MODIS}(x,y)$ for each pixel located at the image pixel coordinates (x,y), and we estimate the actual local hour angle $\omega_\mathrm{h}^\mathrm{EPIC}(x,y)$ by using Eq. 4.11 and the solar angles available within the given EPIC dataset. The pixel value is discarded if:

$$|\omega_\mathrm{h}^\mathrm{EPIC}(x,y) - \omega_\mathrm{h}^\mathrm{MODIS}(x,y)| \geq 15 \text{ deg}, \tag{4.13}$$

   which permits to compare measurements with a maximum time difference of about 1 h. Note that the solar azimuth angle φ_0 in the EPIC datasets uses a reverse sign convention, so a sign change is required before using Eq. 4.11.

Figure 4.11 presents the daily radiometric cloud fraction product from the EPIC sensor using the OCRA algorithm (Figure 4.11**a**) and the daily geometric cloud fraction product from the MODIS/Terra Level 3 collection (Figure 4.11**b**) on 3 November 2017. We see that the main cloud patterns are shared between both products and, as expected, the EPIC product presents in general lower values than the MODIS product (Figure 4.11**c**). The discrepancy between the EPIC and MODIS products is higher under the presence of thin clouds, i.e. clouds with optical thickness lower than 5 (Figure 4.11**d**), while the agreement is higher under the presence of optically thick clouds and under clear-sky conditions. This difference between the radiometric cloud products and the geometric cloud fraction products is expected and has been already reported in previous studies [79–81]: the OCRA algorithm makes use of the EPIC UV–VIS channels, which are not sensitive to optically thin clouds that can be detected by the shortwave infrared (SWIR) channels from MODIS.

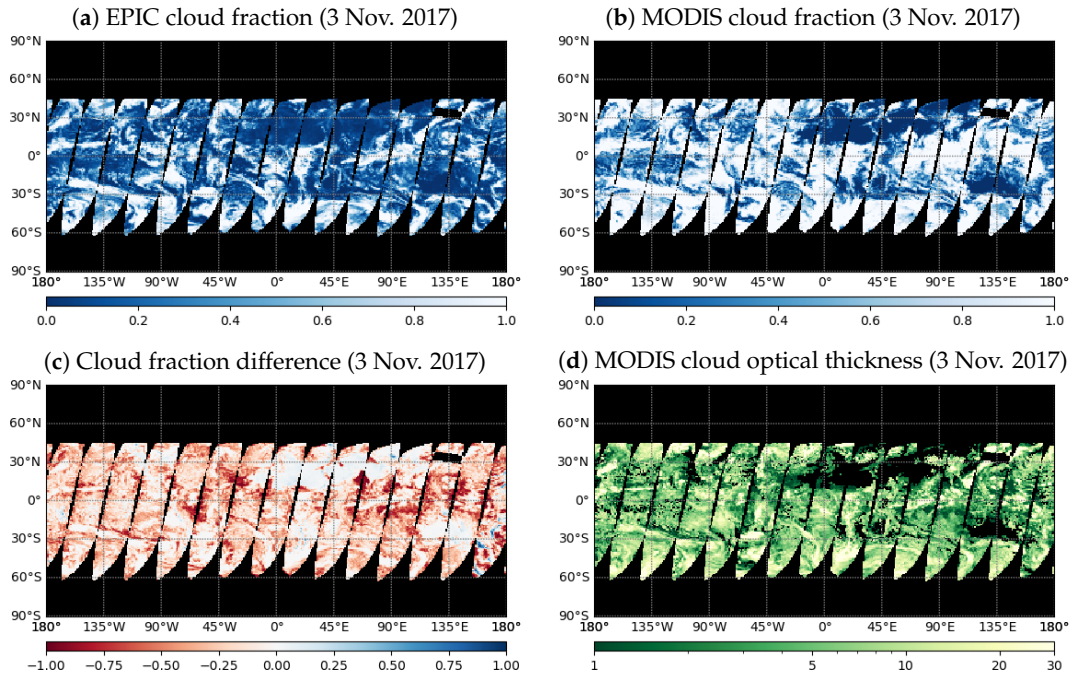


Figure 4.11: Daily comparison on 3 November 2017 between (**a**) EPIC radiometric cloud fraction and (**b**) MODIS geometric cloud fraction. (**c**) Difference between EPIC and MODIS cloud fractions, and (**d**) MODIS cloud optical thickness. OCRA and MODIS results agree very well for optically thick clouds and in clear-sky scenes.

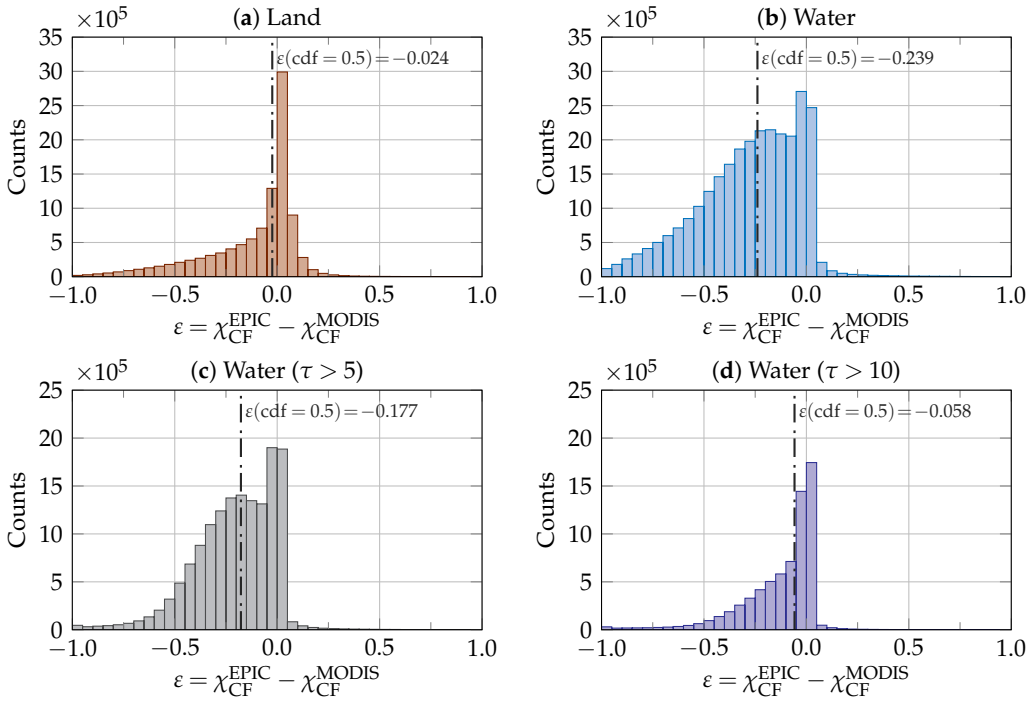


Figure 4.12: Difference between EPIC and MODIS daily cloud fractions for the period from 12 June 2015 to 27 June 2019 (**a**) over land, (**b**) over water, and over water only when the cloud optical thickness τ is (**c**) greater than 5 or is (**d**) greater than 10.

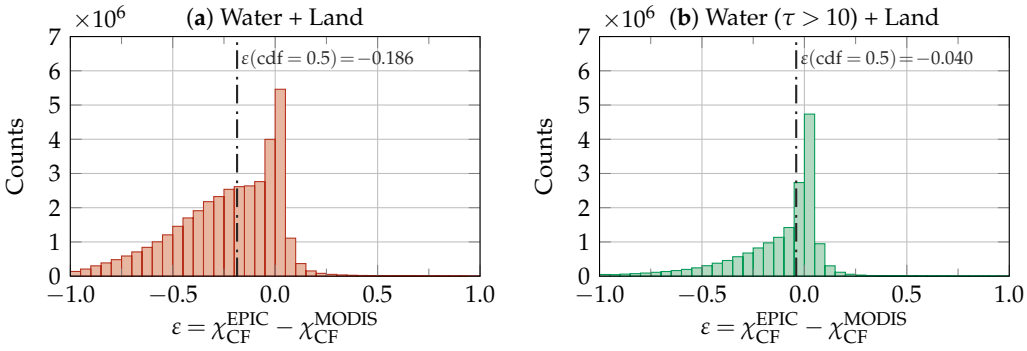


Figure 4.13: Difference between daily EPIC radiometric cloud fractions and MODIS geometric cloud fractions for the time period from 12 June 2015 to 27 June 2019 (**a**) over water plus land, and (**b**) over water locations with optically thick clouds with $\tau > 10$ plus land.

Figure 4.12 shows the histograms of differences between the EPIC daily radiometric cloud fraction and the MODIS daily geometric cloud fraction for the time period from 12 June 2015 to 27 June 2019. Figure 4.12**a** shows the difference over land locations, for which we observe a good agreement, with the histogram median value at -0.024. Figure 4.12**b** presents the difference over water locations; the histogram is displaced to the left of the zero value, showing a global underestimation of the EPIC cloud fraction with respect to the MODIS cloud fraction with a histogram median value of -0.239. The difference between cloud fractions over water can be better analysed after excluding optically thin clouds (i.e. clouds with an optical thickness τ lower than a small threshold). In Figure 4.12**c**, we observe the difference over water locations when only considering clouds for which $\tau > 5$. This histogram still has a similar behaviour to that of Figure 4.12**b**, although the histogram median value is reduced to -0.177. Similarly, in Figure 4.12**d** we see the difference over water locations only when $\tau > 10$. In this case, the histogram shape changes significantly and turns more similar to that of Figure 4.12**a**, with a histogram median value of -0.058.

Figure 4.13 presents the histograms of differences when aggregating land and water locations (Figure 4.13**a**), and when aggregating land locations and those water locations for which the cloud optical thickness τ is greater than 10 (Figure 4.13**b**). The global median difference is located at -0.186, and it decreases to -0.040 after discarding thin clouds over water. These histograms support our findings about the lack of sensitivity of the EPIC UV channels to the presence of optically thin clouds.

References

[72] A. Doicu, D. S. Efremenko, D. Loyola, T. Trautmann, Discrete ordinate method with matrix exponential for stochastic radiative transfer in broken clouds, Journal of Quantitative Spectroscopy and Radiative Transfer 138 (2014) 1–16. doi:10.1016/j.jqsrt.2014.01.011.

[73] EPIC Science Team, DSCOVR/EPIC L2 Cloud HDF-5 File - Version 1 (2018). doi:10.5067/EPIC/DSCOVR/L2_CLOUD_01.

[74] Y. Yang, K. Meyer, G. Wind, Y. Zhou, A. Marshak, S. Platnick, Q. Min, A. B. Davis, J. Joiner, A. Vasilkov, D. Duda, W. Su, Cloud products from the Earth Polychromatic Imaging Camera (EPIC): Algorithms and initial evaluation, Atmospheric Measurement Techniques 12 (3) (2019) 2019–2031. doi:10.5194/amt-12-2019-2019.

[75] D. Loyola, T. Ruppert, A new PMD cloud-recognition algorithm for GOME, ESA Earth Observation Quarterly (58) (1998) 45–47.

[76] A. Delgado-Bonal, A. Marshak, Y. Yang, L. Oreopoulos, Daytime variability of cloud fraction from DSCOVR/EPIC observations, Journal of Geophysical Research: Atmospheres 125 (10) (2020). doi:10.1029/2019jd031488.

[77] D. G. Loyola, J. Xu, K.-P. Heue, 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) (2020) 985–999. doi:10.5194/amt-13-985-2020.

[78] V. Molina García, S. Sasi, D. S. Efremenko, D. Loyola, Improvement of EPIC/DSCOVR image registration by means of automatic coastline detection, Remote Sensing 11 (15) (2019) 1747. doi:10.3390/rs11151747.

[79] D. Loyola, W. Thomas, Y. Livschitz, T. Ruppert, P. Albert, R. Hollmann, Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, IEEE Transactions on Geoscience and Remote Sensing 45 (9) (2007) 2747–2758. doi:10.1109/tgrs.2007.901043.

[80] R. Lutz, D. Loyola, S. Gimeno García, F. Romahn, OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B, Atmospheric Measurement Techniques 9 (5) (2016) 2357–2379. doi:10.5194/amt-9-2357-2016.

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

[82] N. Lomb, Least-squares frequency analysis of unequally spaced data, Astrophysics and Space Science 39 (2) (1976) 447–462. doi:10.1007/bf00648343.

[83] J. Scargle, Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data, The Astrophysical Journal 263 (1982) 835–853. doi:10.1086/160554.

[84] K. Hocke, Phase estimation with the Lomb-Scargle periodogram method, Annales Geophysicae 16 (1998) 356–358.

[85] S. Platnick, et al., MODIS Atmosphere L3 Daily Product, NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA (2015). doi:10.5067/MODIS/MOD08_D3.061.

[86] O. Schüssler, D. G. Loyola Rodriguez, A. Doicu, R. Spurr, Information content in the oxygen A-band for the retrieval of macrophysical cloud parameters, IEEE Transactions on Geoscience and Remote Sensing 52 (6) (2014) 3246–3255. doi:10.1109/tgrs.2013.2271986.

[87] A. B. Davis, G. Merlin, C. Cornet, L. C. Labonnote, J. Riédi, N. Ferlay, P. Dubuisson, Q. Min, Y. Yang, A. Marshak, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: An optimal estimation approach, Journal of Quantitative Spectroscopy and Radiative Transfer 216 (2018) 6–16. doi:10.1016/j.jqsrt.2018.05.007.

[88] A. B. Davis, N. Ferlay, Q. Libois, A. Marshak, Y. Yang, Q. Min, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: A physics-based approach, Journal of Quantitative Spectroscopy and Radiative Transfer 220 (2018) 84–96. doi:10.1016/j.jqsrt.2018.09.006.

[89] M. Gao, P.-W. Zhai, Y. Yang, Y. Hu, Cloud remote sensing with EPIC/DSCOVR observations: A sensitivity study with radiative transfer simulations, Journal of Quantitative Spectroscopy and Radiative Transfer 230 (2019) 56–60. doi:10.1016/j.jqsrt.2019.03.022.

Chapter 5

EPIC-ROCINN: Estimation of cloud properties from EPIC/DSCOVR

In this chapter, we implement and analyse the Retrieval Of Cloud Information using Neural Networks (ROCINN) to estimate cloud optical thickness and cloud-top height of liquid-water clouds from EPIC/DSCOVR measurements. Our main contributions are the following:

1. The input space definition needs to be addressed for the specific geometries of EPIC. Replacing the viewing zenith angle with the relative zenith angle ensures that the neural networks will be later trained for back-scattering viewing geometries while unfeasible geometries will be skipped, and replacing the relative azimuth angle with the scattering angle allows to keep the handling of perpendicular incident or viewing geometries (i.e. undefined relative azimuth angle) outside of the neural network training process.
2. The radiative transfer model for EPIC can be replaced by two feed-forward neural networks corresponding to the clear-sky and fully-cloudy scenarios, and one single hidden layer is enough to reach errors lower than 0.3% for 80% of the validation dataset when using 64 neurons for the clear-sky neural network and 128 neurons for the fully-cloudy neural network.
3. The sensitivity analysis shows that only the cloud-top height can be retrieved with certain accuracy for most sources of uncertainty. Combining noise of SNR equal to 290 and fixing the cloud effective particle size to 13 µm and the cloud geometrical thickness to 2 km, 80% of the validation dataset allows to retrieve the cloud optical thickness and cloud-top height with absolute errors within $[-26, +38]$ and $[-0.74, +0.92]$ km, respectively. The retrieved cloud optical thicknesses are strongly affected by the selection of a fixed cloud effective particle size and by negative offsets in the cloud fraction.
4. The conclusions arisen from the sensitivity analysis agree with the comparison between the retrieved ROCINN cloud products and the operational EPIC cloud products for one test example.

Conference contributions

V. Molina García, D. Loyola, Retrieval of cloud properties from EPIC/DSCOVR with ROCINN/OCRA: Status report, oral presentation at DSCOVR Science Team Meeting, 6–8 October 2020, Greenbelt, MD (United States), National Aeronautics and Space Administration, 2020.

V. Molina García, D. S. Efremenko, A. del Águila, Automatic differentiation for Jacobian computations in radiative transfer problems, 21st European Workshop on Automatic Differentiation, 19–20 November 2018, Jena (Germany), Friedrich Schiller University Jena, 2018.

5.1 Introduction

The Retrieval Of Cloud Information using Neural Networks (ROCINN) is an algorithm for retrieving cloud macrophysical parameters from remote sensing radiance measurements [90]. In its more general approach, ROCINN solves an optimisation problem in which the original (and time-expensive) radiative transfer model (RTM) assumed as forward model is replaced by a weighted linear combination of two trained feed-forward neural networks:

$$\boldsymbol{y} = \mathrm{RTM}(\boldsymbol{x}, \boldsymbol{p}) \approx (1 - \chi_{\mathrm{CF}})\,\mathrm{NN}_{\text{clear-sky}}(\boldsymbol{p}_{\mathrm{sg}}) + \chi_{\mathrm{CF}}\,\mathrm{NN}_{\text{cloud}}(\boldsymbol{x}, \boldsymbol{p}_{\mathrm{sg}}, \boldsymbol{p}_{\mathrm{c}}) = \mathrm{RTM}_{\mathrm{NN}}(\boldsymbol{x}, \boldsymbol{p}), \quad (5.1)$$

where $\boldsymbol{y}$ is the measurements vector, $\boldsymbol{x}$ is the state vector (i.e. the cloud properties allowed as retrieval variables), $\boldsymbol{p}$ denotes a generic data vector, $\boldsymbol{p}_{\mathrm{sg}}$ is the data vector of surface and geometry parameters, $\boldsymbol{p}_{\mathrm{c}}$ is the data vector consisting of the non-retrieved cloud properties, and χ_{CF} is the cloud fraction, which acts as weight of the weighted linear combination.

The optimiser can be any standard non-linear least squares solver, such as Gauss-Newton [91], Levenberg-Marquardt [92] or Tikhonov [93]. Throughout this chapter, we will utilise the Levenberg-Marquardt algorithm as ROCINN optimiser, which balances between the Gauss-Newton and the gradient descent methods [94] and is more robust than the standard Gauss-Newton even if the initial solution is far from the final minimum. The objective function is:

$$\Phi(\boldsymbol{x}, \boldsymbol{p}) = ||\boldsymbol{y}^{\delta} - \mathrm{RTM}_{\mathrm{NN}}(\boldsymbol{x}, \boldsymbol{p})||^2, \quad (5.2)$$

and the solution is found with successive updates of $\boldsymbol{x}$ by an update step $\Delta\boldsymbol{x}$ obtained from:

$$[J^{\mathsf{T}} J + \alpha\,\mathrm{diag}(J^{\mathsf{T}} J)]\Delta\boldsymbol{x} = J^{\mathsf{T}}[\boldsymbol{y} - \mathrm{RTM}_{\mathrm{NN}}(\boldsymbol{x}, \boldsymbol{p})], \quad (5.3)$$

where J is the Jacobian of the model function with respect to the state vector components, and $\alpha > 0$ is a damping factor adjusted at each iteration, normally by reducing its value if $\Phi(\boldsymbol{x}, \boldsymbol{p})$ reduces fast, and by increasing it if $\Phi(\boldsymbol{x}, \boldsymbol{p})$ increases. For $\alpha \to 0$, the method is equivalent to Gauss-Newton, whereas for $\alpha \to \infty$, the method is equivalent to gradient descent. In this chapter, the initial value of α is set to $0.001\ \max\{\mathrm{diag}(J^{\mathsf{T}} J)\}$. For the purpose of this research, a standard feed-forward neural network class has been implemented in Python. The derivatives of this neural network implementation (and, thus, the elements of J) are obtained by automatic differentiation with the self-implemented Python library `dnumpy` [95], which extends a subset of `numpy` [96] functionality with derivative propagation using function decorators.

The description of the neural network input/output space and its hyperparameter analysis is provided in Section 5.2. In Section 5.3, we present the sensitivity analysis in the retrieval of two cloud macrophysical parameters (cloud optical thickness and cloud-top height) based on different sources of uncertainty. Finally, in Section 5.4 we show a comparison between the retrieved ROCINN cloud products with the operational EPIC/DSCOVR cloud products for one test example and provide a discussion of the results.

5.2 Neural network configuration

5.2.1 Input space definition

The first step in the neural network training process is the generation of an appropriate input sample, i.e. the set of points distributed through the input space of the forward model to be approximated by the neural network. A proper selection of the input points is crucial for the neural network training and validation processes to succeed [97], and it should consider:

- the selection and number N of input variables,
- the distribution of the input sample points within the input space, and
- the minimal number of input sample points n which is enough to describe the features of the forward model being approximated.

The set of input variables for radiative transfer modelling consist of—at least—two surface parameters (surface altitude h_s and surface albedo A_s) and three geometric parameters (relative azimuth angle $\Delta\varphi$, solar zenith angle θ_0, viewing zenith angle θ) to model the clear-sky scenes. This set of input variables is increased with the set of cloud properties (cloud effective particle size r_{eff}, cloud optical thickness τ_c, cloud-top height h_t and cloud geometrical thickness Δh_c) to model the cloudy scenes. To sum up, the input space for the neural network modelling the clear-sky scenes has a number of dimensions $N_{\text{clear}} = 5$, and the input space for the neural network modelling the cloudy scenes has a number of dimensions $N_{\text{cloudy}} = 9$.

Given the number of input dimensions N, the generation of each input sample point is divided in three stages:

1. The generation of a point within the $\mathbb{R}^N$ unit hypercube by means of a low-discrepancy sequence generator. An example of this kind is the Halton generator [98], which creates sample points relying on a set of N van der Corput generators [99] in which the sequence bases are coprime numbers.

2. The rescaling of the point components into the valid ranges for every input dimension and, if required, the remapping of the point components by means of an importance function [97] that mimics the probability density function of the input variables, which may not be necessarily uniform.

3. The validation of the input point based on additional criteria, such as physical or geometrical constraints. If these restrictions are not held by the input point, it is discarded.

It is common practice to substitute the input geometric parameters $[\Delta\varphi, \theta_0, \theta]^\mathsf{T}$ with their corresponding cosines $[\cos\Delta\varphi, \mu_0, \mu,]^\mathsf{T}$, where $\mu_0 = \cos\theta_0$ and $\mu = \cos\theta$, since this is the actual dependency of the radiative transfer model on the geometric parameters. The selection of

the input geometric parameters for our problem is modified from this standard setup because of two additional aspects that need to be taken into consideration:

1. The input variables are selected under the assumption that they are independent from each other. Due to the viewing geometry of the EPIC instrument, this assumption does not hold for the pair $[\mu_0, \mu]^\mathsf{T}$, since the relative zenith angle $\Delta\theta = \theta - \theta_0$ is always lower than approximately 12 deg in absolute value (Figure 5.1). This means that a sampling method that considers $[\mu_0, \mu]^\mathsf{T}$ as independent input variables will generate a significant number of points of the input space that are not representative of the EPIC viewing geometries, e.g. combinations of small solar zenith angles and high viewing zenith angles (see Figure 5.2). From this fact, the input geometric parameters are modified from $[\cos\Delta\varphi, \mu_0, \mu]^\mathsf{T}$ into $[\cos\Delta\varphi, \mu_0, \Delta\theta]^\mathsf{T}$, from which we may recalculate μ as

$$\mu = \cos\theta = \cos\left[\arccos\mu_0 + \Delta\theta\right],$$

   and which requires to add a geometric restriction in order to consider an input point as valid, since θ must be restricted to the range $[0, 90)$ deg:

$$0 \text{ deg} \leq \arccos\mu_0 + \Delta\theta < 90 \text{ deg}. \tag{5.4}$$

2. From the nature of the spherical coordinates system, $\Delta\varphi$ is undefined (and so is $\cos\Delta\varphi$) whenever θ_0 or θ are equal to 0 (i.e. whenever μ_0 or μ are equal to 1). In such situations, variations in $\cos\Delta\varphi$ should not cause variation in the neural network output values. Since this behaviour cannot be captured well enough by the neural networks, it is recommended to replace $\cos\Delta\varphi$ with the cosine of the scattering angle $\cos\Theta$ to describe the geometric configuration, since $\cos\Theta$ is well-defined when θ_0 or θ are equal to 0 and $\cos\Delta\varphi$ can always be recovered from the relation:

$$\cos\Theta = -\mu_0\mu + [(1-\mu_0^2)(1-\mu^2)]^{1/2}\cos\Delta\varphi. \tag{5.5}$$

In summary, the *free* input geometric parameters used for the generation of the input sample points should be $[\cos\Theta, \mu_0, \Delta\theta]^\mathsf{T}$, since they allow us to generate valid $[\Delta\varphi, \theta_0, \theta]^\mathsf{T}$ points with appropriate geometric constraints for the EPIC instrument while avoiding the configurations with undefined $\Delta\varphi$ for normal incident Sun light or nadir view.

The importance functions for the two surface properties (i.e. altitude and albedo) need to be generated from auxiliary database sources. In the case of the surface elevation, the importance function is obtained from the statistics of ETOPO1 data [100] regridded to a resolution of 0.2 deg (see Figure 5.3**a**). For the surface albedo, the importance function is computed from the statistics of MERIS black-sky albedo climatology [101] interpolated to the four EPIC channels in the oxygen bands (see Figure 5.3**b**). The use of these importance functions during the input space sample generation will increase the concentration of sample points with surface altitudes close to 0 km and also smaller surface albedos ($A_s < 0.2$), which correspond to compatible configurations over ocean.

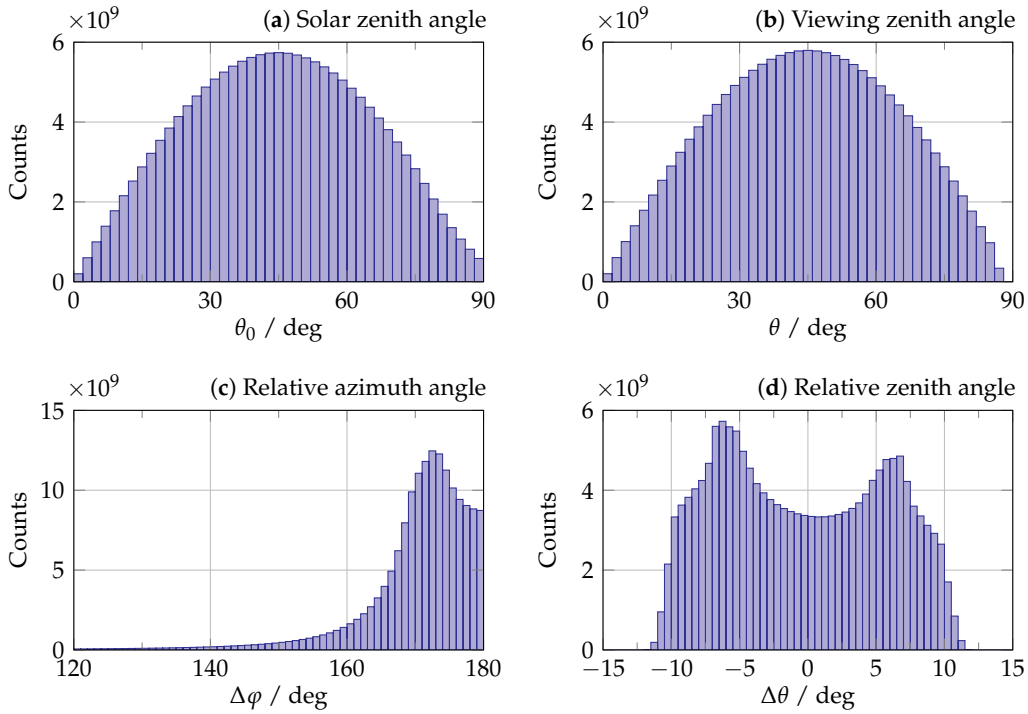


Figure 5.1: Distribution of geometric variables for the EPIC mission during the period from 12 June 2015 to 27 June 2019: (**a**) solar zenith angle θ_0, (**b**) viewing zenith angle θ, (**c**) relative azimuth angle $\Delta\varphi$ and (**d**) relative zenith angle $\Delta\theta$. We observe that the distributions of the solar and viewing zenith angles are very similar, and the distribution of their relative differences is always below 12 deg in absolute value.

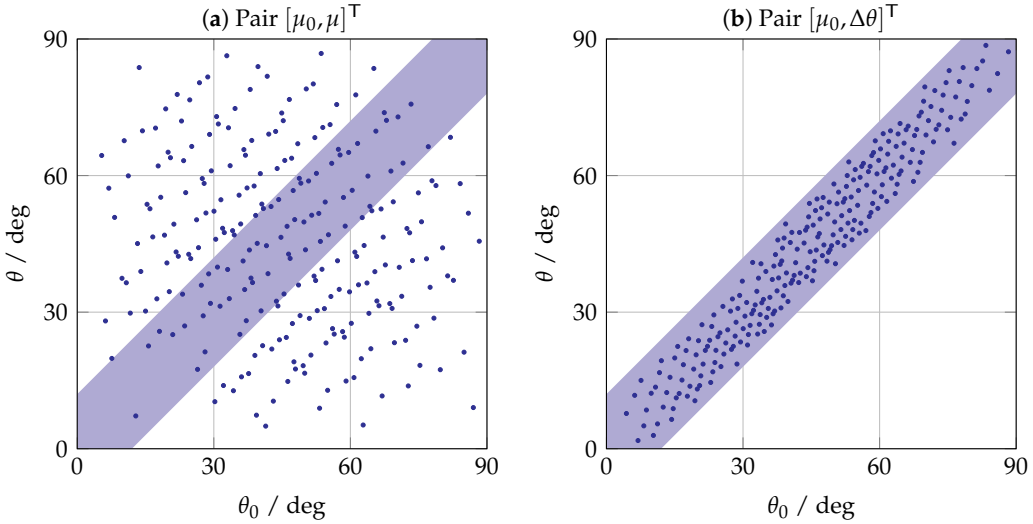


Figure 5.2: Distribution of 250 input sample points in the projected space of $[\theta_0, \theta]^\mathsf{T}$ when the sample generator relies on (**a**) the pair $[\mu_0, \mu]^\mathsf{T}$ and (**b**) the pair $[\mu_0, \Delta\theta]^\mathsf{T}$. The light blue area is the region where $|\Delta\theta| < 12$ deg, and which corresponds to the real EPIC viewing geometries.

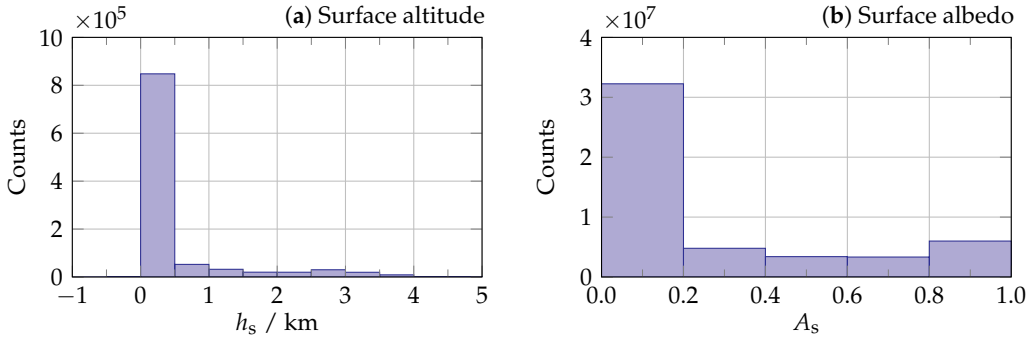


Figure 5.3: Distribution of surface variables for the EPIC mission as obtained from the auxiliary databases (i.e. ETOPO1 for the surface altitude, MERIS climatology for the surface albedo).

The input space sample generation does not apply any importance function to the components corresponding to the cloud properties, in order to cover the domain of definition for the potential retrieved parameters as uniformly as possible. However, we need to apply physical restrictions, e.g. to avoid unfeasible configurations in which the cloud geometrical thickness is bigger than the cloud-top height. The domains of definition were chosen as $[4, 20]$ µm for r_{eff}, $(0, 150]$ for τ_c, $[1, 15]$ km for h_c and $(0, 4]$ km for Δh_c. Figure 5.4 shows the resulting distributions for the four cloud variables when generating the first 20000 sample points. While r_{eff} and τ_c follow a quasi-uniform distribution, the physical restriction $\Delta h_c \leq h_c$ causes a decrease of points for $h_c < 4$ km, and a global negative slope in the distribution of Δh_c.

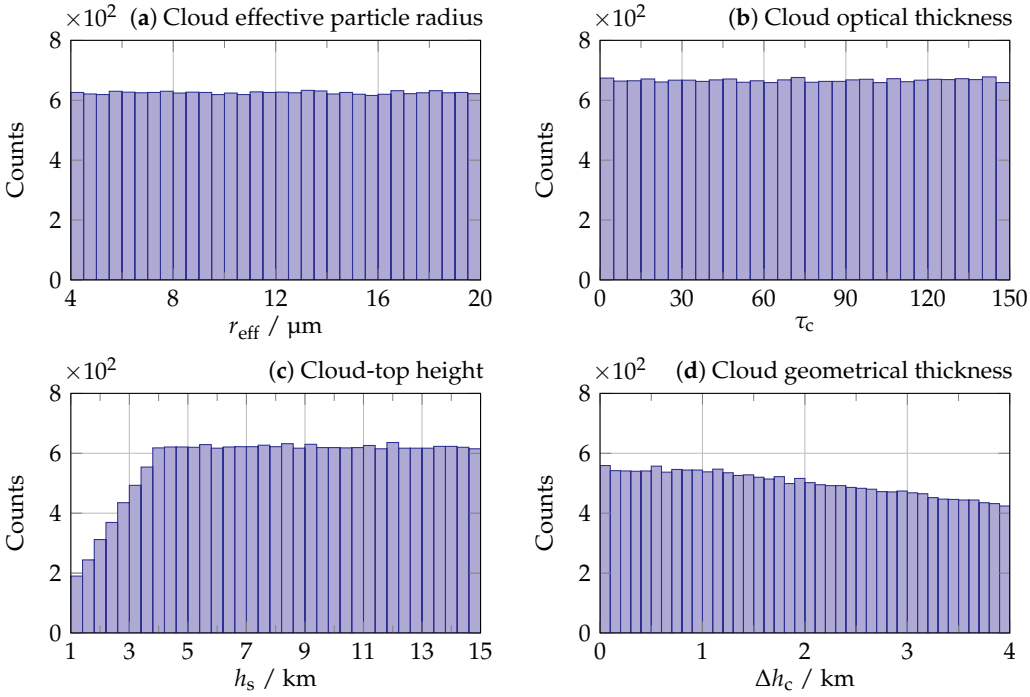


Figure 5.4: Distributions of the cloud variables as generated by a Halton sequence of 20000 points without importance sampling but applying physical constraints.

5.2.2 Validation dataset generation

The neural network training processes require to generate training and validation datasets, consisting of vectors of input parameters (five if dealing with clear-sky scenes, nine for cloudy scenes) and the corresponding vectors of output values (four radiances). The training data is employed by an optimiser to find the sets of biases and weights that allow the neural network to better reproduce the underlying physical model. The validation data is used to ensure that the neural network does not overfit the training data.

To obtain a first rough estimate of the minimum sample size that is representative of the population of atmospheric scenes, we analysed basic statistics of the radiative transfer output space as a function of the sample size. Figure 5.5 and Figure 5.6 show the evolution of the radiance mean and standard deviation, respectively, as a function of the input sample size for the clear-sky and cloudy scenes. We see that the convergence of the radiance mean and standard deviation is not reached at least until considering 10000 sample points. For validation purposes, next sections will therefore use a validation sample consisting of 20000 points.

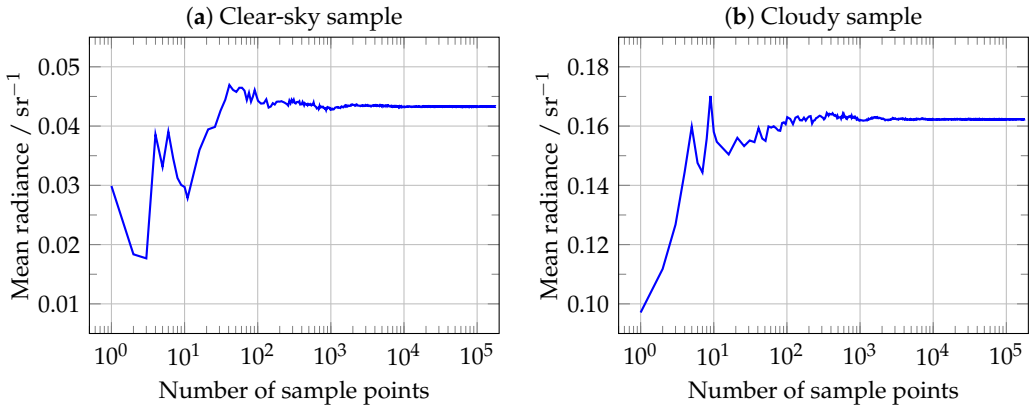


Figure 5.5: Sun-normalised radiance mean as a function of the number of sample points for (**a**) clear-sky and (**b**) cloudy scenes.

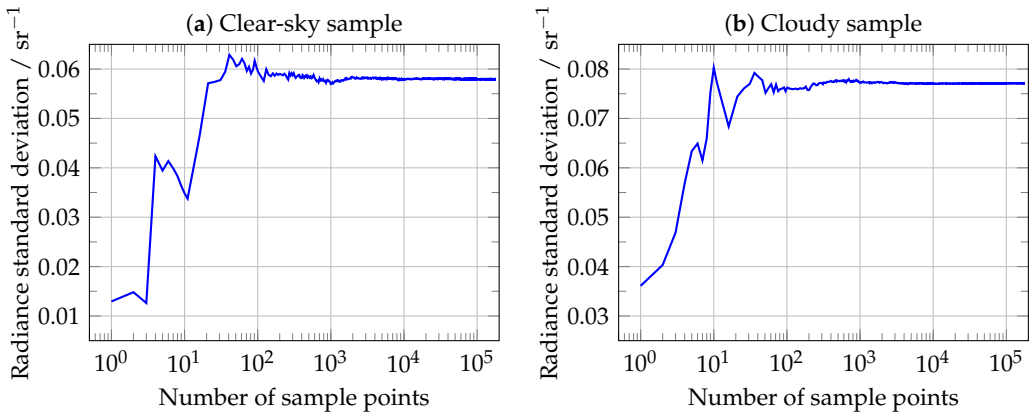


Figure 5.6: Sun-normalised radiance standard deviation as a function of the number of sample points for (**a**) clear-sky and (**b**) cloudy scenes.

5.2.3 Hyperparameter analysis

The hyperparameter analysis in machine learning studies the influence of the model hyperparameters on the performance of a trained machine learning model [102]. By model hyperparameters, we refer to those parameters that are not computed or optimised as part of the training process of the machine learning model. For instance, neural network hyperparameters include the topology and number of neurons per hidden layer, the activation functions, the optimiser, the tolerance convergence value to consider for the residuals and the gradient update, and the maximum number of iterations.

Our hyperparameter analysis focused on the influence of the number of neurons and the training dataset size on the performance of the trained neural networks modelling clear-sky and cloudy scenes as seen by EPIC. The other hyperparameters were fixed: the topology was set to multi-layer perceptron; the number of hidden layers was set to one; the activation functions were set to hyperbolic tangent for the connection between input and hidden layers, and to linear for the connection between hidden and output layers; the optimiser was chosen to L-BFGS-B [103, 104], the tolerance convergence was set to 10^{-12} for the residuals and 10^{-9} for the gradient update; the maximum number of iterations was set to a high value to ensure that the training processes stop due to the former tolerance values. Specific to the L-BFGS-B optimiser is the maximum number of variable metric corrections utilised to define the limited memory matrix, which was fixed to 64. All the neural networks were trained with `scikit-learn` [105], and both training dataset inputs and outputs were scaled with standard scalers (i.e. scalers that convert the training dataset inputs and outputs so that each input and output component will have mean equal to 0 and standard deviation equal to 1).

In Figure 5.7, we show the performance of the trained neural networks for clear-sky and cloudy scenes as a function of the training dataset size (from 600 to 160000) and the number of neurons in the hidden layer (from 8 to 160, trained every 8 neurons). To analyse the neural network performance, we focused on the mean absolute error (MAE) of the neural networks to predict the outputs of a validation dataset of 20000 points, and on an overfitting ratio defined as the quotient of the MAE on the validation dataset over the MAE on the training dataset. A neural network is considered well trained when the validation MAE remains small and the overfitting ratio is close to 1. High values in the validation MAE and an overfitting ratio close to 1 indicate that the neural network suffers from underfitting, i.e. the neural network configuration is too simple to encapsulate the underlying physical model from the training data. High values for the validation MAE and high overfitting ratios indicate that the neural network suffers from overfitting, i.e. the training dataset is not representative of the data population, or the neural network configuration contains too many parameters (biases and weights) with respect to the training dataset size, and the final selection of biases and weights fits the training data too well but do not encapsulate the global behaviour of the underlying physical model.

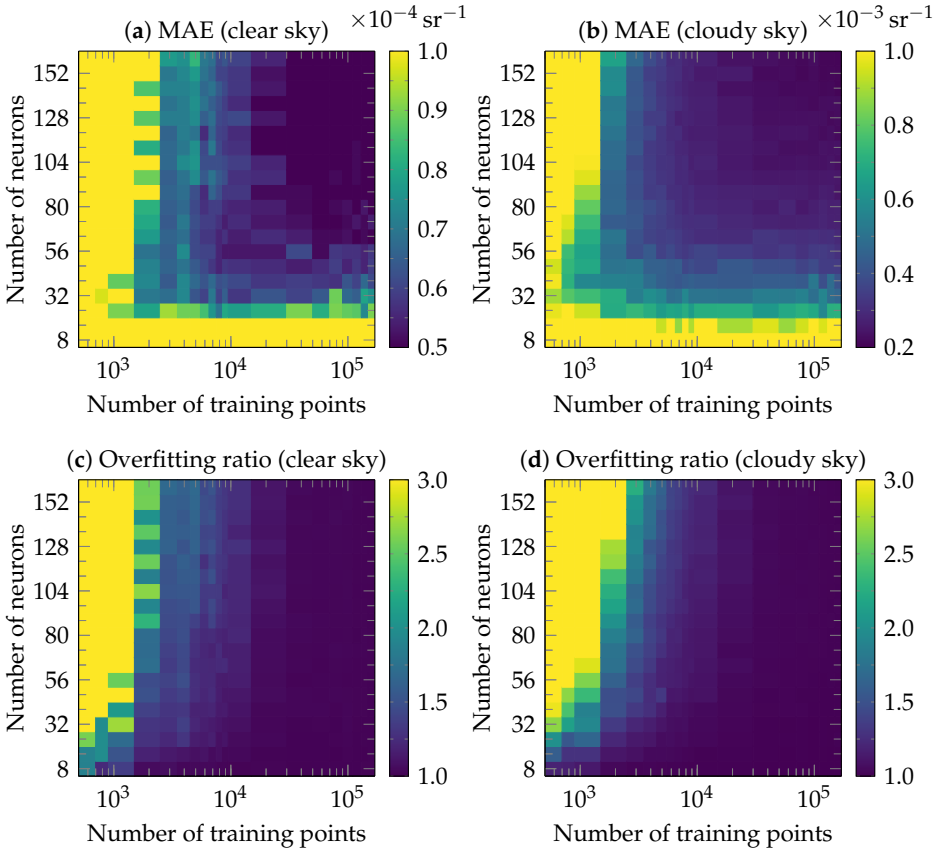


Figure 5.7: Performance of the neural networks simulating the clear-sky and cloudy scenes as a function of the training dataset size and the number of neurons in the hidden layer. (**a**) and (**b**) show the MAE on the validation dataset, while (**c**) and (**d**) show the overfitting ratio.

From Figure 5.7**a** and Figure 5.7**c**, at least 100000 training points and 64 neurons are required to ensure a validation MAE lower than 5.3×10^{-5} sr^{-1} and an overfitting ratio below 1.025 for the neural network simulating clear-sky scenes. From Figure 5.7**b** and Figure 5.7**d**, at least 100000 training points and 128 neurons are required to ensure a validation MAE lower than 2.3×10^{-4} sr^{-1} and an overfitting ratio below 1.025 for the neural network simulating cloudy scenes. In general, we see that there is not a unique optimal configuration neural network for each case. After a sufficient training dataset size and number of neurons, the validation MAE tends to stabilise. It is also interesting to point that the overfitting ratio tends to increase with the number of neurons if the number of training points is fixed.

For the clear-sky scenes, we chose the neural network trained with 64 neurons and 120000 training points. For the cloudy scenes, we chose the neural network trained with 128 neurons and 140000 training points. The performance of these networks is shown in Figure 5.8. For both neural networks, 80% of the outputs from the validation dataset can be predicted with a relative error smaller than about 0.3%.

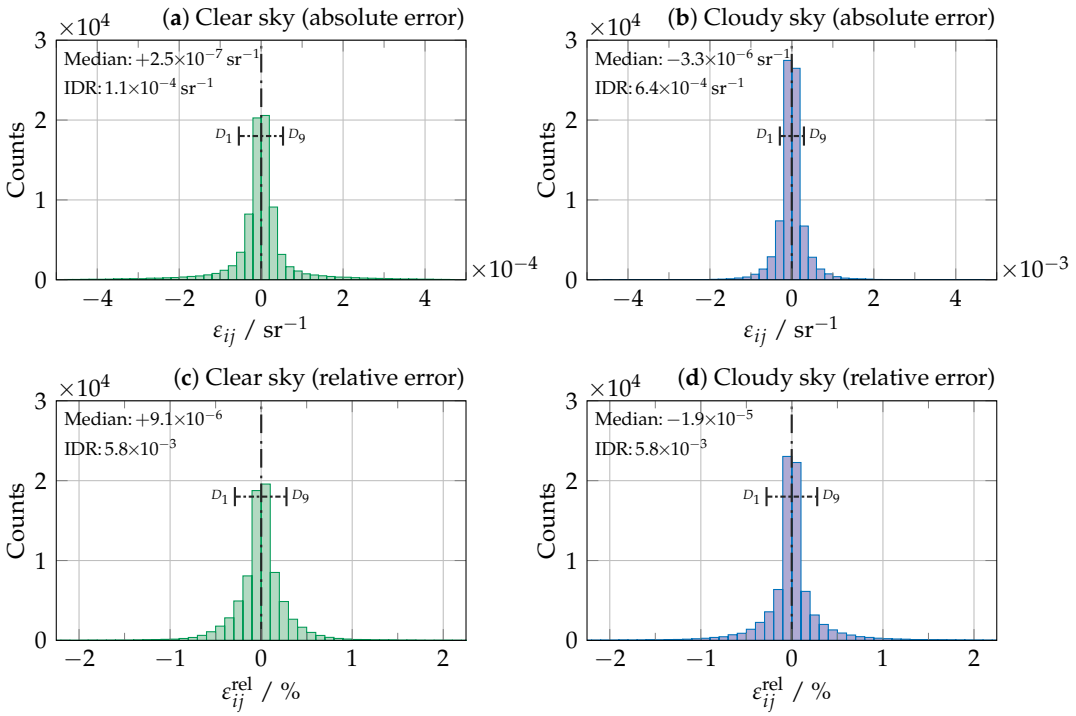


Figure 5.8: Neural network performance (64 neurons and 120000 training points for clear sky; 128 neurons and 140000 training points for cloudy sky) for the validation datasets; (**a**) and (**b**) show the errors on the predicted validation outputs in absolute value, whereas (**c**) and (**d**) show the corresponding relative errors.

5.3 Sensitivity analysis

In this section, we perform a sensitivity analysis on the retrieval of two cloud macrophysical parameters, cloud optical thickness (COT) and cloud-top height (CTH), based on various known sources of uncertainty. The forward model for this analysis is the set of neural networks chosen in the previous section, and replace the original radiative transfer model DOME applied to the EPIC instrument. Unless otherwise stated, the sensitivity analysis is performed on the set of $N = 20000$ examples corresponding to the validation dataset described in the previous section, the scenes under study are fully cloudy, the initial guess is $10^{1.5} \approx 31.6$ for COT and 10 km for CTH, and only one source of uncertainty is addressed at a time.

5.3.1 Sensitivity to instrument measurement noise

The first source of uncertainty in study is the instrument noise, which is described by means of the signal-to-noise ratio (SNR). For the EPIC instrument, this maximum SNR is 290 [106], while the SNR according to the design specification is about 200 [107], and the value may even be lower (about 40) in dark areas such as clear-sky scenes over ocean (as stated by J. Herman at DSCOVR Science Team Meeting 2019).

We performed the retrieval of cloud macrophysical properties (COT and CTH) over the synthetic validation data sample under the assumption of perfect knowledge of both CER and CGT for each sample scene, and using perturbed Sun-normalised radiances $\{\boldsymbol{y}_i^{\delta}\}_{i\in N}$ given in terms of their exact theoretical values $\{\boldsymbol{y}_i\}_{i\in N}$:

$$y_{ij}^{\delta} = y_{ij} + \delta_{ij}^{\text{noise}}, \qquad j = 7, 8, 9, 10, \tag{5.6}$$

where j is the EPIC channel index and $\delta_{ij}^{\text{noise}}$ is a random variable with mean $\mu_{\text{noise}} = 0$ and standard deviation defined as $\sigma_{\text{noise}} = y_{ij}/\text{SNR} \geq 0$, and which lets us rewrite Eq. 5.6 as:

$$y_{ij}^{\delta} = y_{ij} + \delta_{ij}^{\text{noise}} = y_{ij} + \mathcal{N}(0, \sigma_{\text{noise}}^2) = y_{ij} + \sigma_{\text{noise}} \cdot \mathcal{N}(0,1) = y_{ij}\left(1 + \frac{\mathcal{N}(0,1)}{\text{SNR}}\right). \tag{5.7}$$

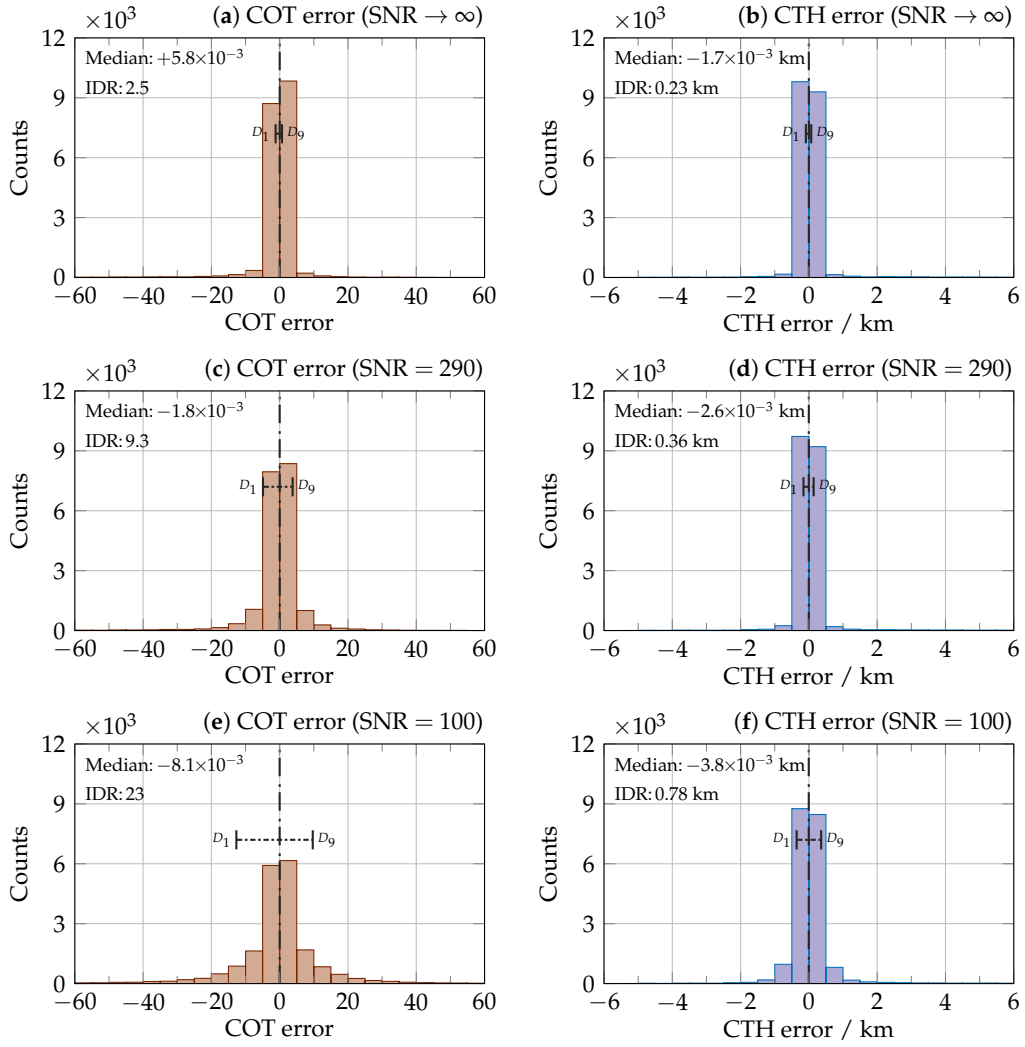


Figure 5.9: Absolute error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the SNR: (**a**) and (**b**) for SNR $\to \infty$ (i.e. no noise), (**c**) and (**d**) for an SNR equal to 290 (maximum instrument value), and (**e**) and (**f**) for an SNR equal to 100.

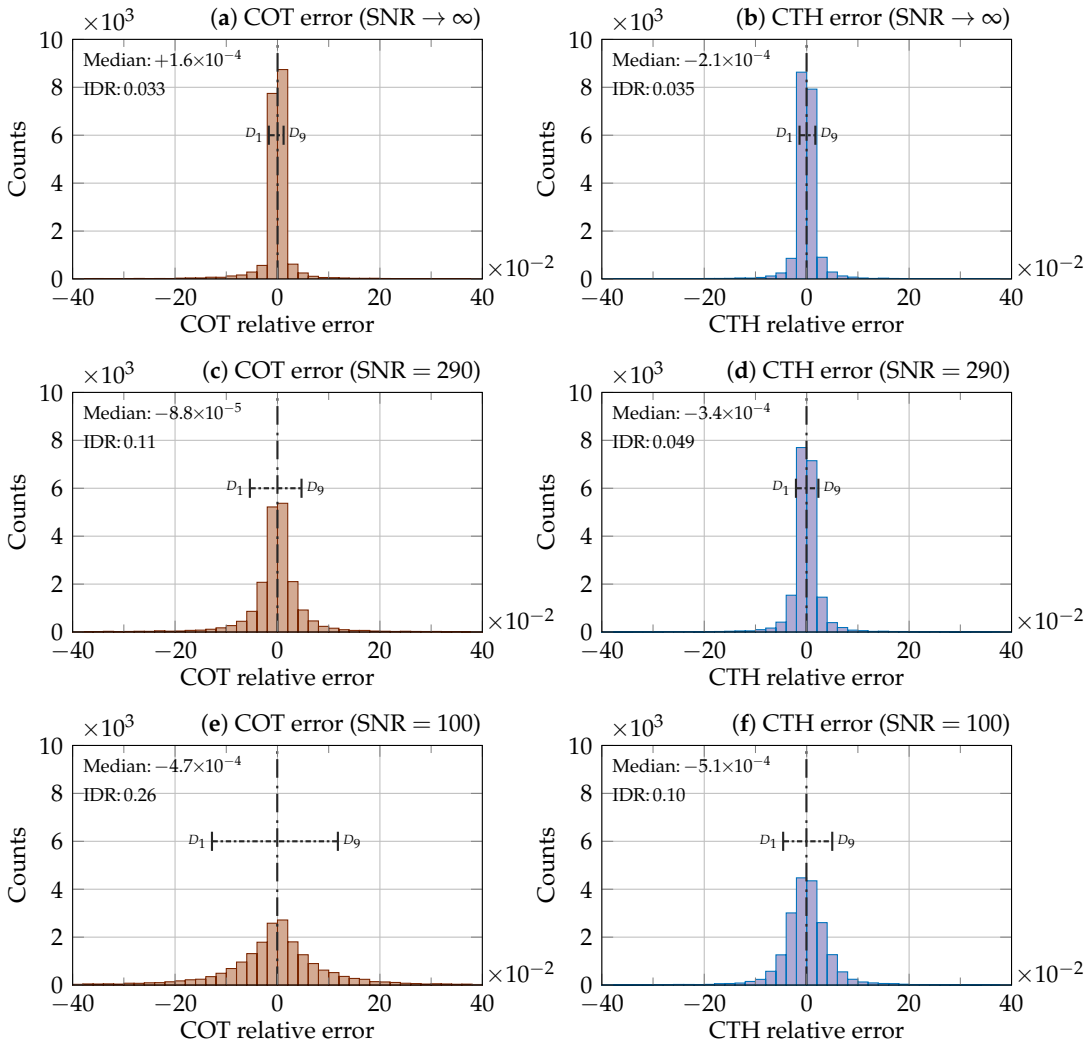


Figure 5.10: Relative error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the SNR: (**a**) and (**b**) for SNR $\to \infty$ (i.e. no noise), (**c**) and (**d**) for an SNR equal to 290 (maximum instrument value), and (**e**) and (**f**) for an SNR equal to 100.

Figure 5.9 shows the absolute retrieval errors in COT and CTH on the synthetic validation sample when three different values of SNR are used (infinity, 290 and 100), whereas the corresponding relative retrieval errors are shown in Figure 5.10. In the absence of instrument noise (Figure 5.9**a** and Figure 5.9**b**), the median absolute errors were $+5.8 \times 10^{-3}$ for the COT and -1.7 m for the CTH, and 80% of the absolute errors were found within $[-1.50, +0.99]$ for the COT and within $[-0.12, +0.10]$ km for the CTH, and 80% of the relative errors were found within $[-1.9\%, +1.4\%]$ for the COT and within $[-1.6\%, +1.9\%]$ for the CTH (Figure 5.10**a** and Figure 5.10**b**); the rate of converged retrievals was 98.5%. This case would correspond to the ideal situation in which no noise would be present in the EPIC measurements and we would also know all the non-retrieved parameters with complete accuracy. If the SNR is decreased

to its maximum value of 290 for EPIC (Figure 5.9**c** and Figure 5.9**d**), the median errors remain negligible in both retrieved parameters, but the histograms of errors are spread so that 80% of the absolute errors are found within $[-5.2, +4.1]$ for the COT and within $[-0.19, +0.17]$ km for the CTH, and 80% of the relative errors are found within $[-5.6\%, +4.9\%]$ for the COT and within $[-2.3\%, +2.5\%]$ for the CTH (Figure 5.10**c** and Figure 5.10**d**); the rate of converged retrievals decreases slightly to 98.3%. Finally, if the SNR is decreased to 100 (Figure 5.9**e** and Figure 5.9**f**), the histograms of errors spread more so that 80% of the absolute errors are found within $[-13, +10]$ for the COT and within $[-0.39, +0.39]$ km for the CTH, and 80% of the relative errors are found within $[-13\%, +12\%]$ for the COT and within $[-4.8\%, +5.2\%]$ for the CTH (Figure 5.10**e** and Figure 5.10**f**); the rate of converged retrievals decreases to 97.5%. In general, we observe that the retrieval of COT is more sensitive to the SNR than the retrieval of CTH. This result is expected, as the COT is closely related to the absolute measurements in the non-absorbing O_2 channels, whereas the CTH is closely related to the absorption-continuum ratio within the corresponding O_2 channels.

In Figure 5.11, we show the dependency of the absolute retrieval error statistics as a function of the SNR. For both COT and CTH, the absolute median error remains close to 0 for any SNR greater than 50 (Figure 5.11**a** and Figure 5.11**b**). The spread of the error histograms can be inferred from the IDR, and it is a decreasing function of the SNR (Figure 5.11**c** and Figure 5.11**d**). Both results are expected, since the Gaussian noise has mean 0 (i.e. unbiased), and the size of the noise is modulated with σ_{noise}, which is inversely proportional to the SNR.

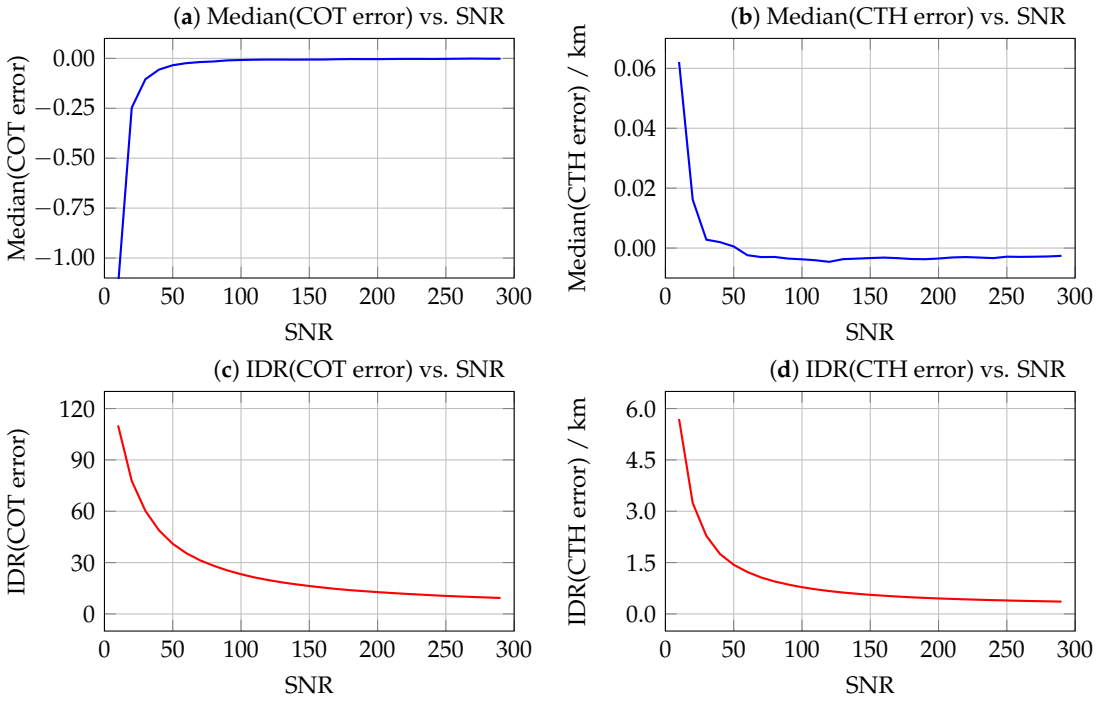


Figure 5.11: Dependency of the absolute retrieval error statistics (median and IDR) as a function of the SNR.

5.3.2 Sensitivity to fixed cloud geometrical thickness

Another source of retrieval uncertainties is the value defined for the cloud geometrical thickness (CGT) in the radiative transfer model. Since the information content in the EPIC channels located at the oxygen bands is limited, only the cloud-top height (CTH) can be retrieved with an acceptable uncertainty [108, 109]. As a consequence, the radiative transfer model needs to define the CGT based on an *a priori* assumption, for example, by setting the CGT to a fixed value. This is the approach followed by the TROPOMI/Sentinel-5P cloud processor, where the CGT is fixed to 1 km [90].

We performed the retrieval of cloud macrophysical properties (COT and CTH) over the synthetic validation data sample under the assumption of measurements without noise, perfect knowledge of CER and different fixed values of CGT (which differ from their actual ones).

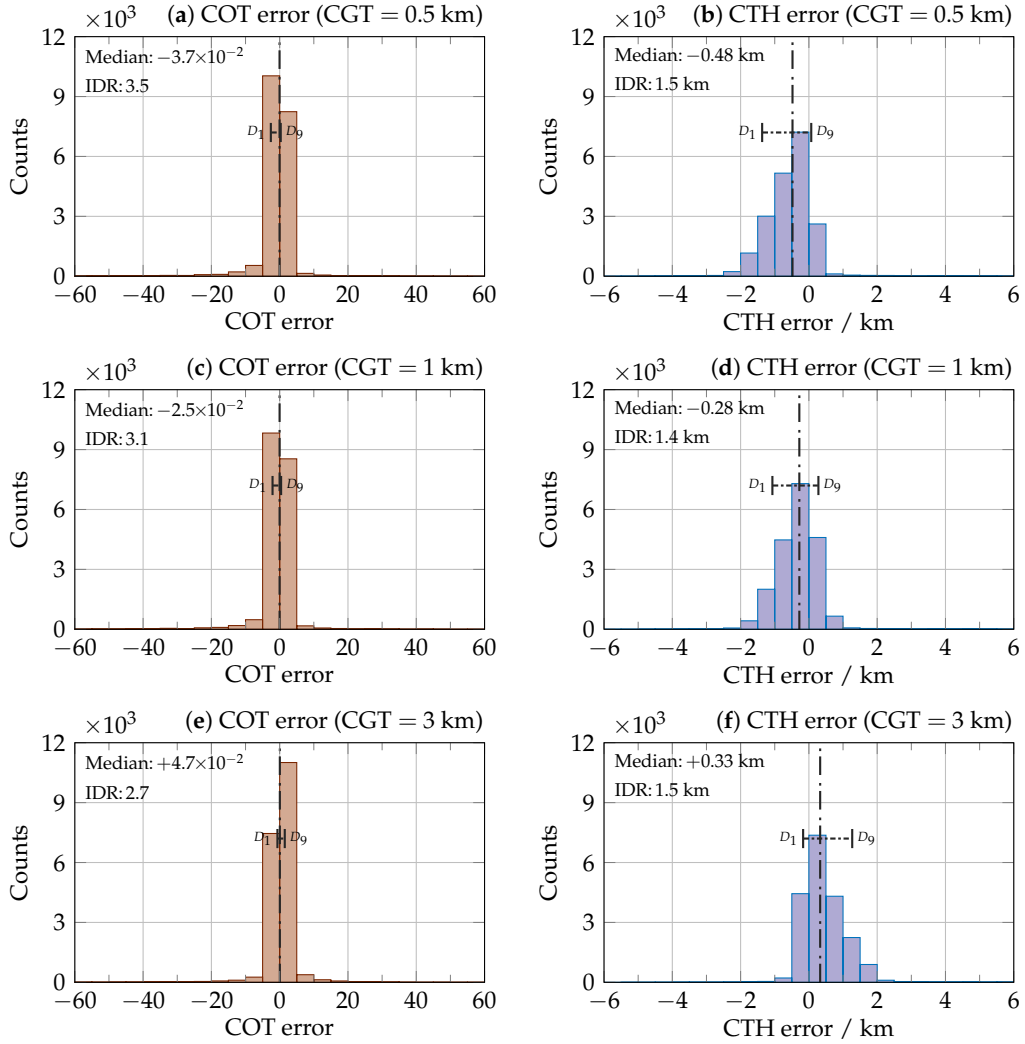


Figure 5.12: Absolute error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the fixed CGT: (**a**) and (**b**) for CGT = 0.5 km, (**c**) and (**d**) for CGT = 1 km, and (**e**) and (**f**) for CGT = 3.0 km.

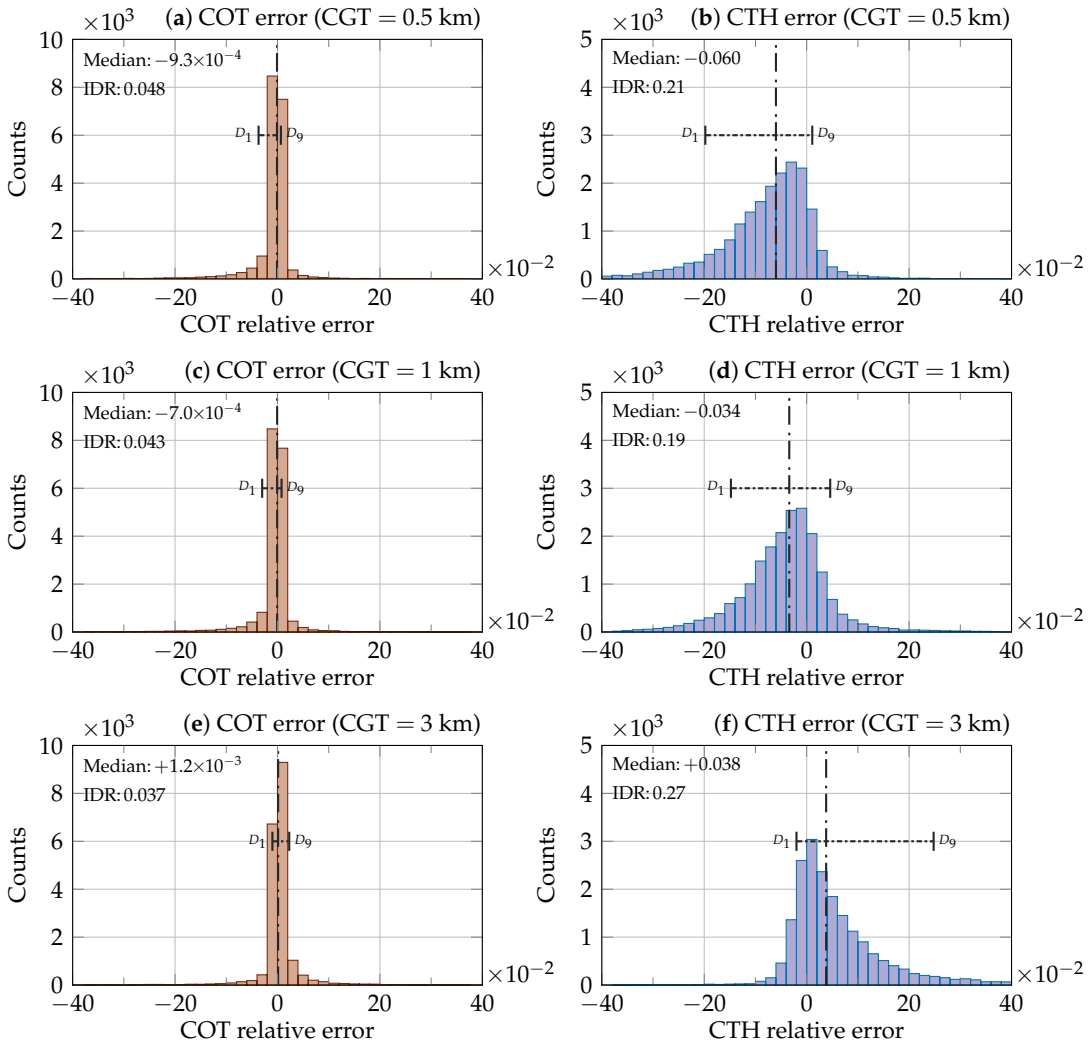


Figure 5.13: Relative error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the fixed CGT: (**a**) and (**b**) for CGT = 0.5 km, (**c**) and (**d**) for CGT = 1 km, and (**e**) and (**f**) for CGT = 3.0 km.

Figure 5.12 presents the absolute retrieval errors in COT and CTH on the synthetic validation sample if the CGT is set to a fixed value (0.5 km, 1 km and 3 km); the corresponding relative retrieval errors are shown in Figure 5.13. We see that setting the CGT to a fixed value does not have a significant impact on the retrieved COT independently from its value (Figure 5.12**a**, Figure 5.12**c** and Figure 5.12**e**). In the three cases, the median absolute errors in COT is close to 0, and at least 80% of the absolute errors are within $[-2.9, +1.8]$, i.e. at least 80% of the relative errors are within $[-3.9\%, +2.5\%]$. The rate of converged retrievals was at least 98.4%.

In Figure 5.12**b**, Figure 5.12**d** and Figure 5.12**f**, we observe the statistics of absolute retrieval errors for CTH on the synthetic validation data sample for the cases in which CGT is fixed to 0.5 km, 1 km and 3 km, respectively. When CGT is fixed to 0.5 km, the absolute errors in the retrieved CTH present a median bias of -0.48 km and 80% of the absolute errors are

found within $[-1.4, +0.10]$ km. When CGT is fixed to 1 km, the median bias as well as the interval for the IDR are shifted to -0.28 km and $[-1.1, +0.31]$ km, respectively. Finally, if CGT is fixed to 3 km, we observe a median bias of $+0.33$ km and the IDR interval is $[-0.20, +1.3]$ km. The histograms of absolute errors in CTH are in general non-symmetric with respect to its median value: they have a negative skew when the fixed CGT tends to underestimate the actual CGT values, and they have a positive skew when the fixed CGT tends to overestimate them.

The statistics of relative retrieval errors in CTH are shown in Figure 5.13**b**, Figure 5.13**d** and Figure 5.13**f**. We see that the bias in the median relative error in CTH is an increasing function of the fixed CGT, moving from -6.0% when CGT is set to 0.5 km, to -3.4% when CGT is set to 1 km, and to $+3.8\%$ when CGT is fixed to 3 km. The IDR interval is not constant as it occurred for the absolute errors in CTH: the relative errors in CTH are within $[-20\%, +1.3\%]$ when CGT is fixed to 0.5 km, it shifts and narrows to $[-15\%, +4.8\%]$ when CGT is fixed to 1 km, and it shifts and stretches to $[-2.2\%, +25\%]$ when CGT is fixed to 3 km. The effect in the skewness is similar to the one observed in the histograms of absolute errors in CTH.

In Figure 5.14, we show the dependency of the absolute retrieval error statistics as a function of the fixed CGT. For both COT and CTH, the absolute median error is an increasing function of the fixed CGT (Figure 5.14**a** and Figure 5.14**b**), and their values are close to 0 when the CGT is fixed to about 2 km, which is the middle point of the CGT domain definition used to train the cloudy-sky neural network. For both COT and CTH, the IDR is a convex function of the fixed CGT (Figure 5.14**c** and Figure 5.14**d**), but their minima are found at different values: at about 2.4 km for the error in COT, and at about 1.6 km for the error in CTH.

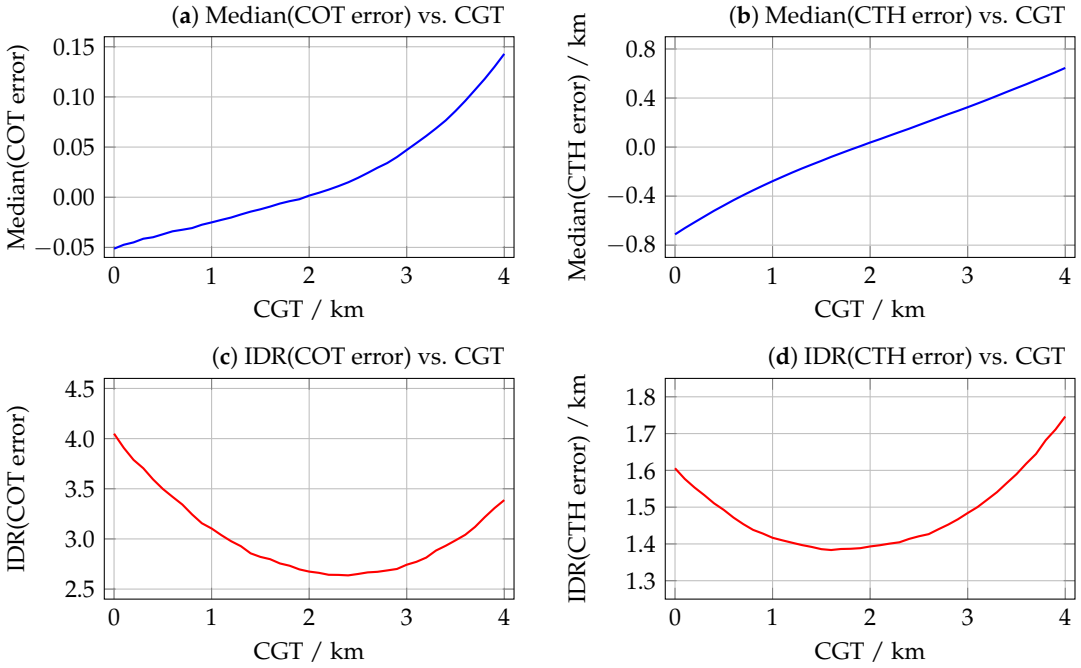


Figure 5.14: Dependency of the absolute retrieval error statistics (median and IDR) as a function of the fixed CGT.

5.3.3 Sensitivity to fixed cloud effective particle radius

As described in Section 3.2, liquid-water clouds can be modelled as a homogeneous layer of Mie water droplets whose radii follow a Gamma size distribution characterised by the cloud effective particle radius (CER) [110]. Several instruments exploit the use of radiance measurements at shortwave infrared spectral windows containing information on CER, e.g. at 1.6, 2.1 and 3.7 µm [111–116]. Since none of these spectral windows are available in the EPIC instrument, the radiative transfer model needs to define the CER based on an *a priori* assumption. Similarly to the CGT, one approach is to fix the CER to a predefined value, as done by the operational EPIC/DSCOVR cloud processor, where the CER is set to 14 µm for liquid-water clouds [117]; and by the TROPOMI/Sentinel-5P cloud processor [90].

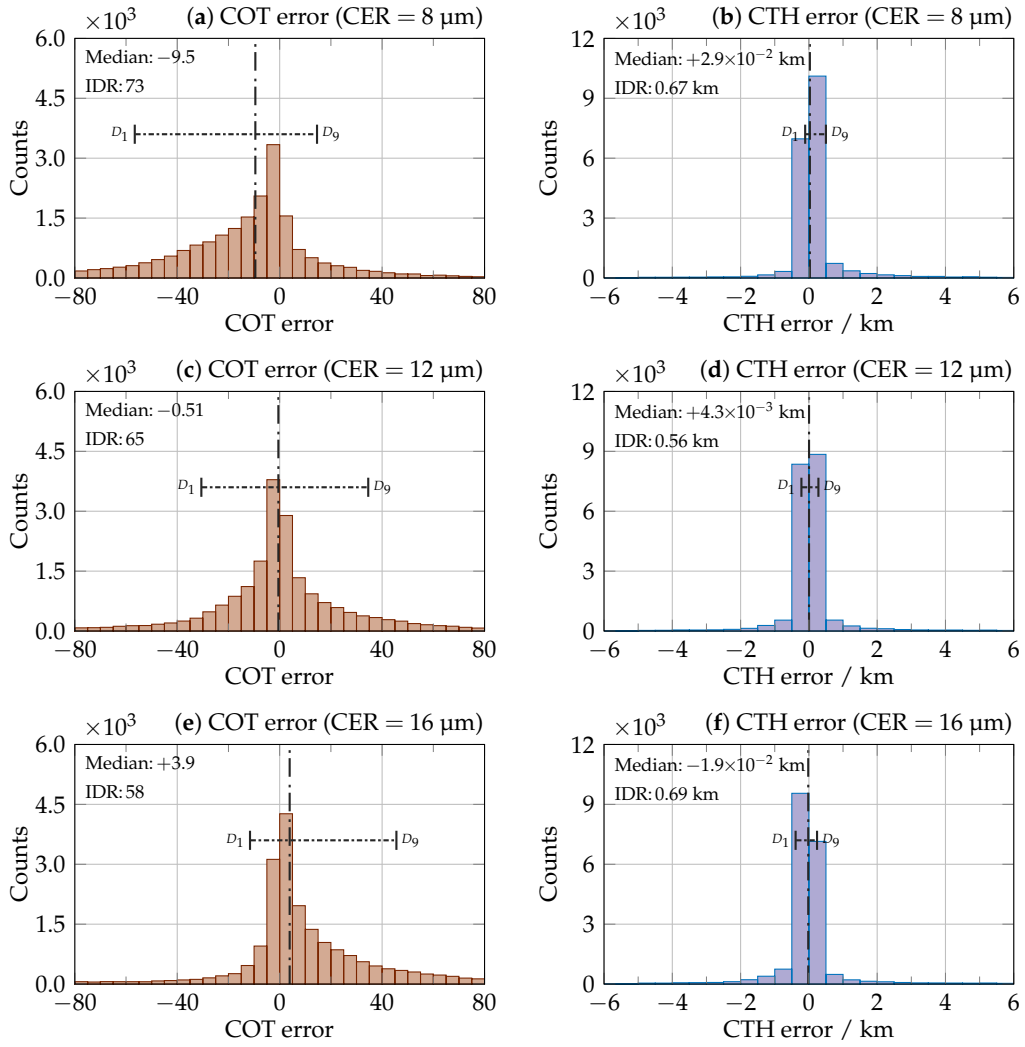


Figure 5.15: Absolute error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the fixed CER: (**a**) and (**b**) for CER = 8 µm, (**c**) and (**d**) for CER = 12 µm, and (**e**) and (**f**) for CER = 16 µm.

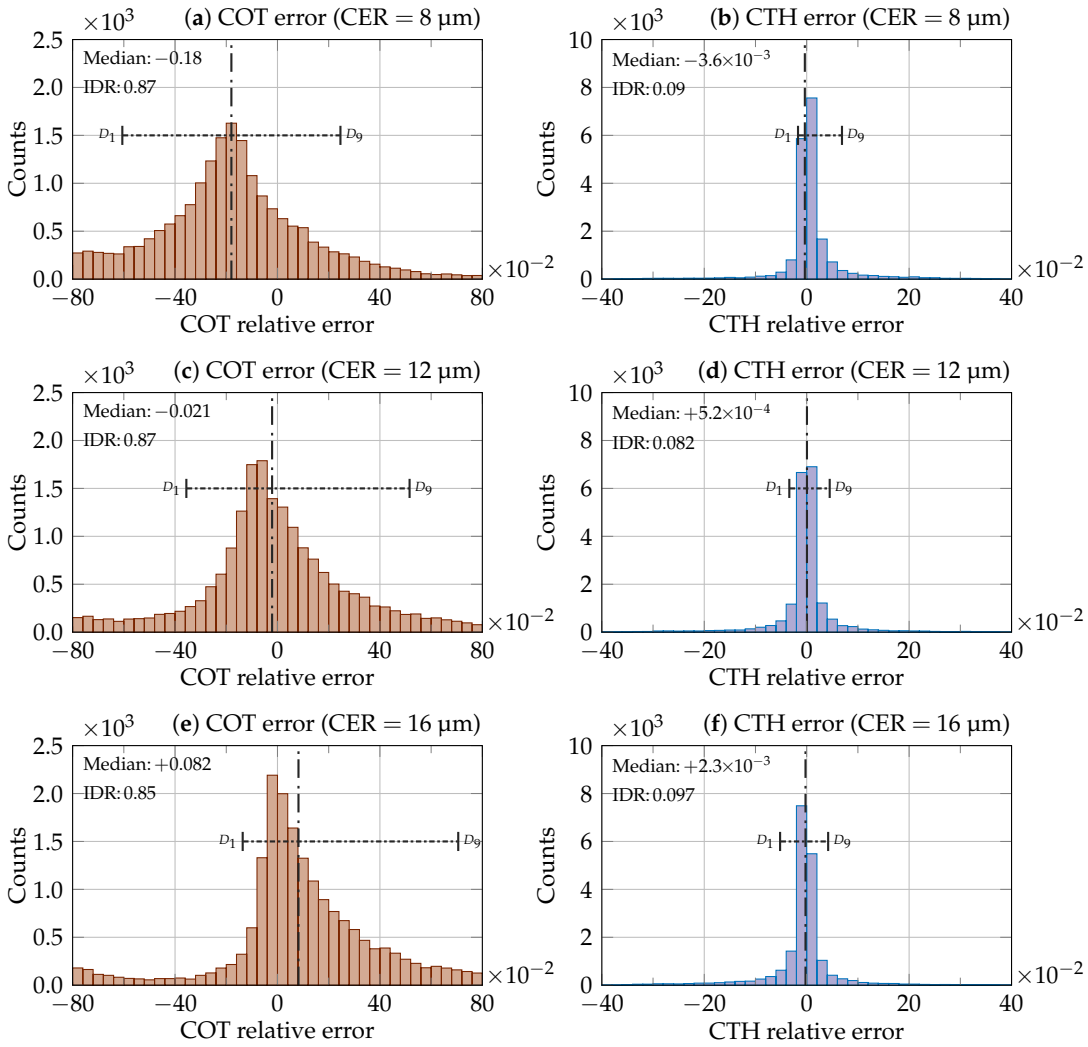


Figure 5.16: Relative error in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample as a function of the fixed CER: (**a**) and (**b**) for CER = 8 µm, (**c**) and (**d**) for CER = 12 µm, and (**e**) and (**f**) for CER = 16 µm.

We simulated the retrieval of cloud macrophysical properties (COT and CTH) over the synthetic validation data sample under the assumption of measurements without noise, perfect knowledge of CGT and different fixed values of CER (differing from the correct values). Figure 5.15 shows the absolute retrieval errors in COT and CTH on the synthetic validation dataset if the CER is fixed to 8, 12 and 16 µm, and Figure 5.16 shows the associated retrieval relative errors. With these configurations, the CTH retrieval is not affected significantly, with absolute errors in CTH whose median is close to 0 km and whose IDR interval is contained within $[-0.42, +0.53]$ km (Figure 5.15**b**, Figure 5.15**d** and Figure 5.15**f**), i.e. within $[-5.4\%, +7.1\%]$ in relative error (Figure 5.16**b**, Figure 5.16**d** and Figure 5.16**f**).

Figure 5.15**a**, Figure 5.15**c** and Figure 5.15**e** present the corresponding absolute retrieval errors in COT caused by fixing the CER. When CER is set to 8 µm (Figure 5.15**a**), the median

of the retrieved absolute errors in COT is biased to -9.5, and 80% of the simulated retrievals have COT errors within $[-57, +15]$, or within $[-61\%, +25\%]$ in relative error (Figure 5.16**a**); the rate of converged retrievals was 93.3%. For CER fixed to 12 µm (Figure 5.15**c**), the COT errors are less biased, with a median absolute error of -0.51 and an IDR interval of $[-31, +35]$, or within $[-36\%, +52\%]$ in relative error (Figure 5.16**c**); the rate of converged retrievals decreased to 90.0%. Finally, when CER is fixed to 16 µm (Figure 5.15**e**), the COT errors are again more biased, with a median absolute error of $+3.9$ and an IDR interval of $[-12, +46]$, or within $[-14\%, +71\%]$ in relative error (Figure 5.16**e**); the rate of converged retrievals also decreased to 85.7%. In general, the histograms of absolute errors in COT are also non-symmetric with respect to its median value: they have a negative skew if the fixed CER tends to underestimate the correct values, and a positive skew if the fixed CER tends to overestimate the correct ones.

In Figure 5.17, we observe the dependency of the absolute retrieval error statistics as a function of the fixed CER. For the COT, the absolute median error is an increasing function of the fixed CER (Figure 5.17**a**), with a value close to 0 when CER is about 13 µm; the IDR is a decreasing function of the fixed CER (Figure 5.17**c**), but it does not let us draw conclusions due to the severe skewness of the histograms of absolute and relative COT errors. For the CTH, the absolute median error is a decreasing function of the fixed CER (Figure 5.17**b**), and it approaches to 0 when CER is also about 13 µm; the IDR is a convex function with a spread minimum area for CER between 9 and 13.5 µm.

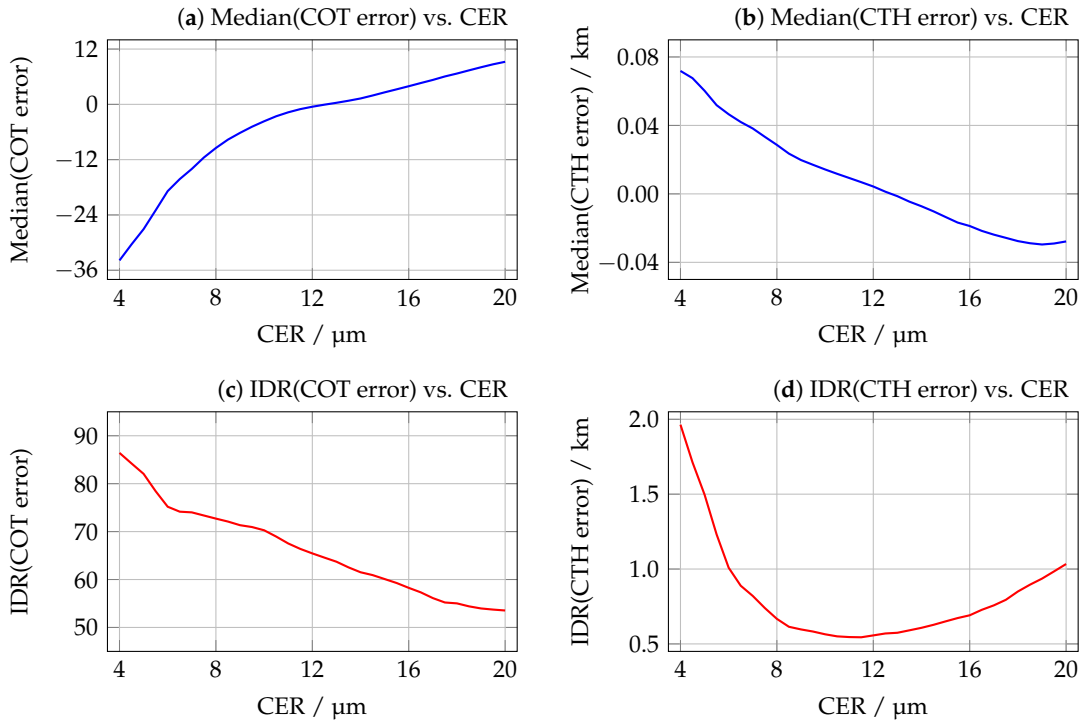


Figure 5.17: Dependency of the absolute retrieval error statistics (median and IDR) as a function of the fixed CER.

5.3.4 Sensitivity to negative offset in surface albedo

Former subsections discussed the sensitivity in the retrieval of cloud macrophysical properties when addressing limitations related to the instrument, e.g. the signal noise or the fixing of model parameters (CGT and CER) due to the lack of enough information content in the EPIC channels. The surface properties (i.e. height and albedo), which are ingested from auxiliary datasets (digital elevation models and climatology maps, respectively), are additional sources of uncertainty in the retrieval process. Current digital elevation models are quite accurate (e.g. ETOPO1 [100]) and the main uncertainties are related to the regridding process to the actual spatial resolution of the target instrument. Standard climatology maps for surface albedo include the black-sky albedo (BSA) database from the MERIS instrument [101], or the recent surface Lambertian-equivalent reflectance (LER) maps for GOME-2 and SCIAMACHY [118]. Since the EPIC instrument channels do not coincide in general with those of the climatology database, rough interpolation with the closest database channels needs to be applied.

We analysed the impact of a surface albedo bias equal to -0.05 for all the oxygen-band channels on the retrieval of cloud macrophysical properties (COT and CTH) over the synthetic validation data sample, assuming that no other sources of uncertainty are present (i.e. no noise, perfect knowledge of CGT and CER). Figure 5.18**a** and Figure 5.18**c** show the histograms of absolute and relative retrieval errors in COT, respectively, whereas Figure 5.18**b**

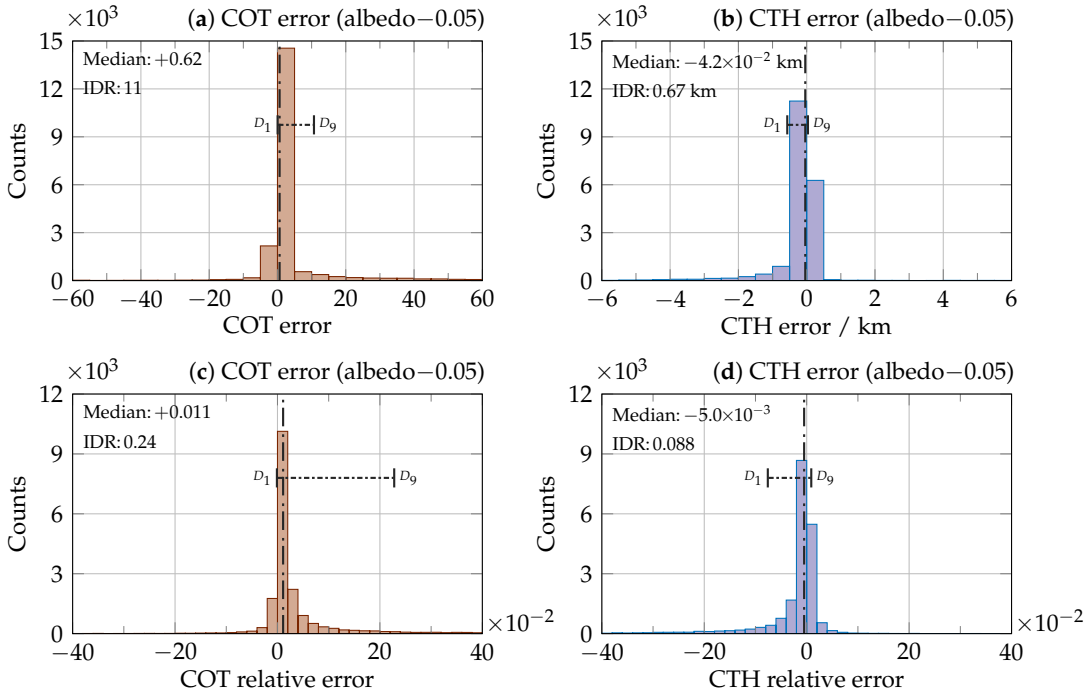


Figure 5.18: Absolute and relative errors in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample after including an offset of -0.05 to the surface albedos.

and Figure 5.18**d** present the histograms of absolute and relative retrieval errors for the CTH, respectively. For the COT, the underestimation in the surface albedo causes a positive skew in the distribution of absolute errors, with a median value of $+0.62$ and an IDR interval of $[-0.27, +11]$; similarly, the distribution of relative errors has a median value of $+1.1\%$ and an IDR interval of $[-0.32\%, +23\%]$. The long right tail in Figure 5.18**c** is caused by the synthetic scenes in which the COT is very small, and for which the negative bias in the surface albedo is compensated with a much higher optical depth for the cloud. For the CTH, the negative bias in the surface albedo is propagated as a slight negative skew in the distribution of absolute errors, with a median value of -0.042 km and an IDR interval of $[-0.60, +0.068]$ km, and the distribution of relative errors has a median value of -0.5% and an IDR interval of $[-7.8\%, +1.1\%]$. Overall, the impact of the surface albedo bias is higher in COT that in CTH, and the rate of converged retrievals (96.3%) was not affected significantly.

5.3.5 Sensitivity to negative offset in cloud fraction

All former sensitivity analyses were performed under the assumption of fully-cloudy scenes. However, real instrument measurements may also correspond to partially-cloudy scenes, for which the cloud fraction (CF) is not exactly 1, but lies within the interval $0 < \text{CF} < 1$. The Optical Cloud Recognition Algorithm (OCRA), whose implementation and application for EPIC was presented in Chapter 4, provides a radiometric cloud fraction based on image analysis, which may be used to feed the scene cloud fraction into ROCINN. In Section 4.5, we showed that the EPIC radiometric cloud fraction has a median difference of about -0.04 with respect to the MODIS geometric cloud fraction over water bodies and high COT and over land for any COT. Based on this result, it is also of interest to discuss the impact that a negative offset in the cloud fraction would show in the retrieved cloud macrophysical parameters.

The synthetic validation data sample for this analysis is an extension of the validation sample used in previous sensitivity analyses, in which we attach an additional pseudo-random column corresponding to the CF, and in which the Sun-normalised radiances are computed as the CF-weighted linear combination of the clear-sky and cloud-sky neural networks. For our analysis, the CF were generated by means of a van der Corput sequence with base equal to 29 (i.e. coprime of the bases used for the other parameters during the input sample generation) and rescaled to the interval $[0.1, 1.0]$. As in previous sensitivity analyses, no other sources of uncertainty were included (i.e. no noise and perfect knowledge of CGT and CER).

Figure 5.19**a** and Figure 5.19**c** show the histograms of absolute and relative retrieval errors in COT, and Figure 5.19**b** and Figure 5.19**d** show the histograms of absolute and relative retrieval errors in CTH. The first important point to mention is that the rate of converged retrievals decreased substantially to just 62.5%. The distribution of absolute errors in the COT suffers from a significant positive skew, with median COT error of $+22$ and IDR interval of $[-19, +78]$ (Figure 5.19**a**). When analysed in relative values, the distribution of COT errors has

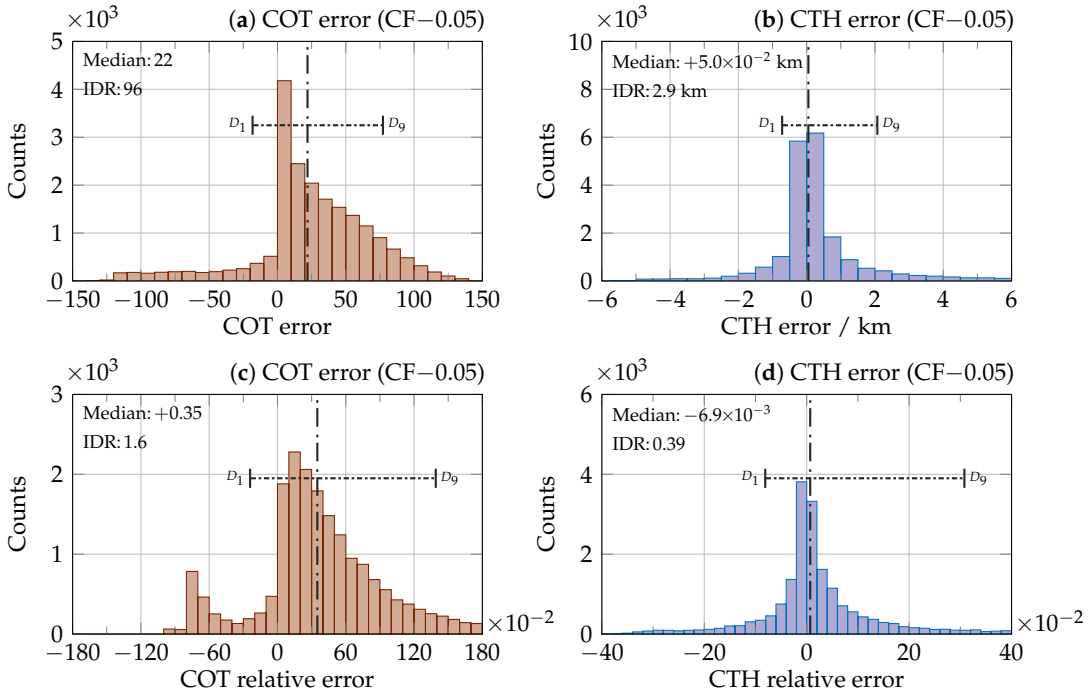


Figure 5.19: Absolute and relative errors in the retrieved cloud macrophysical parameters (COT, CTH) for the partially-cloudy synthetic validation sample after including an offset of -0.05 to the cloud fraction.

a positive median of 35%, and 80% of the results lie within $[-25\%, +140\%]$ (Figure 5.19**c**). We also observe a second mode on the left tail of the histogram, corresponding to non-converged retrievals in which COT barely deviated from the initial guess. Underestimating the CF translates into lower Sun-normalised radiances simulated by the radiative transfer model for the non-absorbing regions of the oxygen bands, which the retrieval process compensates by increasing the COT of the cloud layer; this compensation will be higher when the CF is lower. For the CTH, the negative bias in the CF does not have a significant impact on the median error (absolute or relative), but the IDR interval spreads to $[-0.75, 2.1]$ km in absolute values (Figure 5.19**b**) and to $[-8.3\%, +31\%]$ in relative values (Figure 5.19**d**). The right skew spread is also consequence of the non-converged retrievals that were mentioned for Figure 5.19**c**, and for which CTH barely moved from its initial guess of 10 km.

5.3.6 Sensitivity to combined sources of uncertainty

To conclude the sensitivity analysis, we include one additional study with several of the former sources of uncertainty combined together. We performed the retrieval of cloud macrophysical properties (COT and CTH) over the synthetic validation data sample (i.e. all fully-cloudy scenes) under the assumption of noisy measurements with a SNR equal to 290, a fixed CGT value of 2 km and a fixed CER value of 13 µm.

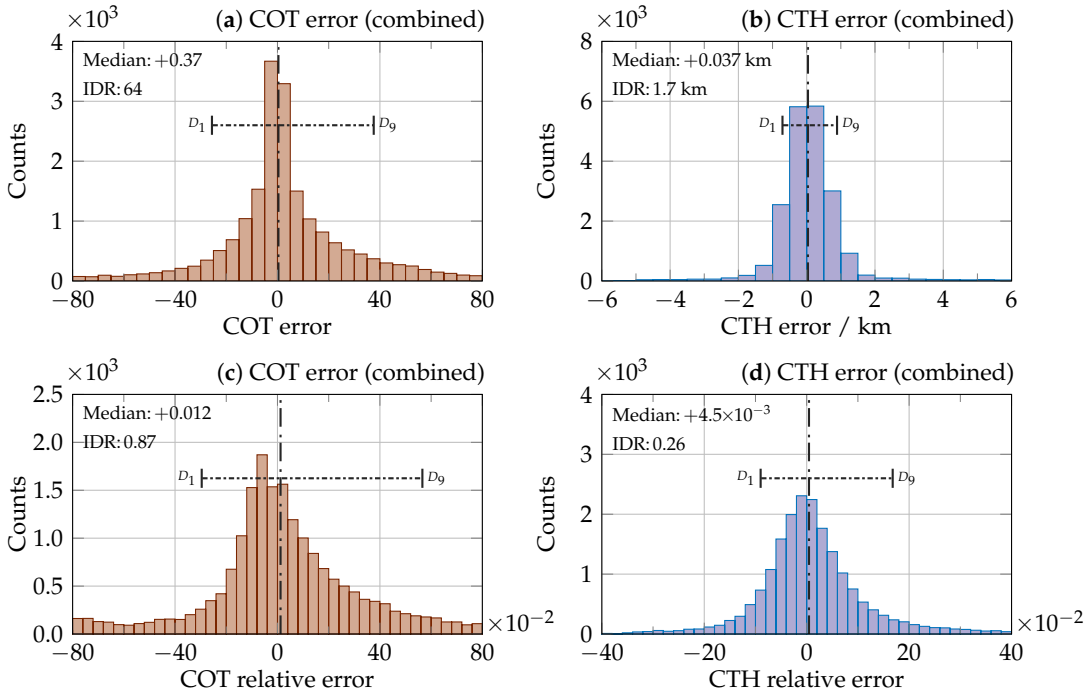


Figure 5.20: Absolute and relative errors in the retrieved cloud macrophysical parameters (COT, CTH) for the synthetic validation sample when at the same time the SNR is set to 290, the CGT is fixed to 2 km and the CER is fixed to 13 µm.

In Figure 5.20**a** and Figure 5.20**b**, the absolute retrieval errors for COT and CTH under these combined sources of uncertainty are shown. In Figure 5.20**c** and Figure 5.20**d**, we display their corresponding relative errors. The rate of converged retrievals was 88.6%.

Regarding the COT, the median absolute error was equal to $+0.37$ (almost unbiased), and 80% of the simulated retrievals found COT with an absolute error within $[-26, +38]$, which translates into an interval of $[-30\%, +57\%]$ in relative terms. Regarding the CTH, the median absolute error was equal $+0.037$ km (almost unbiased too), and the IDR interval for the CTH absolute error was $[-0.74, +0.92]$ km, or $[-9.2\%, +17\%]$ when analysed in relative errors.

To sum up, all the sensitivity analysis point into the same conclusion: the CTH is a much more reliable retrieval parameter than the COT, based on the sources of uncertainty that have been described. The measurement noise plays no role in the retrieval error bias for both CTH and COT: the distribution of retrieval errors spreads if the SNR is decreased, especially for the COT, but this spread is small for the maximum instrument SNR value of 290. Fixing the CGT plays little role in the COT retrieval, whereas it may introduce a bias in the retrieved CTH, as well as a bigger error spread. Fixing the CER does not affect the CTH retrieval, but it strongly affects both COT retrieval bias and error spread. A negative offset in the surface albedos causes an increase in the retrieved COT, but the magnitude is much smaller than the errors due to fixing CER. Finally, a negative offset in the cloud fraction causes strong overestimations in the retrieved COT and decreases the retrieval convergence rate considerably.

5.4 Results

In this section, we provide a comparison of the cloud products as obtained by ROCINN and the corresponding operational EPIC products for one test example. The EPIC Level 1B image corresponds to 2018-10-10 16:39:39 UTC. The associated dataset from the EPIC Level 2 CLOUD Version 1 collection was employed as reference.

To construct the vector of measurements $\boldsymbol{y}$, the image counts per second need to be converted first to reflectances by using the calibration factors corrected from degradation (result from Chapter 2), and from them into Sun-normalised radiances. Because of the geometric constraints for the calibration factors, we restrict our comparison to pixels with solar and viewing zenith angles below 60 deg.

Regarding the ROCINN configuration, we fixed the CER and the CGT to 13 µm and 2 km, respectively, to create the data vector $\boldsymbol{p}_{\mathrm{c}}$. We used the same initial guesses for COT and CTH as in Section 5.3 to define the state vector $\boldsymbol{x}$. The data vector $\boldsymbol{p}_{\mathrm{sg}}$ is constructed with surface altitudes regridded from the ETOPO1 database, surface albedos from the MERIS black-sky albedo database interpolated to the EPIC oxygen band channels, and the geometric variables available in the EPIC Level 1B dataset. The retrieval is restricted to pixels with $\chi_{\mathrm{CF}} \geq 0.1$.

In Figure 5.21**a**, we show the EPIC false-RGB image (channels at 780, 551 and 388 nm) for the test example. The associated OCRA radiometric cloud fraction is shown in Figure 5.21**b** (result from Chapter 4). As expected, global cloud features are properly identified in the radiometric cloud fraction, except for the optically-thin clouds over water.

(**a**) False RGB (**b**) OCRA cloud fraction

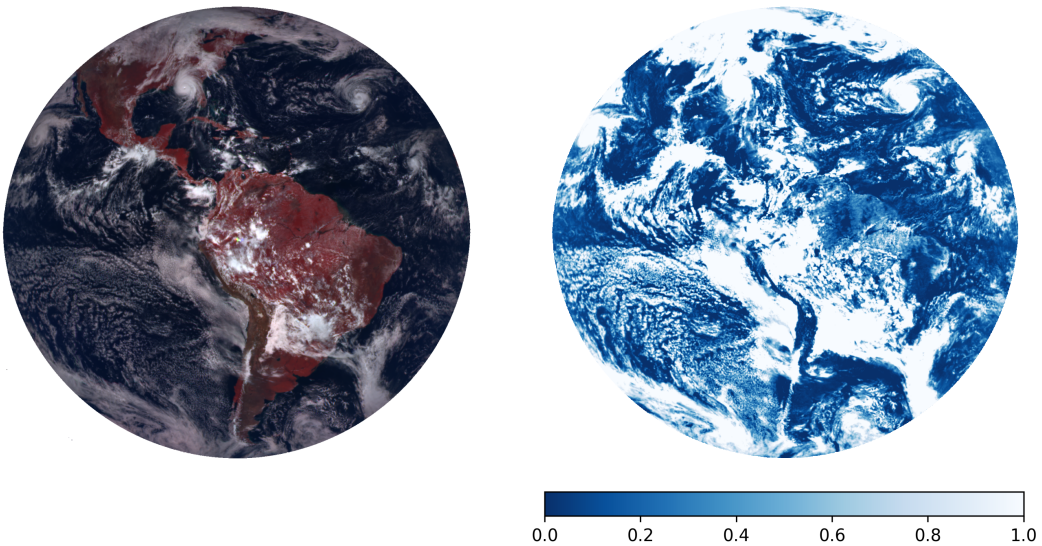
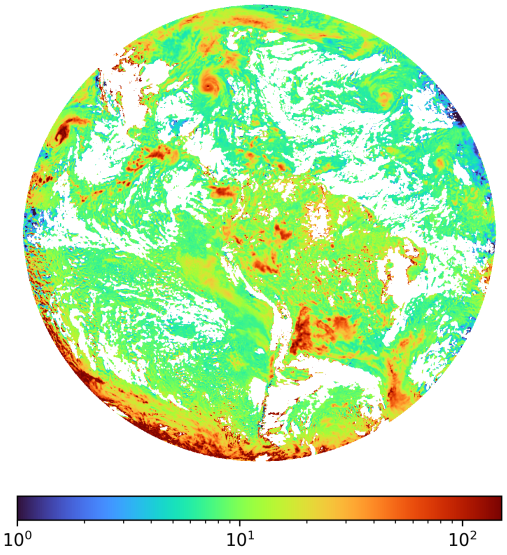
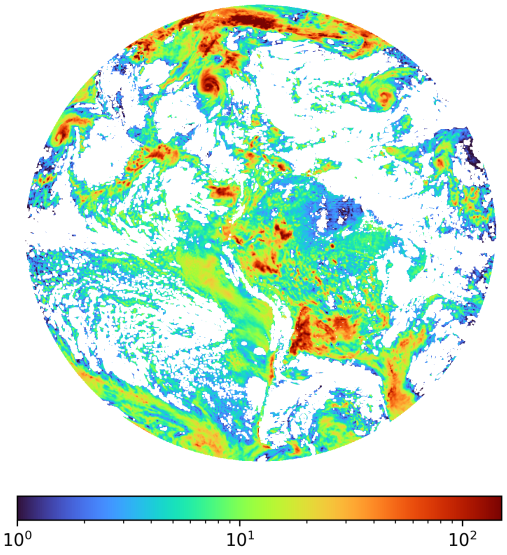


Figure 5.21: (**a**) False-RGB image from the EPIC Level 1B dataset on 2018-10-10 at 16:39:39 UTC using channels at 780, 551 and 388 nm. (**b**) OCRA radiometric cloud fraction for the former dataset.

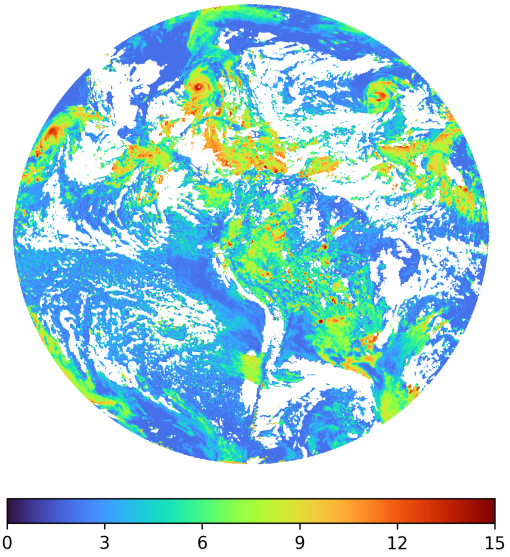
(**a**) ROCINN cloud optical thickness



(**b**) Operational cloud optical thickness



(**c**) ROCINN cloud-top height / km



(**d**) Operational cloud effective height / km

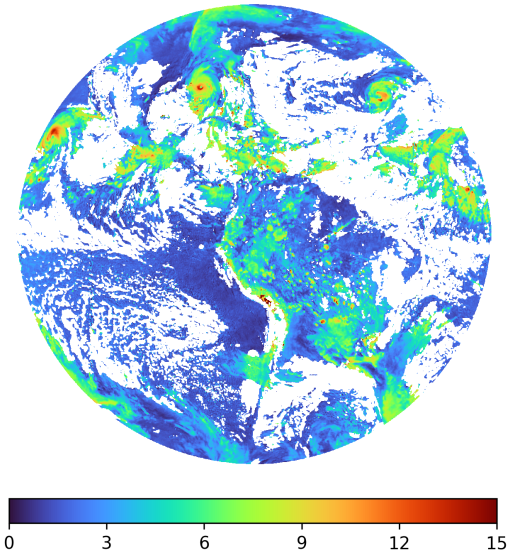


Figure 5.22: Cloud optical thickness (**a**) with ROCINN and (**b**) operational reference, and cloud-top height (**c**) with ROCINN and (**d**) operational reference as retrieved from the oxygen channels of the EPIC test dataset.

Figure 5.22**a** and Figure 5.22**c** show the COT and CTH values as retrieved by ROCINN for the EPIC test dataset. The operational COT and cloud effective height (CH) are shown next to them in Figure 5.22**b** and Figure 5.22**d**.

With respect to the retrieved optical thickness, the agreement is mixed. The thick clouds (COT $>$ 20) are identified as such by both products, although the absolute discrepancy may be considerable. The operational product also shows thin clouds (COT $<$ 5) over ocean (see

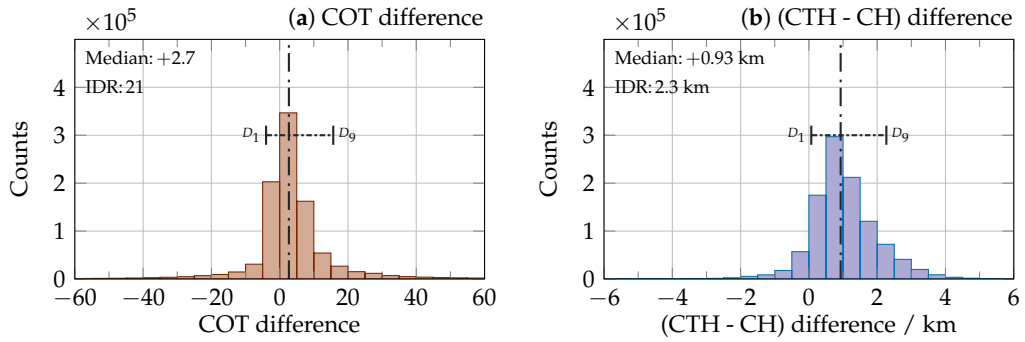


Figure 5.23: Absolute difference between the retrieved cloud macrophysical parameters (COT, CTH) with ROCINN and the operational EPIC cloud product (COT, CH) for the EPIC test dataset.

e.g. over the South Pacific), but in these areas the retrieved COT is higher for ROCINN. We must note that the operational product works using a cloud mask, while ROCINN works using a cloud fraction, and in Section 5.3.5 we saw that negative offsets in the cloud fraction are translated into an increase in the retrieved COT. This phenomenon can also be observed in Amazonia in Figure 5.22**a**: pixels close to the cloud fraction threshold of 0.1 suffer from a considerable increase in the COT. The ROCINN retrievals were also more unstable close to the zenith angle edge of 60 deg.

With respect to the retrieved cloud-top height, the agreement between both products is in general high. There is a slight tendency of bigger height values by ROCINN for the higher clouds (visible e.g. in the hurricane over Florida). For lower clouds over ocean, the operational CH goes also below 2 km, which was not possible for our ROCINN configuration because of the fixed value for the CGT.

Finally, in Figure 5.22 we present the histogram of differences in COT and CTH between the cloud macrophysical parameters retrieved with ROCINN and the EPIC operational cloud products for the EPIC test dataset. The COT difference has a median value of +2.7, and for 80% of the pixels the COT differs within $[-4.3, +16]$. The histogram has a small second local maximum in the bin at $[145, 150]$, which corresponds to the unstable ROCINN retrievals at the zenith angle edge of 60 deg. The difference between the ROCINN CTH and the operational CH has a median value of $+0.93$ km, and for 80% of the pixels this difference is found within $[+0.039, +2.3]$ km. In Section 5.3.2, we showed that the fixed CGT value may cause an overall bias in the retrieved CTH values. The operational EPIC cloud effective height will be therefore equivalent to the ROCINN forward model with a lower fixed CGT.

References

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

[91] Å. Björck, Numerical Methods for Least Squares Problems, Siam Philadelphia, 1996.

[92] J. J. Moré, The Levenberg-Marquardt algorithm: Implementation and theory, in: G. A. Watson (Ed.), Numerical Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 1978, pp. 105–116.

[93] A. N. Tikhonov, V. I. Arsenin, Solutions of ill-posed problems, Scripta series in mathematics, Winston & Sons, Washington, D.C., 1977.

[94] H. B. Curry, The method of steepest descent for non-linear minimization problems, Quarterly of Applied Mathematics 2 (3) (1944) 258–261. doi:10.1090/qam/10667.

[95] V. Molina García, D. S. Efremenko, A. del Águila, Automatic differentiation for Jacobian computations in radiative transfer problems, Oral talk presented at 21st European Workshop on Automatic Differentiation, Jena (Germany), Friedrich Schiller University Jena (2018).

[96] 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 Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy, Nature 585 (7825) (2020) 357–362. doi:10.1038/s41586-020-2649-2.

[97] D. G. Loyola R., M. Pedergnana, S. Gimeno García, Smart sampling and incremental function learning for very large high dimensional data, Neural Networks 78 (2016) 75–87. doi:10.1016/j.neunet.2015.09.001.

[98] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, Numerische Mathematik 2 (1) (1960) 84–90. doi:10.1007/bf01386213.

[99] J. G. van der Corput, Verteilungsfunktionen, in: Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Vol. 38, 1935, pp. 813–821.

[100] C. Amante, B. W. Eakins, ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA technical memorandum NESDIS NGDC-24, Tech. rep. (2009). doi:10.7289/V5C8276M.
URL https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/docs/ETOPO1.pdf

[101] C. Popp, P. Wang, D. Brunner, P. Stammes, Y. Zhou, M. Grzegorski, MERIS albedo climatology for FRESCO+O_2 A-band cloud retrieval, Atmospheric Measurement Techniques 4 (3) (2011) 463–483. doi:10.5194/amt-4-463-2011.

[102] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing 415 (2020) 295–316. doi:https://doi.org/10.1016/j.neucom.2020.07.061.

[103] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM Journal of Scientific Computing 16 (1995) 1190–1208. doi:10.1137/0916069.

[104] C. Zhu, R. H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: FORTRAN subroutines for large-scale bound-constrained optimization, ACM Transactions on Mathematical Software 23 (4) (1997) 550–560. doi:10.1145/279232.279236.

[105] 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, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (85) (2011) 2825–2830. doi:10.48550/arXiv.1201.0490.

[106] J. Herman, L. Huang, R. McPeters, J. Ziemke, A. Cede, K. Blank, Synoptic ozone, cloud reflectivity, and erythemal irradiance from sunrise to sunset for the whole earth as viewed by the DSCOVR spacecraft from the earth–sun lagrange 1 orbit, Atmospheric Measurement Techniques 11 (1) (2018) 177–194. doi:10.5194/amt-11-177-2018.

[107] F. P. J. Valero, J. Herman, P. Minnis, W. D. Collins, R. Sadourny, W. Wiscombe, D. Lubin, K. Ogilvie, Triana - a Deep Space Earth and Solar Observatory, Tech. rep. (1999).
URL https://www-pm.larc.nasa.gov/site/doc-library/references/DSCOVR/NAS.Triana.report.12.99.pdf

[108] A. B. Davis, G. Merlin, C. Cornet, L. C. Labonnote, J. Riédi, N. Ferlay, P. Dubuisson, Q. Min, Y. Yang, A. Marshak, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: An optimal estimation approach, Journal of Quantitative Spectroscopy and Radiative Transfer 216 (2018) 6–16. doi:10.1016/j.jqsrt.2018.05.007.

[109] A. B. Davis, N. Ferlay, Q. Libois, A. Marshak, Y. Yang, Q. Min, Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: A physics-based approach, Journal of Quantitative Spectroscopy and Radiative Transfer 220 (2018) 84–96. doi:10.1016/j.jqsrt.2018.09.006.

[110] J. Hansen, L. Travis, Light scattering in planetary atmospheres, Space Science Reviews 16 (1974) 527–610. doi:10.1007/BF00168069.

[111] S. J. Evans, J. D. Haigh, The retrieval of total optical depth and effective droplet radius of clouds from solar reflection measurements using the Along Track Scanning Radiometer-2 (ATSR-2), Geophysical Research Letters 22 (6) (1995) 695–698. doi:https://doi.org/10.1029/95GL00361.

[112] M. D. King, S.-C. Tsay, S. E. Platnick, M. Wang, K.-N. Liou, Cloud retrieval algorithms for MODIS: Optical thickness, effective particle radius, and thermodynamic phase, Tech. rep. (1997).
URL https://modis.gsfc.nasa.gov/data/atbd/atbd_mod05.pdf

[113] Y. Oku, M. Kajino, H. Ishikawa, Estimation of the cloud effective particle radius using MTSAT-1R data, International Journal of Remote Sensing 31 (20) (2010) 5439–5447. doi:10.1080/01431160903369634.

[114] S. Platnick, S. Twomey, Determining the susceptibility of cloud albedo to changes in droplet concentration with the Advanced Very High Resolution Radiometer, Journal of Applied Meteorology and Climatology 33 (3) (1994) 334–347. doi:10.1175/1520-0450(1994)033<0334:DTSOCA>2.0.CO;2.

[115] S. Platnick, F. P. J. Valero, A validation of a satellite cloud retrieval during ASTEX, Journal of Atmospheric Sciences 52 (16) (1995) 2985–3001. doi:10.1175/1520-0469(1995)052<2985:AVOASC>2.0.CO;2.

[116] S. Platnick, J. Y. Li, M. D. King, H. Gerber, P. V. Hobbs, A solar reflectance method for retrieving the optical thickness and droplet size of liquid water clouds over snow and ice surfaces, Journal of Geophysical Research: Atmospheres 106 (D14) (2001) 15185–15199. doi:https://doi.org/10.1029/2000JD900441.

[117] Y. Yang, K. Meyer, G. Wind, Y. Zhou, A. Marshak, S. Platnick, Q. Min, A. B. Davis, J. Joiner, A. Vasilkov, D. Duda, W. Su, Cloud products from the Earth Polychromatic Imaging Camera (EPIC): Algorithms and initial evaluation, Atmospheric Measurement Techniques 12 (3) (2019) 2019–2031. doi:10.5194/amt-12-2019-2019.

[118] Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, Journal of Geophysical Research: Atmospheres 122 (7) (2017) 4084–4111. doi:https://doi.org/10.1002/2016JD025940.

Chapter 6
Conclusions

The goal of this dissertation was to create a framework for the retrieval of cloud macrophysical properties from measurements of the EPIC/DSCOVR instrument. The implementation of this framework, however, required to expand the scope of the thesis to ensure its success.

First of all, the retrieval required a reliable source of EPIC radiance measurements. At the time in which this dissertation was started, the latest EPIC Level 1B Version 2 collection was unfortunately not reliable enough due to severe misregistration issues that would translate into wrong assignments of surface properties for each pixel scene, and the errors in the surface properties would affect any retrieval of physical parameters in which the surface properties were assumed as *a priori* (and correct) data. Therefore, a solution for this problem had to be developed as shown in Chapter 2. A representative number of EPIC Level 1B datasets was inspected and the effects of common camera issues (wrong camera orientation, problems related to the optical elements of the instrument) were investigated. We found out that a combination of motion (shift plus rotation) and radial distortion correction could fix the majority of the EPIC Level 1B datasets with a registration error of about 1 ground pixel instead of the initial registration error of about 5 ground pixels. The EPIC Level 1B Version 3 collection, whose processing started in late 2019, improves the registration of the EPIC imagery significantly by means of coastline comparison with the MODIS and GOES-17 sensors [119], and the registration issue was reported as the result of inaccuracies in the measurements provided by the star tracker used to control the instrument orientation as well as in the characterisation of the telescope lens (stated by K. Blank at DSCOVR Science Team Meeting 2019). This proves that the registration function described in this work, which was published as a peer-reviewed article [120], was indeed addressing the underlying problems of EPIC.

The second problem to address in this dissertation was the time efficiency of the radiative transfer models. Any retrieval algorithm that uses standard non-linear optimisers will require several calls to the model function. If the model function is time-consuming, then the retrieval computation time will be at least as high as the number of calls to the model function. Since the number of EPIC measurements is found in the order of 10^6 per dataset, and each dataset is generated every one or two hours, it is not possible to use standard radiative transfer models as model functions for operational purposes. In this regard, ROCINN replaces the radiative transfer model with the combination of two neural networks, whose computation time is very small, but the neural networks need to be trained in advance with training and validation data previously obtained by calling the time-consuming radiative transfer model. This means that the time efficiency problem would persist during the training and validation data generation.

These concerns are discussed in Chapter 3 with two peer-reviewed publications [121, 122] and one co-authored review [123]. In this chapter, the possibility to use approximate models to describe the clouds within the radiative transfer model was analysed (e.g. asymptotic models, clouds as Lambertian surfaces), but it was discarded because of the errors propagated from the model simplifications. Alternatively, a combination of two acceleration techniques (correlated k-distribution and PCA) was found to decrease the computation time by two orders of magnitude with respect to the equivalent LBL calculations. Thanks to this speed-up, the creation of training and validation tables for ROCINN would take, instead of weeks, only about 1 day for the clear-sky scenes and about 2.5 days for the cloudy-scenes covering 200000 scenarios and running the radiative transfer simulations on a high-end 44-core server.

The underlying non-linear optimisation within the retrieval of cloud properties needs not only the evaluation of the physical model, but also the derivatives (Jacobian) of the model with respect to the retrieved variables, in order to compute the state vector update at each iteration of the optimisation process. In this regard, Chapter 3 showed that the combination of the former acceleration techniques with the forward-adjoint approach decreases the computation time by three orders of magnitude with respect to the equivalent linearised LBL calculations. This improvement could be exploited to also include the derivatives of interest as outputs of the neural networks trained for ROCINN. This idea was finally discarded for two reasons. First, increasing the number of outputs in the trained neural networks would result into an increase of complexity for the output space, and this would imply that the topology of the neural networks would need to become more complex (e.g. by increasing the number of hidden layers or the neurons per hidden layer). Second, the optimisation algorithms are implemented under the assumption that for a given model function $\boldsymbol{f}$, the state vector update depends on the Jacobian $J_{\boldsymbol{f}}$; for consistency reasons, it was decided to simply use the derivatives of the neural networks with respect to the state vector components since, from the point of view of ROCINN, the neural networks *are* the model function. In the case of the ROCINN feed-forward neural networks, the internal operations reduce to matrix additions, matrix multiplications and evaluations of the hyperbolic tangent, and all these operations can be differentiated easily. Instead of implementing the derivatives only for this specific case, the prototype Python library `dnumpy` was developed [124]. The library is built as a superset of `numpy`, i.e. the public interface of `numpy` is preserved, but `dnumpy` allows to define variables within the standard arrays, and the `numpy` functions overloaded by `dnumpy` keep track of the presence of these variables and apply appropriate rules for the propagation of the derivatives with respect to the found variables (i.e. automatic differentiation in forward mode). As a use case, `dnumpy` provides the derivatives of the ROCINN implementation of this thesis, but the benefit is that any scientific work written in Python with `numpy` as core library can also use `dnumpy` to compute derivatives, as demonstrated for example with the radiative transfer library `pydome` [125].

In Chapter 4, the OCRA algorithm for the retrieval of cloud fraction was implemented for EPIC and applied to the 4 years of data available in the EPIC Level 1B Version 2 collection. The main power of OCRA resides in having a simple retrieval model with fast but at the same time accurate results. From the comparisons of the EPIC radiometric cloud fraction product with the daily geometric cloud fraction product from MODIS, it was seen that EPIC-OCRA has good agreement over land, and also over water as long as the clouds are optically thick. When optically thin clouds are found over water, the performance of EPIC-OCRA is limited, with a tendency to estimate cloud fractions lower than MODIS. All in all, the application of OCRA as described in Chapter 4 allows to obtain a cloud fraction product for EPIC that enhances the operational cloud mask product, which is limited to only four possible values (clear with high confidence, clear with low confidence, cloudy with low confidence, and cloudy with high confidence) [126], and it accomplishes it without using any of the EPIC channels located at the oxygen bands (i.e. the channels usually employed to retrieve cloud properties).

In Chapter 5, the ROCINN algorithm was implemented for EPIC. First, the importance of choosing an appropriate input space for the neural networks was outlined. From the systematic hyperparameter analysis, the main conclusion is that there is *not* a unique optimal neural network topology. The only requirements to train neural networks for ROCINN are to ensure that the neural network topologies are not too simple (to avoid underfitting), and that the neural networks do not overfit the training data after the training process. It is still important to keep in mind that, for a feed-forward neural network with one hidden layer, the total number of neural network parameters (i.e. biases and weights) increases with the number of neurons. Therefore, the number of neurons should not be increased unless the accuracy enhancement is worthwhile. Once the final configurations for the ROCINN neural networks are chosen, they act as extremely fast emulators of the radiative transfer model and provide in addition the derivatives of first order with respect to the input parameters.

After all these considerations, the final ROCINN neural networks for clear-sky and cloudy-sky scenes were built with 64 and 128 neurons, respectively, and they provide high accuracies, with less than 0.3% error for 80% of the validation dataset. These neural networks were used in order to perform a sensitivity analysis on the retrieval of cloud optical thickness and cloud-top height for liquid-water clouds, based on several sources of uncertainty. From this analysis, it was shown that the EPIC cloud-top height retrieval remains stable for most of the sources of uncertainty, except for the case in which the cloud geometrical thickness is fixed to a constant value, where a negative bias (resp. positive bias) is found if the fixed cloud geometrical thickness underestimates (resp. overestimates) its real value. Combining the main sources of uncertainty, the cloud-top height is retrieved within ± 1 km for 80% of the cases. However, the cloud optical thickness is significantly affected by several sources of uncertainty, such as fixing the cloud effective particle size or the existence of a negative offset in the surface albedo or the cloud fraction, and the uncertainty spread is too wide to guarantee reliable results.

Finally, the OCRA and ROCINN algorithms were used together to retrieve cloud fraction, cloud optical thickness and cloud-top height on one EPIC Level 1B test dataset, and the retrieved results were compared with the corresponding operational cloud products. The cloud optical thickness shows a mixed agreement, while the cloud-top height agrees well in general, and the existence of a positive bias in the differences between the ROCINN cloud-top height and the operational cloud effective height are explained from the selection of the fixed cloud geometrical thickness.

Future lines of work could further improve the retrieval of cloud properties from EPIC measurements. The current radiative transfer theory only considers the modelling of liquid-water clouds. However, a significant amount of clouds are found in ice phase. Therefore, the EPIC radiative transfer model presented in this thesis could be extended so that both types of cloud phases can be modelled. The migration to the EPIC Level 1B Version 3 collection, which aims to provide better image registration, could also be addressed, and for which the OCRA support data (clear-sky maps, and the scaling and offset parameters) would need to be recomputed. The EPIC OCRA/ROCINN results could also be further compared with the complete operational EPIC cloud product collection from NASA, as well as with other missions, like MODIS, TROPOMI and CALIPSO. In addition, improvements to the ROCINN configurations could be further investigated. One improvement is related to the setup of the initial guesses for the state vector, which in this dissertation were always set to the same pair of fixed values. The ROCINN implementation for TROPOMI, for example, uses the solutions obtained under the assumption of clouds as Lambertian surfaces (CRB) as initial guesses for a second optimisation process with clouds as homogeneous layers (CAL). And since the fixing of the cloud geometrical thickness translates into biases in the retrieved cloud-top height, the possibility to parameterise the cloud geometrical thickness as a function of the cloud-top height at each iteration of the retrieval process could be investigated.

References

[119] A. Kostinski, A. Marshak, T. Várnai, Deep space observations of terrestrial glitter, Earth and Space Science 8 (2) (2021). doi:https://doi.org/10.1029/2020EA001521.

[120] V. Molina García, S. Sasi, D. S. Efremenko, D. Loyola, Improvement of EPIC/DSCOVR image registration by means of automatic coastline detection, Remote Sensing 11 (15) (2019) 1747. doi:10.3390/rs11151747.

[121] V. Molina García, S. Sasi, D. Efremenko, A. Doicu, D. Loyola, Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, Journal of Quantitative Spectroscopy and Radiative Transfer 123 (2018) 228–240. doi:10.1016/j.jqsrt.2018.03.014.

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

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

[124] V. Molina García, D. S. Efremenko, A. del Águila, Automatic differentiation for Jacobian computations in radiative transfer problems, Oral talk presented at 21st European Workshop on Automatic Differentiation, Jena (Germany), Friedrich Schiller University Jena (2018).

[125] D. S. Efremenko, V. Molina García, B. Pflug, T. Trautmann, PYDOME: Python library for radiative transfer computations, Poster presented at ESA Living Planet Symposim 2019, Milan (Italy), European Space Agency (2019).

[126] Y. Yang, K. Meyer, G. Wind, Y. Zhou, A. Marshak, S. Platnick, Q. Min, A. B. Davis, J. Joiner, A. Vasilkov, D. Duda, W. Su, Cloud products from the Earth Polychromatic Imaging Camera (EPIC): Algorithms and initial evaluation, Atmospheric Measurement Techniques 12 (3) (2019) 2019–2031. doi:10.5194/amt-12-2019-2019.

Appendix A

List of abbreviations

Abbreviation	Description
BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm
BRDF	Bidirectional Reflectance Distribution Function
BRIEF	Binary Robust Independent Elementary Features
BSA	Black-Sky Albedo
CAL	Clouds As Layers
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CCD	Charged-Coupled Device
CER	Cloud Effective particle Radius
CGT	Cloud Geometrical Thickness
CH	Cloud effective Height
COT	Cloud Optical Thickness
cps	counts per second
CPU	Central Processing Unit
CRB	Clouds as Reflective Boundaries
CTH	Cloud-Top Height
DISORT	DIScrete Ordinate Radiative Transfer
DOAS	Differential Optical Absorption Spectroscopy
DOME	Discrete Ordinates with Matrix Exponential
DSCOVR	Deep Space Climate Observatory
ENVISAT	ENVIronmental SATellite
EPIC	Earth Polychromatic Imaging Camera
ERS	European Remote sensing Satellite
ETOPO1	Earth TOPOgraphy 1 arc minute
FAST	Features from Accelerated Segment Test
FDOME	Forward-Adjoint DOME
GOES	Geostationary Operational Environmental Satellite
GOME	Global Ozone Monitoring Experiment
IDR	InterDecile Range
ISRF	Instrument Spectral Response Function
L-BFGS-B	Limited-memory BFGS with Bounds
LBL	Line-By-Line
LDOME	Linearised DOME
LER	Lambertian-Equivalent Reflectance
LIDORT	LInearized Discrete Ordinate Radiative Transfer
LMOME	Linearised MOME
MAE	Mean Absolute Error
MERIS	MEdium Resolution Imaging Spectrometer

Abbreviation	**Description**
MetOp	Meteorological Operational satellite
MODIS	MODerate-resolution Imaging Spectroradiometer
MOME	Matrix Operator with Matrix Exponential
NASA	National Aeronautics and Space Administration
NGU	Near infrared-Green-Ultraviolet
NPP	National Polar-orbiting Partnership
OCRA	Optical Cloud Recognition Algorithm
OMI	Ozone Monitoring Instrument
OMPS	Ozone Mapping and Profiler Suite
OpenCV	Open Computer Vision
ORB	Oriented FAST and Rotated BRIEF
PCA	Principal Component Analysis
POSIX	Portable Operating System Interface
RGB	Red-Green-Blue
ROCINN	Retrieval Of Cloud Information using Neural Networks
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY
SNR	Signal-to-Noise Ratio
SWIR	ShortWave InfraRed
TOA	Top Of Atmosphere
TROPOMI	TROPOspheric Monitoring Instrument
UTC	Universal Time Coordinated
UV	UltraViolet
VIS	VISible
VNIR	Visible Near-InfraRed

Appendix B

List of publications

This dissertation is presented in cumulative form based on three peer-reviewed scientific articles as main author and one peer-reviewed scientific review as co-author:

1. V. Molina García, S. Sasi, D. S. Efremenko, D. Loyola, Improvement of EPIC/DSCOVR image registration by means of automatic coastline detection, *Remote Sensing* 11 (15) (2019) 1747.
   doi: 10.3390/rs11151747

2. D. S. Efremenko, V. Molina García, S. Gimeno García, A. Doicu, A review of the matrix-exponential formalism in radiative transfer, *Journal of Quantitative Spectroscopy and Radiative Transfer* 196 (2017) 17–45.
   doi: 10.1016/j.jqsrt.2017.02.015

3. V. Molina García, S. Sasi, D. Efremenko, A. Doicu, D. Loyola, Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, *Journal of Quantitative Spectroscopy and Radiative Transfer* 123 (2018) 228–240.
   doi: 10.1016/j.jqsrt.2018.03.014

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

Copies of the manuscripts are provided in the subsequent pages.

remote sensing

MDPI

Article

Improvement of EPIC/DSCOVR Image Registration by Means of Automatic Coastline Detection

Víctor Molina García *, Sruthy Sasi, Dmitry S. Efremenko and Diego Loyola

Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany
* Correspondence: Victor.MolinaGarcia@dlr.de

Received: 27 May 2019; Accepted: 23 July 2019; Published: 25 July 2019

Abstract: In this work, we address the image geolocation issue that is present in the imagery of EPIC/DSCOVR (Earth Polychromatic Imaging Camera/Deep Space Climate Observatory) Level 1B version 2. To solve it, we develop an algorithm that automatically computes a registration correction consisting of a motion (translation plus rotation) and a radial distortion. The correction parameters are retrieved for every image by means of a regularised non-linear optimisation process, in which the spatial distances between the theoretical and actual locations of chosen features are minimised. The actual features are found along the coastlines automatically by using computer vision techniques. The retrieved correction parameters show a behaviour that is related to the period of DSCOVR orbiting around the Lagrangian point L_1. With this procedure, the EPIC coastlines are collocated with an accuracy of about 1.5 pixels, thus significantly improving the original registration of about 5 pixels from the imagery of EPIC/DSCOVR Level 1B version 2.

Keywords: EPIC; registration; geolocation; computer vision; regularisation

1. Introduction

The spacecraft DSCOVR (Deep Space Climate Observatory) provides a unique view of the Earth from its Lissajous orbit around the Earth–Sun Lagrangian point L_1, at a distance of 1.5 million kilometres from the Earth. DSCOVR carries a range of sensors on board, including EPIC (Earth Polychromatic Imaging Camera) for climate science applications [1].

EPIC consists of a CCD (Charge-Coupled Device) camera that monitors the sunlit half of the Earth in 10 spectral channels ranging in the ultraviolet, visible and near infrared—approximately from 317 to 780 nm. Every two hours, EPIC measures its 10 channel images with a shape of 2048 × 2048 detector pixels, and a varying ground pixel size of approximately 12 × 12 km^2 at the center of the images. These available channels make EPIC a suitable candidate for several applications, such as monitoring the vegetation condition [2], the synoptic ozone [3], the sulphur dioxide content from volcanic eruptions [4,5], the aerosol layer height and optical depth [6,7], and different cloud properties [8]. Extensive work has been performed to estimate the cloud information content in the EPIC oxygen bands [9,10], as well as its sensitivity to liquid-phase cloud microphysical parameters [11].

Standard techniques for retrieval of atmospheric properties, such as the layer height and the optical depth of clouds and aerosols, are non-linear least squares fitting [12] and Tikhonov regularisation [13], which match the results of forward simulations (i.e., synthetic spectra) with the actual measurements (i.e., real spectra). The forward simulations are performed on the base of radiative transfer models. Due to the particular viewing geometry of EPIC, with scattering angles ranging between 168 and 176 degrees, the radiative transfer modelling faces challenges in terms of computational efficiency and accuracy [14,15], which can be solved by using various acceleration techniques [16–18]. Furthermore, this approach needs to assume that the surface properties required as input for the radiative transfer model, e.g., the surface altitude and the spectral surface albedo,

are already known from external sources. The assignment of surface altitude and surface albedos must be performed for every image pixel based on their corresponding latitude and longitude values allocated in the EPIC Level 1B (L1B) arrays of Earth coordinates, which are provided together with the measurement arrays.

Since the evaluation of these surface properties for each image pixel is based on its geolocation information, any incorrect registration of the images can lead into an invalid assignment of these surface properties and, thus, an erroneous retrieval of the atmospheric properties. In the case of EPIC, the geolocation arrays, including the latitude and longitude locations as well as the Sun and instrument viewing angles, are computed by means of a complete navigation algorithm. Essentially, it determines the spacecraft location and orientation and maps the 3D-model coordinates into the 2D image coordinates [19]. This mapping includes an optical correction modelled as a small barrel distortion. Although this navigation algorithm provides a good first estimate for the image geolocation, the imagery from EPIC L1B version 2 presents a misregistration that is especially noticeable when inspecting the expected and actual locations of the land bodies (Figure 1). A similar problem was already mentioned and briefly described for the previous release L1B version 1 [20].

In this article, we develop an automatic registration algorithm for EPIC that relies on the expected location of the Earth land bodies. Assuming that the geolocation error is smaller than 10 detector pixels, we state that it is enough to apply a transformation on the measurement pixel indices, consisting of a motion (translation plus rotation of the Earth's disk) together with a small optical correction, in order to improve the registration of the EPIC imagery. The optimal transformation parameters must be computed for every EPIC L1B dataset with the help of reference pixels taken from the expected and actual locations of the land bodies, which can be determined by using conventional pattern recognition techniques that are of common use in computer vision.

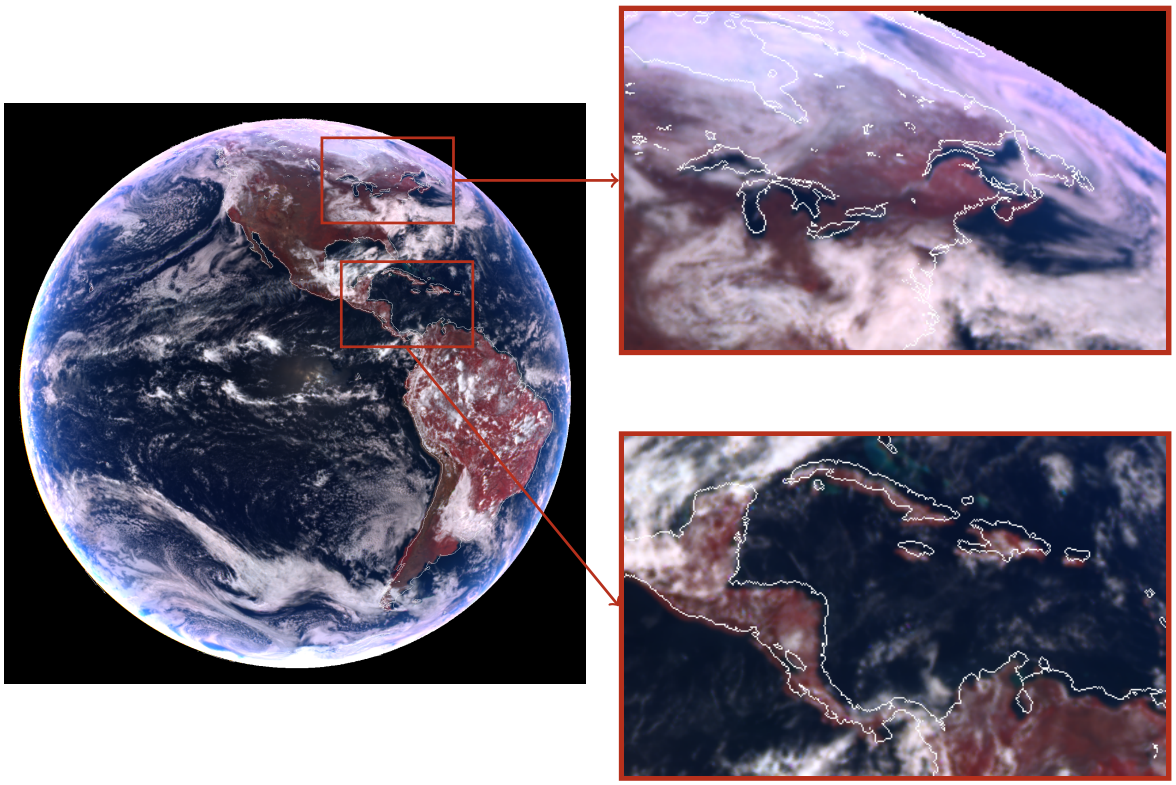


Figure 1. Example of reflectance image from EPIC L1B version 2 on 20 March 2016 at 18:36:56 UTC, with the coastline shape vector drawn in white line. On the left side, the full Earth's disk is shown. On the right side, we zoom into two regions where the misregistration is noticeable.

2. Methodology

In this section, we present a mathematical model that describes the misregistration of the EPIC L1B images, and we also describe the procedure to estimate the optimal transformation parameters that improve their registration quality. The original EPIC L1B images are accessible through NASA Earth Data portal [21]. In the following, we use the word *image* to refer to false-RGB images containing the reflectance values from the EPIC channels of 779.5, 551.0 and 443.0 nm, with the northern polar region of the Earth's disk located at the top of the image (as the example shown in Figure 1). The original counts per second measured by EPIC can be converted into reflectance by using the calibration factors estimated by Geogdzhayev and Marshak [22] and dividing then by the solar zenith-directional cosine. Although the transformation parameters were obtained for these false-RGB images, they can be used for all the EPIC L1B channels.

2.1. Mathematical Description

Given an original image pixel located at indices $[x_\mathrm{d}, y_\mathrm{d}]^\intercal$ (hereinafter the superscript $^\intercal$ stands for "transpose") which is not aligned with the proper pixels from the arrays of Earth coordinates, we model its transformation into a registered pixel at indices $[x_\mathrm{r}, y_\mathrm{r}]^\intercal$ by the following expression:

$$\begin{bmatrix} x_\mathrm{r} \\ y_\mathrm{r} \end{bmatrix} = \begin{bmatrix} x_\mathrm{s} \\ y_\mathrm{s} \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_\mathrm{u} - x_0 \\ y_\mathrm{u} - y_0 \end{bmatrix}, \tag{1}$$

where $[x_\mathrm{s}, y_\mathrm{s}]^\intercal$ is the shift vector, θ is the rotation angle around the center of rotation $[x_0, y_0]^\intercal$, and $[x_\mathrm{u}, y_\mathrm{u}]^\intercal$ are the pixel indices after radial distortion correction by means of the single-parameter Fitzgibbon division model [23,24]:

$$\begin{bmatrix} x_\mathrm{u} \\ y_\mathrm{u} \end{bmatrix} = \begin{bmatrix} x_\mathrm{c} \\ y_\mathrm{c} \end{bmatrix} + g(\lambda, r) \begin{bmatrix} x_\mathrm{d} - x_\mathrm{c} \\ y_\mathrm{d} - y_\mathrm{c} \end{bmatrix}, \qquad g(\lambda, r) = \frac{1}{1 + \lambda r^2}, \tag{2}$$

where λ is the first radial distortion coefficient, $[x_\mathrm{c}, y_\mathrm{c}]^\intercal$ is the center of distortion and r is the distance from the distorted point to the center of distortion:

$$r = \sqrt{(x_\mathrm{d} - x_\mathrm{c})^2 + (y_\mathrm{d} - y_\mathrm{c})^2}. \tag{3}$$

The alignment transformation (Equation (1)) depends on eight different parameters which are unknown and need to be estimated. Before proceeding further, and based on empirical observations, we need to apply two restrictions in order to reduce the complexity of the problem:

1. The center of distortion and the center of rotation are assumed identical, i.e., $x_0 = x_\mathrm{c}$ and $y_0 = y_\mathrm{c}$.
2. The center of distortion/rotation is assumed to be known and located at the center of the image.

Under these assumptions, the transformation (Equation (1)) from $[x_\mathrm{d}, y_\mathrm{d}]^\intercal$ to $[x_\mathrm{r}, y_\mathrm{r}]^\intercal$ can be simplified into:

$$\begin{bmatrix} x_\mathrm{r} \\ y_\mathrm{r} \end{bmatrix} = \begin{bmatrix} x_\mathrm{s} \\ y_\mathrm{s} \end{bmatrix} + \begin{bmatrix} x_\mathrm{c} \\ y_\mathrm{c} \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} g(\lambda, r)\,(x_\mathrm{d} - x_\mathrm{c}) \\ g(\lambda, r)\,(y_\mathrm{d} - y_\mathrm{c}) \end{bmatrix}, \tag{4}$$

which depends on four parameters (i.e., x_s, y_s, θ, and λ), as the point $[x_\mathrm{c}, y_\mathrm{c}]^\intercal$ is fixed now. To compute the optimal transformation parameters, we adapt Equation (4) for $n > 2$ known pairs of distorted and registered points $\{([x_{\mathrm{d}i}, y_{\mathrm{d}i}]^\intercal, [x_{\mathrm{r}i}, y_{\mathrm{r}i}]^\intercal)\}_{i=1}^{n}$, i.e.,

$$\begin{bmatrix} x_{\mathrm{r}1} \\ y_{\mathrm{r}1} \\ \vdots \\ x_{\mathrm{r}n} \\ y_{\mathrm{r}n} \end{bmatrix} = \begin{bmatrix} x_\mathrm{s} + x_\mathrm{c} \\ y_\mathrm{s} + y_\mathrm{c} \\ \vdots \\ x_\mathrm{s} + x_\mathrm{c} \\ y_\mathrm{s} + y_\mathrm{c} \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta & \cdots & 0 & 0 \\ -\sin\theta & \cos\theta & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cos\theta & \sin\theta \\ 0 & 0 & \cdots & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} g(\lambda, r_1)\,(x_{\mathrm{d}1} - x_\mathrm{c}) \\ g(\lambda, r_1)\,(y_{\mathrm{d}1} - y_\mathrm{c}) \\ \vdots \\ g(\lambda, r_n)\,(x_{\mathrm{d}n} - x_\mathrm{c}) \\ g(\lambda, r_n)\,(y_{\mathrm{d}n} - y_\mathrm{c}) \end{bmatrix}, \tag{5}$$

rewritten in compact form as

$$\boldsymbol{z}_\mathrm{r} = \boldsymbol{f}(\boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{c}, \boldsymbol{p}) \tag{6}$$

where $\boldsymbol{f}$ denotes the transformation of the distorted data vector $\boldsymbol{z}_\mathrm{d} = [x_{\mathrm{d}1}, y_{\mathrm{d}1}, \ldots, x_{\mathrm{d}n}, y_{\mathrm{d}n}]^\intercal$ into the registered data vector $\boldsymbol{z}_\mathrm{r} = [x_{\mathrm{r}1}, y_{\mathrm{r}1}, \ldots, x_{\mathrm{r}n}, y_{\mathrm{r}n}]^\intercal$ by means of a known center of distortion/rotation $\boldsymbol{z}_\mathrm{c} = [x_\mathrm{c}, y_\mathrm{c}]^\intercal$ and an unknown state vector $\boldsymbol{p} = [x_\mathrm{s}, y_\mathrm{s}, \theta, \lambda]^\intercal$. The state vector $\boldsymbol{p}_\alpha$ that optimally transforms $\boldsymbol{z}_\mathrm{d}$ into $\boldsymbol{z}_\mathrm{r}$ can be computed by least squares minimisation of the Tikhonov function $\mathcal{T}(\boldsymbol{p})$ defined as

$$\mathcal{T}(\boldsymbol{p}) = \frac{1}{2}\left(||\boldsymbol{f}(\boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{c}, \boldsymbol{p}) - \boldsymbol{z}_\mathrm{r}||^2 + \alpha||\boldsymbol{L}(\boldsymbol{p} - \boldsymbol{p}_\mathrm{a})||^2\right), \tag{7}$$

where α is the regularisation parameter, $\boldsymbol{L}$ is the regularisation matrix, and $\boldsymbol{p}_\mathrm{a}$ is the a priori state vector. The procedure to minimise $\mathcal{T}(\boldsymbol{p})$ is iterative; starting from an initial guess $\boldsymbol{p}_{(0)}$, the kth iteration consists of four steps [25]:

1. Compute the vector $\boldsymbol{y}_{(k)}$ at iteration k as

$$\boldsymbol{y}_{(k)} = \boldsymbol{z}_\mathrm{r} - \boldsymbol{f}(\boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{c}, \boldsymbol{p}_{(k)}) + \boldsymbol{J}_{(k)}(\boldsymbol{p}_{(k)} - \boldsymbol{p}_\mathrm{a}), \tag{8}$$

   where $\boldsymbol{J}_{(k)}$ denotes the Jacobian matrix of $\boldsymbol{f}$ with respect to the state vector $\boldsymbol{p}$ at iteration k:

$$\boldsymbol{J}_{(k)} = \left[\frac{\partial f_i}{\partial p_j}\right]_{\boldsymbol{p}=\boldsymbol{p}_{(k)}}, \qquad i = 1, \ldots, 2n, \quad j = 1, \ldots, 4. \tag{9}$$

2. Compute the regularised generalised inverse $\boldsymbol{J}^\dagger$ at iteration k by means of generalised singular value decomposition:

$$\boldsymbol{J}^\dagger_{(k)} = (\boldsymbol{J}^\intercal_{(k)}\boldsymbol{J}_{(k)} + \alpha\boldsymbol{L}^\intercal\boldsymbol{L})^{-1}\boldsymbol{J}^\intercal_{(k)}, \tag{10}$$

3. Compute the state vector $\boldsymbol{p}$ for the next iteration $k+1$:

$$\boldsymbol{p}_{(k+1)} = \boldsymbol{p}_\mathrm{a} + \boldsymbol{J}^\dagger_{(k)}\boldsymbol{y}_{(k)}. \tag{11}$$

4. Check the convergence criteria. If any is passed, set $\boldsymbol{p}_\alpha = \boldsymbol{p}_{(k+1)}$ and exit; otherwise, go to Step 1 for iteration $k+1$. We use the following convergence criteria:

   - X-convergence criterion: $||\boldsymbol{p}_{(k+1)} - \boldsymbol{p}_{(k)}|| < \delta$, where δ is a predefined tolerance value.
   - S-convergence criterion: $|s_{(k+1)} - s_{(k)}| < \varepsilon$, where ε is a predefined tolerance value and $s_{(k)} = ||\boldsymbol{f}(\boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{c}, \boldsymbol{p}_{(k)}) - \boldsymbol{z}_\mathrm{r}||^2$ is the squared residual sum at iteration k.

Our fitting problem is much more sensitive to the parameters $[\theta, \lambda]^\intercal$, i.e., small perturbations in these parameters cause large transformations, especially for points far from the center of distortion/rotation $[x_\mathrm{c}, y_\mathrm{c}]^\intercal$. For this reason, the computation of the optimal state vector $\boldsymbol{p}_\alpha$ is performed in two steps:

1. Define an a priori state vector $\boldsymbol{p}_{\mathrm{a}0} = [0, 0, \theta_\mathrm{a}, \lambda_\mathrm{a}]^\intercal$ with no shift, and compute the optimal state vector $\boldsymbol{p}_{\alpha 0}$ by applying Tikhonov least squares minimisation with $\theta = \theta_\mathrm{a}$ and $\lambda = \lambda_\mathrm{a}$ as fixed parameters.
2. Use $\boldsymbol{p}_{\alpha 0}$ from the previous step as the new a priori state vector $\boldsymbol{p}_\mathrm{a} = \boldsymbol{p}_{\alpha 0}$, and compute the optimal state vector $\boldsymbol{p}_\alpha$ by applying Tikhonov least squares minimisation with no fixed parameters.

The configuration for the second-step Tikhonov least squares minimisation is summarised as follows:

1. The analysis of the Jacobian matrices at the first iteration by means of the generalised singular value decomposition shows that $\alpha \approx 100$ is an appropriate value for the regularisation parameter.
2. The regularisation matrix $\boldsymbol{L}$ is defined as a diagonal matrix,

$$\boldsymbol{L} = \operatorname{diag}\left[\frac{w_{x_\mathrm{s}}}{\varepsilon_{x_\mathrm{s}}}, \frac{w_{y_\mathrm{s}}}{\varepsilon_{y_\mathrm{s}}}, \frac{w_\theta}{\varepsilon_\theta}, \frac{w_\lambda}{\varepsilon_\lambda}\right],$$

where, for every parameter, we denote by ε its expected dispersion, and we denote by $w \in [0, +\infty)$ a weighting factor that determines its freedom in the retrieval, so that $w \to 0^+$ implies total freedom and $w \to +\infty$ fixes the parameter to the a priori solution [25]. Based on the expected dispersions of the state vector components, which are set as

$$\varepsilon_{x_\mathrm{s}} = 10, \qquad \varepsilon_{y_\mathrm{s}} = 10, \qquad \varepsilon_\theta = 0.1 \text{ deg}, \qquad \varepsilon_\lambda = 1 \times 10^{-8},$$

the optimal weights have to be found empirically by performing a set of retrievals with synthetic data, ensuring that the state vector components stay within the valid ranges around the a priori solution. For our case, the following optimal values are found:

$$w_{x_\mathrm{s}} = 0, \qquad w_{y_\mathrm{s}} = 0, \qquad w_\theta = 10, \qquad w_\lambda = 10,$$

thereby implying total freedom for $[x_\mathrm{s}, y_\mathrm{s}]^\intercal$ and reduced freedom for $[\theta, \lambda]^\intercal$.
3. The first a priori state vector $\boldsymbol{p}_{\mathrm{a}0}$ is set with the following initial guesses for the rotation angle and the radial distortion parameter:

$$\theta_\mathrm{a} = 0.5 \text{ deg}, \qquad \lambda_\mathrm{a} = -5 \times 10^{-9}.$$

The workflow for the complete non-linear optimisation process is summarised in Figure 2a.

2.2. Detection of Matching Coastline Features

The registration of an image by means of the model described in Section 2.1 assumes that the data vectors $\boldsymbol{z}_\mathrm{d}$ and $\boldsymbol{z}_\mathrm{r}$ are known. In our case, they can be built with a sufficient number of pairs of incorrect points $[x_\mathrm{d}, y_\mathrm{d}]^\intercal$ and correct points $[x_\mathrm{r}, y_\mathrm{r}]^\intercal$ within the image. The coastline that is visible in this kind of images is a source of these points of interest, as the correct coastline can be determined with the arrays of Earth coordinates, which are available. With this consideration, we compute the required pairs of incorrect and correct points as follows (see Figure 2b):

1. Create a mask with the theoretical coastline, i.e., that inferred from the arrays of Earth coordinates.
2. Create a mask with the radiometric coastline, i.e., that inferred from the actual image.
3. Find pairs of common features from both coastline masks by using computer vision techniques.

2.2.1. Computation of the Theoretical Coastline

The computation of the theoretical coastline, i.e., the coastline that is assumed to be correct, can be divided into three steps:

1. Create the theoretical land mask by checking for every image pixel if its (latitude, longitude) pair is contained inside a land polygon from the low-resolution GSHHG (Global Self-consistent, Hierarchical, High-resolution Geography database) [26].
2. Apply one morphological binary erosion to the theoretical land mask with a flat diamond shaped structuring element of dimensions 3×3 [27].
3. Compute the theoretical coastline mask as the result of the bitwise operator XOR on the original and eroded theoretical land masks.

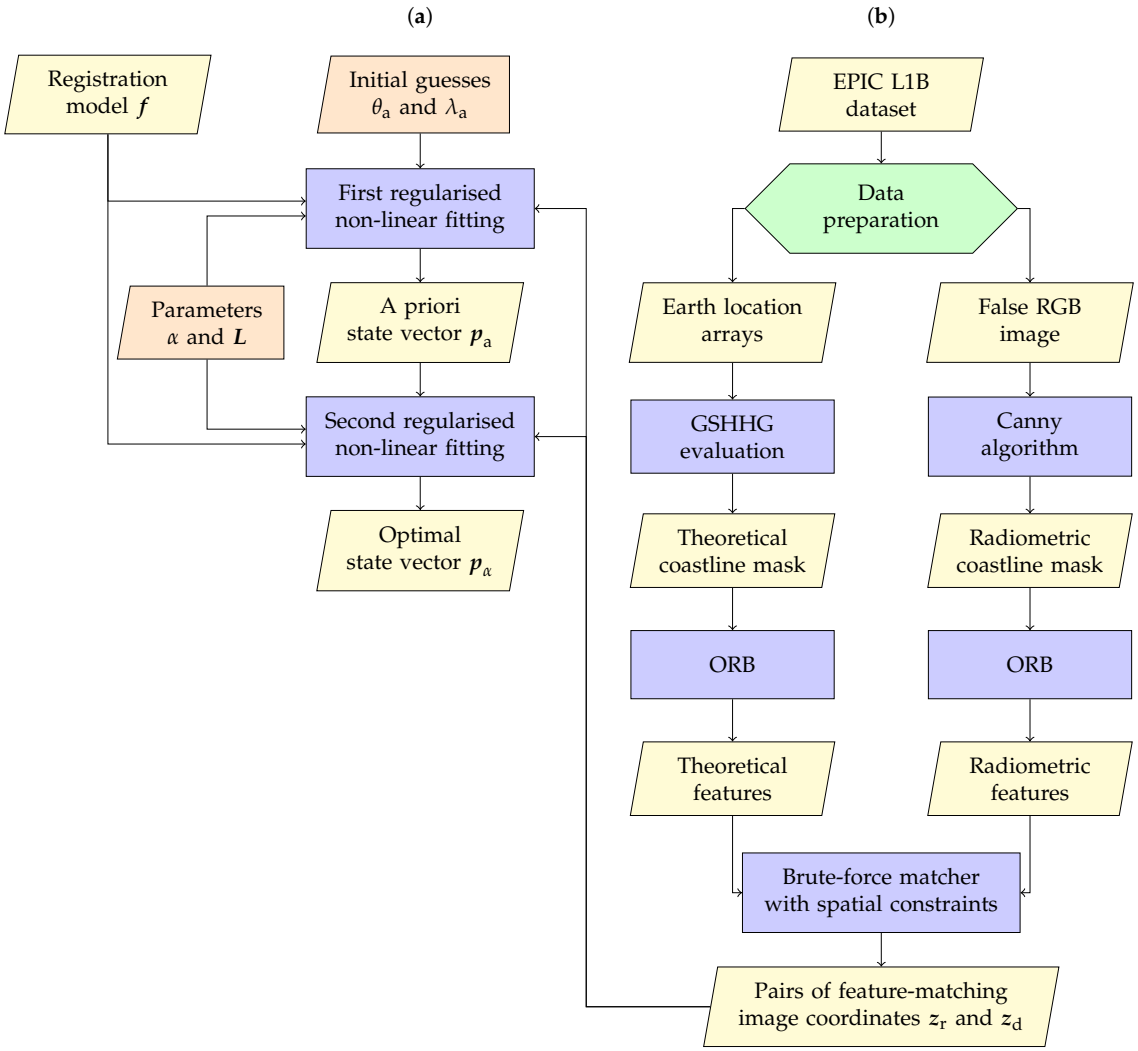


Figure 2. Workflow chart for the computation of the optimal transformation parameters for a given image. (**a**) Main workflow chain in which the non-linear least squares fitting with regularisation is performed. (**b**) Secondary workflow chain in which the known data vectors z_r and z_d are prepared.

2.2.2. Computation of the Radiometric Coastline

The computation of the radiometric coastline, i.e., the coastline that is visible in the image, is a common edge detection problem in image analysis, and can be performed as follows:

1. Convert the image channel with the highest contrast between land and water into 8-bit form (i.e., the false-red channel, which corresponds to the 779.5 nm EPIC channel).
2. Compute the median value v from the Earth pixels of this 8-bit image.
3. Compute the radiometric coastline mask by applying the Canny edge detection algorithm [28] on the previous 8-bit image with hysteresis thresholding parameters given by

$$t_{\text{lower}} = \max[0, (1-\sigma)v], \qquad t_{\text{upper}} = \min[255, (1+\sigma)v], \tag{12}$$

where $\sigma \geq 0$ is an argument that controls the separation between both thresholds [29] and its value is found empirically: if σ is small, only the strong borders are preserved; if σ is big, both strong and weak borders are kept. For our purpose, we found that $\sigma = 0.33$ is a good compromise.

With this procedure, the radiometric coastline will be polluted in general with other edges also present in the image, e.g., cloud edges (see Figure 3). However, they do not need to be filtered as long as the detected matching features are restricted to the neighbourhood of the theoretical coastline.

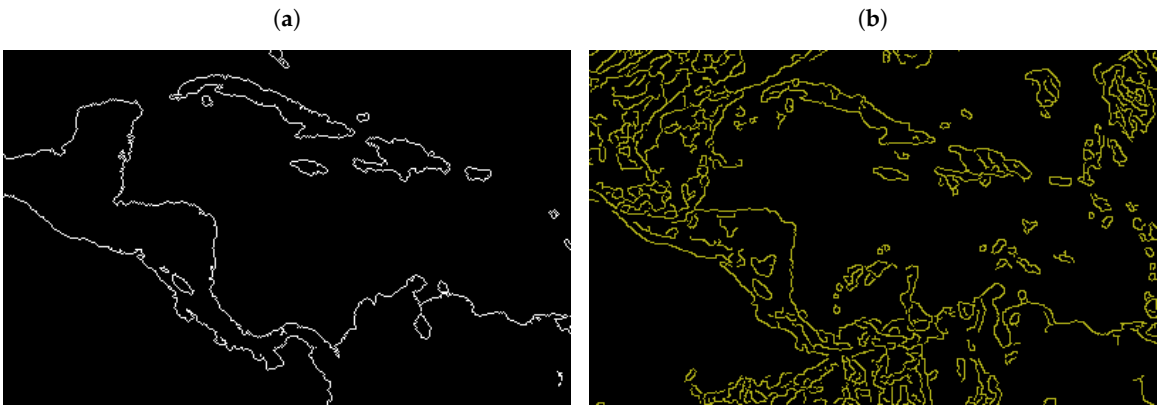


Figure 3. Comparison of the (**a**) theoretical and (**b**) radiometric coastlines for one of the regions shown in Figure 1. The radiometric coastline is partially polluted with other type of borders detected by the Canny algorithm (e.g., clouds or interfaces between two different land covers).

2.2.3. Matching of Coastline Features

Once the theoretical and radiometric coastline masks are computed, it is possible to find features that are common to both masks, as long as the misalignment and distortion of the radiometric coastline is not severe. Such image matching problem is one of the fundamental research topics in computer vision [30]; the procedure to find common features between two images consists of three steps:

1. Detect keypoints (e.g., edges, corners, and regions of interest) in the compared images.
2. Describe every keypoint by a descriptor vector with information from its neighbourhood.
3. Match keypoint pairs based on the similarity of their descriptor vectors.

Examples of successful detectors and descriptors are SIFT (Scale-Invariant Feature Transform) [31], SURF (Speeded-Up Robust Features) [32] and ORB (Oriented FAST and Rotated BRIEF) [33]. OpenCV [34] already provides a complete interface to these algorithms. For our purpose, we use ORB because it is not patented, its time performance is better, and the amount of detected keypoints is enough. ORB consists of two routines:

1. The ORB detector is a modified version of the keypoint detector FAST (Features from Accelerated Segment Test) [35]. In addition to the original FAST, it also computes the orientation angle of the detected keypoints.
2. The ORB descriptor is a modified version of the descriptor BRIEF (Binary Robust Independent Elementary Features) [36]. The binary descriptor vector generated by the original BRIEF shows problems when identifying matching keypoints under rotation conditions; ORB fixes the issue taking into consideration the orientation computed by the ORB detector.

The keypoints detected and described with ORB for both the theoretical and radiometric coastline masks are matched by using a brute-force matcher with Hamming distance as measurement (see [37] for further details). Given a set of theoretical keypoints $\{z_{1k_1}\}_{k_1=1}^{n_1}$ with descriptor vectors $\{\boldsymbol{d}_{1k_1}\}_{k_1=1}^{n_1}$ and a set of radiometric keypoints $\{z_{2k_2}\}_{k_2=1}^{n_2}$ with descriptor vectors $\{\boldsymbol{d}_{2k_2}\}_{k_2=1}^{n_2}$, the matching algorithm proceeds for every theoretical keypoint z_{1i}, $1 \leq i \leq n_1$, as follows:

1. Compute the Hamming distance (i.e., the number of positions at which the corresponding values of arrays are different [38]) h_{ik_2} from z_{1i} to every radiometric keypoint z_{2k_2}, $1 \leq k_2 \leq n_2$, as

$$h_{ik_2} = \sum_{m=1}^{n_{\text{bits}}} (\boldsymbol{d}_{1i})_m \oplus (\boldsymbol{d}_{2k_2})_m, \tag{13}$$

where n_{bits} is the length of the binary descriptor vectors, $\oplus$ is the XOR operator, and $(\boldsymbol{d}_{1i})_m$ and $(\boldsymbol{d}_{2k_2})_m$ denote the mth components of the descriptor vectors $\boldsymbol{d}_{1i}$ and $\boldsymbol{d}_{2k_2}$, respectively.

2. Select the radiometric keypoint z_{2j} with minimum Hamming distance to z_{1i} as the matching candidate for the theoretical keypoint z_{1i}, where

$$j = \arg\min\{h_{ik_2}\}_{k_2=1}^{n_2}. \tag{14}$$

3. For the given radiometric keypoint z_{2j}, compute its Hamming distance h_{k_1j} to every theoretical keypoint z_{1k_1}, $1 \leq k_1 \leq n_1$, as

$$h_{k_1j} = \sum_{m=1}^{n_{\text{bits}}} (\boldsymbol{d}_{1k_1})_m \oplus (\boldsymbol{d}_{2j})_m. \tag{15}$$

4. The pair (z_{1i}, z_{2j}) is a valid matching pair only if z_{1i} is the theoretical keypoint with minimum Hamming distance to z_{2j}, otherwise the pair is discarded, i.e., it is valid only if

$$i = \arg\min\{h_{k_1j}\}_{k_1=1}^{n_1}. \tag{16}$$

The previous procedure ensures that all the valid pairs of matching keypoints hold the minimum Hamming distance criterion reciprocally. Note that this criterion, however, does not consider any spatial restriction, but only the similarity between the descriptor vectors. Due to the spatial constraints of our problem, the valid pairs need to be filtered based on two additional spatial criteria (Figure 4):

1. The spatial distance between a theoretical keypoint z_{1i} and the theoretical coastline.
2. The spatial distance between the keypoints of a matching pair (z_{1i}, z_{2j}).

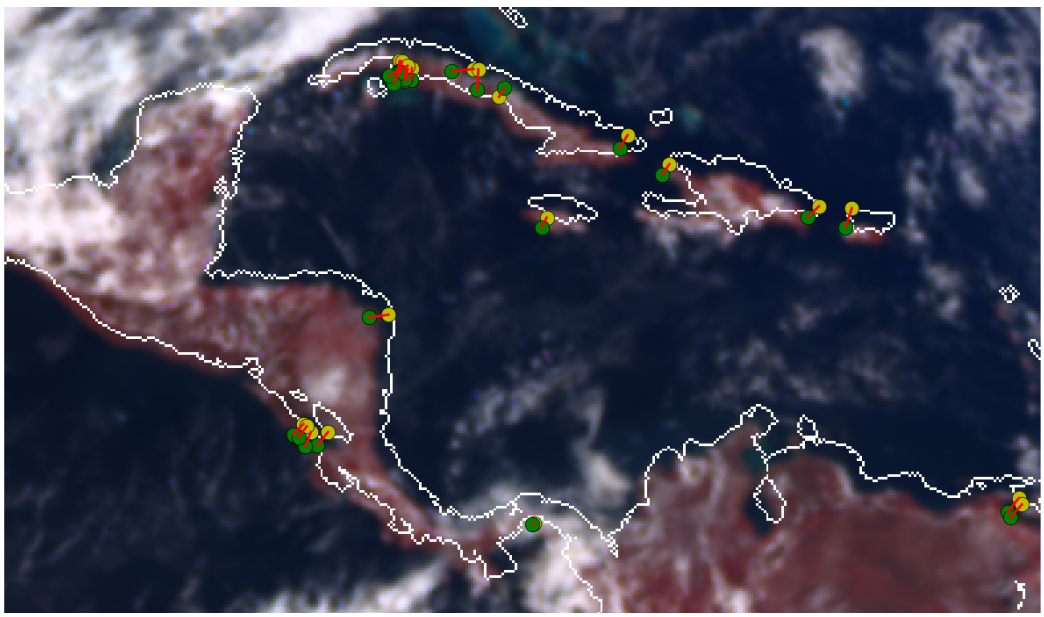


Figure 4. Pairs of matching keypoints found within one of the regions shown in Figure 1 after applying the outlier rejection step based on spatial criteria. The keypoints detected in the theoretical coastline (white lines) are shown as yellow points, while the corresponding keypoints detected in the radiometric coastline are shown as green points. Every matching pair is connected with a red segment.

If any of these spatial distances is greater than a threshold (e.g., 10 pixels), the pair is discarded. After the application of this filter, the remaining pairs are used to build the registered data vector z_r with the theoretical keypoints and the distorted data vector z_d with the radiometric keypoints. Alternatively, an approach based on RANSAC (RANdom SAmple Consensus) [39] could be added afterwards in the retrieval procedure to generate z_r and z_d. We do not use it in this article because the improvement in our registration problem was small and it increased the computation time in excess.

3. Results

We performed the estimation of the optimal transformation parameters for all the datasets from EPIC L1B version 2 within the time period from 13 June 2015 to 31 July 2018. The parameters were computed with an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz processor using a single thread, and it required approximately 15 s of computation time for every image (i.e., roughly between 18 and 36 h for every year of datasets, assuming 12–24 available datasets per day). To assess the behaviour of the obtained results, we analysed the following aspects:

1. The performance of the non-linearised fitting procedure in reducing the spatial distance between the image coordinates from matching theoretical and radiometric features.
2. The global impact of this correction procedure on the image registration quality.
3. The behaviour of the retrieved transformation parameters as a function of time.

In Figure 5, we show the result of computing the spatial distances between the points of every matching pair before and after applying the registration procedure described in the previous section. The histograms were calculated for all the matching pairs that were found in all the EPIC L1B images within the complete time period under study. In Figure 5a, we can observe a normal distribution of spatial distances spread within the interval of $[0, 10]$ pixels, with a higher density of occurrences in the subinterval of $[3, 7]$ pixels. We can find 50% of the values in the range of $[0, 5]$ pixels. The histogram is cut for values higher than 10 pixels due to the spatial constraints defined in Section 2.2.3. In Figure 5b, we see that the distribution of spatial distances is turned into a Gamma distribution whose mode is found in the bin of $[1.25, 1.50]$ pixels, and where approximately 50% of the values are located within the range of $[0, 1.75]$ pixels. The distribution also shows a long decreasing tail on its right side, meaning that a small percentage of values are now greater than 5 pixels.

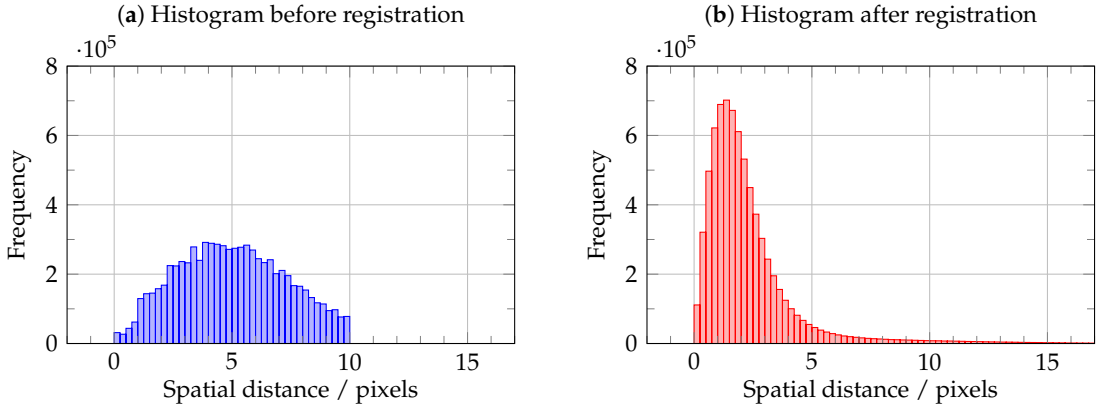


Figure 5. Histograms of spatial distances (in pixels) between the points of matching pairs: (**a**) before applying registration, i.e., as they are originally located; and (**b**) after applying registration, i.e., after transforming the distorted points into registered points with the model described in Section 2. The histograms were computed for all the matching pairs found in three years of EPIC L1B images.

From these histograms, we conclude that the transformation significantly reduces the spatial distance between pairs of matching points. As long as these matching pairs are representative of the distortion that the original images suffer, this fact translates into an improvement in the registration

of the images. The existence of the right tail in the histogram in Figure 5b also suggests that some matching pairs do not reduce its spatial distance after the transformation, or they might even increase it. Indeed, this means that there are matching pairs whose points do not correspond to the same feature, but they could still pass the quality filters applied in our procedure. As the amount of wrong matching pairs is not substantial, they do not significantly influence in the retrieval of the transformation parameters, and for every wrong matching pair the points are moved to their new positions determined by the optimal transformation parameters. The location of the histogram mode in Figure 5b is above one pixel due to the necessary discretisation of the coastline together with the actual ground pixel size. The theoretical coastlines were always computed as the last contour of land pixels for every land body based on the GSHHG database. The radiometric coastlines were obtained by finding intensity borders within the false-RGB images, which coincide in general with the actual land borders (see Figure 6), but in some cases they can be found one pixel inside land (because the land–ocean interface pixels are a mixture of land and ocean information) or one pixel outside land (due to water turbidity).

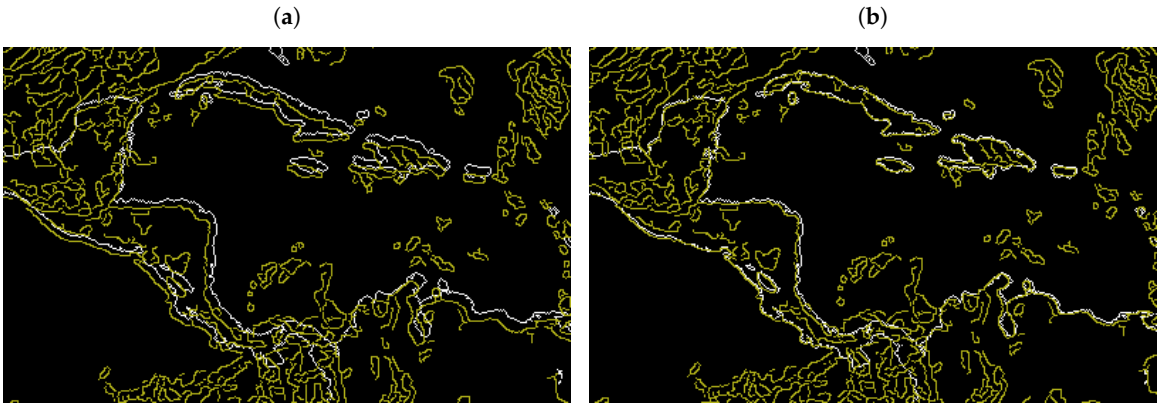


Figure 6. Comparison of the theoretical coastline (white lines) and the radiometric coastline (yellow lines) for one of the regions shown in Figure 1 (**a**) before and (**b**) after applying the registration.

In Figure 7, we show the application of the registration procedure to four examples of EPIC L1B image regions. On the left side, we observe the original L1B images with the corresponding theoretical coastline shape vector superposed. On the right side, the image pixels are relocated by means of our registration model with the transformation parameters automatically retrieved for every image. It can be observed that there is an improvement in the collocation of the radiometric coastline and the theoretical coastline with respect to their original situation. Because the features were detected along the coastlines, the density of matching features is sensitive to the percentage of land that is visible on every image. As a consequence, the procedure may fail for those images in which the majority of the Earth's disk corresponds to the Pacific Ocean due to the absence of matching features with enough quality. For these situations, we must rely on the retrieved parameters from other images that are close in time.

In Figure 8, we show the time evolution of the transformation parameters for the complete time period under study. The time series was computed for the daily mean values of each parameter. We observed that the horizontal shift x_s has an oscillatory behaviour around its mean value of 2.5 pixels, while the vertical shift y_s also presents a periodic response around its mean value of -0.2 pixels. The rotation angle θ and the distortion parameter λ present noisier time series, and their values remain closer to the mean values of 0.498 deg and -4.958×10^{-9}, respectively. Indeed, this means that these two parameters tend to stay close to their a priori solutions—an expected result based on the selection of the regularisation matrix $\boldsymbol{L}$. The negative sign of λ indicates that the original EPIC L1B images still suffer from a slight barrel distortion.

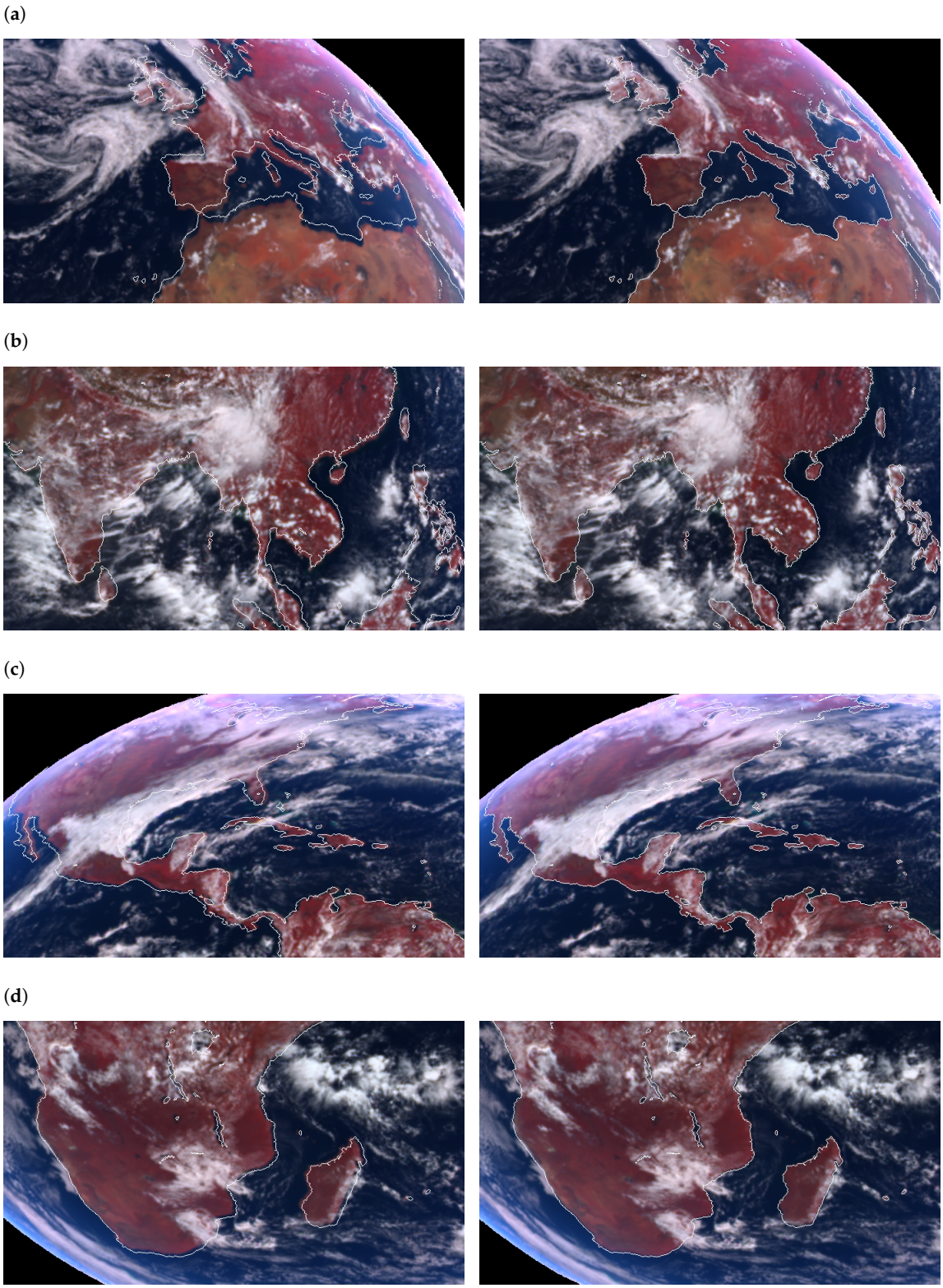


Figure 7. Examples of different reflectance images from EPIC L1B version 2 before (**left**) and after (**right**) applying the new registration: (**a**) Europe and North Africa on 4 August 2015 14:37:27 UTC; (**b**) East Asia on 24 July 2016 07:09:19 UTC; (**c**) Caribbean Sea and Gulf of Mexico on 15 December 2017 17:17:03 UTC; and (**d**) the south of Africa and Madagascar on 16 May 2018 09:10:54 UTC.

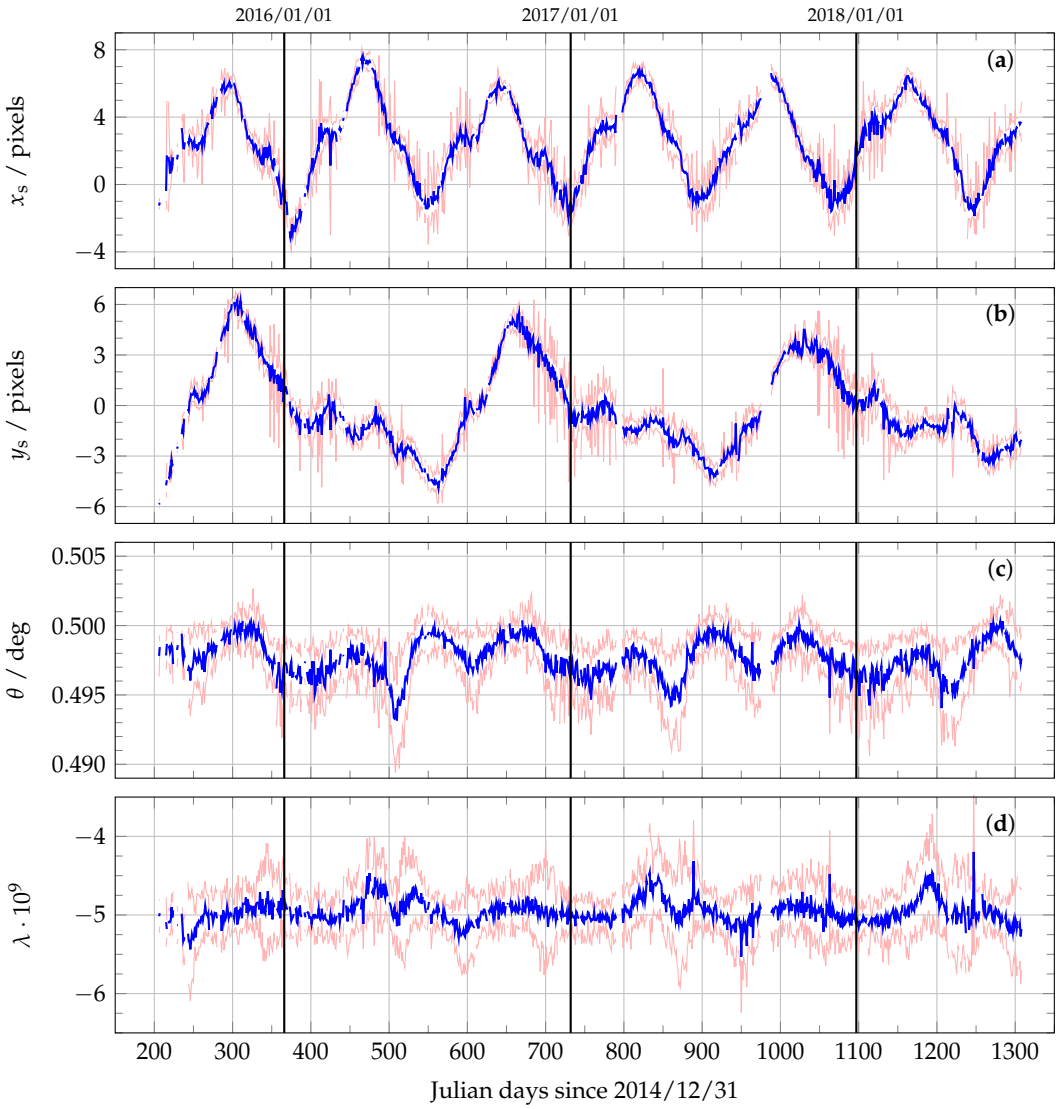


Figure 8. Time series plot for every transformation parameter: (**a**) the horizontal shift x_s; (**b**) the vertical shift y_s; (**c**) the rotation angle θ; and (**d**) the radial distortion parameter λ. The blue lines represent the evolution of the daily mean value for every parameter, while the light red curves represent the daily dispersion of each parameter given as the daily mean value plus/minus the daily standard deviation.

The time series in Figure 8 can be also interpreted as noisy signals with unevenly sampled data due to the days where there are not enough images to be processed. To estimate the dominant frequencies for every time series, we computed their periodograms by means of the Lomb–Scargle method [40–42]. Before this computation, the signals need to be adapted to have zero mean (subtracting their mean values) and standard deviation of unity (dividing by their standard deviations). In Figure 9, we show the resulting normalised periodogram amplitudes for every transformation parameter as a function of the time period. It can be noticed that three parameters (x_s, y_s, and λ) share one common frequency whose period T_1 is located in the interval of $[173, 180]$ days, and two parameters (y_s and λ) share a second frequency whose period T_2 is located in the interval of $[363, 366]$ days. The first period T_1 corresponds to the 180-day orbit period of DSCOVR's spacecraft around the Lagrangian point L_1 [43]. The rotation angle θ shows different dominant frequencies; however, we do not consider them relevant because the behaviour of its time series is indeed flat around a constant value.

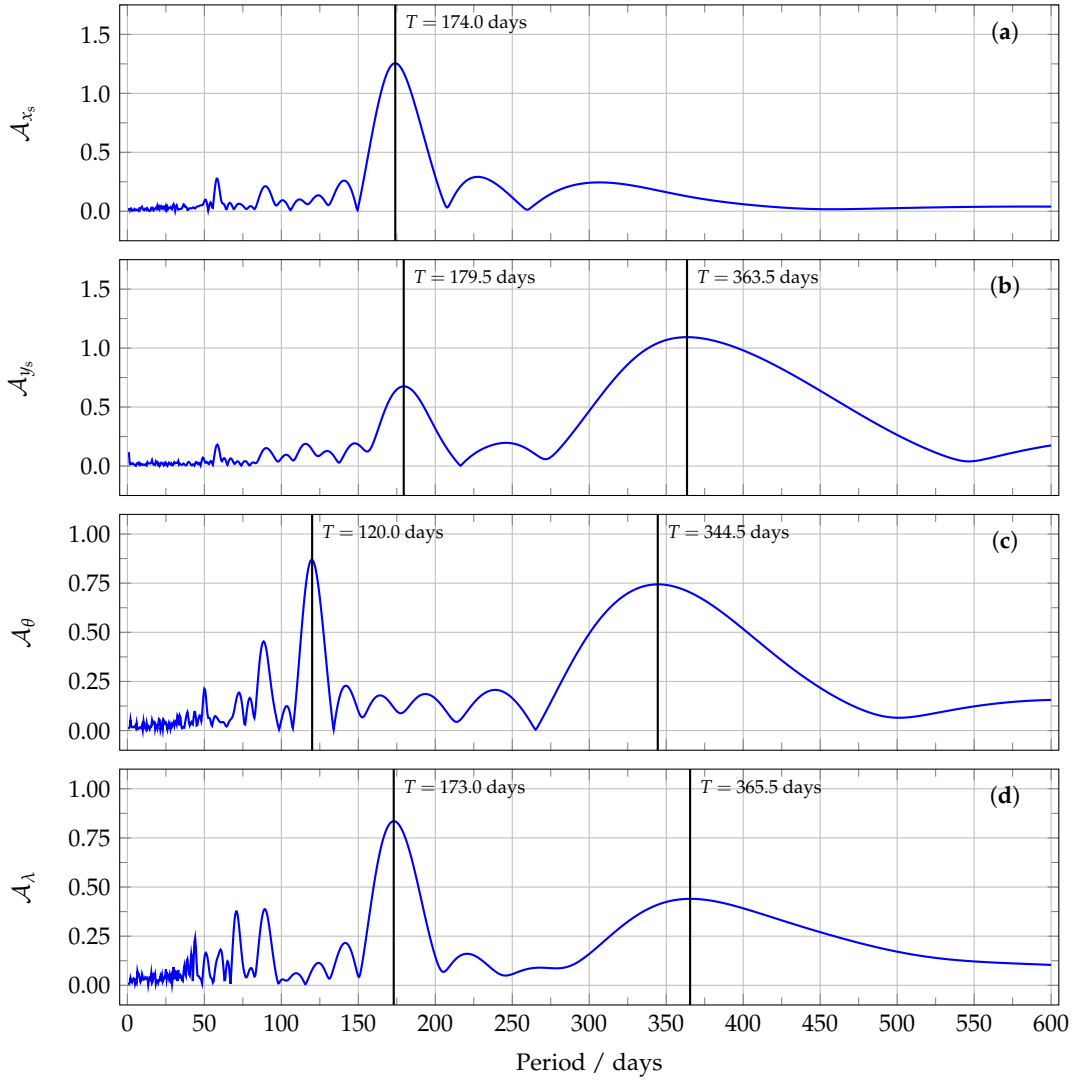


Figure 9. Normalised periodogram amplitude $\mathcal{A}$ as a function of the time period in days for every transformation parameter: (**a**) the horizontal shift x_s; (**b**) the vertical shift y_s; (**c**) the rotation angle θ; and (**d**) the radial distortion parameter λ.

4. Conclusions

In this study, we addressed the geolocation issue that is present in the imagery from EPIC L1B version 2, and to solve it we developed a fast automatic image registration scheme consisting of a motion plus a radial distortion correction. The optimal transformation parameters were computed through Tikhonov least squares minimisation, in which the reference pairs of distorted and registered pixel coordinates were determined automatically by matching features from the theoretical and radiometric coastlines.

We showed that the shift parameters $[x_s, y_s]^\intercal$ present an oscillatory behaviour whose period is close to the period of DSCOVR's Lissajous orbit, while the rotation angle θ and the distortion parameter λ were found to be more stable and close to the a priori solution that we provided. After the proposed registration, the spatial distances between the common features found in the theoretical and radiometric coastlines are concentrated around the interval of $[1.25, 1.50]$ pixels, a considerable improvement compared to the interval of $[3, 7]$ pixels from the original L1B images. Therefore, we conclude that this procedure enhances the registration of the images from EPIC L1B version 2 with their corresponding arrays of Earth coordinates. This enhancement will have a positive impact in climate science applications from EPIC measurements, as it will reduce the errors in the

retrieval of atmospheric properties due to the inappropriate selection of the surface properties of every pixel. The algorithm developed in this work could be adapted for improving the registration of the future geostationary UVN (Ultraviolet-Visible-Near-infrared) sensors Sentinel-4, GEMS and TEMPO.

This study was based on the OpenCV library, which contains other matching algorithms apart from the brute-force matching of corresponding keypoints, including template matching (e.g., Chamfer matching [44]). In principle, it is interesting to compare their efficiencies for solving the registration problem considered in this article. In addition, the effect of using different robust estimators (e.g., RANSAC) can be analysed in more detail. These topics will be investigated in our future research.

Author Contributions: Conceptualisation, V.M.G.; methodology, V.M.G. and D.S.E.; software, V.M.G. and S.S.; visualisation, V.M.G. and S.S.; validation, V.M.G. and S.S. and D.S.E.; formal analysis, V.M.G., D.S.E. and D.L.; writing—original draft preparation, V.M.G. and S.S.; writing—review and editing, V.M.G., S.S., D.S.E. and D.L.; supervision, D.L.; and funding acquisition, D.L.

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 2015 (57186656), with reference numbers 91613528 and 91627488.

Acknowledgments: V.M.G. thanks Adrian Doicu for his support on questions related to regularised non-linear least squares minimisation processes.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BRIEF	Binary Robust Independent Elementary Features
CCD	Charge-Coupled Device
DSCOVR	Deep Space Climate Observatory
EPIC	Earth Polychromatic Imaging Camera
FAST	Features from Accelerated Segment Test
GEMS	Geostationary Environment Monitoring Spectrometer
GSHHG	Global Self-consistent Hierarchical, High-resolution Geography database
L1B	Level 1B
MERIS	MEdium Resolution Imaging Spectrometer
NASA	National Aeronautics and Space Administration
ORB	Oriented FAST and Rotated BRIEF
RANSAC	RANdom SAmple Consensus
RGB	Red Green Blue
SIFT	Scale-Invariant Feature Transform
SURF	Speeded-Up Robust Features
TEMPO	Tropospheric Emissions: Monitoring of POllution
UVN	Ultraviolet-Visible-Near-infrared

References

1. Marshak, A.; Herman, J.; Szabo, A.; Blank, K.; Carn, S.; Cede, A.; Geogdzhayev, I.; Huang, D.; Huang, L.K.; Knyazikhin, Y.; et al. Earth Observations from DSCOVR EPIC Instrument. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 1829–1850. [CrossRef] [PubMed]
2. Marshak, A.; Knyazikhin, Y. The spectral invariant approximation within canopy radiative transfer to support the use of the EPIC/DSCOVR oxygen B-band for monitoring vegetation. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *191*, 7–12. [CrossRef] [PubMed]
3. Herman, J.; Huang, L.; McPeters, R.; Ziemke, J.; Cede, A.; Blank, K. Synoptic ozone, cloud reflectivity, and erythemal irradiance from sunrise to sunset for the whole earth as viewed by the DSCOVR spacecraft from the earth–sun Lagrange 1 orbit. *Atmos. Meas. Tech.* **2018**, *11*, 177–194. [CrossRef]

4. Carn, S.A.; Krotkov, N.A.; Fisher, B.L.; Li, C.; Prata, A.J. First observations of volcanic eruption clouds from the L1 Earth-Sun Lagrange point by DSCOVR/EPIC. *Geophys. Res. Lett.* **2018**, *45*, 11456–11464. [CrossRef]
5. Efremenko, D.S.; Loyola R., 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. [CrossRef]
6. Xu, X.; Wang, J.; Wang, Y.; Zeng, J.; Torres, O.; Yang, Y.; Marshak, A.; Reid, J.; Miller, S. Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point. *Geophys. Res. Lett.* **2017**, *44*, 7544–7554. [CrossRef]
7. Natraj, V.; Jiang, J.H.; Doicu, A.; Loyola, D.; Kopparla, P.; Yung, Y.L. Aerosol retrievals from DSCOVR measurements. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6026–6028. [CrossRef]
8. Yang, Y.; Meyer, K.; Wind, G.; Zhou, Y.; Marshak, A.; Platnick, S.; Min, Q.; Davis, A.B.; Joiner, J.; Vasilkov, A.; et al. Cloud products from the Earth Polychromatic Imaging Camera (EPIC): Algorithms and initial evaluation. *Atmos. Meas. Tech.* **2019**, *12*, 2019–2031. [CrossRef]
9. Davis, A.B.; Merlin, G.; Cornet, C.; Labonnote, L.C.; Riédi, J.; Ferlay, N.; Dubuisson, P.; Min, Q.; Yang, Y.; Marshak, A. Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: An optimal estimation approach. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *216*, 6–16. [CrossRef]
10. Davis, A.B.; Ferlay, N.; Libois, Q.; Marshak, A.; Yang, Y.; Min, Q. Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: A physics-based approach. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *220*, 84–96. [CrossRef]
11. Gao, M.; Zhai, P.W.; Yang, Y.; Hu, Y. Cloud remote sensing with EPIC/DSCOVR observations: A sensitivity study with radiative transfer simulations. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *230*, 56–60. [CrossRef]
12. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]
13. Tikhonov, A.N.; Arsenin, V.I. *Solutions of Ill-Posed Problems*; Scripta Series in Mathematics; Winston & Sons: Washington, DC, USA, 1977.
14. 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]
15. 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]
16. 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]
17. Efremenko, D.S.; Doicu, A.; Loyola, D.; Trautmann, T. Fast stochastic radiative transfer models for trace gas and cloud property retrievals under cloudy conditions. In *Springer Series in Light Scattering*; Springer International Publishing: Basel, Switzerland, 2018; pp. 231–277. [CrossRef]
18. 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]
19. Blank, K. EPIC Geolocation and Color Imagery Algorithm Revision 5. Technical Report. 2017. Available online: https://eosweb.larc.nasa.gov/project/dscovr/DSCOVR_EPIC_Geolocation_V02.pdf (accessed on 23 July 2019).
20. Haney, C.; Doelling, D.; Minnis, P.; Bhatt, R.; Scarino, B.; Gopalan, A. The calibration of the DSCOVR EPIC multiple visible channel instrument using MODIS and VIIRS as a reference. In Proceedings of the Earth Observing Systems XXI, San Diego, CA, USA, 28 August–1 September 2016. [CrossRef]
21. Available online: https://search.earthdata.nasa.gov (accessed on 27 May 2019).
22. Geogdzhayev, I.V.; Marshak, A. Calibration of the DSCOVR EPIC visible and NIR channels using MODIS Terra and Aqua data and EPIC lunar observations. *Atmos. Meas. Tech.* **2018**, *11*, 359–368. [CrossRef]
23. Fitzgibbon, A.W. Simultaneous linear estimation of multiple view geometry and lens distortion. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, 8–14 December 2001; pp. 125–132. [CrossRef]
24. Wang, A.; Qiu, T.; Shao, L. A simple method of radial distortion correction with centre of distortion estimation. *J. Math. Imaging Vis.* **2009**, *35*, 165–172. [CrossRef]

25. Doicu, A.; Trautmann, T.; Schreier, F. *Numerical Regularization for Atmospheric Inverse Problems*; Springer: Berlin/Heidelberg, Germany, 2010. [CrossRef]
26. Wessel, P.; Smith, W.H.F. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res. Solid Earth* **1996**, *101*, 8741–8743. [CrossRef]
27. Serra, J. Introduction to mathematical morphology. *Comput. Vis. Graph. Image Process.* **1986**, *35*, 283–305. [CrossRef]
28. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 679–698. [CrossRef]
29. Available online: https://www.pyimagesearch.com/2015/04/06/zero-parameter-automatic-canny-edge-detection-with-python-and-opencv (accessed on 11 July 2019).
30. Remondino, F.; Spera, M.G.; Nocerino, E.; Menna, F.; Nex, F. State of the art in high density image matching. *Photogramm. Rec.* **2014**, *29*, 144–166. [CrossRef]
31. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
32. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
33. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571. [CrossRef]
34. Kaehler, A.; Bradski, G. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*; O'Reilly Media: Sebastopol, CA, USA, 2016.
35. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]
36. Calonder, M.; Lepetit, V.; Ozuysal, M.; Trzcinski, T.; Strecha, C.; Fua, P. BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1281–1298. [CrossRef]
37. Available online: https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_matcher/py_matcher.html (accessed on 11 July 2019).
38. Hamming, R. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147–160. [CrossRef]
39. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
40. Lomb, N. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **1976**, *39*, 447–462. [CrossRef]
41. Scargle, J. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **1982**, *263*, 835–853. [CrossRef]
42. Hocke, K. Phase estimation with the Lomb-Scargle periodogram method. *Annal. Geophys.* **1998**, *16*, 356–358.
43. Jiang, J.H.; Zhai, A.J.; Herman, J.; Zhai, C.; Hu, R.; Su, H.; Natraj, V.; Li, J.; Xu, F.; Yung, Y.L. Using Deep Space Climate Observatory measurements to study the Earth as an exoplanet. *Astron. J.* **2018**, *156*, 26. [CrossRef]
44. Van Herk, M. Image registration using Chamfer matching. In *Handbook of Medical Image Processing and Analysis*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 591–603. [CrossRef]

The paper was published in Journal of Quantitaitve Spectroscopy & Radiative Transfer, **196**(2017)17-45.
DOI: 10.1016/j.jqsrt.2017.02.015

A review of the matrix-exponential formalism in radiative transfer

Dmitry S. Efremenko*, Víctor Molina García, Sebastián Gimeno García[1], Adrian Doicu

Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung (IMF), 82234 Oberpfaffenhofen, Germany

Abstract

This paper outlines the matrix exponential description of radiative transfer. The eigendecomposition method which serves as a basis for computing the matrix exponential and for representing the solution in a discrete ordinate setting is considered. The mathematical equivalence of the discrete ordinate method, the matrix operator method, and the matrix Riccati equations method is proved rigorously by means of the matrix exponential formalism. For optically thin layers approximate solution methods relying on the Padé and Taylor series approximations to the matrix exponential, as well as on the matrix Riccati equations are presented. For optically thick layers, the asymptotic theory with higher-order corrections is derived, and parametrizations of the asymptotic functions and constants for a water-cloud model with a Gamma size distribution are obtained.

Keywords: matrix-exponential, discrete ordinate method, matrix operator method, matrix Riccati equations

Contents

1. Introduction

The radiative transfer is an important issue for astrophysics, atmospheric physics, meteorology and engineering sciences. A wide range of solution methods of the radiative transfer equation (RTE) have been proposed (see, e.g., [1–11] and references therein for a general review). The discrete ordinate method [6, 12–14] and the matrix operator method [15–18] involve replacing the continuous dependence of the radiance on direction by a dependence on a discrete set of directions. For a homogeneous layer, the discretized radiative transfer equation then takes the form of a system of linear first-order differential equations. In the classical discrete ordinate method of Chandrasekhar, the solution of the system of equations is expressed as a linear combination of characteristic solutions of the discretized problem, while the matrix operator method is primarily oriented toward numeri-

*Corresponding author
Email address: dmitry.efremenko@dlr.de (Dmitry S. Efremenko)
[1]Now at EUMETSAT, Eumetsat Allee 1, D-64295 Darmstadt, Germany.

cal computations of the reflection and transmission matrices. Another group of methods are based on the concept of invariant imbedding, which is due to Ambarzumian [19]. Ambarzumian derived an equation for the reflection function of a semi-infinite atmosphere by noting that the reflection function remains unchanged upon addition of a new layer. This technique was further generalized by Chandrasekhar [13] to a finite layer, while Bellman et al. [20] showed that the reflection function derived by using the invariant imbedding satisfies the Riccati equation.

The system of differential equations of the discretized radiative transfer equation can be solved by using a classical mathematical procedure involving the matrix exponential operator, in which the system matrix appears in the exponent. Waterman [21] was the first who provided a matrix exponential description of radiative transfer. Mathematical elegance aside, he showed its practical value in radiative transfer computations from both the analytical and purely numerical point of view. In particular, Waterman related the matrix exponential to the extinction matrix incorporating the reflection and transmission matrices of a homogeneous layer, provided an eigenvector representation of the matrix exponential, derived analytical expressions for the reflection and transmission matrices in the limit of small and large optical thicknesses, showed that the matrix exponential can be used to generate starting values for the doubling method, and applied the matrix exponential formalism to conservative scattering. Flatau and Stephens [22] extended the concept of matrix exponential of a homogeneous layer to an inhomogeneous atmosphere by introducing the so-called propagator (matrix) operator. As Waterman, Flatau and Stephens related the propagator to the extinction matrix of a homogeneous layer, notified the similarity between the matrix exponential solution and Chandrasekhar's discrete ordinate solution, established various properties of the propagator and used them to derive the Riccati matrix equations for an inhomogenous atmosphere, as well as the adding and doubling formulas. Although in both papers [21, 22] an eigendecomposition method for computing the matrix exponential is considered, explicit and stable representations of the reflection and transmission matrices are not given. This problem has been solved by Nakajima and Tanaka [18] by using a system of characteristic solutions of the discretized problem, and by Budak et al. [23, 24] by using the matrix exponential formalism. It should be also mentioned that Doicu and Trautmann [25, 26] designed the so-called discrete ordinate method with matrix exponential to compute the radiance field in a multilayered atmosphere.

The purpose of this paper is to provide a consistent overview of the matrix exponential description of radiative transfer. We mainly focus on a mathematical rigorous and self-contained analysis based on the results given in [21, 22, 27] and our own results [25, 26, 28, 29]. The final goals are to prove the mathematical equivalence of the discrete ordinate method, matrix operator method, and the matrix Riccati equations method, on the one hand, and to derive efficient computations formulas for the reflection and transmission matrices in the limit of small and large optical thicknesses, on the other hand.

The rest of the paper is organized as follows. In Section 2, we present the discrete ordinate setting in which the matrix exponential method is applied, while in Section 3, we discuss the eigendecomposition method for computing the matrix exponential. Section 4 is devoted the discrete ordinate method with matrix exponential. In Section 5, dealing with the matrix operator method with matrix exponential, we derive several representations of the reflection and transmission matrices for arbitrary optical thickness, as well as, for small and large optical thicknesses. In Section 6 we establish the matrix Riccati equations, prove the mathematical equivalence between the matrix Riccati equations method and the matrix exponential method in computing the reflection and transmission matrices of a homogeneous layer, and discuss some approximation solution methods for small values of the optical thickness and/or single scattering albedo. Finally, Section 7 contains some concluding remarks. Additional results dealing with a justification of the Gaussian quadrature in the discrete ordinate method, a review of eigendecomposition methods for computing the matrix exponential, and an extension of the analytical results to conservative scattering are presented in appendices.

2. Matrix formulation of the radiative transfer equation

For a given solar direction $\mathbf{\Omega}_0 = (-\mu_0, \varphi_0)$, with $\mu_0 > 0$ being the cosine of the solar zenith angle and φ_0 the solar azimuthal angle, the equation describing the radiative transfer in a plane-parallel homogeneous layer of optical thickness $\overline{\tau}$ is

$$\begin{aligned}\mu \frac{\mathrm{d}I_{\mathrm{d}}(\tau,\mu,-\mu_0,\varphi-\varphi_0)}{\mathrm{d}\tau} \\ = I_{\mathrm{d}}(\tau,\mu,-\mu_0,\varphi-\varphi_0) - \frac{\omega}{4\pi}F_0 p(\mu,-\mu_0,\varphi-\varphi_0)\mathrm{e}^{-\tau/\mu_0} \\ - \frac{\omega}{4\pi}\int_0^{2\pi}\int_{-1}^{1} p(\mu,\mu',\varphi-\varphi')I_{\mathrm{d}}(\tau,\mu',-\mu_0\varphi'-\varphi_0)\mathrm{d}\mu'\mathrm{d}\varphi', \quad (1)\end{aligned}$$

where $I_{\mathrm{d}}(\tau,\mu,-\mu_0,\varphi-\varphi_0)$ is the diffuse radiance at optical depth τ along the direction specified by the cosine of the zenith angle μ and the azimuthal angle φ, $p(\mu,\mu',\varphi-\varphi')$ is the scattering phase function for the radiation scattered from the direction $\mathbf{\Omega}' = (\mu',\varphi')$ into the direction $\mathbf{\Omega} = (\mu,\varphi)$, ω is the single scattering albedo, and F_0 is the solar flux. For simplicity, the thermal emission term is neglected in equation (1). If the homogeneous layer is placed in a multi-layered atmosphere at optical depth τ_0, the radiative transfer equation for the diffuse radiance e $I_{\mathrm{d}}(\tau_0+\tau,\mu,-\mu_0,\varphi-\varphi_0)$ contains the direct transmission term $\exp[-(\tau_0+\tau)/\mu_0]$ instead of $\exp(-\tau/\mu_0)$. The total radiance, defined in terms of the diffuse and direct radiances by

$$I(\tau,\mu,-\mu_0,\varphi-\varphi_0) = I_{\mathrm{d}}(\tau,\mu,-\mu_0,\varphi-\varphi_0) + I_{\odot}(\tau,\mu,-\mu_0,\varphi-\varphi_0)$$

solves the radiative transfer equation (1) without the single scattering source term. Note that for the direct radiance $I_{\odot}$,

we have

$$I_\odot(\tau,\mu,-\mu_0,\varphi-\varphi_0)=F_0\delta(\mu+\mu_0)\delta(\varphi-\varphi_0)\mathrm{e}^{-\tau/\mu_0},$$

and

$$\mu\frac{\mathrm{d}I_\odot}{\mathrm{d}\tau}(\tau,\mu,-\mu_0,\varphi-\varphi_0)=I_\odot(\tau,\mu,-\mu_0,\varphi-\varphi_0).$$

Considering the Fourier cosine expansions for the phase function

$$p(\mu,\mu',\varphi-\varphi')=\sum_{m=0}^{M_{\max}}(2-\delta_{m0})\,p_m(\mu,\mu')\cos[m(\varphi-\varphi')], \tag{2}$$

and the diffuse radiance

$$I_\mathrm{d}(\tau,\mu,-\mu_0,\varphi-\varphi_0)=\sum_{m=0}^{M_{\max}}I_{\mathrm{d}m}(\tau,\mu,-\mu_0)\cos[m(\varphi-\varphi_0)], \tag{3}$$

where $M_{\max}$ is the number of azimuthal modes in the expansions, yields the following radiative transfer equation for the individual azimuthal components of the radiance:

$$\begin{aligned}\mu\frac{\mathrm{d}I_{\mathrm{d}m}(\tau,\mu,-\mu_0)}{\mathrm{d}\tau}&=I_{\mathrm{d}m}(\tau,\mu,-\mu_0).\\&-(2-\delta_{m0})\frac{\omega}{4\pi}F_0\,p_m(\mu,-\mu_0)\mathrm{e}^{-\tau/\mu_0}\\&-\frac{1}{2}\omega\int_{-1}^{1}p_m(\mu,\mu')I_{\mathrm{d}m}(\tau,\mu',-\mu_0)\mathrm{d}\mu'.\end{aligned} \tag{4}$$

To simplify notations, hereafter the index m will be suppressed with the dependence on azimuthal mode assumed.

In order to deal with (4) we replace the integral by a symmetric quadrature rule with $2N$ nodes and weights, i.e., if μ_k with $k=1,\ldots,N$, is a node associated with the weight w_k, then $-\mu_k$ is also a node associated with the same weight. Usually, the quadrature is chosen to be (double) Gaussian, in which case the number of azimuthal modes is $M_{\max}\leq 2N-1$. This result is discussed in Appendix 1. In the discrete ordinate space, the radiative transfer equation for the diffuse radiance vector $\mathbf{i}_\mathrm{d}=[\mathbf{i}_\mathrm{d}^+,\mathbf{i}_\mathrm{d}^-]^T$ with $\mathbf{i}_\mathrm{d}^\pm=[I_{\mathrm{d}m}(\pm\mu_k,-\mu_0)]$, $k=1,\ldots,N$, reads as

$$\frac{\mathrm{d}\mathbf{i}_\mathrm{d}}{\mathrm{d}\tau}(\tau)=-\mathbf{A}\mathbf{i}_\mathrm{d}(\tau)-\mathrm{e}^{-\tau/\mu_0}\mathbf{b},\ 0\leq\tau\leq\overline{\tau}. \tag{5}$$

The entries of the layer matrix

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_{11}&\mathbf{A}_{12}\\-\mathbf{A}_{12}&-\mathbf{A}_{11}\end{bmatrix}, \tag{6}$$

are

$$\begin{aligned}\mathbf{A}_{11}&=\mathbf{M}\mathbf{S}_+\mathbf{W}-\mathbf{M},\\\mathbf{A}_{12}&=\mathbf{M}\mathbf{S}_-\mathbf{W},\end{aligned} \tag{7}$$

while the entries of the layer vector

$$\mathbf{b}=\begin{bmatrix}\mathbf{b}_1\\\mathbf{b}_2\end{bmatrix}, \tag{8}$$

are

$$\begin{aligned}\mathbf{b}_1&=\mathbf{M}\mathbf{b}_+,\\\mathbf{b}_2&=-\mathbf{M}\mathbf{b}_-,\end{aligned} \tag{9}$$

where

$$[\mathbf{S}_\pm]_{kl}=\frac{1}{2}\omega p_m(\mu_k,\pm\mu_l), \tag{10}$$

$$[\mathbf{W}]_{kl}=w_k\delta_{kl}, \tag{11}$$

$$[\mathbf{M}]_{kl}=\frac{1}{\mu_k}\delta_{kl}, \tag{12}$$

$$[\mathbf{b}_\pm]_k=(2-\delta_{m0})\frac{F_0}{4\pi}\omega p_m(\pm\mu_k,-\mu_0), \tag{13}$$

for $k,l=1,\ldots,N$. Here, δ_{kl} is the Kronecker symbol.

In order to reduce the eigenvalue decomposition of $\mathbf{A}$ from a general to a symmetric problem, we define the scaled diffuse radiance vector $\widehat{\mathbf{i}}_\mathrm{d}=[\widehat{\mathbf{i}}_\mathrm{d}^+,\widehat{\mathbf{i}}_\mathrm{d}^-]^T$, through the relation

$$\widehat{\mathbf{i}}_\mathrm{d}^\pm=\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}\mathbf{i}_\mathrm{d}^\pm. \tag{14}$$

Hereafter, the "hat" symbol on vectors and matrices refers to scaled quantities. For the scaled diffuse radiance vector, the radiative transfer equation is

$$\frac{\mathrm{d}\widehat{\mathbf{i}}_\mathrm{d}}{\mathrm{d}\tau}(\tau)=-\widehat{\mathbf{A}}\,\widehat{\mathbf{i}}_\mathrm{d}(\tau)-\mathrm{e}^{-\tau/\mu_0}\,\widehat{\mathbf{b}},\ 0\leq\tau\leq\overline{\tau}, \tag{15}$$

where the expressions of the scaled layer matrix $\widehat{\mathbf{A}}$ and the layer vector $\widehat{\mathbf{b}}$ are as in (6) and (8), respectively, with

$$\begin{aligned}\widehat{\mathbf{A}}_{11}&=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}(\mathbf{S}_+-\mathbf{W}^{-1})\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}},\\\widehat{\mathbf{A}}_{12}&=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}\mathbf{S}_-\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}},\end{aligned} \tag{16}$$

and

$$\begin{aligned}\widehat{\mathbf{b}}_1&=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}\mathbf{b}_+,\\\widehat{\mathbf{b}}_2&=-\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}\mathbf{b}_-.\end{aligned} \tag{17}$$

From the principle of reciprocity of the phase function, it follows that $\mathbf{S}_+$ and $\mathbf{S}_-$, and so, that $\widehat{\mathbf{A}}_{11}$ and $\widehat{\mathbf{A}}_{12}$ are symmetric matrices. The scaling procedure (14), which is equivalent to the application of a similarity transformation to $\mathbf{A}$ with the diagonal block matrices $\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}$, is standard in radiative transfer and has been used by Waterman [21], Nakajima and Tanaka [18], and Stamnes and Swanson [14].

In the framework of the matrix exponential approach, the solution of the initial value problem consisting in the vector differential equation (15) and the initial condition $\widehat{\mathbf{i}}_\mathrm{d}(0)$, is given by

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau)=\mathrm{e}^{-\widehat{\mathbf{A}}\tau}\,\widehat{\mathbf{i}}_\mathrm{d}(0)-\int_0^\tau\mathrm{e}^{-\widehat{\mathbf{A}}(\tau-\tau')}\mathrm{e}^{-\tau'/\mu_0}\,\widehat{\mathbf{b}}\,\mathrm{d}\tau'. \tag{18}$$

Let us give an interpretation of the matrix exponential solution (18). Making use of a spectral decomposition of the matrix $\widehat{\mathbf{A}}$, it can be shown that

$$\int_0^\tau\mathrm{e}^{-\widehat{\mathbf{A}}(\tau-\tau')}\mathrm{e}^{-\tau'/\mu_0}\,\widehat{\mathbf{b}}\,\mathrm{d}\tau'=(\mathrm{e}^{-\widehat{\mathbf{A}}\tau}-\mathrm{e}^{-\tau/\mu_0}\mathbf{I})\mu_0(\mathbf{I}-\mu_0\widehat{\mathbf{A}})^{-1}\,\widehat{\mathbf{b}}, \tag{19}$$

whence, setting

$$\widehat{\mathbf{c}} = \mu_0(\mathbf{I} - \mu_0\widehat{\mathbf{A}})^{-1}\,\widehat{\mathbf{b}}, \tag{20}$$

we express (18) as

$$\widehat{\mathbf{i}}_{\mathrm{d}}(\tau) = \mathrm{e}^{-\widehat{\mathbf{A}}\tau}\,[\widehat{\mathbf{i}}_{\mathrm{d}}(0) - \widehat{\mathbf{c}}] + \mathrm{e}^{-\tau/\mu_0}\,\widehat{\mathbf{c}}. \tag{21}$$

The classical approach for solving the vector differential equation (15) is to express the general solution as the sum of a homogeneous and a particular solution, i.e.,

$$\widehat{\mathbf{i}}_{\mathrm{d}}(\tau) = \widehat{\mathbf{i}}_{\mathrm{h}}(\tau) + \widehat{\mathbf{i}}_{\mathrm{p}}(\tau). \tag{22}$$

The particular integral solving (15), is

$$\widehat{\mathbf{i}}_{\mathrm{p}}(\tau) = \mathrm{e}^{-\tau/\mu_0}\,\widehat{\mathbf{c}}, \tag{23}$$

where $\widehat{\mathbf{c}}$ is given by (20). The homogeneous or the fundamental solution solves the equation

$$\frac{\mathrm{d}\widehat{\mathbf{i}}_{\mathrm{h}}}{\mathrm{d}\tau}(\tau) = -\widehat{\mathbf{A}}\,\widehat{\mathbf{i}}_{\mathrm{h}}(\tau), \tag{24}$$

and is given by

$$\widehat{\mathbf{i}}_{\mathrm{h}}(\tau) = \mathrm{e}^{-\widehat{\mathbf{A}}\tau}\,\widehat{\mathbf{c}}_{\mathrm{h}}. \tag{25}$$

The integration vector $\widehat{\mathbf{c}}_{\mathrm{h}}$ is obtained from (22) and the initial condition $\widehat{\mathbf{i}}_{\mathrm{d}}(0)$; the result is $\widehat{\mathbf{c}}_{\mathrm{h}} = \widehat{\mathbf{i}}_{\mathrm{d}}(0) - \widehat{\mathbf{c}}$. Then, accounting of (22), (23), and (25) we find that an equivalent representation of (21) is

$$\widehat{\mathbf{i}}_{\mathrm{d}}(\tau) = \mathrm{e}^{-\widehat{\mathbf{A}}\tau}\,[\widehat{\mathbf{i}}_{\mathrm{d}}(0) - \widehat{\mathbf{i}}_{\mathrm{p}}(0)] + \widehat{\mathbf{i}}_{\mathrm{p}}(\tau). \tag{26}$$

If the incident direction μ_0 coincides with a discrete ordinate direction, say $\mu_0 = \mu_l$ for some l, then the homogeneous and particular solutions can be interpreted as the total and direct radiance vectors, respectively. To show this, let us define the vector $\widehat{\mathbf{i}}_{\odot}(0) = [\widehat{\mathbf{i}}^{+}_{\odot}(0), \widehat{\mathbf{i}}^{-}_{\odot}(0)]^T$ by $\widehat{\mathbf{i}}^{+}_{\odot}(0) = 0$ and

$$[\widehat{\mathbf{i}}^{-}_{\odot}(0)]_k = (2 - \delta_{m0})\frac{F_0}{2\pi}\sqrt{\frac{\mu_k}{w_k}}\delta_{kl},\ \ k = 1, \ldots, N.$$

By straightforward calculation it can be shown that $\widehat{\mathbf{i}}_{\odot}(0) = -\widehat{\mathbf{c}}$. As a result, the particular solution $\widehat{\mathbf{i}}_{\mathrm{p}}(\tau)$ given by (23) can be identified with the direct radiance vector $\widehat{\mathbf{i}}_{\odot}(\tau)$, defined by $\widehat{\mathbf{i}}_{\odot}(\tau) = \mathrm{e}^{-\tau/\mu_0}\widehat{\mathbf{i}}_{\odot}(0)$, i.e., $\widehat{\mathbf{i}}_{\odot}(\tau) = -\widehat{\mathbf{i}}_{\mathrm{p}}(\tau)$, and so, the total radiance vector $\widehat{\mathbf{i}}(\tau)$, defined by $\widehat{\mathbf{i}}(\tau) = \widehat{\mathbf{i}}_{\mathrm{d}}(\tau) + \widehat{\mathbf{i}}_{\odot}(\tau)$, can be identified with the homogeneous solution $\widehat{\mathbf{i}}_{\mathrm{h}}(\tau)$. Note that both $\widehat{\mathbf{i}}_{\odot}$ and $\widehat{\mathbf{i}}_{\mathrm{p}}$ solve the differential equation $\mu_0\mathrm{d}\widehat{\mathbf{i}}_{\odot}(\tau)/\mathrm{d}\tau = -\widehat{\mathbf{i}}_{\odot}(\tau)$, while both $\widehat{\mathbf{i}}$ and $\widehat{\mathbf{i}}_{\mathrm{h}}$ solve the differential equation (24). It should be pointed out that in a continuous setting, the total radiance is a generalized function, or a distribution, while in a discrete setting and under the above assumption, the total radiance, regarded as a function μ_k, has a jump at $\mu_k = \mu_l (= \mu_0)$.

In [22], the matrix exponential $\exp(-\widehat{\mathbf{A}}\tau)$, reflecting the internal properties of the homogeneous medium, is called propagator and is denoted by $\mathbf{P}(\tau)$, i.e., $\mathbf{P}(\tau) = \exp(-\widehat{\mathbf{A}}\tau)$. If the initial condition is given, then the solution deeper in the medium can be recovered (propagated) down from the upper boundary by applying this propagator. However, the obtained solution has no physical meaning, as long as the radiative transfer equation cannot be treated as an initial value problem. The initial condition means that both sets of upward and downward radiances at the upper boundary are known, a fact which typically does not occur in atmospheric radiative transfer. This by no means reduces the usefulness of matrix exponential (propagator), as it will be demonstrated in the course of our analysis.

3. Eigendecomposition method for computing the matrix exponential

The matrix exponential can be computed by using an eigendecomposition of the matrix $\widehat{\mathbf{A}}$. Exploiting the block symmetry of $\widehat{\mathbf{A}}$, we find

$$\widehat{\mathbf{A}} = \widehat{\mathbf{V}}\begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda} \end{bmatrix}\widehat{\mathbf{V}}^{-1}, \tag{27}$$

with

$$\widehat{\mathbf{V}} = \begin{bmatrix} \widehat{\mathbf{V}}_{+} & \widehat{\mathbf{V}}_{-} \\ \widehat{\mathbf{V}}_{-} & \widehat{\mathbf{V}}_{+} \end{bmatrix} \tag{28}$$

and (the abbreviation 'not' stands for notation)

$$\Lambda = \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \overset{\mathrm{not}}{=} [\lambda_k]. \tag{29}$$

The spectral decomposition of the matrix $\widehat{\mathbf{A}}$ can be obtained by one of the following methods: direct decomposition of an asymmetric matrix [14], square-root decomposition [18], and Cholesky decomposition [30]. These approaches are summarized in Appendix 2. In (28), the matrices $\widehat{\mathbf{V}}_{\pm}$ are of the form

$$\widehat{\mathbf{V}}_{\pm} = [\widehat{\mathbf{v}}^{\pm}_1, \ldots, \widehat{\mathbf{v}}^{\pm}_N] \overset{\mathrm{not}}{=} [\widehat{\mathbf{v}}^{\pm}_k],$$

where $\begin{bmatrix} \widehat{\mathbf{v}}^{+}_k \\ \widehat{\mathbf{v}}^{-}_k \end{bmatrix}$ are the right eigenvectors of $\widehat{\mathbf{A}}$ corresponding to λ_k, and $\begin{bmatrix} \widehat{\mathbf{v}}^{-}_k \\ \widehat{\mathbf{v}}^{+}_k \end{bmatrix}$ are the right eigenvectors of $\widehat{\mathbf{A}}$ corresponding to $-\lambda_k$. The matrix exponential is then given by

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} = \widehat{\mathbf{V}}\begin{bmatrix} \mathbf{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}(-\tau) \end{bmatrix}\widehat{\mathbf{V}}^{-1}, \tag{30}$$

with

$$\mathbf{\Gamma}(\tau) = [\mathrm{e}^{-\lambda_k\tau}]. \tag{31}$$

From (30) it is apparent that the computation of the matrix exponential requires the computation of the inverse of the right eigenvectors matrix $\widehat{\mathbf{V}}$. In Waterman's approach, the inverse $\widehat{\mathbf{V}}^{-1}$ is computed by using the following result: For any matrix $\mathbf{A}$, which has a complete set of linearly independent eigenvectors, the inverse of the right eigenvector matrix is the transpose of the left eigenvector matrix. Indeed, let $\mathbf{A}$ be an

$n \times n$ matrix with distinct eigenvalues, $\mathbf{x}_k$ be a right eigenvector of $\mathbf{A}$ corresponding to λ_k, i.e., $\mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{x}_k$, and $\mathbf{y}_l$ be a left eigenvector of $\mathbf{A}$ corresponding to λ_l, i.e., $\mathbf{A}^T\mathbf{y}_l = \lambda_l \mathbf{y}_l$. Then, from $\mathbf{y}_l^T\mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{y}_l^T\mathbf{x}_k = \lambda_l \mathbf{y}_l^T\mathbf{x}_k$, we see that for $\lambda_k \neq \lambda_l$, we have $\mathbf{y}_l^T\mathbf{x}_k = 0$. Moreover, assuming that $\mathbf{x}_k$ and $\mathbf{y}_k$ are normalized in the sense that $\mathbf{y}_k^T\mathbf{x}_k = 1$, $k = 1, ..., n$, we find that $\mathbf{X}^{-1} = \mathbf{Y}^T$, where $\mathbf{X} = [\mathbf{x}_k]$ and $\mathbf{Y} = [\mathbf{y}_k]$. Thus, the spectral decomposition of $\mathbf{A}$ reads as $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1} = \mathbf{X}\mathbf{\Lambda}\mathbf{Y}^T$.

Accounting of (27)-(29), we deduce that the systems of normalized right and left eigenvectors corresponding to λ_k and $-\lambda_k$ are

$$\left(\lambda_k, \frac{1}{\sqrt{|a_k|}}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}, \frac{\mathrm{sign}(a_k)}{\sqrt{|a_k|}}\begin{bmatrix} -\widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}\right) \tag{32}$$

and

$$\left(-\lambda_k, \frac{1}{\sqrt{|a_k|}}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}, \frac{\mathrm{sign}(a_k)}{\sqrt{|a_k|}}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ -\widehat{\mathbf{v}}_k^+ \end{bmatrix}\right), \tag{33}$$

respectively, with $a_k = ||\widehat{\mathbf{v}}_k^-||^2 - ||\widehat{\mathbf{v}}_k^+||^2$. To simplify notations we put

$$\frac{1}{\sqrt{|a_k|}}\widehat{\mathbf{v}}_k^\pm \to \widehat{\mathbf{v}}_k^\pm \text{ and } \overline{\mathbf{v}}_k^\pm = \mathrm{sign}(a_k)\widehat{\mathbf{v}}_k^\pm, \tag{34}$$

so that

$$\left(\lambda_k, \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}, \begin{bmatrix} -\overline{\mathbf{v}}_k^+ \\ \overline{\mathbf{v}}_k^- \end{bmatrix}\right) \tag{35}$$

and

$$\left(-\lambda_k, \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}, \begin{bmatrix} \overline{\mathbf{v}}_k^- \\ -\overline{\mathbf{v}}_k^+ \end{bmatrix}\right) \tag{36}$$

are the systems of normalized right and left eigenvectors corresponding to λ_k and $-\lambda_k$, respectively. Thus, we have $\widehat{\mathbf{V}}^{-1} = \overline{\mathbf{V}}^T$, with

$$\overline{\mathbf{V}} = \begin{bmatrix} -\overline{\mathbf{V}}_+ & \overline{\mathbf{V}}_- \\ \overline{\mathbf{V}}_- & -\overline{\mathbf{V}}_+ \end{bmatrix}, \tag{37}$$

$$\overline{\mathbf{V}}_\pm = [\overline{\mathbf{v}}_k^\pm]. \tag{38}$$

The spectral decomposition of $\widehat{\mathbf{A}}$ is then

$$\widehat{\mathbf{A}} = \widehat{\mathbf{V}}\begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda} \end{bmatrix}\overline{\mathbf{V}}^T, \tag{39}$$

or explicitly,

$$\widehat{\mathbf{A}} = \sum_{k=1}^{N} \lambda_k \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}\begin{bmatrix} -\overline{\mathbf{v}}_k^+ \\ \overline{\mathbf{v}}_k^- \end{bmatrix}^T - \lambda_k \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}\begin{bmatrix} \overline{\mathbf{v}}_k^- \\ -\overline{\mathbf{v}}_k^+ \end{bmatrix}^T, \tag{40}$$

while the matrix exponential is

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} = \widehat{\mathbf{V}}\begin{bmatrix} \mathbf{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}(-\tau) \end{bmatrix}\overline{\mathbf{V}}^T, \tag{41}$$

or explicitly,

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} = \sum_{k=1}^{N} \mathrm{e}^{-\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}\begin{bmatrix} -\overline{\mathbf{v}}_k^+ \\ \overline{\mathbf{v}}_k^- \end{bmatrix}^T + \mathrm{e}^{\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}\begin{bmatrix} \overline{\mathbf{v}}_k^- \\ -\overline{\mathbf{v}}_k^+ \end{bmatrix}^T. \tag{42}$$

A short comment is in order. In the absence of scattering ($\omega = 0$), we have $\widehat{\mathbf{A}}_{11} = -\widehat{\mathbf{A}}_{22} = -\mathbf{M}$, and $\widehat{\mathbf{A}}_{12} = \widehat{\mathbf{A}}_{21} = 0$. As a result, we obtain $\widehat{\mathbf{V}}_+ = 0$, $\widehat{\mathbf{V}}_- = \mathbf{I}$, and $\mathbf{\Lambda} = \mathbf{M}$; thus, the eigenvalues are the inverse of the discrete ordinates, i.e., $\lambda_k = 1/\mu_k$. The matrix exponential is the diagonal matrix $\exp(\widehat{\mathbf{A}}\tau) = [\mathrm{e}^{\lambda_k\tau}; \mathrm{e}^{-\lambda_k\tau}]$ and the homogeneous solution at $\tau = \overline{\tau}$, given by $\widehat{\mathbf{i}}_\mathrm{h}(\overline{\tau}) = [\mathrm{e}^{\lambda_k\overline{\tau}}; \mathrm{e}^{-\lambda_k\overline{\tau}}]\widehat{\mathbf{i}}_\mathrm{h}(0)$, is a representation of the Beer-Lambert attenuation law for the downward and upward radiances, i.e., $[\widehat{\mathbf{i}}_\mathrm{h}^-(\overline{\tau})]_k = \exp(-\lambda_k\overline{\tau})[\widehat{\mathbf{i}}_\mathrm{h}^-(0)]_k$ and $[\widehat{\mathbf{i}}_\mathrm{h}^+(0)]_k = \exp(-\lambda_k\overline{\tau})[\widehat{\mathbf{i}}_\mathrm{h}^+(\overline{\tau})]_k$, respectively. If scattering is present, the Beer-Lambert law is still valid but for the downward and upward radiances $\boldsymbol{\xi}_-(\overline{\tau})$ and $\boldsymbol{\xi}_-(0)$ corresponding to the transformed radiance vector

$$\boldsymbol{\xi}(\tau) = \begin{bmatrix} \boldsymbol{\xi}_+(\tau) \\ \boldsymbol{\xi}_-(\tau) \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix}\overline{\mathbf{V}}^T\widehat{\mathbf{i}}_\mathrm{h}(\tau).$$

We proceed now to derive some matrix identities which will be frequently used in the following. In terms of block matrices, the orthogonality relation $\widehat{\mathbf{V}}\overline{\mathbf{V}}^T = \mathbf{I}_{2N}$, where $\mathbf{I}_{2N}$ is the identity matrix of dimension $2N \times 2N$, reads as

$$\begin{bmatrix} \widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_- \\ \widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+ \end{bmatrix}\begin{bmatrix} -\overline{\mathbf{V}}_+ & \overline{\mathbf{V}}_- \\ \overline{\mathbf{V}}_- & -\overline{\mathbf{V}}_+ \end{bmatrix}^T = \mathbf{I}_{2N}, \tag{43}$$

from which we infer that

$$\widehat{\mathbf{V}}_-\overline{\mathbf{V}}_-^T - \widehat{\mathbf{V}}_+\overline{\mathbf{V}}_+^T = \mathbf{I}_N, \tag{44}$$

$$\widehat{\mathbf{V}}_+\overline{\mathbf{V}}_-^T = \widehat{\mathbf{V}}_-\overline{\mathbf{V}}_+^T. \tag{45}$$

Similarly, from $\overline{\mathbf{V}}^T\widehat{\mathbf{V}} = \mathbf{I}_{2N}$, we obtain

$$\overline{\mathbf{V}}_-^T\widehat{\mathbf{V}}_- - \overline{\mathbf{V}}_+^T\widehat{\mathbf{V}}_+ = \mathbf{I}_N, \tag{46}$$

$$\overline{\mathbf{V}}_-^T\widehat{\mathbf{V}}_+ = \overline{\mathbf{V}}_+^T\widehat{\mathbf{V}}_-. \tag{47}$$

Accounting of (44)-(47), the following matrix identities readily follow:

$$\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T} = \widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+, \tag{48}$$

$$\overline{\mathbf{V}}_-^{-T}\overline{\mathbf{V}}_+^T = \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1}, \tag{49}$$

$$\overline{\mathbf{V}}_-^{-T} = \widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+. \tag{50}$$

On the other hand, from (34) and (38), we see that

$$\overline{\mathbf{V}}_\pm = \widehat{\mathbf{V}}_\pm\mathbb{S}, \tag{51}$$

where $\mathbb{S} = [\mathrm{sign}(a_k)]$ is a diagonal matrix of plus and minus ones. As $\mathbb{S}\mathbb{S} = \mathbf{I}_N$, we find that the matrices $\widehat{\mathbf{V}}_+^T\widehat{\mathbf{V}}_-$ and $\widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1}$ are symmetric, i.e.,

$$(\widehat{\mathbf{V}}_+^T\widehat{\mathbf{V}}_-)^T = \widehat{\mathbf{V}}_+^T\widehat{\mathbf{V}}_-, \tag{52}$$

$$(\widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1})^T = \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1}, \tag{53}$$

but $\widehat{\mathbf{V}}_-\widehat{\mathbf{V}}_+^T$ and $\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+$ are not; in particular, we have

$$(\widehat{\mathbf{V}}_-\mathbb{S}\widehat{\mathbf{V}}_+^T)^T = \widehat{\mathbf{V}}_-\mathbb{S}\widehat{\mathbf{V}}_+^T, \tag{54}$$

$$(\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+)^T = \mathbb{S}(\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+)\mathbb{S}. \tag{55}$$

Solution (18) with the matrix exponential as in (41) is the starting point in our analysis. This solution, in the form of the propagator $\mathbf{P}(\tau) = \exp(-\widehat{\mathbf{A}}\tau)$, is a combination of both growing and decaying exponentials. For large optical thicknesses, growing exponentials will dominate the solution given in this form, and the direct application of the propagator leads to numerical instability. This behavior is referred to as dichotomic [22]. However, by appropriate manipulations of (18) and by introducing scaling transformations, computationally stable equations, the so-called layer equation and the interaction principle equation can be derived. These equations, which are the quintessence of the discrete ordinate and matrix operator method with matrix exponential, are discussed in Sections 4 and 5, respectively.

4. Discrete ordinate method with matrix exponential

In the framework of the discrete ordinate method with matrix exponential, the layer equation is a computationally stable relation connecting the layer-top radiance vector $\widehat{\mathbf{i}}_\mathrm{d}(0)$ and the layer-bottom radiance vector $\widehat{\mathbf{i}}_\mathrm{d}(\tau)$ [25, 26]. For a multi-layered atmosphere, each layer equation is assembled into the system matrix of the entire atmosphere. By imposing appropriate boundary conditions at the top and the bottom of the atmosphere, the system of equations is solved for the level values of the radiances. Thus, the method avoids computing an explicit solution for each layer by imposing boundary conditions for the entire atmosphere, as well as, the continuity condition for the radiances across the layer interfaces.

The layer equation is derived by inserting the matrix exponential representation (41) in (18), and by multiplying the resulting equation with an appropriate scaling matrix as in [31]. The result is

$$\mathbf{D}_1(\overline{\tau})\overline{\mathbf{V}}^T\widehat{\mathbf{i}}_\mathrm{d}(\overline{\tau}) = \mathbf{D}_0(\overline{\tau})\overline{\mathbf{V}}^T\widehat{\mathbf{i}}_\mathrm{d}(0) - \mathbf{D}_\mathrm{b}(\overline{\tau})\overline{\mathbf{V}}^T\widehat{\mathbf{b}}, \tag{56}$$

where

$$\mathbf{D}_1(\overline{\tau}) = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}, \; \mathbf{D}_0(\overline{\tau}) = \begin{bmatrix} \boldsymbol{\Gamma}(\overline{\tau}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix},$$
$$\mathbf{D}_\mathrm{b}(\overline{\tau}) = \begin{bmatrix} \frac{\mathrm{e}^{-\lambda_k\overline{\tau}} - \mathrm{e}^{-\overline{\tau}/\mu_0}}{1/\mu_0 - \lambda_k} & \mathbf{0} \\ \mathbf{0} & \frac{1-\mathrm{e}^{-\overline{\tau}(\lambda_k+1/\mu_0)}}{1/\mu_0+\lambda_k} \end{bmatrix}. \tag{57}$$

If the level values of the radiances $\widehat{\mathbf{i}}_\mathrm{d}(0)$ and $\widehat{\mathbf{i}}_\mathrm{d}(\overline{\tau})$ are known, the radiance at an internal point τ, with $0 \le \tau \le \overline{\tau}$, is computed as

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau) = \widehat{\mathbf{V}}\mathbf{E}(\tau,\overline{\tau}) \begin{bmatrix} \boldsymbol{\xi}_+(0) \\ \boldsymbol{\xi}_-(\overline{\tau}) \end{bmatrix} - \widehat{\mathbf{V}}\mathbf{E}_\mathrm{b}(\tau,\overline{\tau})\boldsymbol{\eta}, \tag{58}$$

where now $\boldsymbol{\xi}(\tau) = [\boldsymbol{\xi}_+(\tau), \boldsymbol{\xi}_-(\tau)]^T = \overline{\mathbf{V}}^T\widehat{\mathbf{i}}_\mathrm{d}(\tau)$, $\boldsymbol{\eta} = \overline{\mathbf{V}}^T\widehat{\mathbf{b}}$, and

$$\mathbf{E}(\tau,\overline{\tau}) = \begin{bmatrix} \boldsymbol{\Gamma}(\tau) & 0 \\ 0 & \boldsymbol{\Gamma}(\overline{\tau}-\tau) \end{bmatrix},$$
$$\mathbf{E}_\mathrm{b}(\tau,\overline{\tau}) = \begin{bmatrix} \frac{\mathrm{e}^{-\lambda_k\tau} - \mathrm{e}^{-\tau/\mu_0}}{1/\mu_0 - \lambda_k} & \mathbf{0} \\ \mathbf{0} & -\mathrm{e}^{-\tau/\mu_0}\frac{1-\mathrm{e}^{-(\overline{\tau}-\tau)(\lambda_k+1/\mu_0)}}{1/\mu_0+\lambda_k} \end{bmatrix}. \tag{59}$$

The matrix exponential representation of the solution as given by (18) is mathematically equivalent to the classical Chandrasekhar's representation in terms of the characteristic solutions

$$\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} \text{ and } \mathrm{e}^{\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}. \tag{60}$$

To show this equivalences, we consider (20), i.e.,

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau) = \mathrm{e}^{-\widehat{\mathbf{A}}\tau}[\widehat{\mathbf{i}}_\mathrm{d}(0) - \widehat{\mathbf{i}}_\mathrm{p}(0)] + \widehat{\mathbf{i}}_\mathrm{p}(\tau), \tag{61}$$

for $0 \le \tau \le \overline{\tau}$. Using (30) and writing

$$\begin{bmatrix} \boldsymbol{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(-\tau) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(-\tau) \end{bmatrix} \times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}^{-1}, \tag{62}$$

we obtain

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau) = \begin{bmatrix} \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\tau) & \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau}-\tau) \\ \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\tau) & \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau}-\tau) \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \widehat{\mathbf{i}}_\mathrm{p}(\tau), \tag{63}$$

where the N-dimensional vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ do not depend on τ, and are given by

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau}) \\ \widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}^{-1}[\widehat{\mathbf{i}}_\mathrm{d}(0) - \widehat{\mathbf{i}}_\mathrm{p}(0)]. \tag{64}$$

The explicit form of (63), i.e.,

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau) = \sum_{k=1}^{N}\alpha_k\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} + \beta_k\mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix} + \widehat{\mathbf{i}}_\mathrm{p}(\tau), \tag{65}$$

is the solution representation in the Chandrasekhar's discrete ordinate method. Another representation can be obtained by using the relation

$$\begin{bmatrix} \boldsymbol{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(-\tau) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(-\tau) \end{bmatrix} \times \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ \boldsymbol{\Gamma}(\overline{\tau}) & \boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}\begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ \boldsymbol{\Gamma}(\overline{\tau}) & \boldsymbol{\Gamma}(\overline{\tau}) \end{bmatrix}^{-1}, \tag{66}$$

which yields

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau) = \begin{bmatrix} \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau}-\tau) + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\tau) & \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau}-\tau) - \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\tau) \\ \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau}-\tau) + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\tau) & \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau}-\tau) - \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\tau) \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \widehat{\mathbf{i}}_\mathrm{p}(\tau), \tag{67}$$

with

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau})+\widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_-\boldsymbol{\Gamma}(\overline{\tau})-\widehat{\mathbf{V}}_+ \\ \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau})+\widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}(\overline{\tau})-\widehat{\mathbf{V}}_- \end{bmatrix}^{-1} [\widehat{\mathbf{i}}_{\mathrm{d}}(0)-\widehat{\mathbf{i}}_{\mathrm{p}}(0)]. \quad (68)$$

From (67), we get

$$\widehat{\mathbf{i}}_{\mathrm{d}}(\tau) = \sum_{k=1}^{N} \alpha_k \Big(\mathrm{e}^{-\lambda_k \tau} \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)} \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix} \Big) + \beta_k \Big(-\mathrm{e}^{-\lambda_k \tau} \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)} \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix} \Big) + \hat{\mathbf{i}}_{\mathrm{p}}(\tau), \quad (69)$$

where as before, the integration constants α_k and β_k do not depend on τ.

Equations (63)-(64) and (67)-(68) are equivalent solution representations in the matrix exponential method and will be used in the next section to derive the reflection and transmission matrices. In both representations, the radiance vector at optical depth τ is a superposition of eigenfields propagating from the upper and lower boundaries, with the attenuation factors $\exp(-\lambda_k\tau)$ and $\exp[-\lambda_k(\overline{\tau}-\tau)]$, respectively.

Although the classical and the matrix exponential version of the discrete ordinate method are very similar, several differences can be emphasized:

1. In the classical discrete ordinate method, the expansion coefficients α_k and β_k are the unknowns of the discretized radiative transfer problem and are computed by imposing the continuity condition for the radiances across the layer interfaces. In the discrete ordinate method with matrix exponential, the unknowns are the level values of the radiances. Once they are computed, the integration constants can be obtained from (64) or (68).
2. In the classical discrete ordinate method, the computation of the particular solution requires the computation of the inverse $(\mathbf{I}-\mu_0\widehat{\mathbf{A}})^{-1}$. This inversion step is not present in the discrete ordinate method with matrix exponential. However, if the systems of normalized right and left eigenvectors stay at our disposal, the constant vector $\widehat{\mathbf{c}}$ which enters in (23), can be calculated with a less computational effort as

$$\widehat{\mathbf{c}} = \mu_0 \widehat{\mathbf{V}} \begin{bmatrix} (\mathbf{I}-\mu_0\boldsymbol{\Lambda})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}+\mu_0\boldsymbol{\Lambda})^{-1} \end{bmatrix} \overline{\mathbf{V}}^T \widehat{\mathbf{b}} \quad (70)$$

for $\mu_0 \neq \lambda_k$, $k = 1, \ldots, N$.

5. Matrix operator method with matrix exponential

In a continuous setting, the interaction principle equation, which is the central feature of the matrix operator method, relates the outcoming radiances at the layer top $I_{\mathrm{d}}(0,\mu,-\mu_0,\varphi-\varphi_0)$ and layer bottom $I_{\mathrm{d}}(\overline{\tau},-\mu,-\mu_0,\varphi-\varphi_0)$ to the incoming radiances $I_{\mathrm{d}}(0,-\mu,-\mu_0,\varphi-\varphi_0)$ and $I_{\mathrm{d}}(\overline{\tau},\mu,-\mu_0,\varphi-\varphi_0)$ through the reflection, total transmission and diffuse transmission functions $R(\mu,\mu',\varphi-\varphi';\overline{\tau})$, $T(\mu,\mu',\varphi-\varphi';\overline{\tau})$ and $T_{\mathrm{d}}(\mu,\mu',\varphi-\varphi';\overline{\tau})$, respectively, where $\mu,\mu' > 0$. The transmission function for the diffuse radiance T_{d} is related to the transmission function for the total radiance by the relation

$$T_{\mathrm{d}}(\mu,\mu',\varphi-\varphi';\overline{\tau}) = T(\mu,\mu',\varphi-\varphi';\overline{\tau}) - \frac{\pi}{\mu'}\delta(\mu-\mu')\delta(\varphi-\varphi')\mathrm{e}^{-\overline{\tau}/\mu'}. \quad (71)$$

Considering the Fourier cosine expansions

$$I_{\mathrm{d}}(\overline{\tau},\mu,-\mu_0,\varphi-\varphi_0) = \sum_{m=0}^{M_{\max}} I_{\mathrm{d}m}(\overline{\tau},\mu,-\mu_0)\cos[m(\varphi-\varphi_0)], \quad (72)$$

$$X(\mu,\mu',\varphi-\varphi';\overline{\tau}) = \sum_{m=0}^{M_{\max}} (2-\delta_{m0}) X_m(\mu,\mu';\overline{\tau})\cos[m(\varphi-\varphi')], \quad (73)$$

where X stands for R, T, and T_{d}, and noting that

$$T_m(\mu,\mu';\overline{\tau}) = T_{\mathrm{d}m}(\mu,\mu';\overline{\tau}) + \frac{1}{2\mu'}\delta(\mu-\mu')\mathrm{e}^{-\overline{\tau}/\mu'}, \quad (74)$$

for any azimuthal mode m, yields the following representations of the interaction principle in the Fourier space:

$$I_{\mathrm{d}m}(0,\mu,-\mu_0) = (2-\delta_{m0})\frac{1}{\pi}\mu_0 F_0 R_m(\mu,\mu_0;\overline{\tau}) + 2\int_0^1 R_m(\mu,\mu';\overline{\tau}) I_{\mathrm{d}m}(0,-\mu',-\mu_0)\mu'\mathrm{d}\mu' + 2\int_0^1 T_m(\mu,\mu';\overline{\tau}) I_{\mathrm{d}m}(\overline{\tau},\mu',-\mu_0)\mu'\mathrm{d}\mu' \quad (75)$$

and

$$I_{\mathrm{d}m}(\overline{\tau},-\mu,-\mu_0) = (2-\delta_{m0})\frac{1}{\pi}\mu_0 F_0 T_{\mathrm{d}m}(\mu,\mu_0;\overline{\tau}) + 2\int_0^1 T_m(\mu,\mu';\overline{\tau}) I_{\mathrm{d}m}(0,-\mu',-\mu_0)\mu'\mathrm{d}\mu' + 2\int_0^1 R_m(\mu,\mu';\overline{\tau}) I_{\mathrm{d}m}(\overline{\tau},\mu',-\mu_0)\mu'\mathrm{d}\mu'. \quad (76)$$

The plane albedo r, total transmission t, and the spherical albedo r_s of the layer are given by

$$r(\mu,\overline{\tau}) = 2\int_0^1 R_0(\mu,\mu';\overline{\tau})\mu'\mathrm{d}\mu' = 2\int_0^1 R_0(\mu',\mu;\overline{\tau})\mu'\mathrm{d}\mu', \quad (77)$$

$$t(\mu,\overline{\tau}) = 2\int_0^1 T_0(\mu,\mu';\overline{\tau})\mu'\mathrm{d}\mu' = 2\int_0^1 T_0(\mu',\mu;\overline{\tau})\mu'\mathrm{d}\mu', \quad (78)$$

$$r_s(\overline{\tau}) = 4\int_0^1\int_0^1 R_0(\mu,\mu';\overline{\tau})\mu'\mu\mathrm{d}\mu'\mathrm{d}\mu, \quad (79)$$

while for a homogeneous layer with an underlying Lambertian surface of albedo A, the interaction principle is

$$I_{\mathrm{d}m}(0,\mu,-\mu_0) = (2-\delta_{m0})\frac{1}{\pi}\mu_0 F_0 R_{Am}(\mu,\mu_0;\overline{\tau}) + 2\int_0^1 R_{Am}(\mu,\mu';\overline{\tau}) I_{\mathrm{d}m}(0,-\mu',-\mu_0)\mu'\mathrm{d}\mu' \quad (80)$$

and

$$I_{\mathrm{d}m}(\overline{\tau},-\mu,-\mu_0) = (2-\delta_{m0})\frac{1}{\pi}\mu_0 F_0\, T_{\mathrm{A}\mathrm{d}m}(\mu,\mu_0;\overline{\tau}) + 2\int_0^1 T_{Am}(\mu,\mu';\overline{\tau})\, I_{\mathrm{d}m}(0,-\mu',-\mu_0)\mu'\,\mathrm{d}\mu', \tag{81}$$

where

$$R_{Am}(\mu,\mu';\overline{\tau}) = R_m(\mu,\mu';\overline{\tau}) + \delta_{m0}\frac{A}{1-Ar_s}\,t(\mu,\overline{\tau})\,t(\mu',\overline{\tau}), \tag{82}$$

$$T_{Am}(\mu,\mu';\overline{\tau}) = T_m(\mu,\mu';\overline{\tau}) + \delta_{m0}\frac{A}{1-Ar_s}\,r(\mu,\overline{\tau})\,t(\mu',\overline{\tau}), \tag{83}$$

$$T_{\mathrm{A}\mathrm{d}m}(\mu,\mu_0;\overline{\tau}) = T_{\mathrm{d}m}(\mu,\mu_0;\overline{\tau}) + \delta_{m0}\frac{A}{1-Ar_s}\,r(\mu,\overline{\tau})\,t(\mu_0,\overline{\tau}). \tag{84}$$

For a pertinent and mathematical elegant description of the radiative transfer in a continuous setting including the definitions of the reflection and transmission functions, as well as the derivation of the interaction principle equation, we refer to [32].

In a discrete setting, the interaction principle equation relates the outcoming radiances $\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0)$ and $\widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau})$ to the incoming radiances $\widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0)$ and $\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau})$ through the reflection and transmission matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$, respectively. Transforming the solution representations of the matrix exponential method into a form which resembles the interaction principle equation, equivalent expressions for the reflection and transmission matrices can be obtained. This derivation can be regarded as a conversion of the initial value problem of the matrix exponential method into a two-point boundary value problem (the incoming radiances $\widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0)$ and $\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau})$ are specified). In the framework of the matrix exponential formalism, the reflection and transmission matrices are introduced in a natural way, and as they are well behaved and bounded, the numerical instability is avoided. In the matrix operator method and for a multi-layered atmosphere, the reflection matrix of the entire atmosphere is computed recursively from the reflection and transmission matrices of each layer by using the adding algorithm.

5.1. *Reflection and transmission matrices of a homogeneous layer*

First representations of the reflection and transmission matrices can be obtained from layer equation (56). Expressing (56) in terms of the upward and downward radiance vectors $\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0)$ and $\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau})$, we obtain the interaction principle equation

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{T}} \\ \widehat{\mathbf{T}} & \widehat{\mathbf{R}} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau}) \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{\Sigma}}_{+} \\ \widehat{\mathbf{\Sigma}}_{-} \end{bmatrix}, \tag{85}$$

where $\widehat{\mathbf{R}} = \widehat{\mathbf{R}}(\overline{\tau})$, $\widehat{\mathbf{T}} = \widehat{\mathbf{T}}(\overline{\tau})$, and

$$\begin{bmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{T}} \\ \widehat{\mathbf{T}} & \widehat{\mathbf{R}} \end{bmatrix} = \begin{bmatrix} -\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{+}^{T} & -\overline{\mathbf{V}}_{-}^{T} \\ \overline{\mathbf{V}}_{-}^{T} & \mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{+}^{T} \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{-}^{T} & -\overline{\mathbf{V}}_{+}^{T} \\ \overline{\mathbf{V}}_{+}^{T} & \mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{-}^{T} \end{bmatrix}, \tag{86}$$

$$\begin{bmatrix} \widehat{\mathbf{\Sigma}}_{+} \\ \widehat{\mathbf{\Sigma}}_{-} \end{bmatrix} = \begin{bmatrix} -\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{+}^{T} & -\overline{\mathbf{V}}_{-}^{T} \\ \overline{\mathbf{V}}_{-}^{T} & \mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_{+}^{T} \end{bmatrix}^{-1} \mathbf{D}_{\mathrm{b}}\overline{\mathbf{V}}^{T}\,\widehat{\mathbf{b}}. \tag{87}$$

By inspection of (86) it is apparent that the computation of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ requires an inversion and a multiplication of matrices of dimension $2N\times 2N$. Similar expressions for $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$, which however do not use the right- and left-eigenvectors technique, can be found in [23, 24]. The $2N\times 2N$ matrix in the left-hand side of (86) is called extinction matrix. The extinction matrix is expressed in terms of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$, and as the propagator, it depends only on the internal properties of the homogeneous layer [22].

The computation of the reflection and transmission matrices can be halved in order. These representations, corresponding to the interaction principle equation

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{T}} \\ \widehat{\mathbf{T}} & \widehat{\mathbf{R}} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0) - \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau}) - \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(\overline{\tau}) \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(\overline{\tau}) \end{bmatrix}, \tag{88}$$

can be derived from (63)-(64), and (67)-(68), and do not necessarily require the use of the systems of normalized right and left eigenvectors. In the first case, we use (63) with $\tau = \overline{\tau}$ and (64) to express $\widehat{\mathbf{i}}_{\mathrm{d}}(\overline{\tau}) = [\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau}), \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau})]^{T}$ and $\widehat{\mathbf{i}}_{\mathrm{d}}(0) = [\widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0), \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0)]^{T}$, respectively, in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; from these representations, we get

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_{-} & \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}(\overline{\tau}) \\ \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}(\overline{\tau}) & \widehat{\mathbf{V}}_{-} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(\overline{\tau}) \end{bmatrix} \tag{89}$$

and

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_{+} & \widehat{\mathbf{V}}_{-}\mathbf{\Gamma}(\overline{\tau}) \\ \widehat{\mathbf{V}}_{-}\mathbf{\Gamma}(\overline{\tau}) & \widehat{\mathbf{V}}_{+} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(\overline{\tau}) \end{bmatrix}, \tag{90}$$

and further,

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_{+} & \widehat{\mathbf{V}}_{-}\mathbf{\Gamma}(\overline{\tau}) \\ \widehat{\mathbf{V}}_{-}\mathbf{\Gamma}(\overline{\tau}) & \widehat{\mathbf{V}}_{+} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{V}}_{-} & \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}(\overline{\tau}) \\ \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}(\overline{\tau}) & \widehat{\mathbf{V}}_{-} \end{bmatrix}^{-1} \times \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{d}}^{-}(0) - \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{d}}^{+}(\overline{\tau}) - \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(\overline{\tau}) \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(\overline{\tau}) \end{bmatrix}. \tag{91}$$

Employing now the matrix identity

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{C} \\ -\mathbf{A}^{-1}\mathbf{B}\mathbf{C} & \mathbf{C} \end{bmatrix}, \tag{92}$$

with

$$\mathbf{C} = (\mathbf{A} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B})^{-1}, \tag{93}$$

we end up with

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{V}}_{+} - \widehat{\mathbf{V}}_{-}\mathbf{\Gamma}\widehat{\mathbf{V}}_{-}^{-1}\widehat{\mathbf{V}}_{+}\mathbf{\Gamma})(\widehat{\mathbf{V}}_{-} - \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}\widehat{\mathbf{V}}_{-}^{-1}\widehat{\mathbf{V}}_{+}\mathbf{\Gamma})^{-1}, \tag{94}$$

$$\widehat{\mathbf{T}} = (\widehat{\mathbf{V}}_{-}\mathbf{\Gamma} - \widehat{\mathbf{V}}_{+}\widehat{\mathbf{V}}_{-}^{-1}\widehat{\mathbf{V}}_{+}\mathbf{\Gamma})(\widehat{\mathbf{V}}_{-} - \widehat{\mathbf{V}}_{+}\mathbf{\Gamma}\widehat{\mathbf{V}}_{-}^{-1}\widehat{\mathbf{V}}_{+}\mathbf{\Gamma})^{-1}, \tag{95}$$

where $\mathbf{\Gamma} = \mathbf{\Gamma}(\overline{\tau})$. In the second case, we proceed analogously and use the matrix identity

$$\begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{A} & \mathbf{B} \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \mathbf{A}^{-1} & \mathbf{A}^{-1} \\ -\mathbf{B}^{-1} & \mathbf{B}^{-1} \end{bmatrix}, \tag{96}$$

to conclude that

$$\widehat{\mathbf{R}} = \frac{1}{2}[(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\mathbf{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\mathbf{\Gamma})^{-1} + (\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-\mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\mathbf{\Gamma})^{-1}], \tag{97}$$

$$\widehat{\mathbf{T}} = \frac{1}{2}[(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\mathbf{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\mathbf{\Gamma})^{-1} - (\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-\mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\mathbf{\Gamma})^{-1}]. \tag{98}$$

By making use on fundamental matrix identities it can be shown that (94)-(95) and (97)-(98) are identical. On the other hand, it is apparent that (94)-(95) require 2 matrix inversions and 5 matrix multiplications, while (97)-(98) require 2 matrix inversions and 2 matrix multiplications. In this regard, the quantities which enter in the interaction principle equation (85) may be computed as follows:

1. calculate the matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ from (97) and (98), respectively,
2. calculate the particular solution by means of (23) and (70), and the particular solution source vector according to

$$\begin{bmatrix} \widehat{\mathbf{\Sigma}}_+ \\ \widehat{\mathbf{\Sigma}}_- \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(\overline{\tau}) \end{bmatrix} - \begin{bmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{T}} \\ \widehat{\mathbf{T}} & \widehat{\mathbf{R}} \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{p}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{p}}^{+}(\overline{\tau}) \end{bmatrix}.$$

Depending on the choice of the method for computing the spectral decomposition of the layer matrix $\widehat{\mathbf{A}}$, specific representations of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ can be derived. In the square-root method, the matrices $\widehat{\mathbf{Q}}_+$ and $\widehat{\mathbf{Q}}_-$, defined by $\widehat{\mathbf{V}}_+ = (\widehat{\mathbf{Q}}_+ + \widehat{\mathbf{Q}}_-)/2$ and $\widehat{\mathbf{V}}_- = (\widehat{\mathbf{Q}}_+ - \widehat{\mathbf{Q}}_-)/2$, are related through the relation (cf. (344) of Appendix 1) $\widehat{\mathbf{Q}}_- = -\widehat{\mathbf{Q}}_+^{-T}\mathbf{\Lambda}$, and it can bee shown that $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ can be expressed in terms of $\widehat{\mathbf{Q}}_+^{-T}$ only. On the other hand, in the Cholesky method, the identity (cf. (351) of Appendix 1) $\widehat{\mathbf{Q}}_+^{-T} = -\widehat{\mathbf{Q}}_-\mathbf{\Lambda}$ can be used to express $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ in terms of $\widehat{\mathbf{Q}}_-$ only. These representations which play an important role in the asymptotic theory will be derived in Section 5.4.

We conclude this section by presenting Waterman's derivation of the reflection and transmission matrices. Considering for simplicity, the interaction principle equation for the homogeneous solution (cf. (88))

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{T}} \\ \widehat{\mathbf{T}} & \widehat{\mathbf{R}} \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(0) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\overline{\tau}) \end{bmatrix}, \tag{99}$$

we obtain

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\overline{\tau}) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{T}}^{-1} & -\widehat{\mathbf{T}}^{-1}\widehat{\mathbf{R}} \\ \widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1} & \widehat{\mathbf{T}} - \widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1}\widehat{\mathbf{R}} \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(0) \end{bmatrix}. \tag{100}$$

On the other hand, from (cf. (26))

$$\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\overline{\tau}) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\overline{\tau}) \end{bmatrix} = \mathrm{e}^{-\widehat{\mathbf{A}}\overline{\tau}}\begin{bmatrix} \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(0) \\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(0) \end{bmatrix}, \tag{101}$$

it is apparent that the matrix exponential is identified as

$$\mathrm{e}^{-\widehat{\mathbf{A}}\overline{\tau}} = \begin{bmatrix} \widehat{\mathbf{T}}^{-1} & -\widehat{\mathbf{T}}^{-1}\widehat{\mathbf{R}} \\ \widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1} & \widehat{\mathbf{T}} - \widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1}\widehat{\mathbf{R}} \end{bmatrix}. \tag{102}$$

Flatau and Stephens [22] called (102) the fundamental relationship connecting the propagator (matrix exponential) and the extinction matrix. So basically, what we have to do is to evaluate $\exp(-\widehat{\mathbf{A}}\overline{\tau})$, to invert its upper-left-hand block to obtain $\widehat{\mathbf{T}}$, and finally, to postmultiply its lower-left-hand block by $\widehat{\mathbf{T}}$ in order to get $\widehat{\mathbf{R}}$. By means of (41) in conjunction with (28) and (37), we find

$$\widehat{\mathbf{T}}^{-1} = -\widehat{\mathbf{V}}_+\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_+^T + \widehat{\mathbf{V}}_-\mathbf{\Gamma}(-\overline{\tau})\overline{\mathbf{V}}_-^T, \tag{103}$$

$$\widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1} = -\widehat{\mathbf{V}}_-\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_+^T + \widehat{\mathbf{V}}_+\mathbf{\Gamma}(-\overline{\tau})\overline{\mathbf{V}}_-^T. \tag{104}$$

Equations (103)-(104) have been used by Waterman as a starting point for deriving the expressions of the reflection and transmission matrices for small and large values of the optical thickness. Here, we use a different approach in order to show the equivalences with (94)-(95), and so, with (97)-(98). Let us define the quantity

$$\widehat{\mathbf{T}}_0 = \overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\overline{\tau})\widehat{\mathbf{V}}_-^{-1}, \tag{105}$$

which is computationally stable, and let us construct the matrix product

$$\widehat{\mathbf{T}}^{-1}\widehat{\mathbf{T}}_0 = \mathbf{I} - \widehat{\mathbf{V}}_+\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\overline{\tau})\widehat{\mathbf{V}}_-^{-1}. \tag{106}$$

Then, we obtain

$$\widehat{\mathbf{T}}_0^{-1}\widehat{\mathbf{T}} = [\mathbf{I} - \widehat{\mathbf{V}}_+\mathbf{\Gamma}(\overline{\tau})\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\overline{\tau})\widehat{\mathbf{V}}_-^{-1}]^{-1} \tag{107}$$

and so,

$$\widehat{\mathbf{T}} = \overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})^{-1}, \tag{108}$$

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})^{-1}, \tag{109}$$

where, as before, $\mathbf{\Gamma} = \mathbf{\Gamma}(\overline{\tau})$. Equations (108)-(109) seems to be new. However, employing (48) and (50) in (108)-(109), yields (94)-(95), and the mathematical equivalence between the various representations of the reflection and transmission matrices is proved.

The analytical formulas derived so far are valid for non-conservative scattering ($\omega \neq 1$). The case of conservative scattering, which is merely of theoretical interest, is treated in Appendix 3.

5.2. Discrete approximations of the reflection and transmission functions of a homogeneous layer

There are several applications, e.g., asymptotic theory, in which discrete approximations of the reflection and transmission functions of a homogeneous layer are of particular interest. Considering the interaction principle (88) and passing from the scaled diffuse radiance vector $\widehat{\mathbf{i}}_{\mathrm{d}}$ to the diffuse

radiance vector $\mathbf{i}_{\mathtt{d}}$ according to the transformation rule (cf. (14)) $\widehat{\mathbf{i}}_{\mathtt{d}}^{\pm} = \mathbf{W}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}\mathbf{i}_{\mathtt{d}}^{\pm}$, we obtain

$$\begin{bmatrix} \mathbf{i}_{\mathtt{d}}^{+}(0) \\ \mathbf{i}_{\mathtt{d}}^{-}(\overline{\tau}) \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{T} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{i}_{\mathtt{d}}^{-}(0) - \mathbf{c}_2 \\ \mathbf{i}_{\mathtt{d}}^{+}(\overline{\tau}) - \mathrm{e}^{-\overline{\tau}/\mu_0}\mathbf{c}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{c}_1 \\ \mathrm{e}^{-\overline{\tau}/\mu_0}\mathbf{c}_2 \end{bmatrix}, \tag{110}$$

where

$$\mathbf{R} = \mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{R}}\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}, \tag{111}$$

$$\mathbf{T} = \mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{T}}\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}, \tag{112}$$

and $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2]^T = \mu_0(\mathbf{I} - \mu_0\mathbf{A})^{-1}\mathbf{b}$, or blockwise, $\mathbf{c}_{1,2} = \mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\widehat{\mathbf{c}}_{1,2}$. Further, from (75), (76) and (110) we find the relations

$$[R_m(\mu_k, \mu_l)] = \mathbb{R}, \tag{113}$$

$$[T_m(\mu_k, \mu_l)] = \mathbb{T}, \tag{114}$$

$$[R_m(\mu_k, \mu_0)] = \mathsf{c}_1 - \mathbf{R}\mathsf{c}_2 - \mathrm{e}^{-\overline{\tau}/\mu_0}\mathbf{T}\mathsf{c}_1, \tag{115}$$

$$[T_{\mathtt{d}m}(\mu_k, \mu_0)] = \mathrm{e}^{-\overline{\tau}/\mu_0}\mathsf{c}_2 - \mathbf{T}\mathsf{c}_2 - \mathrm{e}^{-\overline{\tau}/\mu_0}\mathbf{R}\mathsf{c}_1, \tag{116}$$

where the matrices $\mathbb{R}$ and $\mathbb{T}$, given by

$$\mathbb{R} = \frac{1}{2}\mathbf{R}\mathbf{M}\mathbf{W}^{-1} = \frac{1}{2}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{R}}\mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}, \tag{117}$$

$$\mathbb{T} = \frac{1}{2}\mathbf{T}\mathbf{M}\mathbf{W}^{-1} = \frac{1}{2}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{T}}\mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}, \tag{118}$$

are the discrete approximations of the reflection and transmission functions, respectively, and the vector $\mathsf{c} = [\mathsf{c}_1, \mathsf{c}_2]^T$ is defined through the relation $\mathbf{c} = (2 - \delta_{m0})(F_0/\pi)\mu_0\mathsf{c}$. In addition, the matrix

$$\mathbb{T}_{\mathtt{d}} = \mathbb{T} - \mathbf{D}, \tag{119}$$

with

$$[\mathbf{D}]_{kl} = \frac{1}{2w_k\mu_k}\mathrm{e}^{-\overline{\tau}/\mu_k}\delta_{kl}, \tag{120}$$

is the discrete approximation of the diffuse transmission function, i.e.,

$$[T_{\mathtt{d}m}(\mu_k, \mu_l)] = \mathbb{T}_{\mathtt{d}}. \tag{121}$$

Equations (117) and (118) show how to convert the scaled reflection and transmission matrices into physical functions. As the interaction principle has been formulated for the diffuse radiance, relations (115) and (116) corresponding to the incident direction can be used for checking the reflection and transmission matrix calculations. The computational process of some reflective and transmissive characteristics of the layer involves the following steps:

1. For the azimuthal mode $m = 0$, compute the plane albedo vector $\mathbf{r}$ and the transmission vector $\mathbf{t}$ of the layer

$$\mathbf{r} = 2\mathbb{R}\mathbf{W}\mathbf{M}^{-1}\mathbf{1} = \mathbf{R}\mathbf{1}, \tag{122}$$

$$\mathbf{t} = 2\mathbb{T}\mathbf{W}\mathbf{M}^{-1}\mathbf{1} = \mathbf{T}\mathbf{1}, \tag{123}$$

together with the spherical albedo

$$r_s = 4\mathbf{v}^T\mathbb{R}\mathbf{v} = 2\mathbf{v}^T\mathbf{R}\mathbf{1}, \tag{124}$$

where $\mathbf{v} = \mathbf{W}\mathbf{M}^{-1}\mathbf{1}$ and $\mathbf{1} = [1, \ldots, 1]^T$.

2. Compute the reflection and transmission matrices of the layer with an underlying Lambertian surface

$$\mathbb{R}_A = \mathbb{R} + \frac{A}{1 - Ar_s}\delta_{m0}\mathbf{t}\mathbf{t}^T, \tag{125}$$

$$\mathbb{T}_A = \mathbb{T} + \frac{A}{1 - Ar_s}\delta_{m0}\mathbf{r}\mathbf{t}^T. \tag{126}$$

3. For the azimuthal mode $m = 0$, compute the plane albedo vector of the homogeneous layer with an underlying Lambertian surface

$$\mathbf{r}_A = 2\mathbb{R}_A\mathbf{W}\mathbf{M}^{-1}\mathbf{1} \tag{127}$$

and the spherical albedo

$$r_{sA} = 4\mathbf{v}^T\mathbb{R}_A\mathbf{v}. \tag{128}$$

According to (74), the total transmission function is a generalized function, or a distribution. In a discrete setting this means that in contrast to the diffuse transmission matrix $[T_{\mathtt{d}m}(\mu_k, \mu_l; \overline{\tau})]$, the total transmission matrix $[T_m(\mu_k, \mu_l; \overline{\tau})]$, regarded as a function of μ_l, has a jump at $\mu_l = \mu_k$.

The discrete approximations of the reflection and transmission functions $R(\mu, \mu', \varphi - \varphi'; \overline{\tau})$, $T(\mu, \mu', \varphi - \varphi'; \overline{\tau})$ and $T_{\mathtt{d}}(\mu, \mu', \varphi - \varphi'; \overline{\tau})$ are obtained by summing up the Fourier series in the azimuth, i.e.,

$$\mathbb{X} = \sum_{m=0}^{M_{\max}} (2 - \delta_{m0})\mathbb{X}_m \cos[m(\varphi - \varphi')], \tag{129}$$

where $\mathbb{X}$ stands for $\mathbb{R}$, $\mathbb{T}$, and $\mathbb{T}_{\mathtt{d}}$. The number of azimuthal modes $M_{\max}$ is whether $M_{\max} = 2N - 1$, or $M_{\max} < 2N - 1$. In the second case, an azimuthal convergence test over the diagonal elements of $\mathbb{R}$ and $\mathbb{T}_{\mathtt{d}}$ is performed. A stronger test may involve the convergence of the Frobenius norms of $\mathbb{R}$ and $\mathbb{T}_{\mathtt{d}}$.

For highly peaked phase functions the delta-M method [33] can be used. In this case, the matrices $\mathbf{S}_{\pm}$, and hence, the matrices $\widehat{\mathbf{V}}_{\pm}$ in (97)-(98) are altered by modifying the Legendre expansion coefficients χ_n of the phase function $p(\mu, \mu', \varphi - \varphi')$ and the single scattering albedo ω according to

$$\chi_n^{\star} = \frac{1}{1 - f}(\chi_n - 2f) \tag{130}$$

and

$$\omega^{\star} = \frac{1 - f}{1 - f\omega}\omega, \tag{131}$$

respectively, while the diagonal matrix $\mathbf{\Gamma} = [\mathrm{e}^{-\lambda_k\overline{\tau}}]$ in (97)-(98) is altered by modifying the optical thickness $\overline{\tau}$ according to $\overline{\tau}^{\star} = (1 - f\omega)\overline{\tau}$. Here, the truncation factor f is defined by $f = (1/2)\chi_{2N}$. The delta-M method enhances the convergence of $\mathbb{R}$ and $\mathbb{T}_{\mathtt{d}}$ when compared with the corresponding results obtained in the absence of truncation.

The delta-M method can be used in conjunction with the truncated-plus-single-scattering (TMS) method [34]. The application of the TMS correction to the reflection matrix of a semi-infinite atmosphere was discussed in [35]. For a layer of finite optical thickness, the TMS corrections of the reflection and transmission matrices use the corresponding functions in the single-scattering approximation

$$R_{\mathtt{ss}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) = \frac{1}{4(\mu+\mu_0)}\omega p(\mu,-\mu_0,\varphi-\varphi_0) \times\left[1-\mathrm{e}^{-(1/\mu_0+1/\mu)\overline{\tau}}\right], \tag{132}$$

$$T_{\mathtt{ssd}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) = \frac{1}{4(\mu-\mu_0)}\omega p(-\mu,-\mu_0,\varphi-\varphi_0) \times\left(\mathrm{e}^{-\overline{\tau}/\mu}-\mathrm{e}^{-\overline{\tau}/\mu_0}\right), \tag{133}$$

For example, the TMS correction of the reflection function is

$$\triangle R(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) = R^{\mathtt{e}}_{\mathtt{ss}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) - R^{\mathtt{t}}_{\mathtt{ss}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}), \tag{134}$$

where,

$$R^{\mathtt{e}}_{\mathtt{ss}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) = \frac{1}{4(\mu+\mu_0)}\frac{\omega}{1-f\omega}p(\mu,-\mu_0,\varphi-\varphi_0) \times\left[1-\mathrm{e}^{-(1/\mu_0+1/\mu)\overline{\tau}^\star}\right], \tag{135}$$

$$p(\mu,-\mu_0,\varphi-\varphi_0) = \sum_{n=0}^{N_{\mathtt{max}}}\sqrt{\frac{2n+1}{2}}\chi_n P_n(\cos\Theta),$$

and

$$R^{\mathtt{t}}_{\mathtt{ss}}(\mu,\mu_0,\varphi-\varphi_0;\overline{\tau}) = \frac{1}{4(\mu+\mu_0)}\omega^\star p^\star(\mu,-\mu_0,\varphi-\varphi_0) \times\left[1-\mathrm{e}^{-(1/\mu_0+1/\mu)\overline{\tau}^\star}\right], \tag{136}$$

correspond to the "exact" and truncated phase functions, respectively. In (135), P_n are the normalized Legendre polynomials, $\cos\Theta = -\mu\mu_0 + \sqrt{1-\mu^2}\sqrt{1-\mu_0^2}\cos(\varphi-\varphi_0)$, $N_{\mathtt{max}} > 2N-1$ is the number of all expansion coefficients, and $\overline{\tau}^\star$ is the delta-M scaled optical thickness. The truncated phase function $p^\star(\Theta)$ in (136) can be computed in two different ways. If an azimuthal convergence test is performed ($M_{\mathtt{max}} < 2N-1$), $p^\star(\Theta)$ is computed as

$$p^\star(\mu,-\mu_0,\varphi-\varphi_0) = \sum_{m=0}^{M_{\mathtt{max}}}(2-\delta_{m0})p^\star_m(\mu,-\mu_0)\cos[m(\varphi-\varphi_0)], \tag{137}$$

$$p^\star_m(\mu,-\mu_0) = \sum_{n=m}^{2N-1}\chi^\star_n P^m_n(\mu)P^m_n(-\mu_0),$$

where P^m_n are the normalized Legendre functions, and otherwise, as

$$p^\star(\Theta) = \sum_{n=0}^{2N-1}\sqrt{\frac{2n+1}{2}}\chi^\star_n P_n(\cos\Theta). \tag{138}$$

To analyze the accuracy of the TMS method in computing the reflection and transmission functions we perform numerical simulations of the Earth Polychromatic Imaging Camera (EPIC) measurements in the oxygen A-band absorption channel at 764 nm (Channel 9). Note that this channel is used for cloud parameters retrieval, and that the EPIC instrument, onboard the Deep Space Climate Observatory (DSCOVR), measures from the L1 Lagrangian point, at which the satellite remains near the Sun–Earth line. For an atmospheric layer consisting of oxygen molecules and a water cloud, we assume that the radiative transfer in the layer involves, in addition to cloud scattering and absorption, oxygen absorption. More precisely, we neglect the molecular Rayleigh scattering, that is, for the homogenized layer, we use the approximations $\sigma_{\mathtt{ext}} \approx \sigma^{\mathtt{c}}_{\mathtt{ext}} + \sigma^{\mathtt{g}}_{\mathtt{abs}}$, $\sigma_{\mathtt{sct}} \approx \sigma^{\mathtt{c}}_{\mathtt{sct}}$, $\chi_n \approx \chi^{\mathtt{c}}_n$, where $\sigma^{\mathtt{c}}_{\mathtt{ext}}$, $\sigma^{\mathtt{c}}_{\mathtt{sct}}$ and $\chi^{\mathtt{c}}_n$ are the extinction coefficient, scattering coefficient and the phase function expansion coefficients of the cloud, while $\sigma^{\mathtt{g}}_{\mathtt{abs}}$ is the absorption coefficient of the oxygen gas. Perhaps it should be pointed out that the high oxygen absorption (computed by line-by-line calculations), yields a large interval of variation of the single scattering albedo of the homogenized layer (between 0.2 and 0.999) As the Sun is in the back of the instrument, the observed scattering angle is close to the backscattering direction. To model the scattering in the backward region, we take $\mu = \mu_0$, and as in [36], we choose $\varphi-\varphi_0 = 176^\circ$. In Table 1 we show the relative errors in R and $T_{\mathtt{d}}$, defined by

$$\varepsilon_{\mathtt{R}} = \sqrt{\frac{\int_0^1[R(\mu,\mu,\varphi-\varphi_0;\overline{\tau}) - R_{\mathtt{ref}}(\mu,\mu,\varphi-\varphi_0;\overline{\tau})]^2\mathrm{d}\mu}{\int_0^1 R^2_{\mathtt{ref}}(\mu,\mu,\varphi-\varphi_0;\overline{\tau})\mathrm{d}\mu}} \tag{139}$$

and similarly for $\varepsilon_{\mathtt{T}}$, for the delta-M method and the delta-M method with TMS correction. The optical thickness is $\overline{\tau} = 5$, the single scattering albedo is $\omega = 0.85$, and the phase function corresponds to a water-cloud model with a Gamma size distribution

$$P(a) \propto a^\alpha \exp\left[-\alpha\left(\frac{a}{a_{\mathtt{mod}}}\right)\right] \tag{140}$$

of parameters $a_{\mathtt{mod}} = 10$ μm and $\alpha = 6$. The droplet size ranges between 0.02 and 50.0 μm, and the reference values $R_{\mathtt{ref}}$ and $T_{\mathtt{dref}}$ correspond to $N_{\mathtt{rank}} = 565$. The results demonstrate that the TMS method requires less discrete ordinates N as the standard delta-M method. This observation implies that the TMS method improves the accuracy of reflection and transmission functions calculations.

This model based on an eigendecomposition method for computing the matrix exponential is used as a reference for testing the approximate representations of the reflection and transmission matrices in the cases of thin and thick layers. In particular, for the EPIC instrument, we compute the relative errors

$$\varepsilon_{\mathtt{X}} = \sqrt{\frac{\sum_{k=1}^N([\mathbb{X}]_{kk} - [\mathbb{X}_{\mathtt{ref}}]_{kk})^2}{\sum_{k=1}^N[\mathbb{X}_{\mathtt{ref}}]^2_{kk}}}, \tag{141}$$

where as before, $\mathbb{X}$ stands for $\mathbb{R}$ and $\mathbb{T}$, and $\varphi-\varphi_0 = 176^\circ$.

Table 1: Relative errors ε_R and ε_T for the delta-M method and the delta-M method with TMS correction, and for different values of the number of discrete ordinates per hemisphere N.

N	Delta-M		Delta-M and TMS	
	ε_R	ε_T	ε_R	ε_T
128	3.80e-4	4.44e-4	3.53e-4	4.44e-4
64	1.64e-2	2.06e-3	1.41e-3	1.91e-3
56	1.97e-2	2.81e-3	1.61e-3	2.56e-3
48	2.28e-2	4.24e-3	2.21e-3	3.70e-3
32	4.34e-2	8.63e-3	3.76e-3	7.73e-3
16	6.12e-2	2.85e-2	2.42e-2	2.66e-2
8	1.95e-1	1.05e-1	6.71e-2	8.97e-2

5.3. *Reflection and transmission matrices of a homogeneous thin layer*

For optically thin layers, the Padé and Taylor series approximations to the matrix exponential can be used for computing the reflection and transmission matrices and so, to avoid the solution of an eigenvalue problem. The interest in such small values of $\overline{\tau}$ stems from the possibility of using these approximations to generate starting values for the doubling method. Parenthetically, we note that the basic recursion relations of the doubling method can be obtained by using the matrix exponential formalism. Indeed, setting $\widehat{\mathbf{R}}_1 = \widehat{\mathbf{R}}(\overline{\tau})$ and $\widehat{\mathbf{T}}_1 = \widehat{\mathbf{T}}(\overline{\tau})$ for the layer of optical thickness $\overline{\tau}$, and $\widehat{\mathbf{R}}_2 = \widehat{\mathbf{R}}(2\overline{\tau})$ and $\widehat{\mathbf{T}}_2 = \widehat{\mathbf{T}}(2\overline{\tau})$ for the layer of optical thickness $2\overline{\tau}$, we express (102) for the layers $\overline{\tau}$ and $2\overline{\tau}$, as

$$\mathrm{e}^{-\widehat{\mathbf{A}}\overline{\tau}} = \begin{bmatrix} \widehat{\mathbf{T}}_1^{-1} & -\widehat{\mathbf{T}}_1^{-1}\widehat{\mathbf{R}}_1 \\ \widehat{\mathbf{R}}_1\widehat{\mathbf{T}}_1^{-1} & \widehat{\mathbf{T}}_1 - \widehat{\mathbf{R}}_1\widehat{\mathbf{T}}_1^{-1}\widehat{\mathbf{R}}_1 \end{bmatrix}, \tag{142}$$

and

$$\mathrm{e}^{-2\widehat{\mathbf{A}}\overline{\tau}} = \begin{bmatrix} \widehat{\mathbf{T}}_2^{-1} & -\widehat{\mathbf{T}}_2^{-1}\widehat{\mathbf{R}}_2 \\ \widehat{\mathbf{R}}_2\widehat{\mathbf{T}}_2^{-1} & \widehat{\mathbf{T}}_2 - \widehat{\mathbf{R}}_2\widehat{\mathbf{T}}_2^{-1}\widehat{\mathbf{R}}_2 \end{bmatrix}, \tag{143}$$

respectively, and use the identity $\exp(-2\widehat{\mathbf{A}}\overline{\tau}) = [\exp(-\widehat{\mathbf{A}}\overline{\tau})]^2$, to obtain

$$\begin{bmatrix} \widehat{\mathbf{T}}_2^{-1} & -\widehat{\mathbf{T}}_2^{-1}\widehat{\mathbf{R}}_2 \\ \widehat{\mathbf{R}}_2\widehat{\mathbf{T}}_2^{-1} & \widehat{\mathbf{T}}_2 - \widehat{\mathbf{R}}_2\widehat{\mathbf{T}}_2^{-1}\widehat{\mathbf{R}}_2 \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{T}}_1^{-1} & -\widehat{\mathbf{T}}_1^{-1}\widehat{\mathbf{R}}_1 \\ \widehat{\mathbf{R}}_1\widehat{\mathbf{T}}_1^{-1} & \widehat{\mathbf{T}}_1 - \widehat{\mathbf{R}}_1\widehat{\mathbf{T}}_1^{-1}\widehat{\mathbf{R}}_1 \end{bmatrix}^2. \tag{144}$$

Solving for $\widehat{\mathbf{R}}_2$ and $\widehat{\mathbf{T}}_2$, we get

$$\widehat{\mathbf{R}}_2 = \widehat{\mathbf{R}}_1 + \widehat{\mathbf{T}}_1\widehat{\mathbf{R}}_1(\mathbf{I} - \widehat{\mathbf{R}}_1^2)^{-1}\widehat{\mathbf{T}}_1, \tag{145}$$

$$\widehat{\mathbf{T}}_2 = \widehat{\mathbf{T}}_2(\mathbf{I} - \widehat{\mathbf{R}}_1^2)^{-1}\widehat{\mathbf{T}}_1, \tag{146}$$

which are the recursion relations of the doubling method.

5.3.1. *Padé approximation*

In radiative transfer, the Padé approximation has been suggested by Flatau and Stephens [22] for computing the extinction matrix and the source function integral of a layer. More recently, McGararagh and Gabriel [37] used this approximation in connection with the matrix operator method. Essentially, the nth diagonal Padé approximation to the exponential of the matrix $\widehat{\mathbf{A}}\tau$ is defined as [38]

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} \approx [\mathbf{D}_n(\widehat{\mathbf{A}}\tau)]^{-1}\mathbf{N}_n(\widehat{\mathbf{A}}\tau), \tag{147}$$

where $\mathbf{D}_n(\widehat{\mathbf{A}}\tau)$ and $\mathbf{N}_n(\widehat{\mathbf{A}}\tau)$ are polynomials in $\widehat{\mathbf{A}}\tau$ of degree n, given by

$$\mathbf{D}_n(\widehat{\mathbf{A}}\tau) = \sum_{k=0}^{n} c_k \tau^k \widehat{\mathbf{A}}^k, \tag{148}$$

$$\mathbf{N}_n(\widehat{\mathbf{A}}\tau) = \sum_{k=0}^{n} (-1)^k c_k \tau^k \widehat{\mathbf{A}}^k, \tag{149}$$

and

$$c_k = \frac{(2n-k)!n!}{(2n)!k!(n-k)!}. \tag{150}$$

From (148) and (149) it is readily seen that $\mathbf{N}_n(\widehat{\mathbf{A}}\tau) = \mathbf{D}_n(-\widehat{\mathbf{A}}\tau)$. To compute $\mathbf{D}_n(\widehat{\mathbf{A}}\tau)$ and $\mathbf{N}_n(\widehat{\mathbf{A}}\tau)$ we have to compute powers of $\widehat{\mathbf{A}}$. By taking advantage of the block symmetries within $\widehat{\mathbf{A}}$ (cf. (6)), we find

$$\widehat{\mathbf{A}}^k = \begin{bmatrix} \mathbf{X}_k & \mathbf{Y}_k \\ (-1)^k\mathbf{Y}_k & (-1)^k\mathbf{X}_k \end{bmatrix}, \tag{151}$$

where for $k \geq 2$, the matrices $\mathbf{X}_k$ and $\mathbf{Y}_k$ are computed recursively as

$$\mathbf{X}_k = \mathbf{X}_{k-1}\widehat{\mathbf{A}}_{11} - \mathbf{Y}_{k-1}\widehat{\mathbf{A}}_{12}, \tag{152}$$

$$\mathbf{Y}_k = \mathbf{X}_{k-1}\widehat{\mathbf{A}}_{12} - \mathbf{Y}_{k-1}\widehat{\mathbf{A}}_{11}, \tag{153}$$

with the initial values

$$\mathbf{X}_1 = \widehat{\mathbf{A}}_{11}, \tag{154}$$

$$\mathbf{Y}_1 = \widehat{\mathbf{A}}_{12}. \tag{155}$$

The coefficients c_k are also computed recursively by means of

$$c_k = \frac{n-k+1}{k(2n-k+1)}c_{k-1}, \tag{156}$$

with the initial value $c_1 = 1/2$. Accounting of (151), the matrices $\mathbf{D}_n(\widehat{\mathbf{A}}\tau)$ and $\mathbf{N}_n(\widehat{\mathbf{A}}\tau)$ of (148) and (149), respectively, can be written as

$$\mathbf{D}_n(\widehat{\mathbf{A}}\tau) = \begin{bmatrix} \mathbf{I}+\mathbf{X}_n^+ & \mathbf{Y}_n^+ \\ \mathbf{Y}_n^- & \mathbf{I}+\mathbf{X}_n^- \end{bmatrix}, \tag{157}$$

$$\mathbf{N}_n(\widehat{\mathbf{A}}\tau) = \begin{bmatrix} \mathbf{I}+\mathbf{X}_n^- & \mathbf{Y}_n^- \\ \mathbf{Y}_n^+ & \mathbf{I}+\mathbf{X}_n^+ \end{bmatrix}, \tag{158}$$

where

$$\mathbf{X}_n^{\pm} = \sum_{k=1}^{n} (\pm 1)^k c_k \tau^k \mathbf{X}_k, \tag{159}$$

$$\mathbf{Y}_n^{\pm} = \sum_{k=1}^{n} (\pm 1)^k c_k \tau^k \mathbf{Y}_k. \tag{160}$$

Inserting (147) together with (157) and (158) into the homogeneous solution representation (26) yields

$$\widehat{\mathbf{R}}_n = \mathbf{H}_n(\mathbf{I}+\mathbf{X}_n^+) - \mathbf{G}_n\mathbf{Y}_n^-, \tag{161}$$

$$\widehat{\mathbf{T}}_n = \mathbf{G}_n(\mathbf{I}+\mathbf{X}_n^+) - \mathbf{H}_n\mathbf{Y}_n^-, \tag{162}$$

with

$$\mathbf{G}_n = (\mathbf{E}_n - \mathbf{F}_n\mathbf{E}_n^{-1}\mathbf{F}_n)^{-1}, \tag{163}$$

$$\mathbf{H}_n = \mathbf{E}_n^{-1}\mathbf{F}_n\mathbf{G}_n, \tag{164}$$

and

$$\mathbf{E}_n = \mathbf{I}+\mathbf{X}_n^-, \tag{165}$$

$$\mathbf{F}_n = \mathbf{Y}_n^+. \tag{166}$$

Equations (161) and (162) give the nth-order Padé approximations to the reflection and transmission matrices. In the case $n=1$, we find $\mathbf{X}_1^\pm = \pm(\overline{\tau}/2)\widehat{\mathbf{A}}_{11}$ and $\mathbf{Y}_1^\pm = \pm(\overline{\tau}/2)\widehat{\mathbf{A}}_{12}$. As a result, we obtain $\mathbf{E}_1 = \mathbf{I} - (\overline{\tau}/2)\widehat{\mathbf{A}}_{11}$ and $\mathbf{F}_1 = (\overline{\tau}/2)\widehat{\mathbf{A}}_{12}$, and so,

$$\widehat{\mathbf{R}}_1 = \mathbf{H}_1(\mathbf{I}+\frac{\overline{\tau}}{2}\widehat{\mathbf{A}}_{11}) + \frac{\overline{\tau}}{2}\mathbf{G}_1\widehat{\mathbf{A}}_{12}, \tag{167}$$

$$\widehat{\mathbf{T}}_1 = \mathbf{G}_1(\mathbf{I}+\frac{\overline{\tau}}{2}\widehat{\mathbf{A}}_{11}) + \frac{\overline{\tau}}{2}\mathbf{H}_1\widehat{\mathbf{A}}_{12}. \tag{168}$$

Further, approximating $\mathbf{E}_1^{-1} \approx \mathbf{I} + (\overline{\tau}/2)\widehat{\mathbf{A}}_{11}$, and retaining only the first-order terms in the Neumann series of the inverse in (163), gives $\mathbf{G}_1 = \mathbf{I} + (\overline{\tau}/2)\widehat{\mathbf{A}}_{11}$ and $\mathbf{H}_1 = (\overline{\tau}/2)\widehat{\mathbf{A}}_{12}$, and ultimately

$$\widehat{\mathbf{R}}_1 = \overline{\tau}\widehat{\mathbf{A}}_{12} \text{ and } \widehat{\mathbf{T}}_1 = \mathbf{I}+\overline{\tau}\widehat{\mathbf{A}}_{11}. \tag{169}$$

This is the infinitesimal generator initialization scheme of Grant and Hunt [39] (see (181) below).

5.3.2. *Taylor series approximation*

The Taylor series approximation uses the definition of the matrix exponential, namely

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} \approx \mathbf{I} + \sum_{k=1}^{n}(-1)^k\frac{1}{k!}\tau^k\widehat{\mathbf{A}}^k. \tag{170}$$

Accounting of (151), we obtain

$$\mathrm{e}^{-\widehat{\mathbf{A}}\overline{\tau}} \approx \mathbf{I} + \sum_{k=1}^{n}(-1)^k\frac{1}{k!}\overline{\tau}^k\begin{bmatrix}\mathbf{X}_k & \mathbf{Y}_k\\ (-1)^k\mathbf{Y}_k & (-1)^k\mathbf{X}_k\end{bmatrix}, \tag{171}$$

while from (142) we get

$$\widehat{\mathbf{T}}^{-1} \approx \widehat{\mathbf{T}}_n^{-1} = \mathbf{I} + \sum_{k=1}^{n}\overline{\tau}^k\mathbf{E}_k, \tag{172}$$

$$\widehat{\mathbf{R}}\widehat{\mathbf{T}}^{-1} \approx \widehat{\mathbf{R}}_n\widehat{\mathbf{T}}_n^{-1} = \sum_{k=1}^{n}\overline{\tau}^k\mathbf{F}_k, \tag{173}$$

where the matrices $\mathbf{E}_k$ and $\mathbf{F}_k$ are now given by

$$\mathbf{E}_k = \frac{(-1)^k}{k!}\mathbf{X}_k \text{ and } \mathbf{F}_k = \frac{1}{k!}\mathbf{Y}_k, \tag{174}$$

respectively. Using (172) and (173), and seeking for expansions of the form

$$\widehat{\mathbf{T}}_n = \mathbf{I} + \sum_{k=1}^{n}\overline{\tau}^k\mathbf{G}_k, \tag{175}$$

$$\widehat{\mathbf{R}}_n = \sum_{k=1}^{n}\overline{\tau}^k\mathbf{H}_k, \tag{176}$$

we find that the matrices $\mathbf{G}_n$ and $\mathbf{H}_n$ in (175) and (176) can be computed recursively as

$$\mathbf{G}_1 = -\mathbf{E}_1, \tag{177}$$

$$\mathbf{G}_n = -\mathbf{E}_n - \sum_{k=1}^{n-1}\mathbf{E}_{n-k}\mathbf{G}_k,\ n\geq 2, \tag{178}$$

and

$$\mathbf{H}_1 = \mathbf{F}_1, \tag{179}$$

$$\mathbf{H}_n = \mathbf{F}_n + \sum_{k=1}^{n-1}\mathbf{F}_{n-k}\mathbf{G}_k,\ n\geq 2, \tag{180}$$

respectively. Equations (175) and (176) give the nth-order Taylor series approximations to the transmission and reflection matrices. In the case $n=1$, we obtain the infinitesimal generator initialization scheme of Grant and Hunt [39],

$$\widehat{\mathbf{T}}_1 = \mathbf{I}+\overline{\tau}\widehat{\mathbf{A}}_{11},\ \widehat{\mathbf{R}}_1 = \overline{\tau}\widehat{\mathbf{A}}_{12}, \tag{181}$$

in the case $n=2$, we obtain the expanded diamond initialization scheme of Wiscombe [40],

$$\widehat{\mathbf{T}}_2 = \widehat{\mathbf{T}}_1 + \frac{1}{2}\overline{\tau}^2(\widehat{\mathbf{A}}_{11}^2+\widehat{\mathbf{A}}_{12}^2), \tag{182}$$

$$\widehat{\mathbf{R}}_2 = \widehat{\mathbf{R}}_1 + \frac{1}{2}\overline{\tau}^2(\widehat{\mathbf{A}}_{11}\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{12}\widehat{\mathbf{A}}_{11}), \tag{183}$$

and finally, in the case $n=3$, we obtain the scheme of Waterman [21],

$$\widehat{\mathbf{T}}_3 = \widehat{\mathbf{T}}_2 + \frac{1}{6}\overline{\tau}^3(\widehat{\mathbf{A}}_{11}^3+2\widehat{\mathbf{A}}_{12}^2\widehat{\mathbf{A}}_{11}+2\widehat{\mathbf{A}}_{11}\widehat{\mathbf{A}}_{12}^2+\widehat{\mathbf{A}}_{12}\widehat{\mathbf{A}}_{11}\widehat{\mathbf{A}}_{12}), \tag{184}$$

$$\widehat{\mathbf{R}}_3 = \widehat{\mathbf{R}}_2 + \frac{1}{6}\overline{\tau}^3(2\widehat{\mathbf{A}}_{12}^3+\widehat{\mathbf{A}}_{11}^2\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{12}\widehat{\mathbf{A}}_{11}^2+2\widehat{\mathbf{A}}_{11}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{A}}_{11}). \tag{185}$$

The Padé and Taylor series approximations are based on the computation of powers of $\widehat{\mathbf{A}}$, for which we used (151)-(155). An alternative approach for computing $\widehat{\mathbf{A}}^k$, which exploits more efficiently the symmetries of the matrix $\widehat{\mathbf{A}}$, has been proposed by Waterman [21], and can also be found in Flatau and Stephens [22], and McGarragh and Gabriel [37]. The idea is to consider the similarity transformation $\widetilde{\mathbf{A}} = \mathbf{L}^{-1}\widehat{\mathbf{A}}\mathbf{L}$, with

$$\mathbf{L} = \sqrt{\frac{1}{2}}\begin{bmatrix}\mathbf{I} & \mathbf{I}\\ \mathbf{I} & -\mathbf{I}\end{bmatrix} \text{ and } \mathbf{L}^{-1} = \sqrt{\frac{1}{2}}\begin{bmatrix}\mathbf{I} & \mathbf{I}\\ \mathbf{I} & -\mathbf{I}\end{bmatrix}, \tag{186}$$

yielding $\exp(-\widehat{\mathbf{A}}\tau) = \mathbf{L}\exp(-\widetilde{\mathbf{A}}\tau)\mathbf{L}^{-1}$, $\mathbf{D}_n(\widehat{\mathbf{A}}\tau) = \mathbf{L}\mathbf{D}_n(\widetilde{\mathbf{A}}\tau)\mathbf{L}^{-1}$, and $\mathbf{N}_n(\widehat{\mathbf{A}}\tau) = \mathbf{L}\mathbf{N}_n(\widetilde{\mathbf{A}}\tau)\mathbf{L}^{-1}$. Thus, by this similarity transformation, we have to compute powers of $\widetilde{\mathbf{A}}$, for which we find

$$\widetilde{\mathbf{A}}^{2k} = \begin{bmatrix}\mathbf{U} & 0\\ 0 & \mathbf{U}^T\end{bmatrix}^k = \begin{bmatrix}\mathbf{U}^k & 0\\ 0 & (\mathbf{U}^T)^k\end{bmatrix}, \tag{187}$$

$$\widetilde{\mathbf{A}}^{2k+1} = \widetilde{\mathbf{A}}^{2k}\widetilde{\mathbf{A}} = \begin{bmatrix}0 & \mathbf{U}^k\widehat{\mathbf{A}}_-\\ (\mathbf{U}^T)^k\widehat{\mathbf{A}}_+ & 0\end{bmatrix}, \tag{188}$$

Table 2: Relative errors ε_R and ε_T for the series approximations (Padé and Taylor approximations with $n = 5$) and the exponential infinitesimal generator initialization (EIGI).

		Series approximations		EIGI	
$\overline{\tau}$	$\|\|\widehat{\mathbf{A}}\overline{\tau}\|\|_2$	ε_R	ε_T	ε_R	ε_T
5.0e-4	0.30	2.12e-3	1.85e-3	1.16e-2	1.18e-2
1.0e-3	0.66	4.38e-3	3.87e-3	2.15e-2	2.29e-2
2.0e-3	1.22	9.47e-3	1.14e-2	3.65e-2	4.48e-2
3.0e-3	1.83	1.54e-2	2.85e-2	4.58e-2	6.91e-2
4.0e-3	2.44	2.23e-2	6.15e-2	5.12e-2	9.79e-2
5.0e-3	3.05	2.98e-2	9.18e-2	5.45e-2	1.31e-1

where $\mathbf{U} = \widehat{\mathbf{A}}_-\widehat{\mathbf{A}}_+$, and $\widehat{\mathbf{A}}_\pm = \widehat{\mathbf{A}}_{11} \pm \widehat{\mathbf{A}}_{12}$.

The accuracy of the reflection and transmission matrices computed by the Padé and Taylor series approximations is high if $||\widehat{\mathbf{A}}\overline{\tau}|| \le 1$; thus, for small values of $\overline{\tau}$. Therefore, the series approximations can be used for the initialization of the doubling method in radiative transfer. The doubling method is equivalent to the so-called scaling and squaring technique for reducing the norm of a matrix and exploits a fundamental property unique to the exponential function, namely

$$e^{-\widehat{\mathbf{A}}\overline{\tau}} = (e^{-\widehat{\mathbf{A}}\overline{\tau}/2^s})^{2^s}.$$

In practice, we choose s as the smallest power of 2 for which $||\widehat{\mathbf{A}}\overline{\tau}||/2^s \le 1$, evaluate the reflection and transmission matrices of the layer of optical thickness $\overline{\tau}/2^s$ by means of series approximations, and use the doubling equations to build up the reflection and transmission matrices of the layer of optical thickness $\overline{\tau}$. As compared to other initialization methods based on the single scattering approximation, e.g., the exponential infinitesimal generator initialization (EIGI) given by [41]

$$[\widehat{\mathbf{R}}_{\mathrm{EIGI}}]_{kl} = \mu_k(1 - e^{-\frac{1}{\mu_k}\overline{\tau}})[\widehat{\mathbf{A}}_{12}]_{kl},$$
$$[\widehat{\mathbf{T}}_{\mathrm{EIGI}}]_{kl} = \mu_k(1 - e^{-\frac{1}{\mu_k}\overline{\tau}})[\widehat{\mathbf{A}}^0_{11}]_{kl} + \delta_{kl}e^{-\frac{1}{\mu_k}\overline{\tau}},$$

with $\widehat{\mathbf{A}}^0_{11} = \widehat{\mathbf{A}}_{11} + \mathbf{M}$ (see (292) below), the series approximations are more accurate and so, reduce the number of doubling needed. This result can be inferred from Table 2, in which the relative errors in the discrete approximations to the reflection and transmission functions for different values of $||\widehat{\mathbf{A}}\overline{\tau}||_2$ are illustrated. Note that the Padé and Taylor approximations yield the same relative errors, and that the relative errors are of about 1% for $||\widehat{\mathbf{A}}\overline{\tau}||_2 \le 1$.

It is a simple exercise to combine the Padé approximation with the Taylor series approximation to derive an approximation of the layer equation in the discrete ordinate method with matrix exponential. Starting from the solution representation (18) and using the Padé approximation for $\exp(-\widehat{\mathbf{A}}\overline{\tau})$ and the Taylor series approximation for $\exp[-\widehat{\mathbf{A}}(\overline{\tau} - \tau)]$, gives the nth-order approximation of the layer equation (compare to (56))

$$\mathbf{D}_n(\widehat{\mathbf{A}}\overline{\tau})\,\widehat{\mathbf{i}}_{\mathrm{d}}(\overline{\tau}) = \mathbf{N}_n(\widehat{\mathbf{A}}\overline{\tau})\,\widehat{\mathbf{i}}_{\mathrm{d}}(0) - \mathbf{B}_n(\overline{\tau})\,\widehat{\mathbf{b}}, \tag{189}$$

where

$$\begin{aligned}\mathbf{B}_n(\overline{\tau}) &= \mathbf{D}_n(\widehat{\mathbf{A}}\overline{\tau})\int_0^{\overline{\tau}} e^{-\widehat{\mathbf{A}}(\overline{\tau}-\tau)}e^{-\tau/\mu_0}\mathrm{d}\tau \\ &= I_0(\overline{\tau})\mathbf{I} + \sum_{k=1}^{n}\Big[c_k\overline{\tau}^k I_0(\overline{\tau}) + \frac{(-1)^k}{k!}I_k(\overline{\tau}) \\ &+ \sum_{l=1}^{\min(k-1,n)}(1-\delta_{k1})\frac{(-1)^{k-l}}{(k-l)!}c_l\overline{\tau}^l I_{k-l}(\overline{\tau})\Big]\widehat{\mathbf{A}}^k,\end{aligned} \tag{190}$$

and

$$I_0(\overline{\tau}) = \int_0^{\overline{\tau}} e^{-\tau/\mu_0}\mathrm{d}\tau = \frac{1 - e^{-\overline{\tau}/\mu_0}}{(1/\mu_0)}, \tag{191}$$

$$I_k(\overline{\tau}) = \int_0^{\overline{\tau}} (\overline{\tau} - \tau)^k e^{-\tau/\mu_0}\mathrm{d}\tau' = \frac{\overline{\tau}^k}{(1/\mu_0)} - \frac{k e^{-\overline{\tau}/\mu_0}}{(1/\mu_0)}I_{k-1}(\overline{\tau}),\ k \ge 1. \tag{192}$$

Note that because of

$$I_k(\overline{\tau}) = O\Big(\frac{\overline{\tau}^{k+1}}{k+1}\Big),\ \overline{\tau} < 1,$$

each term in the sum (190) behaves as $O(\overline{\tau}^{k+1}\widehat{\mathbf{A}}^k)$. The first-order approximation of (189) is equivalent to the finite-difference method described by Lenoble [5], and used in atmospheric remote sensing by Rozanov et al. [42].

5.4. *Reflection and transmission matrices of a homogeneous thick layer*

When the optical thickness is sufficiently large, the reflection and transmission matrices can be expressed by simple analytical expressions known as the asymptotic theory of thick layers. This analytical model is much faster and more convenient for theoretical considerations than numerical models based on discrete ordinate schemes.

5.4.1. *Asymptotic theory*

In the classical asymptotic theory, the reflection and transmission functions for optically thick atmospheres are given by [32]:

$$R_m(\mu,\mu';\overline{\tau}) = R_{\infty m}(\mu,\mu') - \delta_{m0}\frac{\mathtt{m}\mathtt{l}e^{-2\mathtt{k}\overline{\tau}}}{1 - \mathtt{l}^2e^{-2\mathtt{k}\overline{\tau}}}K(\mu)K(\mu'), \tag{193}$$

$$T_m(\mu,\mu';\overline{\tau}) = \delta_{m0}\frac{\mathtt{m}e^{-\mathtt{k}\overline{\tau}}}{1 - \mathtt{l}^2e^{-2\mathtt{k}\overline{\tau}}}K(\mu)K(\mu'). \tag{194}$$

Here, $\mathtt{k}$ is the diffusion exponent describing the attenuation of the radiation in the diffusion domain and being defined as the smallest positive eigenvalue of the equation

$$(1 - \mathtt{k}\mu)\,i(\mu) = \frac{\omega}{2}\int_{-1}^{1} p_0(\mu,\mu')\,i(\mu')\mathrm{d}\mu', \tag{195}$$

while $i(\mu)$ is the corresponding eigenfunction, or the diffusion pattern, satisfying the Sobolev-van de Hulst relation

$$i(-\mu) = 2\int_0^1 R_{\infty 0}(\mu,\mu')\,i(\mu')\mu'\mathrm{d}\mu' \tag{196}$$

and the normalization condition

$$\frac{1}{2}\int_{-1}^{1} i(\mu)\mathrm{d}\mu = 1. \tag{197}$$

From (195), (197), and the normalization condition of the phase functions $\int_{-1}^{1} p_0(\mu,\mu')\mathrm{d}\mu' = 2$, we find the following expression for the diffusion exponent,

$$\mathtt{k} = \frac{2(1-\omega)}{\int_{-1}^{1} i(\mu)\mu\mathrm{d}\mu}. \tag{198}$$

The escape function $K(\mu)$ is given by the relation

$$\mathtt{m}K(\mu) = i(\mu) - 2\int_0^1 R_{\infty 0}(\mu,\mu')i(-\mu')\mu'\mathrm{d}\mu' \tag{199}$$

and satisfies the normalization condition

$$2\int_0^1 K(\mu)i(\mu)\mu\mathrm{d}\mu = 1, \tag{200}$$

where the constant $\mathtt{m}$ is defined by

$$\mathtt{m} = 2\int_{-1}^{1} i^2(\mu)\mu\mathrm{d}\mu. \tag{201}$$

Finally, the constant $\mathtt{l}$, also known as the negative internal reflection coefficient, is computed as

$$\mathtt{l} = 2\int_0^1 K(\mu)i(-\mu)\mu\mathrm{d}\mu. \tag{202}$$

Relations (193) and (194) show that R depends on the azimuthal angle through the reflection function of a semi-infinite atmosphere R_∞, and that T is azimuthally independent. For a layer with an underlying Lambertian surface of albedo A, the reflection and transmission functions $R_{Am}(\mu,\mu';\overline{\tau})$ and $T_{Am}(\mu,\mu';\overline{\tau})$ are given by (82) and (83), respectively, with

$$r(\mu,\overline{\tau}) = r_\infty(\mu) - \frac{\mathtt{mnl}}{1-\mathtt{l}^2\mathrm{e}^{-2\mathtt{k}\overline{\tau}}}\mathrm{e}^{-2\mathtt{k}\overline{\tau}}K(\mu), \tag{203}$$

$$t(\mu,\overline{\tau}) = \frac{\mathtt{mn}}{1-\mathtt{l}^2\mathrm{e}^{-2\mathtt{k}\overline{\tau}}}\mathrm{e}^{-\mathtt{k}\overline{\tau}}K(\mu), \tag{204}$$

$$r_s(\overline{\tau}) = r_{s\infty} - \frac{\mathtt{mn}^2\mathtt{l}}{1-\mathtt{l}^2\mathrm{e}^{-2\mathtt{k}\overline{\tau}}}\mathrm{e}^{-2\mathtt{k}\overline{\tau}}. \tag{205}$$

Here, r_∞ and $r_{s\infty}$ are the plane albedo and spherical albedo of a semi-infinite atmosphere**,** respectively, and $\mathtt{n}$ is the μ-weighted mean of the escape function

$$\mathtt{n} = 2\int_0^1 K(\mu)\mu\mathrm{d}\mu. \tag{206}$$

In a discrete ordinate setting, the reflection matrix of a semi-infinite atmosphere can be obtained by solving the Ambartsumian nonlinear integral equation by simple iteration [35], while the diffusion pattern and the diffusion exponent can be obtained by solving the integral equation (195) in conjunction with (197) and (198) [43]. The constant $\mathtt{m}$ is then computed from (201), while $K(\mu)$, $\mathtt{l}$, and $\mathtt{n}$ follow from (199), (202), and (206), respectively. A different discrete ordinate approach, which is based on an eigendecomposition method for computing the reflection and transmission matrices has been proposed by Nakajima and King [27]. Here, we apply this approach to estimate $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$, given respectively by (97) and (98), in the limit of large $\overline{\tau}$. Choosing the square-root method for computing the spectral decomposition of the layer matrix $\widehat{\mathbf{A}}$, we begin by estimating the matrix product $(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1}$ for large values of the optical thickness. Using the identity (cf. (344)) $\widehat{\mathbf{Q}}_- = -\widehat{\mathbf{Q}}_+^{-T}\boldsymbol{\Lambda}$, where $\widehat{\mathbf{Q}}_+ = \widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-$, $\widehat{\mathbf{Q}}_- = \widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-$, and $\boldsymbol{\Lambda} = [\lambda_k]$, and setting

$$\widehat{\mathbf{Q}}_+^{-T} \overset{\text{not}}{=} \mathbf{Q} = [\mathbf{q}_1,\ldots,\mathbf{q}_N], \tag{207}$$

we obtain

$$\begin{aligned}&(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1}\\ &= (\mathbb{A}^{-1} - \mathbf{I}) + \mathbb{A}^{-1}\mathbf{Q}[\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Pi}_+\mathbb{B}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Pi}_+]\mathbf{Q}^T\mathbb{A}^{-1},\end{aligned} \tag{208}$$

where

$$\mathbb{A} = \frac{1}{2}(\mathbf{I} + \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T), \tag{209}$$

$$\mathbb{B} = \mathbf{Q}^T\mathbb{A}^{-1}\mathbf{Q}, \tag{210}$$

and

$$\boldsymbol{\Pi}_+ = \frac{1}{2}[\mathbf{I} - (\mathbf{I}-\boldsymbol{\Gamma})(\mathbf{I}+\boldsymbol{\Gamma})^{-1}] = [\pi_k^+], \tag{211}$$

$$\pi_k^+ = \frac{\mathrm{e}^{-\lambda_k\overline{\tau}}}{1+\mathrm{e}^{-\lambda_k\overline{\tau}}},\ k = 1,\ldots,N. \tag{212}$$

Note that all information about the eigenvectors of $\widehat{\mathbf{A}}$ are encapsulated in the matrix $\mathbf{Q}$. Two comments are in order:

1. As $\mathbb{A}$ and $\mathbb{B}$ are symmetric, we see that

$$\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Pi}_+\mathbb{B}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Pi}_+ = [(\boldsymbol{\Lambda}\boldsymbol{\Pi}_+)^{-1} - \mathbb{B}]^{-1} \tag{213}$$

and so, that $(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1}$ are symmetric.

2. If the eigendecomposition of $\widehat{\mathbf{A}}$ is computed by the Cholesky method, the key quantity is the matrix $\widehat{\mathbf{Q}}_-$; we use the identity (cf. (351)) $\widehat{\mathbf{Q}}_+^{-T} = -\widehat{\mathbf{Q}}_-\boldsymbol{\Lambda}$, set

$$\widehat{\mathbf{Q}}_- \overset{\text{not}}{=} \mathbf{Q} = [\mathbf{q}_1,\ldots,\mathbf{q}_N], \tag{214}$$

and obtain

$$\begin{aligned}&(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1}\\ &= (\mathbb{A}^{-1} - \mathbf{I}) + \mathbb{A}^{-1}\mathbf{Q}[(\mathbf{I} - \boldsymbol{\Pi}_+\boldsymbol{\Lambda}\mathbb{B})^{-1}\boldsymbol{\Pi}_+\boldsymbol{\Lambda}]\mathbf{Q}^T\mathbb{A}^{-1}\end{aligned} \tag{215}$$

Then, from (213) and the relation

$$(\mathbf{I} - \boldsymbol{\Pi}_+\boldsymbol{\Lambda}\mathbb{B})^{-1}\boldsymbol{\Pi}_+\boldsymbol{\Lambda} = [(\boldsymbol{\Lambda}\boldsymbol{\Pi}_+)^{-1} - \mathbb{B}]^{-1}, \tag{216}$$

we deduce that both methods are equivalent.

Considering (211) and (212), we assume that for large $\overline{\tau}$, we may approximate

$$\mathbf{\Pi}_+ \approx \pi_N^+ \begin{bmatrix} 0 & \cdots & 0 \\ & & \\ 0 & \cdots & 1 \end{bmatrix}, \tag{217}$$

where π_N^+ corresponds to the smallest eigenvalue λ_N in the set $\{\lambda_1, \lambda_2, ..., \lambda_N\}$. As a result, we obtain

$$(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_- \mathbf{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+ \mathbf{\Gamma})^{-1}$$
$$\approx (\mathbb{A}^{-1} - \mathbf{I}) + \frac{\mathtt{k} e^{-\mathtt{k}\overline{\tau}}}{1 + \mathtt{l} e^{-\mathtt{k}\overline{\tau}}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T \text{ as } \tau \to \infty, \tag{218}$$

where $\mathtt{k} = \lambda_N$ is the diffusion exponent,

$$\mathtt{l} = 1 - \lambda_N b_{NN} = 1 - \mathtt{k} q, \tag{219}$$

is the negative internal reflection coefficient, $q = b_{NN} = [\mathbb{B}]_{NN}$ is the extrapolation length, and

$$\widehat{\mathbf{k}}_N = \mathbb{A}^{-1} \mathbf{q}_N, \tag{220}$$

is a discrete approximation to the scaled escape function. The matrix product $(\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_- \mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+ \mathbf{\Gamma})^{-1}$ is estimated in a similar manner. We get

$$(\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_- \mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+ \mathbf{\Gamma})^{-1}$$
$$= (\mathbb{A}^{-1} - \mathbf{I}) + \mathbb{A}^{-1} \mathbf{Q} \left[\mathbf{\Lambda} (\mathbf{I} - \mathbf{\Pi}_- \mathbb{B} \mathbf{\Lambda})^{-1} \mathbf{\Pi}_- \right] \mathbf{Q}^T \mathbb{A}^{-1}, \tag{221}$$

with

$$\mathbf{\Pi}_- = \frac{1}{2} [\mathbf{I} - (\mathbf{I} + \mathbf{\Gamma})(\mathbf{I} - \mathbf{\Gamma})^{-1}] = [\pi_k^-], \tag{222}$$

$$\pi_k^- = -\frac{e^{-\lambda_k \overline{\tau}}}{1 - e^{-\lambda_k \overline{\tau}}}, \; k = 1, ..., N, \tag{223}$$

whence, under the assumption

$$\mathbf{\Pi}_- \approx \pi_N^- \begin{bmatrix} 0 & \cdots & 0 \\ & & \\ 0 & \cdots & 1 \end{bmatrix} \text{ as } \tau \to \infty, \tag{224}$$

we end up with

$$(\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_- \mathbf{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+ \mathbf{\Gamma})^{-1}$$
$$\approx (\mathbb{A}^{-1} - \mathbf{I}) - \frac{\mathtt{k} e^{-\mathtt{k}\overline{\tau}}}{1 - \mathtt{l} e^{-\mathtt{k}\overline{\tau}}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T \text{ as } \tau \to \infty. \tag{225}$$

Inserting (218) and (225) in (97) and (98), we obtain

$$\widehat{\mathbf{R}} = \widehat{\mathbf{R}}_\infty - \widehat{\mathtt{r}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T, \tag{226}$$
$$\widehat{\mathbf{T}} = \widehat{\mathtt{t}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T, \tag{227}$$

where

$$\widehat{\mathbf{R}}_\infty = \mathbb{A}^{-1} - \mathbf{I}, \tag{228}$$

is the scaled reflection matrix of a semi-infinite atmosphere, and

$$\widehat{\mathtt{r}} = \frac{\mathtt{k} \mathtt{l} e^{-2\mathtt{k}\overline{\tau}}}{1 - \mathtt{l}^2 e^{-2\mathtt{k}\overline{\tau}}}, \tag{229}$$
$$\widehat{\mathtt{t}} = \frac{\mathtt{k} e^{-\mathtt{k}\overline{\tau}}}{1 - \mathtt{l}^2 e^{-2\mathtt{k}\overline{\tau}}}. \tag{230}$$

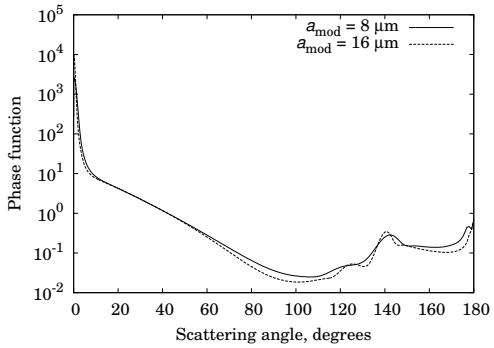


Figure 1: The phase functions for $a_{\mathrm{mod}} = 8$ μm and $a_{\mathrm{mod}} = 16$ μm.

By numerical simulation it can be shown that the diffusion exponent $\mathtt{k} = \lambda_N$ increases as the azimuthal mode m increases. In this limit, the scalars $\widehat{\mathtt{r}}$ and $\widehat{\mathtt{t}}$ become very small, and the contributions of the terms $\widehat{\mathtt{r}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T$ and $\widehat{\mathtt{t}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T$ in the expressions of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$, respectively, are damped. In the asymptotic theory, these terms are neglected for $m > 0$, in which case, (226) and (227) simplifies to

$$\widehat{\mathbf{R}} = \widehat{\mathbf{R}}_\infty - \delta_{m0} \widehat{\mathtt{r}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T, \tag{231}$$
$$\widehat{\mathbf{T}} = \delta_{m0} \widehat{\mathtt{t}} \widehat{\mathbf{k}}_N \widehat{\mathbf{k}}_N^T. \tag{232}$$

Thus, the diffusion exponent $\mathtt{k}$ and the constant $\mathtt{l}$ in (193) and (194) correspond to the azimuthal mode $m = 0$. As a final step, we compute the discrete approximations of the reflection and transmission functions given by (193) and (194), respectively, by means of (117) and (118), respectively. The remainder functions and constants that occur in the asymptotic theory can be obtained by following strictly the derivation of Nakajima and King [27]. Their expressions are given in Appendix 1.

As an application of the discrete ordinate model of the asymptotic theory we derive parameterizations of the escape function K and the asymptotic constants $\mathtt{m}$, $\mathtt{l}$, $\mathtt{n}$, $r_{s\infty}$ and $\mathtt{k}$. Such kind of parameterizations, which speed up the computations, are used in the MODIS algorithm for the retrieval of cloud optical thickness and the droplet/crystal size [44], and in the SemiAnalytical CloUd Retrieval Algorithm (SACURA) for the retrieval of cloud top height and cloud geometrical thickness from measurements in the oxygen A band [45, 46]. For the EPIC instrument, we consider a discrete set of water-cloud models characterized by a Gamma size distribution of parameter $\alpha = 6$ and different modal radii a_{mod}. The phase functions for $a_{\mathrm{mod}} = 8$ μm and $a_{\mathrm{mod}} = 16$ μm are illustrated in Figure 1.

For the escape function, the parameterization parameters are ω and a_{mod}. We look for polynomial parameterizations of the form $K(\mu, \omega, a_{\mathrm{mod}}) = P_{\mathtt{K}}(\mu, \omega, a_{\mathrm{mod}}) K_0(\mu)$ with $P_{\mathtt{K}}(\mu, \omega, a_{\mathrm{mod}}) = \sum_{k=0}^{N_{\mathtt{K}}} H_k(\omega, a_{\mathrm{mod}}) \mu^k$, and $K_0(\mu) = \sum_{k=0}^{N_{\mathtt{K}}} H_k^0 \mu^k$, where K_0 is the escape function for conservative scattering ($\omega = 1$), and $N_{\mathtt{K}}$ is the order of the approximation polynomial. The parameterizations of $K(\mu, \omega, a_{\mathrm{mod}})$ do not change significantly with a_{mod}; the coefficients $H_k(\omega, a_{\mathrm{mod}})$, which are stored in a look-up table, are illustrated in Figure 2 for

Table 3: Coefficients H_k^0

k	0	1	2	3	4
H_k^0	0.362	1.196	-0.5352	0.349	-0.09368

$a_{\text{mod}} = 8$ μm, and $a_{\text{mod}} = 16$ μm, while the coefficients H_k^0 are given in Table 3. The approximations error is smaller than 10^{-3} for $K(\mu, \omega, a_{\text{mod}})$, and smaller than 10^{-5} for $K_0(\mu)$. Coming to the asymptotic constants $\mathtt{m}$, $\mathtt{l}$, $\mathtt{n}$, $r_{s\infty}$ and $\mathtt{k}$, we note that for a Henyey-Greenstein phase function, King [47] derived parameterizations of these quantities in terms of the similarity parameter

$$s = \sqrt{\frac{1-\omega}{1-\omega g}},$$

where g is the asymmetry parameter. In this regard, we take s and a_{mod} as parameterization parameters (s reproduces the variability in ω for a given a_{mod}). As for the escape function, we found that the parameterizations of the asymptotic constant are almost insensitive to a_{mod}, but are slightly different from those of King [47], especially for $\mathtt{m}$, $\mathtt{n}$, and $\mathtt{k}$. The results in Figure 3, corresponding to $a_{\text{mod}} = 8$ μm ($g = 0.853$) and $a_{\text{mod}} = 16$ μm ($g = 0.867$), certify this statement. For $a_{\text{mod}} = 8$ μm, the parameterizations read as

$$\mathtt{m}(s) = (1 + 1.268s)\ln\left[\frac{1 + 1.31s - 5.694s^2 + 3.73s^3}{(1 - 1.070s)(1-s)^2}\right],$$

$$\mathtt{l}(s) = \frac{(1 - 0.7616s)(1-s)}{1 + 0.5897s},$$

$$\mathtt{n}(s) = \sqrt{\frac{(1 + 0.0031s)(1-s)}{1 + 1.4267s}},$$

$$r_{s\infty}(s) = \frac{(1 - 0.1239s)(1-s)}{1 + 1.1867s},$$

$$\frac{\mathtt{k}(s)}{1-g} = (1 + 0.3977s)\ln\left[\frac{1 + 8.4924s}{(1 + 7.2543s)(1 - 1.0265s)}\right].$$

5.4.2. Higher-order corrections

The azimuthal independent parts of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ have been derived by neglecting the azimuthal modes $m > 0$, and the contributions of the terms corresponding to the eigenvalues larger than λ_N. Higher-order corrections can be obtained by considering all azimuthal modes, and by approximating

$$\mathbf{\Pi}_{\pm} \approx \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \cdots & \pi^{\pm}_{N-K} & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \cdots & \pi^{\pm}_{N-1} & 0 \\ 0 & \cdots & 0 & \cdots & 0 & \pi^{\pm}_{N} \end{bmatrix}, \tag{233}$$

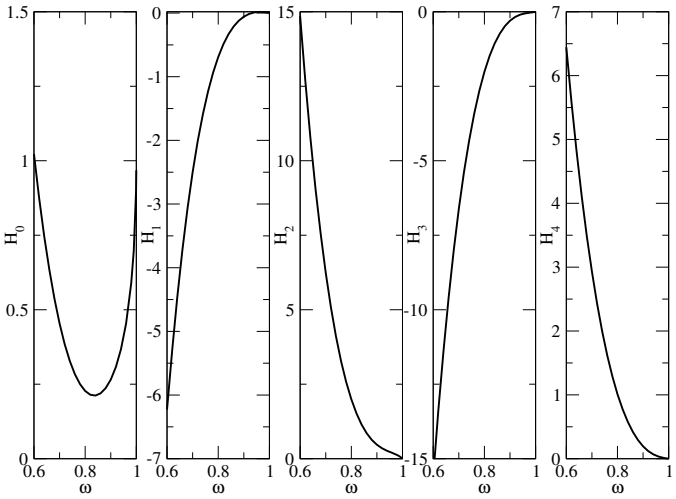


Figure 2: Coefficients $H_k(\omega, a_{\text{mod}})$ for $a_{\text{mod}} = 8$ μm and $a_{\text{mod}} = 16$ μm.

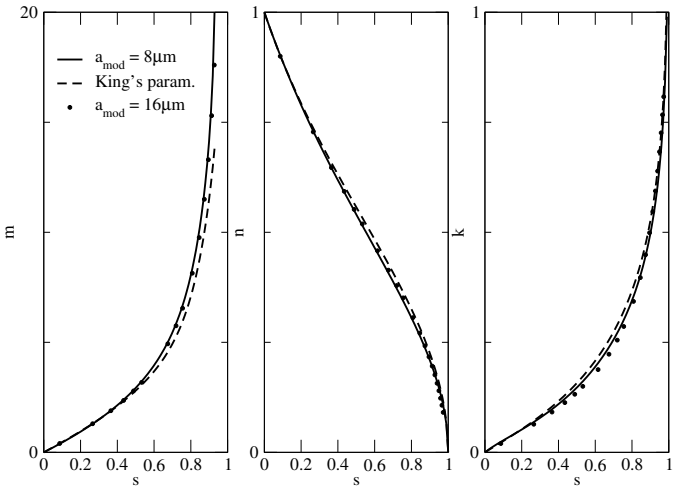


Figure 3: Constants $\mathtt{m}$ and $\mathtt{n}$, and the diffusion exponent $\mathtt{k}$. The curves correspond to $a_{\text{mod}} = 8$ μm, $a_{\text{mod}} = 16$ μm, and King's parametrization.

where π_k^+ and π_k^- are given by (212) and (223), respectively, and the integer $K \geq 0$ (not to be confused with the escape function $K(\mu)$) gives the order of the approximation. For $\mathbf{\Pi}_{\pm}$ of (233), the inverse $(\mathbf{I} - \mathbf{\Pi}_{\pm}\mathbb{B}\mathbf{\Lambda})^{-1}$ cannot be computed analytically, and an additional assumption has to be met. Setting

$$\mathbf{I} - \mathbf{\Pi}_{\pm}\mathbb{B}\mathbf{\Lambda} = \mathbb{C}^0_{\pm} - \mathbb{C}_{\pm},$$

where $\mathbb{C}^0_{\pm}$ are the matrices $\mathbf{I} - \mathbf{\Pi}_{\pm}\mathbb{B}\mathbf{\Lambda}$ of the asymptotic model (with $\mathbf{\Pi}_+$ and $\mathbf{\Pi}_-$ given by (217) and (224), respectively), we approximate the Neumann series of the inverse up to the first order as follows:

$$(\mathbf{I} - \mathbf{\Pi}_{\pm}\mathbb{B}\mathbf{\Lambda})^{-1} = (\mathbb{C}^0_{\pm} - \mathbb{C}_{\pm})^{-1} \approx (\mathbb{C}^0_{\pm})^{-1}[\mathbf{I} + \mathbb{C}_{\pm}(\mathbb{C}^0_{\pm})^{-1}].$$

In this context, the reflection and transmission matrices can be written as

$$\mathbb{R} = \mathbb{R}_0 + \mathbb{R}_K, \tag{234}$$

$$\mathbb{T} = \mathbb{T}_0 + \mathbb{T}_K, \tag{235}$$

where $\mathbb{R}_0$ and $\mathbb{T}_0$ correspond to $K=0$ and are the reflection and transmission matrices of the asymptotic model as given by (231) and (232), respectively. The higher-order correction matrices $\mathbb{R}_K$ and $\mathbb{T}_K$ correspond to $K \geq 1$, and are expressed in terms of the scaled matrices

$$\widehat{\mathbf{R}}_K = \frac{1}{2}\lambda_N \mathrm{e}^{-\lambda_N \overline{\tau}}\Big(\frac{1}{1+1\mathrm{e}^{-\lambda_N \overline{\tau}}}\mathbb{E}^+_{N-K} - \frac{1}{1-1\mathrm{e}^{-\lambda_N \overline{\tau}}}\mathbb{E}^-_{N-K}\Big) + \frac{1}{2}\lambda_N \sum_{k=1}^{K} \mathrm{e}^{-\lambda_{N-k}\overline{\tau}}\Big(\frac{1}{1+\mathrm{e}^{-\lambda_{N-k}\overline{\tau}}}\mathbb{F}^+_{N-k} - \frac{1}{1-\mathrm{e}^{-\lambda_{N-k}\overline{\tau}}}\mathbb{F}^-_{N-k}\Big), \tag{236}$$

$$\widehat{\mathbf{T}}_K = \frac{1}{2}\lambda_N \mathrm{e}^{-\lambda_N \overline{\tau}}\Big(\frac{1}{1+1\mathrm{e}^{-\lambda_N \overline{\tau}}}\mathbb{E}^+_{N-K} + \frac{1}{1-1\mathrm{e}^{-\lambda_N \overline{\tau}}}\mathbb{E}^-_{N-K}\Big) + \frac{1}{2}\lambda_N \sum_{k=1}^{K} \mathrm{e}^{-\lambda_{N-k}\overline{\tau}}\Big(\frac{1}{1+\mathrm{e}^{-\lambda_{N-k}\overline{\tau}}}\mathbb{F}^+_{N-k} + \frac{1}{1-\mathrm{e}^{-\lambda_{N-k}\overline{\tau}}}\mathbb{F}^-_{N-k}\Big), \tag{237}$$

as follows:

$$\mathbb{R}_K = \frac{1}{2}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{R}}_K\mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}, \tag{238}$$

$$\mathbb{T}_K = \frac{1}{2}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{T}}_K\mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}. \tag{239}$$

The derivation of (236) and (237) is lengthy but straightforward. Here, we give only the final expressions. The matrices $\mathbb{E}^+_{N-K}$ and $\mathbb{F}^+_{N-k}$, corresponding to $\mathbf{\Pi}_+$, are computed as

$$\mathbb{E}^+_{N-K} = E^+_{NN}\widehat{\mathbf{k}}_N\widehat{\mathbf{k}}^T_N + \sum_{k=1}^{K} E^+_{N-kN}\widehat{\mathbf{k}}_{N-k}\widehat{\mathbf{k}}^T_N \tag{240}$$

and

$$\mathbb{F}^+_{N-k} = F^+_{NN-k}\widehat{\mathbf{k}}_N\widehat{\mathbf{k}}^T_{N-k} + \sum_{l=1}^{K} F^+_{N-lN-k}\widehat{\mathbf{k}}_{N-l}\widehat{\mathbf{k}}^T_{N-k}, \tag{241}$$

respectively, with

$$\widehat{\mathbf{k}}_{N-k} = \mathbb{A}^{-1}\mathbf{q}_{N-k}, \tag{242}$$

for $k=1,\ldots,K$. The scalars E^+_{N-kN}, E^+_{NN}, F^+_{N-lN-k}, and F^+_{NN-k} are

$$E^+_{N-kN} = f^+_N g_{N-k} X^+_{N-kN}, \quad E^+_{NN} = f^+_N \sum_{k=1}^{K} x^+_{N-k} X^+_{N-kN}, \tag{243}$$

and

$$F^+_{N-lN-k} = g_{N-l} X^+_{N-lN-k}, \quad F^+_{NN-k} = \sum_{l=1}^{K} x^+_{N-l} X^+_{N-lN-k}, \tag{244}$$

respectively, where

$$X^+_{N-kN-l} = \delta_{N-kN-l} + Y^+_{N-kN-l}, \quad X^+_{N-kN} = \delta_{N-kN} + Y^+_{N-kN}, \tag{245}$$

and

$$Y^+_{N-kN-l} = \pi^+_{N-k} b_{N-kN-l}\lambda_{N-l} + x^+_{N-l}\pi^+_{N-k} b_{N-kN}\lambda_N, \quad Y^+_{N-kN} = \frac{\pi^+_{N-k} b_{N-kN}\lambda_N}{f^+_N}, \tag{246}$$

for $k,l = 1,\ldots,K$. In (243)-(246), π^+_{N-k} are given by (212), and

$$x^+_{N-k} = \frac{\pi^+_N b_{NN-k}\lambda_{N-k}}{f^+_N}, \quad f^+_N = 1-\pi^+_N b_{NN}\lambda_N, \quad g_{N-k} = \frac{\lambda_{N-k}}{\lambda_N}, \tag{247}$$

for $k=1,\ldots,K$. The matrices $\mathbb{E}^-_{N-K}$ and $\mathbb{F}^-_{N-k}$, corresponding to $\mathbf{\Pi}_-$, are computed by using (240)-(247), but with π^-_{N-k} of (223) in place of π^+_{N-k}.

Although the discrete ordinate model of the asymptotic theory is still based on an eigendecomposition method, the computation of the reflection and transmission matrices by means of (226), (227), (236), and (237) avoids the matrix inversions and matrix multiplications of (97) and (98).

5.4.3. *Waterman's approximation*

In [21], Waterman derived analytical solutions for the transmission and reflection matrices in the limiting case $\overline{\tau} \gg 1$. The starting point of Waterman's derivation are (108) and (109), which we will write as

$$\widehat{\mathbf{T}} = \widehat{\mathbf{T}}_0(\mathbf{I} - \widehat{\mathbf{V}}_+\mathbf{\Gamma}\overline{\mathbf{V}}^T_+\overline{\mathbf{V}}^{-T}_-\mathbf{\Gamma}\widehat{\mathbf{V}}^{-1}_-)^{-1}, \tag{248}$$

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{V}}_+\mathbf{\Gamma}^{-1}\overline{\mathbf{V}}^T_- - \widehat{\mathbf{V}}_-\mathbf{\Gamma}\overline{\mathbf{V}}^T_+)\widehat{\mathbf{T}}, \tag{249}$$

with

$$\widehat{\mathbf{T}}_0 = \overline{\mathbf{V}}^{-T}_-\mathbf{\Gamma}\widehat{\mathbf{V}}^{-1}_-. \tag{250}$$

Considering the Neuman series for the inverse in (248), we obtain

$$\widehat{\mathbf{T}} = \widehat{\mathbf{T}}_0[\mathbf{I} + \widehat{\mathbf{V}}_+\mathbf{\Gamma}\overline{\mathbf{V}}^T_+\overline{\mathbf{V}}^{-T}_-\mathbf{\Gamma}\widehat{\mathbf{V}}^{-1}_- + \ldots], \tag{251}$$

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{V}}_+\mathbf{\Gamma}^{-1}\overline{\mathbf{V}}^T_- - \widehat{\mathbf{V}}_-\mathbf{\Gamma}\overline{\mathbf{V}}^T_+)\widehat{\mathbf{T}}. \tag{252}$$

Accounting of (50), we set (not to be confused with $\mathbb{A}$ of (209))

$$\mathbb{A} = \overline{\mathbf{V}}^T_+\overline{\mathbf{V}}^{-T}_- = \widehat{\mathbf{V}}^{-1}_-\widehat{\mathbf{V}}_+ \tag{253}$$

and use the identity (cf. (50))

$$\widehat{\mathbf{V}}_+\widehat{\mathbf{V}}^{-1}_-\widehat{\mathbf{V}}_+ = \widehat{\mathbf{V}}_- - \overline{\mathbf{V}}^{-T}_-, \tag{254}$$

to obtain the zeroth- and first-order approximations, namely

$$\widehat{\mathbf{T}}_0 = \overline{\mathbf{V}}^{-T}_-\mathbf{\Gamma}\widehat{\mathbf{V}}^{-1}_-, \tag{255}$$

$$\widehat{\mathbf{R}}_0 = \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}^{-1}_-, \tag{256}$$

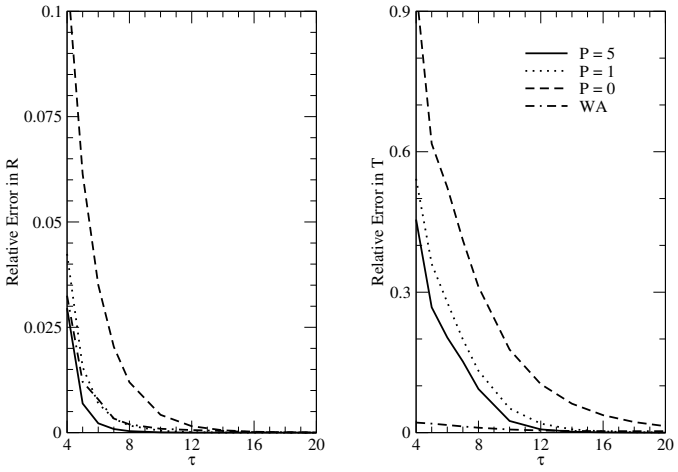


Figure 4: Relative errors in the discrete approximations to the reflection and transmission functions computed by using (234)-(237) with $K = 5$, $K = 1$, and $K = 0$ (asymptotic model), and the first-order Waterman's approximation (WA) given by (257)-(258).

and

$$\widehat{\mathbf{T}}_1 = \overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}\widehat{\mathbf{V}}_-^{-1} + \overline{\mathbf{V}}_-^{-T}(\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma})\widehat{\mathbf{V}}_-^{-1}, \tag{257}$$

$$\widehat{\mathbf{R}}_1 = \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1} - \overline{\mathbf{V}}_-^{-T}(\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma})\widehat{\mathbf{V}}_-^{-1} - \widehat{\mathbf{V}}_-(\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma}\mathbb{A}\mathbf{\Gamma})\widehat{\mathbf{V}}_-^{-1}, \tag{258}$$

respectively. Note that in Waterman's derivation, the last term in the expression of $\widehat{\mathbf{R}}_1$ is not present. In (255)-(258), $\mathbf{\Gamma}$ includes all exponential terms $\exp(-\lambda_k\overline{\tau})$ and not only the dominant one $\exp(-\lambda_N\overline{\tau})$. For large $\overline{\tau}$, relations (257) and (258) are similar to (227) and (226), respectively. To show this, we use the relation $\overline{\mathbf{V}}_- = \widehat{\mathbf{V}}_-\mathbb{S}$, where $\mathbb{S}$ is a diagonal matrix of plus and minus ones, set $\widehat{\mathbf{V}}_-^{-T} = [\widetilde{\mathbf{v}}_1, ..., \widetilde{\mathbf{v}}_N]$, approximate $\mathbf{\Gamma} \approx \mathrm{diag}[0, ..., 0, \exp(-\lambda_N\overline{\tau})]$, and neglect the second term in the expression of $\widehat{\mathbf{T}}_1$ as well as the third term in the expression of $\widehat{\mathbf{R}}_1$. We obtain

$$\widehat{\mathbf{T}}_1 = s_N e^{-\lambda_N\overline{\tau}}\widetilde{\mathbf{v}}_N\widetilde{\mathbf{v}}_N^T, \tag{259}$$

$$\widehat{\mathbf{R}}_1 = \widehat{\mathbf{R}}_\infty - s_N a_{NN} e^{-2\lambda_N\overline{\tau}}\widetilde{\mathbf{v}}_N\widetilde{\mathbf{v}}_N^T, \tag{260}$$

where $s_N = [\mathbb{S}]_{NN}$, $a_{NN} = [\mathbb{A}]_{NN}$, and $\widehat{\mathbf{R}}_\infty = \widehat{\mathbf{V}}_+\widehat{\mathbf{V}}_-^{-1}$. Comparing (259) and (260) with (227) and (226) under the assumptions $\widehat{\mathbf{r}} \approx \mathbf{k}\mathbf{1}\exp(-2\mathbf{k}\overline{\tau})$ and $\widehat{\mathbf{t}} \approx \mathbf{k}\exp(-\mathbf{k}\overline{\tau})$ for $\overline{\tau} \gg 1$ (the Taylor expansions in $x = \exp(-\mathbf{k}\overline{\tau})$ of (229) and (230) for small x), we see that for the azimuthal mode $m = 0$, we must have that $s_N = 1$, $\lambda_N = \mathbf{k}$, $\widetilde{\mathbf{v}}_N = \sqrt{\mathbf{k}}\widehat{\mathbf{k}}_N$, and $a_{NN} = \mathbf{1}$. Thus, by using Waterman's technique we were able to derive new representations for $\widehat{\mathbf{R}}_\infty$, $\widehat{\mathbf{k}}_N$ and $\mathbf{1}$.

In Figure 4 we show the relative errors in the discrete approximations to the reflection and transmission functions computed by using (234)-(237) with $K = 5$, $K = 1$, and $K = 0$ (asymptotic model), and the first-order Waterman's approximation (257)-(258). Obviously, Waterman's approximation yields the smallest errors in the transmission matrix, and sufficiently small errors in the reflection matrix.

6. Matrix Riccati equations

Bellman et al. [20] formulated the radiative transfer problem as an initial value problem via a pair of nonlinear matrix differential equations (matrix Riccati equations) which describe the reflection and transmission matrices in a plane-parallel geometry. The derivation of Bellman et al. [20] is based on the invariant imbedding technique, and for this reason, the invariant imbedding is usually a synonym for the matrix Riccati equations method.

Flatau and Stephens [22], showed that for an inhomogeneous atmosphere, the matrix Riccati equations can be derived by means of an approach based on the propagator operator, that is, by an approach which is close related to the matrix exponential method. The main problem which arises in the case of an inhomogeneous atmosphere is that in view of

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(e^{\mathbf{X}(\tau)}) \neq \mathbf{X}'(\tau)e^{\mathbf{X}(\tau)}, \tag{261}$$

the homogeneous solution of radiative transfer equation

$$\frac{\mathrm{d}\widehat{\mathbf{i}}_\mathrm{h}}{\mathrm{d}\tau}(\tau) = -\widehat{\mathbf{A}}(\tau)\widehat{\mathbf{i}}_\mathrm{h}(\tau), \tag{262}$$

with the initial condition $\widehat{\mathbf{i}}_\mathrm{h}(0)$ and a continuous matrix $\widehat{\mathbf{A}}(\tau)$ on $0 \le \tau \le \overline{\tau}$, cannot be expressed in terms of the matrix exponential, i.e.,

$$\widehat{\mathbf{i}}_\mathrm{h}(\tau) \neq e^{-\int_0^\tau \widehat{\mathbf{A}}(\tau')\mathrm{d}\tau'}\widehat{\mathbf{i}}_\mathrm{h}(0). \tag{263}$$

This fact should not be discouraged because in the matrix Riccati equation method we need only a formal solution representation of (262). The formal solution is [22]

$$\widehat{\mathbf{i}}_\mathrm{h}(\tau) = \mathbf{P}_-(\tau, 0)\widehat{\mathbf{i}}_\mathrm{h}(0), \tag{264}$$

where the downward propagator $\mathbf{P}_-(\tau, 0)$ is a generalization of the propagator $\mathbf{P}(\tau) = \exp(-\widehat{\mathbf{A}}\tau)$ in the case of an inhomogeneous atmosphere. Inserting (264) in (262), we find that the downward propagator solves the differential equation

$$\frac{\mathrm{d}\mathbf{P}_-}{\mathrm{d}\tau}(\tau, 0) = -\widehat{\mathbf{A}}(\tau)\mathbf{P}_-(\tau, 0), \tag{265}$$

with the initial condition $\mathbf{P}_-(0, 0) = \mathbf{I}_{2N}$. The subscript - means that the propagation of the solution occurs from the level 0 (second argument of $\mathbf{P}_-$) downward to the level τ (first argument of $\mathbf{P}_-$). The propagation of the solution from the bottom to the top of the atmosphere is described by the upward propagator $\mathbf{P}_+(\tau, \overline{\tau})$ which is defined by $\widehat{\mathbf{i}}_\mathrm{h}(\tau) = \mathbf{P}_+(\tau, \overline{\tau})\widehat{\mathbf{i}}_\mathrm{h}(\overline{\tau})$, and solve (265) with the initial condition $\mathbf{P}_+(\overline{\tau}, \overline{\tau}) = \mathbf{I}_{2N}$.

In an inhomogeneous atmosphere, the radiation coming from above will be reflected and transmitted differently

than the radiation incident from below, so that in the downward scheme, the interaction principle for an inhomogeneous layer extending from the level 0 downward to the level τ, reads as

$$\left[\begin{array}{c}\widehat{\mathbf{i}}_{\mathrm{h}}^{+}(0)\\ \widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\tau)\end{array}\right]=\left[\begin{array}{cc}\widehat{\mathbf{R}}_{-}(\tau,0) & \widehat{\mathbf{T}}_{+}(\tau,0)\\ \widehat{\mathbf{T}}_{-}(\tau,0) & \widehat{\mathbf{R}}_{+}(\tau,0)\end{array}\right]\left[\begin{array}{c}\widehat{\mathbf{i}}_{\mathrm{h}}^{-}(0)\\ \widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\tau)\end{array}\right]. \tag{266}$$

The notation $\widehat{\mathbf{R}}_{-}(\tau,\tau_0)$ stands for the reflection matrix of a layer of optical thickness τ illuminated from above, and whose top is placed at τ_0. Note that in the upward scheme, the interaction principle equation relates the outcoming radiances $\widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\tau)$ and $\widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\overline{\tau})$ to the incoming radiances $\widehat{\mathbf{i}}_{\mathrm{h}}^{-}(\tau)$ and $\widehat{\mathbf{i}}_{\mathrm{h}}^{+}(\overline{\tau})$ through the reflection and transmission matrices $\widehat{\mathbf{R}}_{\pm}(\overline{\tau}-\tau,\tau)$ and $\widehat{\mathbf{T}}_{\pm}(\overline{\tau}-\tau,\tau)$, respectively. From (264) and (266), we find

$$\begin{aligned}&\mathbf{P}_{-}(\tau,0)\\ &=\left[\begin{array}{cc}\widehat{\mathbf{T}}_{+}^{-1}(\tau,0) & -\widehat{\mathbf{T}}_{+}^{-1}(\tau,0)\widehat{\mathbf{R}}_{-}(\tau,0)\\ \widehat{\mathbf{R}}_{+}(\tau,0)\widehat{\mathbf{T}}_{+}^{-1}(\tau,0) & \widehat{\mathbf{T}}_{-}(\tau,0)-\widehat{\mathbf{R}}_{+}(\tau,0)\widehat{\mathbf{T}}_{+}^{-1}(\tau,0)\widehat{\mathbf{R}}_{-}(\tau,0)\end{array}\right],\end{aligned} \tag{267}$$

which is an extension of the matrix exponential representation (102) to the inhomogeneous case.

Taking the derivative of (267) and accounting of (265), that is,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\tau}\left[\begin{array}{cc}\widehat{\mathbf{T}}_{+}^{-1} & -\widehat{\mathbf{T}}_{+}^{-1}\widehat{\mathbf{R}}_{-}\\ \widehat{\mathbf{R}}_{+}\widehat{\mathbf{T}}_{+}^{-1} & \widehat{\mathbf{T}}_{-}-\widehat{\mathbf{R}}_{+}\widehat{\mathbf{T}}_{+}^{-1}\widehat{\mathbf{R}}_{-}\end{array}\right]\\ &=-\left[\begin{array}{cc}\widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12}\\ -\widehat{\mathbf{A}}_{12} & -\widehat{\mathbf{A}}_{11}\end{array}\right]\left[\begin{array}{cc}\widehat{\mathbf{T}}_{+}^{-1} & -\widehat{\mathbf{T}}_{+}^{-1}\widehat{\mathbf{R}}_{-}\\ \widehat{\mathbf{R}}_{+}\widehat{\mathbf{T}}_{+}^{-1} & \widehat{\mathbf{T}}_{-}-\widehat{\mathbf{R}}_{+}\widehat{\mathbf{T}}_{+}^{-1}\widehat{\mathbf{R}}_{-}\end{array}\right],\end{aligned} \tag{268}$$

gives

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}_{+}=\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{11}\widehat{\mathbf{R}}_{+}+\widehat{\mathbf{R}}_{+}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{R}}_{+}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}_{+}, \tag{269}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}_{-}=\widehat{\mathbf{A}}_{11}\widehat{\mathbf{T}}_{-}+\widehat{\mathbf{R}}_{+}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}_{-}, \tag{270}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}_{-}=\widehat{\mathbf{T}}_{+}\widehat{\mathbf{A}}^{12}\widehat{\mathbf{T}}_{-}, \tag{271}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}_{+}=\widehat{\mathbf{T}}_{+}\widehat{\mathbf{A}}^{11}+\widehat{\mathbf{T}}_{+}\widehat{\mathbf{A}}^{12}\widehat{\mathbf{R}}_{+}, \tag{272}$$

From $\mathbf{P}_{-}(0,0)=\mathbf{I}_{2N}$ and (267), we find the initial conditions

$$\widehat{\mathbf{R}}_{+}(0,0)=\widehat{\mathbf{R}}_{-}(0,0)=\mathbf{0}, \tag{273}$$

$$\widehat{\mathbf{T}}_{-}(0,0)=\widehat{\mathbf{T}}_{+}(0,0)=\mathbf{I}_{N}. \tag{274}$$

By inspection of (269)-(272), it is apparent that equation (269) can be solved independently for $\widehat{\mathbf{R}}_{+}$, equation (270) as well as equation (272), must be solved along with (269) to find $\widehat{\mathbf{T}}_{-}$ and $\widehat{\mathbf{T}}_{+}$, respectively, while equation (271) must be solved together with (269), (270) and (272) to find $\widehat{\mathbf{R}}_{-}$. In summary, for an inhomogeneous atmosphere we started with the interaction principle equation (266), used relation (267) connecting the downward propagator and the extinction matrix, and then, by making use on the differential equation solved by the downward propagator (265), derived the matrix Riccati equations (269)-(272). Essentially, we replaced the problem of computing the downward propagator by the problem of computing the reflection and transmission matrices.

6.1. *Reflection and transmission matrices of a homogeneous layer*

For a homogeneous layer, the propagator $\mathbf{P}_{-}(\tau,0)=\mathbf{P}(\tau)=\exp(-\widehat{\mathbf{A}}\tau)$ solves the differential equation

$$\frac{\mathrm{d}(\mathrm{e}^{-\widehat{\mathbf{A}}\tau})}{\mathrm{d}\tau}=-\widehat{\mathbf{A}}\mathrm{e}^{-\widehat{\mathbf{A}}\tau}, \tag{275}$$

and (269)-(272) simplify to

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}=\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{11}\widehat{\mathbf{R}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}, \tag{276}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}=\widehat{\mathbf{A}}_{11}\widehat{\mathbf{T}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}, \tag{277}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}=\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}, \tag{278}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}=\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}. \tag{279}$$

where $\widehat{\mathbf{R}}_{+}(\tau,0)=\widehat{\mathbf{R}}_{-}(\tau,0)=\widehat{\mathbf{R}}(\tau)$ and $\widehat{\mathbf{T}}_{+}(\tau,0)=\widehat{\mathbf{T}}_{-}(\tau,0)=\widehat{\mathbf{T}}(\tau)$. From (276)-(279), we see that the following identities must hold

$$\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}=\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{11}\widehat{\mathbf{R}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}, \tag{280}$$

$$\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{T}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}=\widehat{\mathbf{A}}_{11}\widehat{\mathbf{T}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}. \tag{281}$$

It should perhaps be pointed out that (276)-(279) may follow directly from (102) and (275).

The solutions of the matrix Riccati equations for a homogeneous layer must be identical with the representations of the reflection and transmission matrices derived by the matrix exponential method of Section 5. Let us prove this result. The initial value problem for the reflection matrix (276) and (273) is equivalent with the following $2n$th order linear initial value problem

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\tau}\left[\begin{array}{c}\mathscr{P}(\tau)\\ \mathscr{Q}(\tau)\end{array}\right]=-\left[\begin{array}{cc}\widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12}\\ -\widehat{\mathbf{A}}_{12} & -\widehat{\mathbf{A}}_{11}\end{array}\right]\left[\begin{array}{c}\mathscr{P}(\tau)\\ \mathscr{Q}(\tau)\end{array}\right],\\ &\left[\begin{array}{c}\mathscr{P}(0)\\ \mathscr{Q}(0)\end{array}\right]=\left[\begin{array}{c}\mathbf{I}_{N}\\ 0\end{array}\right],\end{aligned} \tag{282}$$

where

$$\widehat{\mathbf{R}}(\tau)=\mathscr{Q}(\tau)\mathscr{P}^{-1}(\tau). \tag{283}$$

Making the change of variable

$$\left[\begin{array}{c}\widehat{\mathscr{P}}(\tau)\\ \widehat{\mathscr{Q}}(\tau)\end{array}\right]=\left[\begin{array}{cc}-\overline{\mathbf{V}}_{+}^{T} & \overline{\mathbf{V}}_{-}^{T}\\ \overline{\mathbf{V}}_{-}^{T} & -\overline{\mathbf{V}}_{+}^{T}\end{array}\right]\left[\begin{array}{c}\mathscr{P}(\tau)\\ \mathscr{Q}(\tau)\end{array}\right], \tag{284}$$

the linear differential equation (282) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{bmatrix}\widehat{\mathscr{P}}(\tau)\\ \widehat{\mathscr{Q}}(\tau)\end{bmatrix}=-\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & -\mathbf{\Lambda}\end{bmatrix}\begin{bmatrix}\widehat{\mathscr{P}}(\tau)\\ \widehat{\mathscr{Q}}(\tau)\end{bmatrix}, \tag{285}$$

with the initial condition

$$\begin{bmatrix}\widehat{\mathscr{P}}(0)\\ \widehat{\mathscr{Q}}(0)\end{bmatrix}=\begin{bmatrix}-\overline{\mathbf{V}}_+^T\\ \overline{\mathbf{V}}_-^T\end{bmatrix}. \tag{286}$$

In terms of $\widehat{\mathscr{P}}$ and $\widehat{\mathscr{Q}}$, the expression of the reflection matrix reads as

$$\widehat{\mathbf{R}}(\tau)=[\widehat{\mathbf{V}}_-\widehat{\mathscr{P}}(\tau)\widehat{\mathscr{Q}}^{-1}(\tau)+\widehat{\mathbf{V}}_+][\widehat{\mathbf{V}}_+\widehat{\mathscr{P}}(\tau)\widehat{\mathscr{Q}}^{-1}(\tau)+\widehat{\mathbf{V}}_-]^{-1}. \tag{287}$$

The solution of the differential equation (285) is given by $\widehat{\mathscr{P}}(\tau)=[\mathrm{e}^{-\lambda_k\tau}]\widehat{\mathscr{P}}(0)$ and $\widehat{\mathscr{Q}}(\tau)=[\mathrm{e}^{\lambda_k\tau}]\widehat{\mathscr{Q}}(0)$, so that the matrix product $\widehat{\mathscr{P}}(\tau)\widehat{\mathscr{Q}}^{-1}(\tau)$ in (287) can be expressed as

$$\widehat{\mathscr{P}}(\tau)\widehat{\mathscr{Q}}^{-1}(\tau)=\mathbf{\Gamma}(\tau)\widehat{\mathscr{P}}(0)\,\widehat{\mathscr{Q}}^{-1}(0)\,\mathbf{\Gamma}(\tau)=-\mathbf{\Gamma}(\tau)\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\tau). \tag{288}$$

Then, employing the identity (cf. (48)) $\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}=\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+$, in (288), and inserting the resulting expression in (287), we obtain

$$\widehat{\mathbf{R}}(\tau)=[\widehat{\mathbf{V}}_+-\widehat{\mathbf{V}}_-\mathbf{\Gamma}(\tau)\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+\mathbf{\Gamma}(\tau)][\widehat{\mathbf{V}}_--\widehat{\mathbf{V}}_+\mathbf{\Gamma}(\tau)\widehat{\mathbf{V}}_-^{-1}\widehat{\mathbf{V}}_+\mathbf{\Gamma}(\tau)]^{-1}, \tag{289}$$

which is exactly the reflection matrix representation (94) for $\tau=\overline{\tau}$.

6.2. *Approximations based on matrix Riccati equations*

A benefit of the matrix Riccati equation formalism is that for small values of the optical thickness and/or single scattering albedo, several approximation methods can be derived. These methods are summarized below.

6.2.1. *Successive orders of scattering*

In the matrix Riccati equations for a homogeneous layer

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}=\widehat{\mathbf{A}}_{12}+\widehat{\mathbf{A}}_{11}\widehat{\mathbf{R}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}, \tag{290}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}=\widehat{\mathbf{A}}_{11}\widehat{\mathbf{T}}+\widehat{\mathbf{R}}\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}, \tag{291}$$

we separate the attenuation terms from the multiple-scattering terms, by defining the matrix $\widehat{\mathbf{A}}_{11}^0$ through the relation

$$\widehat{\mathbf{A}}_{11}^0=\widehat{\mathbf{A}}_{11}+\mathbf{M}, \tag{292}$$

and indicate explicitly the dependency of the matrices $\widehat{\mathbf{A}}_{11}^0$ and $\widehat{\mathbf{A}}_{12}$ on the single scattering albedo ω by writing

$$\widehat{\mathbf{A}}_{11}^0=\omega\widetilde{\mathbf{A}}_{11}^0 \text{ and } \widehat{\mathbf{A}}_{12}=\omega\widetilde{\mathbf{A}}_{12}. \tag{293}$$

The resulting matrix Riccati equations, namely

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{R}}+\mathbf{M}\widehat{\mathbf{R}}+\widehat{\mathbf{R}}\mathbf{M}=\omega\widetilde{\mathbf{A}}_{12}+\omega\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{R}}+\omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}_{11}^0+\omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{R}}, \tag{294}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{T}}+\mathbf{M}\widehat{\mathbf{T}}=\omega\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{T}}+\omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{T}}, \tag{295}$$

endowed with the initial conditions $\widehat{\mathbf{R}}(0)=0$ and $\widehat{\mathbf{T}}(0)=\mathbf{I}$, are usually referred to as the discrete invariant imbedding equations. As pointed out by Waterman, the "Padé and Taylor series expansions are fundamentally different from those involving successive orders of scattering; in the later, ω is employed as the expansion parameter". Along this line, we assume the nth-order approximations

$$\widehat{\mathbf{R}}_n(\tau,\omega)=\sum_{k=0}^{n}\omega^k\widehat{\mathbf{H}}_k(\tau), \tag{296}$$

$$\widehat{\mathbf{T}}_n(\tau,\omega)=\sum_{k=0}^{n}\omega^k\widehat{\mathbf{G}}_k(\tau). \tag{297}$$

Inserting (296) and (297) into (294) and (295) and equating the coefficients of ω^n, yields the iterative schemes

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{H}}_0(\tau)+\mathbf{M}\widehat{\mathbf{H}}_0(\tau)+\widehat{\mathbf{H}}_0(\tau)\mathbf{M}=0,\ \widehat{\mathbf{H}}_0(0)=0, \tag{298}$$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{H}}_n(\tau)+\mathbf{M}\widehat{\mathbf{H}}_n(\tau)+\widehat{\mathbf{H}}_n(\tau)\mathbf{M}=\delta_{n1}\widetilde{\mathbf{A}}_{12}+\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{H}}_{n-1}(\tau)\\ &+\widehat{\mathbf{H}}_{n-1}(\tau)\widetilde{\mathbf{A}}_{11}^0+\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{H}}_{n-k-1}(\tau),\ \widehat{\mathbf{H}}_n(0)=0,\ n\geq 1,\end{aligned} \tag{299}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{G}}_0(\tau)+\mathbf{M}\widehat{\mathbf{G}}_0(\tau)=0\ \widehat{\mathbf{G}}_0(0)=\mathbf{I}, \tag{300}$$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\tau}\widehat{\mathbf{G}}_n(\tau)+\mathbf{M}\widehat{\mathbf{G}}_n(\tau)=\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{G}}_{n-1}(\tau)\\ &+\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{G}}_{n-k-1}(\tau),\ \widehat{\mathbf{G}}_n(0)=0,\ n\geq 1.\end{aligned} \tag{301}$$

For $n\geq 1$, the solutions of the initial value problems (299) and (301) are

$$[\widehat{\mathbf{H}}_n(\tau)]_{kl}=\int_0^\tau \mathrm{e}^{-\left(\frac{1}{\mu_k}+\frac{1}{\mu_l}\right)(\tau-\tau')}[\mathbb{H}_n(\tau')]_{kl}\,\mathrm{d}\tau', \tag{302}$$

$$\begin{aligned}\mathbb{H}_n(\tau)&=\delta_{n1}\widetilde{\mathbf{A}}_{12}+\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{H}}_{n-1}(\tau)+\widehat{\mathbf{H}}_{n-1}(\tau)\widetilde{\mathbf{A}}_{11}^0\\ &+\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{H}}_{n-k-1}(\tau),\end{aligned} \tag{303}$$

and

$$[\widehat{\mathbf{G}}_n(\tau)]_{kl}=\int_0^\tau \mathrm{e}^{-\frac{1}{\mu_k}(\tau-\tau')}[\mathbb{G}_n(\tau')]_{kl}\,\mathrm{d}\tau', \tag{304}$$

$$\mathbb{G}_n(\tau)=\widetilde{\mathbf{A}}_{11}^0\widehat{\mathbf{G}}_{n-1}(\tau)+\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{G}}_{n-k-1}(\tau), \tag{305}$$

respectively, with $\widehat{\mathbf{H}}_0(\tau)=0$ and $[\widehat{\mathbf{G}}_0(\tau)]_{kl}=\delta_{kl}\exp(-\tau/\mu_k)$. Accordingly, the zeroth- and first-order scattering solutions are

$$\widehat{\mathbf{R}}_0(\overline{\tau})=0,\ [\widehat{\mathbf{T}}_0(\overline{\tau})]_{kl}=\delta_{kl}\mathrm{e}^{-\frac{1}{\mu_k}\overline{\tau}}, \tag{306}$$

and

$$[\widehat{\mathbf{R}}_1(\overline{\tau})]_{kl} = (\frac{1}{\mu_k} + \frac{1}{\mu_l})^{-1}\left[1 - \mathrm{e}^{-\left(\frac{1}{\mu_k}+\frac{1}{\mu_l}\right)\overline{\tau}}\right][\widehat{\mathbf{A}}_{12}]_{kl}, \tag{307}$$

$$[\widehat{\mathbf{T}}_1(\overline{\tau})]_{kl} = \delta_{kl}\mathrm{e}^{-\frac{1}{\mu_k}\overline{\tau}} + (\frac{1}{\mu_k} - \frac{1}{\mu_l})^{-1}(\mathrm{e}^{-\frac{1}{\mu_l}\overline{\tau}} - \mathrm{e}^{-\frac{1}{\mu_k}\overline{\tau}})[\widehat{\mathbf{A}}^0_{11}]_{kl}, \tag{308}$$

respectively. Analytical formulas for the first three orders of scattering, which for most applications are sufficient, have been derived by Kawabata and Ueno [48]. However, the resulting expressions are too complex, so that even in the case $n = 2$, it is more efficient to use the Gaussian quadrature method to compute $\widehat{\mathbf{H}}_2$ and $\widehat{\mathbf{G}}_2$ by means of (302) and (304), respectively. It should be pointed out that the errors in $\widehat{\mathbf{R}}_2$ are considerably smaller than the errors in $\widehat{\mathbf{T}}_2$. For example, we found that for $\overline{\tau} = 20$ and $\omega = 0.2$, the relative errors in the second-order reflection and transmission functions are $8.14 \cdot 10^{-3}$ and $6.07 \cdot 10^{-2}$, respectively. This can be explained by the fact that in the case of strong absorption, the refection is determined by the (relatively thin) skin layer, while transmission is determined by the whole layer and requires more scattering.

Alternative recurrence relations for $\widehat{\mathbf{H}}_n(\tau)$ and $\widehat{\mathbf{G}}_n(\tau)$, which do not involve an integration over the optical depth, can be obtained by considering the identities (280) and (281) which, by means of (292) and (293), can be written as

$$\begin{aligned}\mathbf{M}\widehat{\mathbf{R}} + \widehat{\mathbf{R}}\mathbf{M} &= \omega\widetilde{\mathbf{A}}_{12} + \omega\widetilde{\mathbf{A}}^0_{11}\widehat{\mathbf{R}} + \omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}^0_{11} \\ &\quad + \omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{R}} - \omega\widehat{\mathbf{T}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{T}},\end{aligned} \tag{309}$$

$$\mathbf{M}\widehat{\mathbf{T}} - \widehat{\mathbf{T}}\mathbf{M} = \omega\widetilde{\mathbf{A}}^0_{11}\widehat{\mathbf{T}} - \omega\widehat{\mathbf{T}}\widetilde{\mathbf{A}}^0_{11} + \omega\widehat{\mathbf{R}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{T}} - \omega\widehat{\mathbf{T}}\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{R}}. \tag{310}$$

Substituting (296) and (297) into (309) and (310), and equating the coefficients of ω^n gives for $n \geq 1$,

$$\begin{aligned}&\mathbf{M}\widehat{\mathbf{H}}_n(\tau) + \widehat{\mathbf{H}}_n(\tau)\mathbf{M} = \delta_{n1}\widetilde{\mathbf{A}}_{12} + \widetilde{\mathbf{A}}^0_{11}\widehat{\mathbf{H}}_{n-1}(\tau) + \widehat{\mathbf{H}}_{n-1}(\tau)\widetilde{\mathbf{A}}^0_{11} \\ &+ \sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{H}}_{n-k-1}(\tau) - \sum_{k=0}^{n-1}\widehat{\mathbf{G}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{G}}_{n-k-1}(\tau)\end{aligned} \tag{311}$$

and

$$\begin{aligned}&\mathbf{M}\widehat{\mathbf{G}}_n(\tau) - \widehat{\mathbf{G}}_n(\tau)\mathbf{M} = \widetilde{\mathbf{A}}^0_{11}\widehat{\mathbf{G}}_{n-1}(\tau) + \widehat{\mathbf{G}}_{n-1}(\tau)\widetilde{\mathbf{A}}^0_{11} \\ &+ \sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{G}}_{n-k-1}(\tau) - \sum_{k=0}^{n-1}\widehat{\mathbf{G}}_k(\tau)\widetilde{\mathbf{A}}_{12}\widehat{\mathbf{H}}_{n-k-1}(\tau),\end{aligned} \tag{312}$$

with $\widehat{\mathbf{H}}_0(\tau) = 0$ and $[\widehat{\mathbf{G}}_0(\tau)]_{kl} = \delta_{kl}\exp(-\tau/\mu_k)$. Equations (311) and (312) have been derived by Hansen and Travis [3] using the invariance principle. Note that (312) is indeterminate for $\mu_k = \mu_l$, and that approximate results for that special case can be obtained by interpolation.

6.2.2. Iterative approximation

For thin layers, the Taylor series approximations to the transmission and reflection matrices (175) and (176), can be rediscover by using the matrix Riccati equation method. In this regard, we mention that Chang and Wu [49] solved (290) and (291) with the initial conditions $\widehat{\mathbf{R}}(0) = 0$ and $\widehat{\mathbf{T}}(0) = \mathbf{I}$ by means of the iterative scheme

$$\widehat{\mathbf{R}}_1(\tau) = \int_0^\tau \widehat{\mathbf{A}}_{12}\,\mathrm{d}\tau', \tag{313}$$

$$\begin{aligned}\widehat{\mathbf{R}}_n(\tau) = \int_0^\tau &[\widehat{\mathbf{A}}_{12} + \widehat{\mathbf{A}}_{11}\widehat{\mathbf{R}}_{n-1}(\tau') \\ &+ \widehat{\mathbf{R}}_{n-1}(\tau')\widehat{\mathbf{A}}_{11} + \widehat{\mathbf{R}}_{n-1}(\tau')\widehat{\mathbf{A}}_{12}\widehat{\mathbf{R}}_{n-1}(\tau')]\,\mathrm{d}\tau',\ n \geq 2,\end{aligned} \tag{314}$$

and

$$\widehat{\mathbf{T}}_1(\tau) = \mathbf{I} + \int_0^\tau \widehat{\mathbf{A}}_{11}\,\mathrm{d}\tau', \tag{315}$$

$$\widehat{\mathbf{T}}_n(\tau) = \mathbf{I} + \int_0^\tau [\widehat{\mathbf{A}}_{11}\widehat{\mathbf{T}}_{n-1}(\tau') + \widehat{\mathbf{R}}_{n-1}(\tau')\widehat{\mathbf{A}}_{12}\widehat{\mathbf{T}}_{n-1}(\tau')]\,\mathrm{d}\tau',\ n \geq 2. \tag{316}$$

It is not hard to see that the first-order solutions $\widehat{\mathbf{T}}_1$ and $\widehat{\mathbf{R}}_1$ correspond to the infinitesimal generator initialization scheme of Grant and Hunt [39], while the second-order solutions $\widehat{\mathbf{T}}_2$ and $\widehat{\mathbf{R}}_2$ (excepting the terms in $\overline{\tau}^3$) correspond to the expanded diamond initialization scheme of Wiscombe [40]. To design a procedure for computing higher-order correction terms, we assume the finite Taylor expansions (175) and (176) for $\widehat{\mathbf{T}}_n$ and $\widehat{\mathbf{R}}_n$, respectively. Then, from (314) and (316) we obtain for $n \geq 2$,

$$\widehat{\mathbf{H}}_n = \frac{1}{n}(\widehat{\mathbf{A}}_{11}\widehat{\mathbf{H}}_{n-1} + \widehat{\mathbf{H}}_{n-1}\widehat{\mathbf{A}}_{11}) + \frac{1}{n}\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k\widehat{\mathbf{A}}_{12}\widehat{\mathbf{H}}_{n-k-1}, \tag{317}$$

$$\widehat{\mathbf{G}}_n = \frac{1}{n}\widehat{\mathbf{A}}_{11}\widehat{\mathbf{G}}_{n-1} + \frac{1}{n}\sum_{k=0}^{n-1}\widehat{\mathbf{H}}_k\widehat{\mathbf{A}}_{12}\widehat{\mathbf{G}}_{n-k-1}, \tag{318}$$

with $\widehat{\mathbf{H}}_0 = 0$, $\widehat{\mathbf{H}}_1 = \widehat{\mathbf{A}}_{12}$, $\widehat{\mathbf{G}}_0 = \mathbf{I}$, and $\widehat{\mathbf{G}}_1 = \widehat{\mathbf{A}}_{11}$. Equations (317) and (318) are the counterparts of (180) and (178) in the matrix Riccati equation formalism.

For optical depths much less than unity, (290) and (291) can be also solved numerically by using an nth-order Runge-Kutta scheme. As shown in [49], the fourth-order Runge-Kutta scheme has a higher computational cost than the fourth-order iterative method but is more accurate.

7. Conclusions

This paper provides a description of the matrix exponential formalism in radiative transfer. The solution of the initial value problem of the discrete radiative transfer is expressed in terms of the matrix exponential. Although, the matrix exponential solution is computationally unstable it is used as a starting point in deriving a computationally stable equation, the layer equation, which is the heart of the discrete ordinate method, the interaction principle equation, which is the heart of the matrix operator method, and finally, the matrix Riccati equations. Thus, the matrix exponential formalism gives the framework for the unification of the discrete ordinate method, the matrix operator method, and the matrix Riccati equations method. In our analysis,

1. we provided an interpretation of the matrix exponential solution in terms of homogeneous and particular solutions, as well as in terms of total and direct radiance vectors,
2. we used the right- and left eigenvector technique of Waterman [21] to compute the matrix exponential, and derived a set of matrix identities for proving the mathematical equivalence between the different solution methods,
3. we established the layer equation of the discrete ordinate method with matrix exponential, compared this method with the classical discrete ordinate method, derived equivalent solution representations, and established the link between the matrix exponential solution and the Chandrasekhar's discrete ordinate solution,
4. we derived equivalent expressions for the reflection and transmission matrices by converting the layer equation and the solution representations of the matrix exponential method into the interaction principle equation,
5. for optically thin layers, we used the nth-order Padé and Taylor series approximations to the matrix exponential to compute the reflection and transmission matrices, and to derive an nth-order approximation to the layer equation,
6. for optically thick layers, we derived the asymptotic expressions of the reflection and transmission matrices by adapting the discrete ordinate approach of Nakajima and King [27] to our framework, obtained higher-order corrections of the reflection and transmission matrices for moderate values of the optical thickness, reconsidered Waterman's approximation by including an additional term in the expression of the reflection matrix, and computed parametrizations of the asymptotic functions and constants for a water-cloud model with a Gamma size distribution,
7. we reviewed the approach of Flatau and Stephens [22] for obtaining the matrix Riccati equations in the case of an inhomogeneous atmosphere, proved the equivalence between the matrix Riccati equations method and the eigendecomposition method in computing the reflection matrix of a homogeneous layer, and discussed the successive order of scattering approximation for small values of the single-scattering albedo, as well as, an nth-order iterative approximation for small values of the optical thickness.

Some additional results are given in appendices. In Appendix 1, we justify the choice of Gaussian quadrature in the discrete ordinate method. In Appendix 2 we review several eigendecomposition methods for computing the matrix exponential in a common framework. In this context it should be pointed out that the direct decomposition method is preferable for numerical implementations, while the square-root method is an important tool for theoretical studies, e.g., conservative scattering and asymptotic theory. In Appendix 3, we extend the analytical formulas for non-conservative scattering to conservative scattering, and prove that the system of characteristic solutions proposed by Nakajima and Tanaka [18] can be used for both non-conservative and conservative scattering.

The practical conclusion of our analysis is that the matrix exponential formalism enables the design of a radiative transfer code incorporating the discrete ordinate method, the matrix operator method and approximate models. For each homogeneous layer of a multi-layered atmosphere, a spectral decomposition of the layer matrix is performed, and then either the layer equation is derived and assembled into the system matrix of the entire atmosphere (discrete ordinate method), or the reflection and transmission matrices of the layer are computed and the adding formula is used to obtain the reflection matrix of the entire atmosphere (matrix operator method). Moreover, a combined model can be also designed.

To speed up the computational process, approximate models can be used. An efficient radiative transfer code should incorporate built-in routines that automatically work when the optical thickness of a homogeneous (sub)layer becomes too small or too large. On the other hand, the nth-order Padé and Taylor series approximations and the nth-order iterative approximation can be used for the initialization of the doubling method in radiative transfer, while the asymptotic form of the reflection function can be used in a cloud parameter retrieval algorithm. In the latter case, the computational process is organized as follows [50]: (I) replace the atmosphere below the cloud bottom by an equivalent Lambertian surface of albedo A, (II) compute the reflection function of a layer with an underlying Lambertian surface of albedo A by means of the asymptotic theory, and (III) use the cloud reflection function as a bidirectional reflection function in a discrete ordinate model to compute the radiance field of the atmosphere above the cloud top.

Appendix 1. Gauss quadrature

In radiative transfer, the phase function p is usually expressed through a finite series of normalized Legendre polynomials P_n, i.e.,

$$p(\mu,\mu',\varphi-\varphi') = \sum_{n=0}^{N_{\max}} \sqrt{\frac{2n+1}{2}}\,\chi_n P_n(\cos\Theta), \tag{319}$$

where $\cos\Theta = \mu\mu' + \sqrt{1-\mu^2}\,\sqrt{1-\mu'^2}\cos(\varphi-\varphi')$, χ_n are the expansion coefficients, and $N_{\max}$ is the number of expansion coefficients. Making use on the addition theorem for the Legendre polynomials, we obtain an expansion in terms of spherical harmonics

$$p(\mu,\mu',\varphi-\varphi') = \sum_{m=0}^{M_{\max}} \sum_{n=m}^{N_{\max}} (2-\delta_{m0})\,\chi_n P_n^m(\mu) P_n^m(\mu')\cos[m(\varphi-\varphi')], \tag{320}$$

with $M_{\max} \le N_{\max}$. We are now in the favorable situation that the kernel of the radiative transfer integral equation is

expanded in an orthogonal and complete system of functions. Further on, we argue in connection with the spherical harmonics method. On the unit sphere, the diffuse radiance $I_{\rm d}(\tau,\mu',-\mu_0,\varphi'-\varphi_0)$ can be also expanded in terms of spherical harmonics, so that after integrating the multiple scattering term with respect to the azimuthal angle, we are led to an integral of the form $\int_{-1}^{1} P_n^m(\mu')P_{n'}^m(\mu')\mathrm{d}\mu'$, with $n,n'=0,\ldots,N_{\max}$. For $n,n'\leq N_{\max}$, $P_n^m(\mu')P_{n'}^m(\mu')=P(\mu')$ is a polynomial of degree at most $2N_{\max}$. This can be seen by making use on the explicit construction of the associated Legendre functions in terms of Jacobi polynomials. The integral then reduces to $\int_{-1}^{1} P(\mu')\mathrm{d}\mu'$, and the task is to find an exact quadrature for this integral. If this is done, the mathematical equivalence between the spherical harmonics and the discrete ordinate method is established. In general, a Gauss quadrature using $N_{\rm q}$ nodes is an exact quadrature for polynomials of degree $2N_{\rm q}-1$ or less [51]. In our case, this condition translates into $N_{\max}<N_{\rm q}-1/2$, and for the choice $N_{\rm q}=2N$, it follows that $N_{\max}<2N-1/2$. Thus, using $2N$ Gaussian nodes and weights, we need $N_{\max}=2N-1$ expansion terms in (319) and $M_{\max}\leq 2N-1$ azimuthal modes. In [52] it is shown that

1. a Gauss quadrature guarantees that the phase function is correctly normalized, i.e., $\int_0^{2\pi}\int_{-1}^{1} p(\mu,\mu',\varphi-\varphi')\mathrm{d}\mu'\mathrm{d}\varphi'=4\pi$, and so, that the energy is conserved in the computation,
2. a double Gauss quadrature, in which the Gaussian formula is applied separateley to the half ranges $(-1,0)$ and $(0,1)$, is preferable that a Gauss quadrature for the complete range $(-1,1)$.

It is a fact that Gauss quadrature has a factor-of-2 advantage in its efficiency as compared to equidistant quadrature methods. A method wich has almost the same performances and can be implemented effortlessly by the fast Fourier transform is the Clenshaw-Curtis scheme [53]. However, as it has been shown in [54], when the number of nodes $N_{\rm q}$ increases, the error of the Clenshaw-Curtis quadrature does not decay to zero evenly but in two distinct stages; for $N_{\rm q}$ smaller than a critical value, the error decreases by the rate $O(\rho^{-2N_{\rm q}})$, where $\rho>1$, and afterwards by the rate $O(\rho^{-N_{\rm q}})$. This means that initially (for small $N_{\rm q}$), Clenshaw-Curtis quadrature converges about as fast as the Gauss quadrature. The outlook for a future work is to focus on the efficiency of the Clenshaw-Curtis quadrature in radiative transfer.

Appendix 2. Spectral decomposition of the layer matrix

Let us consider the layer matrix

$$\widehat{\mathbf{A}}=\begin{bmatrix}\widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12}\\ -\widehat{\mathbf{A}}_{12} & -\widehat{\mathbf{A}}_{11}\end{bmatrix}, \tag{321}$$

with

$$\begin{aligned}\widehat{\mathbf{A}}_{11}&=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}(\mathbf{S}_+-\mathbf{W}^{-1})\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}},\\ \widehat{\mathbf{A}}_{12}&=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}\mathbf{S}_-\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}},\end{aligned} \tag{322}$$

and let us introduce the symmetric matrices

$$\begin{aligned}\widehat{\mathbf{A}}_+&=\widehat{\mathbf{A}}_{11}+\widehat{\mathbf{A}}_{12}=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}(\mathbf{S}_++\mathbf{S}_-)\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}-\mathbf{M},\\ \widehat{\mathbf{A}}_-&=\widehat{\mathbf{A}}_{11}-\widehat{\mathbf{A}}_{12}=\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}(\mathbf{S}_+-\mathbf{S}_-)\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}-\mathbf{M}.\end{aligned} \tag{323}$$

Further, let us define the matrices $\widehat{\mathbf{Q}}_+$ and $\widehat{\mathbf{Q}}_-$ through the factorizations

$$\widehat{\mathbf{A}}_-\widehat{\mathbf{Q}}_-=\widehat{\mathbf{Q}}_+\Lambda, \tag{324}$$

$$\widehat{\mathbf{A}}_+\widehat{\mathbf{Q}}_+=\widehat{\mathbf{Q}}_-\Lambda, \tag{325}$$

yielding

$$(\widehat{\mathbf{A}}_-\widehat{\mathbf{A}}_+)\widehat{\mathbf{Q}}_+=\widehat{\mathbf{Q}}_+\Lambda^2, \tag{326}$$

and

$$(\widehat{\mathbf{A}}_+\widehat{\mathbf{A}}_-)\widehat{\mathbf{Q}}_-=\widehat{\mathbf{Q}}_-\Lambda^2, \tag{327}$$

for $\Lambda=[\lambda_k]$. If $\widehat{\mathbf{Q}}_+$, $\widehat{\mathbf{Q}}_-$ and Λ are known, the desired eigendecomposition is

$$\widehat{\mathbf{A}}=\begin{bmatrix}\widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_-\\ \widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+\end{bmatrix}\begin{bmatrix}\Lambda & \mathbf{0}\\ \mathbf{0} & -\Lambda\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_-\\ \widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+\end{bmatrix}^{-1}, \tag{328}$$

where

$$\begin{aligned}\widehat{\mathbf{V}}_+&=\frac{1}{2}(\widehat{\mathbf{Q}}_++\widehat{\mathbf{Q}}_-),\\ \widehat{\mathbf{V}}_-&=\frac{1}{2}(\widehat{\mathbf{Q}}_+-\widehat{\mathbf{Q}}_-).\end{aligned} \tag{329}$$

If the systems of normalized right and left eigenvectors are required, we first compute $a_k=||\widehat{\mathbf{v}}_k^-||^2-||\widehat{\mathbf{v}}_k^+||^2$, and then apply the transformation rules

$$\frac{1}{\sqrt{|a_k|}}\widehat{\mathbf{v}}_k^{\pm}\to\widehat{\mathbf{v}}_k^{\pm}\text{ and }\overline{\mathbf{v}}_k^{\pm}=\mathrm{sign}(a_k)\widehat{\mathbf{v}}_k^{\pm},$$

to construct the matrices $\widehat{\mathbf{V}}_{\pm}$ and $\overline{\mathbf{V}}_{\pm}$.

The spectral decomposition of $\widehat{\mathbf{A}}$ can be obtained by one of the following methods: direct decomposition of the asymmetric matrix $\widehat{\mathbf{A}}_-\widehat{\mathbf{A}}_+$ [14], the square-root decomposition [18], and the Cholesky decomposition [30]. Before proceeding, we make the change of variables

$$\mathbf{A}_-=-\widehat{\mathbf{A}}_-, \tag{330}$$

$$\mathbf{A}_+=-\widehat{\mathbf{A}}_+, \tag{331}$$

and note that by virtue of (330) and (331), the matrix $\mathbf{A}_-$ is symmetric and positive definite, while the matrix $\mathbf{A}_+$ is symmetric and non-negative definite.

Direct decomposition method

The direct decomposition method involves the following steps:

1. Compute $\mathscr{A}_+=\mathbf{A}_-\mathbf{A}_+$, and determine the eigensystem $\{\mu_k,\widehat{\mathbf{q}}_k^+\}_{k=\overline{1,N}}$ of the matrix $\mathscr{A}_+$, i.e., $\mathscr{A}_+\widehat{\mathbf{q}}_k^+=\mu_k\widehat{\mathbf{q}}_k^+$.

2. If a linearization of the radiative transfer model is required, normalize the vectors $\widehat{\mathbf{q}}_k^+$ for $k = 1, ..., N$.
3. For the nonconservative case ($\mu_N \neq 0$), compute the eigenvectors of the matrix $\mathscr{A}_- = \mathbf{A}_+\mathbf{A}_-$, i.e.,

$$\widehat{\mathbf{q}}_k^- = -\frac{1}{\sqrt{\mu_k}}\mathbf{A}_+\widehat{\mathbf{q}}_k^+, \ k = 1, ..., N. \tag{332}$$

4. Set

$$\Lambda = [\lambda_k], \ \lambda_k = \sqrt{\mu_k}, \tag{333}$$

and

$$\widehat{\mathbf{Q}}_+ = [\widehat{\mathbf{q}}_k^+], \tag{334}$$
$$\widehat{\mathbf{Q}}_- = [\widehat{\mathbf{q}}_k^-]. \tag{335}$$

Square-root method

The square-root method involves the following steps:

1. Compute a singular value decomposition of the symmetric and positive definite matrix $\mathbf{A}_-$, i.e.,

$$\mathbf{A}_- = \mathbf{U}\Sigma_a\mathbf{U}^T, \tag{336}$$

and the square root matrices

$$\mathbf{A}_-^{\frac{1}{2}} = \mathbf{U}\Sigma_a^{\frac{1}{2}}\mathbf{U}^T, \tag{337}$$
$$\mathbf{A}_-^{-\frac{1}{2}} = \mathbf{U}\Sigma_a^{-\frac{1}{2}}\mathbf{U}^T. \tag{338}$$

2. Construct the matrix

$$\mathbf{Z} = \mathbf{A}_-^{\frac{1}{2}}\mathbf{A}_+\mathbf{A}_-^{\frac{1}{2}}, \tag{339}$$

and compute a singular value decomposition of the symmetric and non-negative definite matrix $\mathbf{Z}$, i.e.,

$$\mathbf{Z} = \mathbf{V}\Sigma_z\mathbf{V}^T. \tag{340}$$

3. Set

$$\Lambda = \Sigma_z^{\frac{1}{2}}, \tag{341}$$

and compute

$$\widehat{\mathbf{Q}}_+ = \mathbf{A}_-^{\frac{1}{2}}\mathbf{V}, \tag{342}$$
$$\widehat{\mathbf{Q}}_+^{-T} = \mathbf{A}_-^{-\frac{1}{2}}\mathbf{V}, \tag{343}$$
$$\widehat{\mathbf{Q}}_- = -\widehat{\mathbf{Q}}_+^{-T}\Lambda. \tag{344}$$

To justify this algorithm we note that

$$\mathbf{A}_-\mathbf{A}_+ = \mathbf{A}_-^{\frac{1}{2}}(\mathbf{A}_-^{\frac{1}{2}}\mathbf{A}_+\mathbf{A}_-^{\frac{1}{2}})\mathbf{A}_-^{-\frac{1}{2}} = \mathbf{A}_-^{\frac{1}{2}}\mathbf{V}\Sigma_z\mathbf{V}^T\mathbf{A}_-^{-\frac{1}{2}} = \widehat{\mathbf{Q}}_+\Lambda^2\widehat{\mathbf{Q}}_+^{-1},$$

which is equivalent to (326).

Cholesky method

The Cholesky method involves the following steps:

1. Compute the Cholesky factorization of the symmetric and non-negative definite matrix $\mathbf{A}_+$, i.e.,

$$\mathbf{A}_+ = \mathbf{R}^T\mathbf{R}. \tag{345}$$

2. Construct the matrix

$$\mathbf{Z} = \mathbf{R}\mathbf{A}_-\mathbf{R}^T, \tag{346}$$

and compute a singular value decomposition of the symmetric and non-negative definite matrix $\mathbf{Z}$, i.e.,

$$\mathbf{Z} = \mathbf{V}\Sigma_z\mathbf{V}^T. \tag{347}$$

3. Set

$$\Lambda = \Sigma_z^{\frac{1}{2}}, \tag{348}$$

and for the nonconservative case (Λ^{-1} and $\mathbf{R}^{-1}$ exist), compute

$$\widehat{\mathbf{Q}}_- = -\mathbf{R}^T\mathbf{V}\Lambda^{-1}, \tag{349}$$

and

$$\widehat{\mathbf{Q}}_+ = \mathbf{R}^{-1}\mathbf{V}, \tag{350}$$
$$\widehat{\mathbf{Q}}_+^{-T} = -\widehat{\mathbf{Q}}_-\Lambda. \tag{351}$$

To justify the algorithm, we note that

$$\widehat{\mathbf{Q}}_+\Lambda^2 = \mathbf{R}^{-1}\mathbf{V}\Sigma_z = \mathbf{R}^{-1}\mathbf{Z}\mathbf{V} = \mathbf{R}^{-1}\mathbf{R}\mathbf{A}_-\mathbf{R}^T\mathbf{R}\mathbf{R}^{-1}\mathbf{V} = \mathbf{A}_-\mathbf{A}_+\widehat{\mathbf{Q}}_+,$$

which is equivalent to (326).

In summary,

1. in the direct decomposition method, only one eigenvalue problem for the asymmetric matrix $\widehat{\mathbf{A}}_-\widehat{\mathbf{A}}_+$ is solved,
2. in the square-root method, two eigenvalue problems for the symmetric and positive definite matrix $\mathbf{A}_-$ and the symmetric and non-negative definite matrix $\mathbf{A}_-^{1/2}\mathbf{A}_+\mathbf{A}_-^{1/2}$ are solved, and
3. in the Cholesky method, the matrix $\mathbf{A}_+$ is first factorized as $\mathbf{A}_+ = \mathbf{R}^T\mathbf{R}$, and then, an eigenvalue problem for the symmetric and non-negative definite matrix $\mathbf{R}\mathbf{A}_-\mathbf{R}^T$ is solved.

Although the computation of the eigendecomposition of $\widehat{\mathbf{A}}$ by the square-root method is more time consuming than the other two methods, it is computationally stable in the limit $\omega \to 1$, and even in the conservative case $\omega = 1$.

Appendix 3. Conservative scattering

In the conservative case and the azimuthal mode $m = 0$, the smallest eigenvalues $\pm\lambda_N$ vanish. The two corresponding eigenvectors merge into one, which is not normalizable. The idea is to introduce two additional terms to replace those

that are lost in expansion (40) when $k = N$. In fact, the case of conservative scattering is only of pure theoretical interest. The reason is that in practical numerical simulations, a conservative scattering problem can be modeled by considering the limit $\omega \to 1$, i.e., by setting $\omega = 1 - \varepsilon$, with ε sufficiently small. Even in the case $\omega = 1$, the smallest computed eigenvalue is not exactly zero (due to rounding errors), and the computation of the spectral decomposition of the layer matrix be means of the square root method of Appendix 2 is a stable process. In spite of these practical arguments we decide to include this case in our analysis.

Before proceeding we make some general comments. Let $\mathbf{A}$ be a matrix with $n-2$ real eigenvalues λ_k with the geometric multiplicities $m_k = 1$, $k = 1, ..., n-2$, and let λ_n be an eigenvalue with the geometric multiplicity $m_n = 2$. For the eigenvalue λ_n, the solution of the differential equation

$$\frac{\mathrm{d}\mathbf{i}}{\mathrm{d}\tau}(\tau) = -\mathbf{A}\mathbf{i}(\tau) \tag{352}$$

is seek as a linear combination of characteristic solutions

$$\mathbf{i}(\tau) = \alpha \mathbf{w}_0 \mathrm{e}^{-\lambda_n \tau} + \beta(\mathbf{w}_1 + \mathbf{w}_0 \tau)\mathrm{e}^{-\lambda_n \tau}, \tag{353}$$

where $\mathbf{w}_0$ and $\mathbf{w}_1$ are determined by inserting each characteristic solution into the differential equation. By doing this, we obtain

$$(\mathbf{A} - \lambda_n \mathbf{I}_n)\mathbf{w}_0 = 0, \tag{354}$$

$$(\mathbf{A} - \lambda_n \mathbf{I}_n)\mathbf{w}_1 = -\mathbf{w}_0, \tag{355}$$

$$(\mathbf{A} - \lambda_n \mathbf{I}_n)^2 \mathbf{w}_1 = 0. \tag{356}$$

Defining the null spaces $\mathcal{N}^1_{\lambda_n} = \mathcal{N}(\mathbf{A} - \lambda_n \mathbf{I}_n)$ and $\mathcal{N}^2_{\lambda_n} = \mathcal{N}((\mathbf{A} - \lambda_n \mathbf{I}_n)^2)$, and noting that $\mathcal{N}^1_{\lambda_n} \subset \mathcal{N}^2_{\lambda_n} = \mathcal{M}_{\lambda_n}$, where $\mathcal{M}_{\lambda_n}$ is the main space of the eigenvalue λ_n so that $\mathbb{R}^n = \mathcal{M}_{\lambda_1} \oplus \mathcal{M}_{\lambda_2} ... \mathcal{M}_{\lambda_{n-2}} \oplus \mathcal{M}_{\lambda_n}$, we see that (354) and (356) give

$$\mathbf{w}_0 \in \mathcal{N}(\mathbf{A} - \lambda_n \mathbf{I}_n), \tag{357}$$

$$\mathbf{w}_1 \in \mathcal{N}((\mathbf{A} - \lambda_n \mathbf{I}_n)^2) \setminus \mathcal{N}(\mathbf{A} - \lambda_n \mathbf{I}_n), \tag{358}$$

and further, $\mathcal{N}^1_{\lambda_n} = \mathrm{span}\{\mathbf{w}_0\}$ and $\mathcal{N}^2_{\lambda_n} = \mathrm{span}\{\mathbf{w}_0, \mathbf{w}_1\}$.

Basic results

In the conservative case and the azimuthal mode $m = 0$, we have $\lambda_N = 0$, and so, $\lambda_N = 0$ is an eigenvalue with the geometric multiplicity $m_N = 2$**.** The right and left eigenvectors of the matrix $\widehat{\mathbf{A}}$ are constructed as follows:

1. We assume that the null spaces $\mathcal{N}(\widehat{\mathbf{A}})$ and $\mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}})$ are spanned by the right eigenvectors $\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}$, respectively, i.e.,

$$\mathcal{N}(\widehat{\mathbf{A}}) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}\right\}, \tag{359}$$

$$\mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}}) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}\right\}, \tag{360}$$

and that the null-spaces $\mathcal{N}(\widehat{\mathbf{A}}^T)$ and $\mathcal{N}((\widehat{\mathbf{A}}^2)^T) \setminus \mathcal{N}(\widehat{\mathbf{A}}^T)$ are spanned by the left eigenvectors $\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}$, respectively, i.e.,

$$\mathcal{N}(\widehat{\mathbf{A}}^T) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}\right\}, \tag{361}$$

$$\mathcal{N}((\widehat{\mathbf{A}}^2)^T) \setminus \mathcal{N}(\widehat{\mathbf{A}}^T) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}\right\}. \tag{362}$$

Thus, we have

$$\mathcal{N}(\widehat{\mathbf{A}}^2) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}\right\}, \tag{363}$$

$$\mathcal{N}((\widehat{\mathbf{A}}^2)^T) = \mathrm{span}\left\{\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}\right\}. \tag{364}$$

2. We impose that the vectors $\mathbf{w}_0$ and $\mathbf{w}_1$ are related through the relation (cf. (355))

$$\widehat{\mathbf{A}} \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} = -\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}. \tag{365}$$

Let us normalize the vectors $\mathbf{w}_0$ and $\mathbf{w}_1$ according to the transformations

$$\frac{1}{a_0}\mathbf{w}_0 \to \mathbf{w}_0 \text{ and } \frac{1}{a_0}\mathbf{w}_1 \to \mathbf{w}_1,$$

where the normalization constant a_0 is given by $a_0 = \sqrt{2\mathbf{w}_0^T \mathbf{w}_1}$. Then, the following results hold:

1. The right eigenvector spanning $\mathcal{N}(\widehat{\mathbf{A}})$ is orthogonal to the right eigenvector spanning $\mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}})$. The same result is valid for the left eigenvectors spanning $\mathcal{N}(\widehat{\mathbf{A}}^T)$ and $\mathcal{N}((\widehat{\mathbf{A}}^2)^T) \setminus \mathcal{N}(\widehat{\mathbf{A}}^T)$, i.e.,

$$\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix} = 0. \tag{366}$$

2. The right eigenvector spanning $\mathcal{N}(\widehat{\mathbf{A}})$ is orthogonal to the left eigenvector spanning $\mathcal{N}(\widehat{\mathbf{A}}^T)$. The same result is valid for the right and left eigenvectors spanning $\mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}})$ and $\mathcal{N}((\widehat{\mathbf{A}}^2)^T) \setminus \mathcal{N}(\widehat{\mathbf{A}}^T)$, respectively, i.e.,

$$\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix} = 0. \tag{367}$$

3. The right eigenvector spanning $\mathcal{N}(\widehat{\mathbf{A}})$ is co-linear with the left eigenvector spanning $\mathcal{N}((\widehat{\mathbf{A}}^2)^T) \setminus \mathcal{N}(\widehat{\mathbf{A}}^T)$. The same result is valid for the left and right eigenvectors spanning $\mathcal{N}(\widehat{\mathbf{A}}^T)$ and $\mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}})$, respectively, i.e.,

$$\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} = 1. \tag{368}$$

4. The left eigenvectors $\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}$ spanning $\mathcal{N}((\widehat{\mathbf{A}}^2)^T)$ are orthogonal to the right eigenvectors $\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix}$ corresponding to λ_k, i.e.,

$$\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}^T \begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} = 0, \tag{369}$$

and to the right eigenvectors $\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}$ corresponding to $-\lambda_k$, i.e.,

$$\begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}^T \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}^T \begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix} = 0 \tag{370}$$

Computation of the vectors $\mathbf{w}_0$ *nd* $\mathbf{w}_1$

A method for computing the vectors $\mathbf{w}_0$ and $\mathbf{w}_1$ uses the equations solved by the flux H and the K integral in the case of conservative scattering [32]:

$$\frac{\mathrm{d}H}{\mathrm{d}\tau}(\tau,\mu_0) = 0,$$
$$\frac{\mathrm{d}K}{\mathrm{d}\tau}(\tau,\mu_0) = \left(1-\frac{\chi_1}{2}\right)H(\tau,\mu_0),$$

where

$$H(\tau,\mu_0) = 2\pi\int_{-1}^{1} I_0(\tau,\mu,-\mu_0)\mu\mathrm{d}\mu - F_0\mu_0\mathrm{e}^{-\tau/\mu_0},$$
$$K(\tau,\mu_0) = 2\pi\int_{-1}^{1} I_0(\tau,\mu,-\mu_0)\mu^2\mathrm{d}\mu + F_0\mu_0^2\mathrm{e}^{-\tau/\mu_0}.$$

Another method is the square-root method of Appendix 2, which we will now describe. In the conservative case and the azimuthal mode $m = 0$, the singular value σ_N of the matrix $\mathbf{Z} = \mathbf{V}\Sigma_z\mathbf{V}^T$ is zero. Thus, we have $\lambda_N = \sqrt{\sigma_N} = 0$ and $\mathcal{N}(\mathbf{Z}) = \mathrm{span}\{\mathbf{v}_N\}$, where $\mathbf{v}_N$ is the Nth column vector of $\mathbf{V}$. From the relation $\mathbf{Z}\mathbf{v}_N = 0$ and the definition $\mathbf{Z} = \mathbf{A}_-^{1/2}\mathbf{A}_+\mathbf{A}_-^{1/2}$, we obtain $\mathbf{A}_-^{1/2}\mathbf{A}_+\mathbf{A}_-^{1/2}\mathbf{v}_N = 0$. Since $\mathbf{A}_-$ is positive definite, it follows that $\mathbf{A}_+\mathbf{A}_-^{1/2}\mathbf{v}_N = 0$. Further, setting $\widehat{\mathbf{q}}_N^+ = \mathbf{A}_-^{1/2}\mathbf{v}_N$ (cf. (342)), we get $\mathbf{A}_+\widehat{\mathbf{q}}_N^+ = 0$, or equivalently, $\widehat{\mathbf{q}}_N^+ \in \mathcal{N}(\mathbf{A}_+)$. On the other hand, from $\widehat{\mathbf{Q}}_- = -\mathbf{A}_-^{-1}\widehat{\mathbf{Q}}_+\Lambda$, we obtain $\widehat{\mathbf{q}}_N^- = -\lambda_N\mathbf{A}_-^{-1}\widehat{\mathbf{q}}_N^+ = 0$, yielding

$$\widehat{\mathbf{v}}_N^+ = \widehat{\mathbf{v}}_N^- = \frac{1}{2}\widehat{\mathbf{q}}_N^+. \tag{371}$$

Thus, the right eigenvectors of $\widehat{\mathbf{A}}$ corresponding to λ_N and $-\lambda_N$ merge into one, i.e.,

$$\begin{bmatrix} \widehat{\mathbf{v}}_N^+ \\ \widehat{\mathbf{v}}_N^- \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{v}}_N^- \\ \widehat{\mathbf{v}}_N^+ \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \widehat{\mathbf{q}}_N^+ \\ \widehat{\mathbf{q}}_N^+ \end{bmatrix}. \tag{372}$$

From $\widehat{\mathbf{q}}_N^+ \in \mathcal{N}(\mathbf{A}_+)$, it follows that

$$\widehat{\mathbf{A}}\begin{bmatrix} \widehat{\mathbf{q}}_N^+ \\ \widehat{\mathbf{q}}_N^+ \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_+\widehat{\mathbf{q}}_N^+ \\ \mathbf{A}_+\widehat{\mathbf{q}}_N^+ \end{bmatrix} = 0, \tag{373}$$

and for the first eigenvector

$$\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix} \in \mathcal{N}(\widehat{\mathbf{A}}),$$

we infer that $\mathbf{w}_0$ is given by

$$\mathbf{w}_0 = \widehat{\mathbf{q}}_N^+. \tag{374}$$

For the second eigenvector

$$\begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} \in \mathcal{N}(\widehat{\mathbf{A}}^2) \setminus \mathcal{N}(\widehat{\mathbf{A}}),$$

we compute $\mathbf{w}_1$ such that the equation

$$\widehat{\mathbf{A}}\begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} = -\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix} \tag{375}$$

is fullfield. Accounting of

$$\widehat{\mathbf{A}}\begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_-\mathbf{w}_1 \\ -\mathbf{A}_-\mathbf{w}_1 \end{bmatrix}, \tag{376}$$

and taking into account that $\mathbf{A}_-$ is invertible, we deduce that for

$$\mathbf{w}_1 = \mathbf{A}_-^{-1}\mathbf{w}_0 = \mathbf{A}_-^{-1}\widehat{\mathbf{q}}_N^+, \tag{377}$$

equation (375) holds true. Thus, $\mathbf{w}_0$ and $\mathbf{w}_1$ are given by (374) and (377), respectively.

Analytical formulas for conservative scattering

Using (365)-(370), we find that the Jordan form representation of the layer matrix is

$$\widehat{\mathbf{A}} = \widehat{\mathbf{V}}\begin{bmatrix} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda} \end{bmatrix} & \mathbf{0}_{(2N-2)\times 1} & \mathbf{0}_{(2N-2)\times 1} \\ \mathbf{0}_{1\times(2N-2)} & 0 & -1 \\ \mathbf{0}_{1\times(2N-2)} & 0 & 0 \end{bmatrix}\overline{\mathbf{V}}^T \tag{378}$$

with

$$\widehat{\mathbf{V}} = \left[\begin{bmatrix} \widehat{\mathbf{V}}_+ & \widehat{\mathbf{V}}_- \\ \widehat{\mathbf{V}}_- & \widehat{\mathbf{V}}_+ \end{bmatrix}, \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}\right], \tag{379}$$

$$\overline{\mathbf{V}} = \left[\begin{bmatrix} -\overline{\mathbf{V}}_+ & \overline{\mathbf{V}}_- \\ \overline{\mathbf{V}}_- & -\overline{\mathbf{V}}_+ \end{bmatrix}, \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1 \end{bmatrix}, \begin{bmatrix} \mathbf{w}_0 \\ -\mathbf{w}_0 \end{bmatrix}\right], \tag{380}$$

$\mathbf{\Lambda} = \mathrm{diag}[\lambda_1,...,\lambda_{N-1}] \in \mathbb{R}^{(N-1)\times(N-1)}$, $\widehat{\mathbf{V}}_\pm = [\widehat{\mathbf{v}}_1^\pm,...,\widehat{\mathbf{v}}_{N-1}^\pm] \in \mathbb{R}^{N\times(N-1)}$, and $\overline{\mathbf{V}}_\pm = [\overline{\mathbf{v}}_1^\pm,...,\overline{\mathbf{v}}_{N-1}^\pm] \in \mathbb{R}^{N\times(N-1)}$, while by virtue of the matrix identity

$$\mathrm{e}^{-\tau\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}, \tag{381}$$

the Jordan form representation of the matrix exponential is

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau} = \widehat{\mathbf{V}}\begin{bmatrix} \begin{bmatrix} \mathbf{\Gamma}(\tau) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}(-\tau) \end{bmatrix} & \mathbf{0}_{(2N-2)\times 1} & \mathbf{0}_{(2N-2)\times 1} \\ \mathbf{0}_{1\times(2N-2)} & 1 & \tau \\ \mathbf{0}_{1\times(2N-2)} & 0 & 1 \end{bmatrix}\overline{\mathbf{V}}^T,$$

(382)

with $\mathbf{\Gamma}(\tau)=\mathrm{diag}[\mathrm{e}^{-\lambda_1\tau},...,\mathrm{e}^{-\lambda_{N-1}\tau}]\in\mathbb{R}^{(N-1)\times(N-1)}$.

Insertion of (379) and (380) in (378) and (382), gives a resolution of the layer matrix

$$\widehat{\mathbf{A}}=-\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\begin{bmatrix}\mathbf{w}_0\\ -\mathbf{w}_0\end{bmatrix}^T+\sum_{k=1}^{N-1}\lambda_k\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}\begin{bmatrix}-\overline{\mathbf{v}}_k^+\\ \overline{\mathbf{v}}_k^-\end{bmatrix}^T-\lambda_k\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}\begin{bmatrix}\overline{\mathbf{v}}_k^-\\ -\overline{\mathbf{v}}_k^+\end{bmatrix}^T, \tag{383}$$

and of the matrix exponential

$$\mathrm{e}^{-\widehat{\mathbf{A}}\tau}=\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_1\end{bmatrix}^T+\begin{bmatrix}\mathbf{w}_1\\ -\mathbf{w}_1\end{bmatrix}\begin{bmatrix}\mathbf{w}_0\\ -\mathbf{w}_0\end{bmatrix}^T+\tau\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\begin{bmatrix}\mathbf{w}_0\\ -\mathbf{w}_0\end{bmatrix}^T+\sum_{k=1}^{N-1}\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}\begin{bmatrix}-\overline{\mathbf{v}}_k^+\\ \overline{\mathbf{v}}_k^-\end{bmatrix}^T+\mathrm{e}^{\lambda_k\tau}\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}\begin{bmatrix}\overline{\mathbf{v}}_k^-\\ -\overline{\mathbf{v}}_k^+\end{bmatrix}^T, \tag{384}$$

respectively. Setting $\tau=0$ in (384) we obtain a resolution of the identity matrix

$$\mathbf{I}_{2N}=\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_1\end{bmatrix}^T+\begin{bmatrix}\mathbf{w}_1\\ -\mathbf{w}_1\end{bmatrix}\begin{bmatrix}\mathbf{w}_0\\ -\mathbf{w}_0\end{bmatrix}^T+\sum_{k=1}^{N-1}\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}\begin{bmatrix}-\overline{\mathbf{v}}_k^+\\ \overline{\mathbf{v}}_k^-\end{bmatrix}^T+\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}\begin{bmatrix}\overline{\mathbf{v}}_k^-\\ -\overline{\mathbf{v}}_k^+\end{bmatrix}^T, \tag{385}$$

which further, gives

$$\begin{aligned}&\mathbf{w}_0\mathbf{w}_1^T+\mathbf{w}_1\mathbf{w}_0^T+\sum_{k=1}^{N-1}\widehat{\mathbf{v}}_k^-\overline{\mathbf{v}}_k^{-T}-\widehat{\mathbf{v}}_k^+\overline{\mathbf{v}}_k^{+T}=\mathbf{I}_N,\\ &\mathbf{w}_0\mathbf{w}_1^T-\mathbf{w}_1\mathbf{w}_0^T+\sum_{k=1}^{N-1}\widehat{\mathbf{v}}_k^+\overline{\mathbf{v}}_k^{-T}-\widehat{\mathbf{v}}_k^-\overline{\mathbf{v}}_k^{+T}=0.\end{aligned} \tag{386}$$

Equations (378), (382), (383), and (384) are the counterparts of (27), (30), (40) and (42), respectively, for the conservative case. It should be pointed out that equations (383) and (384) have been derived by Waterman by employing different arguments.

By means of (382) and (384), the analytical formulas for nonconservative scattering can be extended to conservative scattering as follows:

1. The matrices $\mathbf{D}_1$, $\mathbf{D}_0$ and $\mathbf{D}_\mathrm{b}$ which enter in the layer equation (56) are given by

$$\mathbf{D}_1=\begin{bmatrix}\begin{bmatrix}\mathbf{I}_{N-1}&\mathbf{0}\\ \mathbf{0}&\mathbf{\Gamma}(\overline{\tau})\end{bmatrix}&\mathbf{0}_{(2N-2)\times1}&\mathbf{0}_{(2N-2)\times1}\\ \mathbf{0}_{1\times(2N-2)}&1&0\\ \mathbf{0}_{1\times(2N-2)}&0&1\end{bmatrix},$$

$$\mathbf{D}_0=\begin{bmatrix}\begin{bmatrix}\mathbf{\Gamma}(\overline{\tau})&\mathbf{0}\\ \mathbf{0}&\mathbf{I}_{N-1}\end{bmatrix}&\mathbf{0}_{(2N-2)\times1}&\mathbf{0}_{(2N-2)\times1}\\ \mathbf{0}_{1\times(2N-2)}&1&\tau\\ \mathbf{0}_{1\times(2N-2)}&0&1\end{bmatrix},$$

$$\mathbf{D}_\mathrm{b}=\begin{bmatrix}\begin{bmatrix}\frac{\mathrm{e}^{-\lambda_k\overline{\tau}}-\mathrm{e}^{-\overline{\tau}/\mu_0}}{1/\mu_0-\lambda_k}&\mathbf{0}\\ \mathbf{0}&\frac{1-\mathrm{e}^{-\overline{\tau}(\lambda_k+1/\mu_0)}}{1/\mu_0+\lambda_k}\end{bmatrix}&\mathbf{0}_{(2N-2)\times1}&\mathbf{0}_{(2N-2)\times1}\\ \mathbf{0}_{1\times(2N-2)}&\frac{1-\mathrm{e}^{-\overline{\tau}/\mu_0}}{1/\mu_0}&\frac{\overline{\tau}/\mu_0-1+\mathrm{e}^{-\overline{\tau}/\mu_0}}{(1/\mu_0)^2}\\ \mathbf{0}_{1\times(2N-2)}&0&\frac{1-\mathrm{e}^{-\overline{\tau}/\mu_0}}{1/\mu_0}\end{bmatrix}. \tag{387}$$

2. The solution representations (65) and (69) translate into

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau)=\alpha_N\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}+\beta_N\left(\begin{bmatrix}\mathbf{w}_1\\ -\mathbf{w}_1\end{bmatrix}+\tau\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\right)+\sum_{k=1}^{N-1}\alpha_k\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}+\beta_k\mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}+\widehat{\mathbf{i}}_\mathrm{p}(\tau) \tag{388}$$

and

$$\begin{aligned}\widehat{\mathbf{i}}_\mathrm{d}(\tau)=&\alpha_N\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}+\beta_N\left(\begin{bmatrix}\mathbf{w}_1\\ -\mathbf{w}_1\end{bmatrix}+\tau\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_0\end{bmatrix}\right)\\ &+\sum_{k=1}^{N-1}\alpha_k\left(\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}+\mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}\right)\\ &+\beta_k\left(-\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix}\widehat{\mathbf{v}}_k^+\\ \widehat{\mathbf{v}}_k^-\end{bmatrix}+\mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix}\widehat{\mathbf{v}}_k^-\\ \widehat{\mathbf{v}}_k^+\end{bmatrix}\right)+\widehat{\mathbf{i}}_\mathrm{p}(\tau),\end{aligned} \tag{389}$$

respectively.

3. The analogues of (108) and (109) which give the expressions of the transmission and reflection matrices, respectively, are

$$\widehat{\mathbf{T}}=\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma}(\widehat{\mathbf{V}}_-^0-\widehat{\mathbf{V}}_+^1\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})^{-1}, \tag{390}$$

$$\widehat{\mathbf{R}}=(\widehat{\mathbf{V}}_+^0-\widehat{\mathbf{V}}_-^1\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})(\widehat{\mathbf{V}}_-^0-\widehat{\mathbf{V}}_+^1\mathbf{\Gamma}\overline{\mathbf{V}}_+^T\overline{\mathbf{V}}_-^{-T}\mathbf{\Gamma})^{-1}, \tag{391}$$

where now

$$\mathbf{\Gamma}=\begin{bmatrix}\mathrm{e}^{-\lambda_1\overline{\tau}}&...&0&0\\ &...&&\\ 0&...&\mathrm{e}^{-\lambda_{N-1}\overline{\tau}}&0\\ 0&...&0&1\end{bmatrix}, \tag{392}$$

and

$$\begin{aligned}&\widehat{\mathbf{V}}_\pm^0=[\widehat{\mathbf{v}}_1^\pm,...,\widehat{\mathbf{v}}_{N-1}^\pm,\mathbf{w}_0],\\ &\widehat{\mathbf{V}}_\pm^1=[\widehat{\mathbf{v}}_1^\pm,...,\widehat{\mathbf{v}}_{N-1}^\pm,\pm\mathbf{w}_1],\\ &\overline{\mathbf{V}}_+=[\overline{\mathbf{v}}_1^+,...,\overline{\mathbf{v}}_{N-1}^+,-\mathbf{w}_0],\\ &\overline{\mathbf{V}}_-=[\overline{\mathbf{v}}_1^-,...,\overline{\mathbf{v}}_{N-1}^-,\mathbf{w}_1+\tau\mathbf{w}_0].\end{aligned} \tag{393}$$

A special system of characteristic solutions

We conclude this appendix by presenting a system of characteristic solutions which can be used for nonconservative and conservative scattering. It is of the form

$$\widehat{\mathbf{i}}_\mathrm{d}(\tau)=\sum_{k=1}^{N}\alpha_k\mathbf{a}_k(\tau)+\beta_k\mathbf{b}_k(\tau)+\widehat{\mathbf{i}}_\mathrm{p}(\tau), \tag{394}$$

with

$$\mathbf{a}_k(\tau) = \mathrm{e}^{-\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}, \tag{395}$$

$$\mathbf{b}_k(\tau) = \frac{1}{\lambda_k}\Big(-\mathrm{e}^{-\lambda_k\tau}\begin{bmatrix} \widehat{\mathbf{v}}_k^+ \\ \widehat{\mathbf{v}}_k^- \end{bmatrix} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}\begin{bmatrix} \widehat{\mathbf{v}}_k^- \\ \widehat{\mathbf{v}}_k^+ \end{bmatrix}\Big). \tag{396}$$

Obviously, (394)-(396) are equivalent to (69) if we consider the transformation $\beta_k \to (1/\lambda_k)\beta_k$. As a result, the reflection and transmission matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ are given by (97) and (98), respectively. In the conservative case, $\mathbf{b}_N$ is singular and so is $\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\boldsymbol{\Gamma}$ together with $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$. However, these singularities are removable if we compute the eigenvectors $\widehat{\mathbf{v}}_k^{\pm}$ by the square-root method of Appendix 2. To show this, we consider (342)-(344), set $\mathbf{V} = [\mathbf{v}_k]$, $\widehat{\mathbf{Q}}_+ = [\widehat{\mathbf{q}}_k^+]$ and $\mathbf{Q} \overset{\text{not}}{=} \widehat{\mathbf{Q}}_+^{-T} = [\mathbf{q}_k]$, and write in component form $\widehat{\mathbf{q}}_k^+ = \mathbf{A}_-^{1/2}\mathbf{v}_k$, $\mathbf{q}_k = \mathbf{A}_-^{-1}\widehat{\mathbf{q}}_k^+ = \mathbf{A}_-^{-1/2}\mathbf{v}_k$, and $\widehat{\mathbf{q}}_k^- = -\lambda_k\mathbf{q}_k$. As a result, (396) becomes

$$\begin{aligned}
\mathbf{a}_k &= \frac{1}{2}[\mathrm{e}^{-\lambda_k\tau} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}]\begin{bmatrix} \widehat{\mathbf{q}}_k^+ \\ \widehat{\mathbf{q}}_k^+ \end{bmatrix} \\
&\quad - \frac{1}{2}\lambda_k[\mathrm{e}^{-\lambda_k\tau} - \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}]\begin{bmatrix} \mathbf{q}_k \\ -\mathbf{q}_k \end{bmatrix}, \\
\mathbf{b}_k &= -\frac{1}{2\lambda_k}[\mathrm{e}^{-\lambda_k\tau} - \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}]\begin{bmatrix} \widehat{\mathbf{q}}_k^+ \\ \widehat{\mathbf{q}}_k^+ \end{bmatrix} \\
&\quad + \frac{1}{2}[\mathrm{e}^{-\lambda_k\tau} + \mathrm{e}^{-\lambda_k(\overline{\tau}-\tau)}]\begin{bmatrix} \mathbf{q}_k \\ -\mathbf{q}_k \end{bmatrix}.
\end{aligned} \tag{397}$$

In the limit $\lambda_N \to 0$, we have $\widehat{\mathbf{q}}_N^+, \mathbf{q}_N \neq 0$, and $\widehat{\mathbf{q}}_N^- = 0$. Accounting of $\mathbf{w}_0 = \widehat{\mathbf{q}}_N^+$, $\mathbf{w}_1 = \mathbf{A}_-^{-1}\widehat{\mathbf{q}}_N^+ = \mathbf{q}_N$, and

$$\lim_{\lambda_N\to 0}\Big[-\frac{\mathrm{e}^{-\lambda_N\tau} - \mathrm{e}^{-\lambda_N(\overline{\tau}-\tau)}}{2\lambda_N}\Big] = \tau - \frac{\overline{\tau}}{2}, \tag{398}$$

we get

$$\begin{aligned}
\lim_{\lambda_N\to 0}[\alpha_N\mathbf{a}_N(\tau) + \beta_N\mathbf{b}_N(\tau)] &= (\alpha_N - \frac{\overline{\tau}}{2}\beta_N)\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix} \\
&\quad + \beta_N\Big(\tau\begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_0 \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ -\mathbf{w}_1 \end{bmatrix}\Big),
\end{aligned} \tag{399}$$

which is equivalent to the first two terms in (389). Coming to the reflection and transmission matrices, we use (329) and (344) with $\mathbf{Q} = \widehat{\mathbf{Q}}_+^{-T}$, to compute the matrix products which enter in (97) and (98) as

$$\begin{aligned}
&(\widehat{\mathbf{V}}_+ + \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- + \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1} \\
&= [\widehat{\mathbf{Q}}_+(\mathbf{I}+\boldsymbol{\Gamma}) - \mathbf{Q}\Lambda(\mathbf{I}-\boldsymbol{\Gamma})][\widehat{\mathbf{Q}}_+(\mathbf{I}+\boldsymbol{\Gamma}) + \mathbf{Q}\Lambda(\mathbf{I}-\boldsymbol{\Gamma})]^{-1}
\end{aligned} \tag{400}$$

and

$$\begin{aligned}
&(\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})^{-1} \\
&= [(\widehat{\mathbf{V}}_+ - \widehat{\mathbf{V}}_-\boldsymbol{\Gamma})\boldsymbol{\Lambda}^{-1}][(\widehat{\mathbf{V}}_- - \widehat{\mathbf{V}}_+\boldsymbol{\Gamma})\boldsymbol{\Lambda}^{-1}]^{-1} \\
&= [\widehat{\mathbf{Q}}_+(\mathbf{I}-\boldsymbol{\Gamma})\boldsymbol{\Lambda}^{-1} - \mathbf{Q}(\mathbf{I}+\boldsymbol{\Gamma})][\widehat{\mathbf{Q}}_+(\mathbf{I}-\boldsymbol{\Gamma})\boldsymbol{\Lambda}^{-1} + \mathbf{Q}(\mathbf{I}+\boldsymbol{\Gamma})]^{-1}.
\end{aligned} \tag{401}$$

In the limit $\lambda_N \to 0$, the singularity in $(\mathbf{I}-\boldsymbol{\Gamma})\boldsymbol{\Lambda}^{-1}$ is removable, i.e.,

$$\lim_{\lambda_N\to 0}\frac{1-\mathrm{e}^{-\lambda_N\overline{\tau}}}{\lambda_N} = \overline{\tau},$$

so that (97), (98), (400), and (401) give the expressions of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{T}}$ for nonconservative and conservative scattering. The system of characteristic solutions (394)-(396), as well as the reflection and transmission matrices of (97), (98), (400), and (401) have been used by Nakajima and Tanaka in their matrix formulation of the radiative transfer.

Appendix 4. Asymptotic functions and constants

We define the scaled diffusion pattern vectors by

$$\widehat{\boldsymbol{\iota}}_+ = \lambda_N(\mathbf{I} - \widehat{\mathbf{R}}_\infty^2)^{-1}\widehat{\mathbf{k}}_N, \tag{402}$$

$$\widehat{\boldsymbol{\iota}}_- = \widehat{\mathbf{R}}_\infty\widehat{\boldsymbol{\iota}}_+, \tag{403}$$

and note that for

$$\widehat{\mathbf{Q}}_+ = [\widehat{\mathbf{q}}_1^+, ..., \widehat{\mathbf{q}}_N^+], \tag{404}$$

we have the representations

$$\widehat{\boldsymbol{\iota}}_+ = \frac{1}{2}(\widehat{\mathbf{q}}_N^+ + \lambda_N\mathbf{q}_N^+), \tag{405}$$

$$\widehat{\boldsymbol{\iota}}_- = \widehat{\boldsymbol{\iota}}_+ - \lambda_N\mathbf{q}_N^+, \tag{406}$$

and the orthogonality relations

$$\widehat{\mathbf{k}}_N^T\widehat{\boldsymbol{\iota}}_+ = 1, \tag{407}$$

$$\widehat{\mathbf{k}}_N^T\widehat{\boldsymbol{\iota}}_- = 1. \tag{408}$$

To derive the expressions of the functions that occur in the asymptotic theory, we introduce the discrete approximation to the escape function $\mathbf{k}_N$ by

$$\mathbf{k}_N = \sqrt{\frac{\mathtt{k}}{2\mathtt{m}}}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{k}}_N, \tag{409}$$

and the diffusion pattern vectors $\boldsymbol{\iota}_\pm$ by

$$\boldsymbol{\iota}_\pm = \sqrt{\frac{\mathtt{m}}{2\mathtt{k}}}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\boldsymbol{\iota}}_\pm. \tag{410}$$

As a result, the orthogonality relations for $\widehat{\boldsymbol{\iota}}_+$ and $\widehat{\boldsymbol{\iota}}_-$ yield

$$2\mathbf{k}_N^T\mathbf{W}\mathbf{M}^{-1}\boldsymbol{\iota}_+ = 1, \tag{411}$$

$$2\mathbf{k}_N^T\mathbf{W}\mathbf{M}^{-1}\boldsymbol{\iota}_- = 1. \tag{412}$$

Essentially, the vectors $\boldsymbol{\iota}_+$ and $\boldsymbol{\iota}_-$ are the discrete approximations of the diffusion patterns $i(\mu)$ and $i(-\mu)$, $\mu > 0$, respectively. The constant $\mathtt{m}$ in (409) and (410) is obtained by normalizing the diffusion pattern (cf. (197))

$$\frac{1}{2}\int_{-1}^{1} i(\mu)\mathrm{d}\mu = \frac{1}{2}\mathbf{1}^T\mathbf{W}(\boldsymbol{\iota}_+ + \boldsymbol{\iota}_-) = \frac{1}{2}\sqrt{\frac{\mathtt{m}}{2\mathtt{k}}}\mathbf{1}^T\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\widehat{\mathbf{q}}_N^+ = 1, \tag{413}$$

that is,

$$\mathtt{m} = \frac{8\mathtt{k}}{(\mathbf{1}^T\mathbf{W}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\widehat{\mathbf{q}}_N^+)^2}, \tag{414}$$

where $\mathbf{l}$ is the vector of all ones. Obviously, (411) and (412) are the discrete approximations of the normalization conditions (200) and (202), respectively.

In terms of the escape function $\mathbf{k}_N$, the reflection and transmission matrices are given by

$$\mathbb{R} = \mathbb{R}_\infty - \delta_{m0}\mathtt{r}\mathbf{k}_N\mathbf{k}_N^T, \tag{415}$$

$$\mathbb{T} = \delta_{m0}\mathtt{t}\mathbf{k}_N\mathbf{k}_N^T, \tag{416}$$

$$\mathbb{R}_\infty = \frac{1}{2}\mathbf{M}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\widehat{\mathbf{R}}_\infty\mathbf{W}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}, \tag{417}$$

where

$$\mathtt{r} = \widehat{\mathtt{r}}\frac{\mathtt{m}}{\mathtt{k}} = \frac{\mathtt{m}\mathtt{l}\mathrm{e}^{-2\mathtt{k}\bar{\tau}}}{1-\mathtt{l}^2\mathrm{e}^{-2\mathtt{k}\bar{\tau}}}, \tag{418}$$

$$\mathtt{t} = \widehat{\mathtt{t}}\frac{\mathtt{m}}{\mathtt{k}} = \frac{\mathtt{m}\mathrm{e}^{-\mathtt{k}\bar{\tau}}}{1-\mathtt{l}^2\mathrm{e}^{-2\mathtt{k}\bar{\tau}}}. \tag{419}$$

Moreover, for the azimuthal mode $m = 0$, we have

$$\mathbf{r}_\infty = 2\mathbb{R}_\infty\mathbf{v}, \tag{420}$$

$$r_{s\infty} = 4\mathbf{v}^T\mathbb{R}_\infty\mathbf{v}, \tag{421}$$

and

$$\mathbf{r} = \mathbf{r}_\infty - \mathtt{n}\mathtt{r}\mathbf{k}_N, \tag{422}$$

$$\mathbf{t} = \mathtt{n}\mathtt{t}\mathbf{k}_N, \tag{423}$$

$$r_s = 4\mathbf{v}^T\mathbb{R}\mathbf{v}, \tag{424}$$

where $\mathbf{r}_\infty$ and $r_{s\infty}$ are the plane albedo vector and the spherical albedo of the semi-infinite atmosphere, respectively, $\mathbf{r}$, $\mathbf{t}$ and r_s are the plane albedo vector, the transmission vector, and the spherical albedo of the layer, respectively, and $\mathbf{v} = \mathbf{W}\mathbf{M}^{-1}\mathbf{l}$. The μ-weighted mean of the escape function is given by

$$\mathtt{n} = 2\mathbf{k}_N^T\mathbf{v}, \tag{425}$$

which is the discrete approximation of (206).

The reflection and transmission matrices of the homogeneous layer with an underlying Lambertian surface $\mathbb{R}_A$ and $\mathbb{T}_A$, respectively, as well as the plane albedo $\mathbf{r}_A$ and the spherical albedo r_{sA} are computed as

$$\mathbb{R}_A = \mathbb{R} + \frac{A}{1-Ar_s}\delta_{m0}\mathbf{t}\mathbf{t}^T, \tag{426}$$

$$\mathbb{T}_A = \mathbb{T} + \frac{A}{1-Ar_s}\delta_{m0}\mathbf{r}\mathbf{t}^T, \tag{427}$$

$$\mathbf{r}_A = 2\mathbb{R}_A\mathbf{W}\mathbf{M}^{-1}\mathbf{l} \tag{428}$$

$$r_{sA} = 4\mathbf{v}^T\mathbb{R}_A\mathbf{v}. \tag{429}$$

where A is the ground albedo.

References

[1] S. Twomey, H. Jacobowitz, H. B. Howell, Matrix methods for multiple-scattering problems, J. Atmos. Sci. 23 (3) (1966) 289–298. doi:10.1175/1520-0469(1966)023<0289:mmfmsp>2.0.co;2.

[2] V. V. Sobolev, D. ter Haar, Light Scattering in Planetary Atmospheres, Pergamon Press, 1975.

[3] J. E. Hansen, L. D. Travis, Light scattering in planetary atmospheres, Space Science Reviews 16 (4) (1974) 527–610. doi:10.1007/bf00168069.

[4] W. M. Irvine, Multiple scattering in planetary atmospheres, Icarus 25 (2) (1975) 175–204. doi:10.1016/0019-1035(75)90019-6.

[5] J. Lenoble, Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures, Studies in geophysical optics and remote sensing, A. Deepak, 1985.

[6] K. Stamnes, The theory of multiple scattering of radiation in plane parallel atmospheres, Rev. Geophys. 24 (2) (1986) 299. doi:10.1029/rg024i002p00299.

[7] V. Budak, D. Klyuykov, S. Korkin, Convergence acceleration of radiative transfer equation solution at strongly anisotropic scattering, in: A. A. Kokhanovsky (Ed.), Light scattering reviews, Vol. 5, Springer Berlin Heidelberg, 2010, pp. 147–203. doi:10.1007/978-3-642-10336-0_5.

[8] V. P. Afanas'ev, D. S. Efremenko, A. V. Lubenchenko, On the application of the invariant embedding method and the radiative transfer equation codes for surface state analysis, in: A. Kokhanovsky (Ed.), Light Scattering Reviews 8, Springer Science + Business Media, 2013, pp. 363–423. doi:10.1007/978-3-642-32106-1_8.

[9] V. Rozanov, A. Rozanov, A. Kokhanovsky, J. Burrows, Radiative transfer through terrestrial atmosphere and ocean: Software package SCIATRAN, J Quant Spectrosc Radiat Transf 133 (2014) 13–71. doi:10.1016/j.jqsrt.2013.07.004.

[10] V. Afanas'ev, D. Efremenko, P. Kaplya, Analytical and numerical methods for computing electron partial intensities in the case of multilayer systems, Journal of Electron Spectroscopy and Related Phenomena 210 (2016) 16–29. doi:10.1016/j.elspec.2016.04.006.

[11] N. Rogovtsov, F. Borovik, Application of general invariance relations reduction method to solution of radiation transfer problems, J Quant Spectrosc Radiat Transf 183 (2016) 128–153. doi:10.1016/j.jqsrt.2016.06.024.

[12] G. C. Wick, Über ebene Diffusionsprobleme, Zeitschrift für Physik 121 (11-12) (1943) 702–718. doi:10.1007/BF01339167.

[13] S. Chandrasekhar, Radiative Transfer, Dover Publications, Inc., New York, 1950.

[14] K. Stamnes, R. A. Swanson, A New Look at the Discrete Ordinate Method for Radiative Transfer Calculations in Anisotropically Scattering Atmospheres, J Atmos Sci 38 (2) (1981) 387–389. doi:10.1175/1520-0469(1981)038<0387:ANLATD>2.0.CO;2.

[15] G. N. Plass, G. W. Kattawar, F. E. Catchings, Matrix operator theory of radiative transfer 1: Rayleigh scattering, Appl. Opt. 12 (2) (1973) 314. doi:10.1364/ao.12.000314.

[16] G. W. Kattawar, G. N. Plass, F. E. Catchings, Matrix operator theory of radiative transfer 2: Scattering from maritime haze, Appl. Opt. 12 (5) (1973) 1071. doi:10.1364/ao.12.001071.

[17] H. van de Hulst, A new look at multiple scattering, Tech. report, NASA Inst. Space Studies (1963).

[18] T. Nakajima, M. Tanaka, Matrix formulations for the transfer of solar radiation in a plane-parallel scattering atmosphere, J Quant Spectrosc Radiat Transf 35 (1) (1986) 13–21. doi:10.1016/0022-4073(86)90088-9.

[19] V. Ambarzumian, Diffuse reflection of light by a foggy medium, Dokl. Akad. Nauk SSSR 38 (1943) 229–232.

[20] R. Bellman, R. Kalaba, G. Wing, Invariant imbedding and mathematical physics-i: Particle processes, J. Math. Phys. 1 (1960) 280–308. doi:10.1063/1.1703663.

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

[22] P. J. Flatau, G. L. Stephens, On the fundamental solution of the radiative transfer equation, J. Geophys. Res. 93 (D9) (1988) 11037. doi:10.1029/jd093id09p11037.

[23] V. P. Budak, D. A. Klyuykov, S. V. Korkin, Complete matrix solution of radiative transfer equation for PILE of horizontally homogeneous slabs, J Quant Spectrosc Radiat Transf 112 (7) (2011) 1141–1148. doi:10.1016/j.jqsrt.2010.08.028.

[24] V. P. Budak, D. S. Efremenko, O. V. Shagalov, Efficiency of algorithm for solution of vector radiative transfer equation in turbid medium slab, J.

Phys.: Conf. Ser. 369 (2012) 012021. doi:10.1088/1742-6596/369/1/012021.

[25] A. Doicu, T. Trautmann, Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case, J Quant Spectrosc Radiat Transf 110 (1-2) (2009) 146–158. doi:10.1016/j.jqsrt.2008.09.014.

[26] A. Doicu, T. Trautmann, Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Vector case, J Quant Spectrosc Radiat Transf 110 (1-2) (2009) 159–172. doi:10.1016/j.jqsrt.2008.09.013.

[27] T. Nakajima, M. D. King, Asymptotic theory for optically thick layers: application to the discrete ordinates method, Appl. Opt. 31 (36) (1992) 7669. doi:10.1364/ao.31.007669.

[28] D. Efremenko, A. Doicu, D. Loyola, T. Trautmann, Small-angle modification of the radiative transfer equation for a pseudo-spherical atmosphere, J Quant Spectrosc Radiat Transf 114 (2013) 82–90. doi:10.1016/j.jqsrt.2012.08.013.

[29] D. Efremenko, A. Doicu, D. Loyola, T. Trautmann, Acceleration techniques for the discrete ordinate method, J Quant Spectrosc Radiat Transf 114 (2013) 73–81. doi:10.1016/j.jqsrt.2012.08.014.

[30] K. Stamnes, S. C. Tsay, T. Nakajima, Computation of eigenvalues and eigenvectors for the discrete ordinate and matrix operator methods in radiative transfer, J Quant Spectrosc Radiat Transf 39 (5) (1988) 415–419. doi:10.1016/0022-4073(88)90107-0.

[31] A. H. Karp, J. Greenstadt, J. A. Fillmore, Radiative transfer through an arbitrarily thick, scattering atmosphere, J Quant Spectrosc Radiat Transf 24 (5) (1980) 391–406. doi:10.1016/0022-4073(80)90074-6.

[32] E. G. Yanovitskij, Light Scattering in Inhomogeneous Atmospheres, Springer Nature, 1997. doi:10.1007/978-3-642-60465-2.

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

[34] T. Nakajima, M. Tanaka, Algorithms for radiative intensity calculations in moderately thick atmos using a truncation approximation, J Quant Spectrosc Radiat Transf 40 (1) (1988) 51–69. doi:10.1016/0022-4073(88)90031-3.

[35] M. Mishchenko, J. Dlugach, E. Yanovitskij, N. Zakharova, Bidirectional reflectance of flat, optically thick particulate layers: an efficient radiative transfer solution and applications to snow and soil surfaces, J Quant Spectrosc Radiat Transf 63 (2-6) (1999) 409–432. doi:10.1016/s0022-4073(99)00028-x.

[36] Y. Yang, A. Marshak, J. Mao, A. Lyapustin, J. Herman, A method of retrieving cloud top height and cloud geometrical thickness with oxygen a and b bands for the deep space climate observatory (DSCOVR) mission: Radiative transfer simulations, J Quant Spectrosc Radiat Transf 122 (2013) 141–149. doi:10.1016/j.jqsrt.2012.09.017.

[37] G. McGarragh, P. Gabriel, Efficient computation of radiances for optically thin media by Padé approximants, J Quant Spectrosc Radiat Transf 111 (12-13) (2010) 1885–1899. doi:10.1016/j.jqsrt.2010.03.011.

[38] C. V. Loan, Computing integrals involving the matrix exponential, IEEE Transactions on Automatic Control 23 (3) (1978) 395–404. doi:10.1109/tac.1978.1101743.

[39] I. P. Grant, G. E. Hunt, Discrete space theory of radiative transfer. I. Fundamentals, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 313 (1513) (1969) 183–197. doi:10.1098/rspa.1969.0187.

[40] W. Wiscombe, On initialization, error and flux conservation in the doubling method, J Quant Spectrosc Radiat Transf 16 (8) (1976) 637–658. doi:10.1016/0022-4073(76)90056-x.

[41] A. K. Heidinger, C. O'Dell, R. Bennartz, T. Greenwald, The successive-order-of-interaction radiative transfer model. Part I: Model development, J. Appl. Meteor. Climatol. 45 (10) (2006) 1388–1402. doi:10.1175/jam2387.1.

[42] A. V. Rozanov, V. V. Rozanov, J. P. Burrows, Combined differential-integral approach for the radiation field computation in a spherical shell atmosphere: Nonlimb geometry, Journal of Geophysical Research: Atmospheres 105 (D18) (2000) 22937–22942. doi:10.1029/2000jd900378.

[43] X. Sun, Y. Han, X. Shi, Application of asymptotic theory for computing the reflection of optically thick clouds, Journal of Optics A: Pure and Applied Optics 8 (12) (2006) 1074–1079. doi:10.1088/1464-4258/8/12/007.

[44] M. King, S.-C. Tsay, S. Platnick, M. Wang, L. K.-N., Cloud retrieval algorithms for MODIS: Optical thickness, effective particle radius, and thermodynamic phase, Tech. rep. (1997).
URL http://modis.gsfc.nasa.gov/data/atbd/atbd_mod05.pdf

[45] A. A. Kokhanovsky, V. V. Rozanov, The reflection function of optically thick weakly absorbing turbid layers: a simple approximation, J Quant Spectrosc Radiat Transf 77 (2) (2003) 165–175. doi:10.1016/S0022-4073(02)00085-7.

[46] A. A. Kokhanovsky, Reflection of light from nonasbsorbing semi-infinite cloudy media: a simple approximation, J Quant Spectrosc Radiat Transf 85 (1) (2004) 25–33. doi:10.1016/S0022-4073(03)00192-4.

[47] M. D. King, Determination of the scaled optical thickness of clouds from reflected solar radiation measurements, J. Atmos. Sci. 44 (13) (1987) 1734–1751. doi:10.1175/1520-0469(1987)044<1734:dotsot>2.0.co;2.

[48] K. Kawabata, S. Ueno, The first three orders of scattering in vertically inhomogeneous scattering-absorbing media, Astrophys Space Sci 150 (2) (1988) 327–344. doi:10.1007/bf00641728.

[49] H. wen Chang, T.-L. Wu, Numerical solutions of matrix riccati equations for radiative transfer in a plane-parallel geometry, Waves in Random Media 7 (1) (1997) 147–168. doi:10.1088/0959-7174/7/1/010.

[50] T. Kurosu, V. Rozanov, J. Burrows, Parameterization schemes for terrestrial water clouds in the radiative transfer model GOMETRAN, Journal of Geophysical Research: Atmospheres 102 (D18) (1997) 21809–21823. doi:10.1029/97jd01044.

[51] L. Delves, J. Mohamed, Computational methods for integral equations, Cambridge University Press. Cambridge, 1985.

[52] G. E. Thomas, K. Stamnes, Radiative transfer in the Atmosphere and Ocean, Radiative transfer in the Atmosphere and Ocean. Cambridge University press , 1st edition, 1999.

[53] C. Clenshaw, A. Curtis, A method for numerical integration on an automatic computer, Numerische Mathematik 2 (1) (1960) 197–205. doi:10.1007/bf01386223.

[54] J. Weideman, L. Trefethen, The kink phenomenon in fejér and clenshaw–curtis quadrature, Numerische Mathematik 107 (4) (2007) 707–727. doi:10.1007/s00211-007-0101-2.

The paper was published in Journal of Quantitative Spectroscopy and Radiative Transfer **213**(2018)228–240
DOI: 10.1016/j.jqsrt.2018.03.014

Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements

Víctor Molina García*, Sruthy Sasi, Dmitry S. Efremenko, Adrian Doicu, Diego Loyola

Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung (IMF), 82234 Oberpfaffenhofen, Germany

Abstract

In this paper we analyze the accuracy and efficiency of several radiative transfer models for inferring cloud parameters from radiances measured by the Earth Polychromatic Imaging Camera (EPIC) on board the Deep Space Climate Observatory (DSCOVR). The radiative transfer models are the exact discrete ordinate and matrix operator methods with matrix exponential, and the approximate asymptotic and equivalent Lambertian cloud models. To deal with the computationally expensive radiative transfer calculations, several acceleration techniques such as, for example, the telescoping technique, the method of false discrete ordinate, the correlated k-distribution method and the principal component analysis (PCA) are used. We found that, for the EPIC oxygen A-band absorption channel at 764 nm, the exact models using the correlated k-distribution in conjunction with PCA yield an accuracy better than 1.5 % and a computation time of 18 s for radiance calculations at 5 viewing zenith angles.

Keywords: backscattering, PCA, correlated k-distribution, asymptotic theory, EPIC

1. Introduction

The Earth Polychromatic Imaging Camera (EPIC) on board the Deep Space Climate Observatory (DSCOVR) was designed to measure the atmosphere and surface properties over the whole sunlit face of the Earth from the Lagrange point L_1 (a gravity-neutral position at 1.5×10^6 km away from the Earth). DSCOVR is placed in a Lissajous orbit around the L_1 point, and provides a unique angular perspective at almost backward direction with scattering angles approximately between 168° and 176° (Fig. 1). EPIC scans the entire sunlit face of the Earth at a 2048×2048 pixel resolution, with a pixel size of 12×12 km^2 at the image center. The instrument has 10 spectral channels ranging from the ultraviolet to the near-infrared. Four of them are located in the oxygen A- and B- bands: two absorption channels centered at 688 nm and 764 nm with bandwidths of 0.8 nm and 1.0 nm, respectively, and two continuum channels centered at 680 nm and 780 nm with bandwidths of 3.0 nm and 2.0 nm (Table 1). These channels are used for monitoring the vegetation condition [1], the aerosol layer height and optical depth [2], and the cloud height [3].

The radiative transfer for retrieval of cloud parameters involves, in addition to cloud scattering and absorption, gas absorption and molecular Rayleigh scattering. Usually, it is necessary to consider spectral regions containing several overlapping lines with intensities varying over many orders of magnitude. An accurate method for computing the radiative transfer in a molecular atmosphere relies on line-by-line (LBL) calculations. However, LBL calculations are in most cases too computationally expensive to be used directly in online and even sometimes in offline retrieval algorithms. This prompts us to use exact or approximate radiative transfer models endowed with acceleration techniques. As approximate models, asymptotic [4–6] and equivalent Lambertian cloud models [7–9] are frequently used. In the category of acceleration techniques we include here the correlated k-distribution method [10, 11], the radiance sampling method [12], the optimal spectral mapping method [13] and dimensionality reduction techniques. In the latter case, principal component analysis (PCA) is used to map the spectral radiances into a lower-dimensional subspace in which the inversion is performed [14, 15], or to reduce the dimensionality of the optical properties [16–18]. In addition, the telescoping technique [19, 20] and the method of false discrete ordinate [21–24] can be used to speed up radiative transfer calculations.

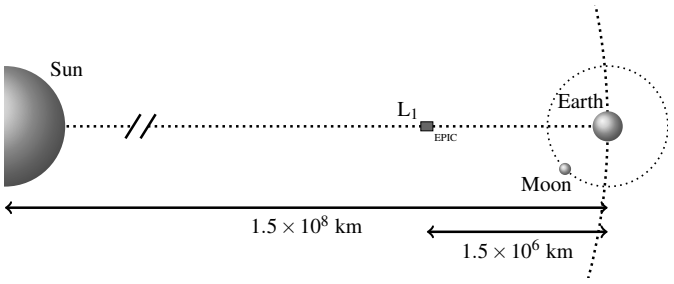


Figure 1: Illustration of EPIC / DSCOVR geometry (distances are not to scale).

For the EPIC instrument, the retrieval is more challenging due to the singular geometry of the radiative transfer problem. Given a particle with a certain size, the scattering phase function shows considerable structure and resonances. When averaged over a size distribution of an ensemble of particles, these features are almost smoothed out, but they are still present in

*Corresponding author
Email address: Victor.MolinaGarcia@dlr.de (Víctor Molina García)

Table 1: Description of EPIC channels, adapted from EPIC's main website (https://epic.gsfc.nasa.gov/epic).

Channel	Central wavelength / nm	Full width / nm	Primary application
1	317.5 ± 0.1	1.0 ± 0.2	Ozone, sulfur dioxide
2	325.0 ± 0.1	2.0 ± 0.2	Ozone
3	340.0 ± 0.3	3.0 ± 0.6	Ozone, aerosols
4	388.0 ± 0.3	3.0 ± 0.6	Aerosols, clouds
5	443.0 ± 1.0	3.0 ± 0.6	Aerosols, clouds
6	551.0 ± 1.0	3.0 ± 0.6	Aerosols
7	680.0 ± 0.2	3.0 ± 0.6	Aerosols, clouds, vegetation
8	687.75 ± 0.20	0.80 ± 0.20	Aerosols, clouds, vegetation
9	764.0 ± 0.2	1.0 ± 0.2	Cloud height
10	779.5 ± 0.3	2.0 ± 0.4	Clouds, vegetation

the backward and forward glories, at scattering angles around 180° and 0°, and in the rainbow region, at around 140° [25]. For an accurate description of the specific features of the single scattering properties in the backward direction, a large number of discrete ordinates is required, even when the delta-M method [26] is used.

The aim of this paper is to analyze the accuracy and efficiency of several radiative transfer models in regard to their applicability to the retrieval of cloud parameters from EPIC measurements. The radiative transfer models, relying on the matrix exponential formalism and endowed with acceleration techniques, include the exact discrete ordinate and matrix operator methods, as well as the approximate asymptotic and equivalent Lambertian cloud models. By "exact" methods, we mean those models used as the starting point to design approximate models by imposing further assumptions.

The paper is organized as follows. In Section 2 we present the exact and approximate radiative transfer models. In Section 3 we summarize the acceleration techniques, focusing on the description of a combined method incorporating the correlated k-distribution and PCA. In Section 4 we analyze the accuracy and efficiency of the radiative transfer models for simulated EPIC measurements in the oxygen A-band absorption channel. Conclusions are formulated in Section 5. Some specific features of the radiative transfer models are outlined in the appendices. Here, we give the main computation formulas of the discrete ordinate and matrix operator methods with matrix exponential, justify the equivalent Lambertian cloud model, and describe the telescoping technique.

2. Radiative transfer models

The radiative transfer equation for the diffuse radiance $I(r,\mathbf{\Omega})$ at point r and in the direction $\mathbf{\Omega} = (\mu,\varphi)$ reads as

$$\mu\frac{\mathrm{d}I}{\mathrm{d}r}(r,\mathbf{\Omega}) = -\sigma_{\mathtt{ext}}(r)I(r,\mathbf{\Omega}) + F_0\frac{\sigma_{\mathtt{sct}}(r)}{4\pi}P(r,\mathbf{\Omega},\mathbf{\Omega}_0)\mathrm{e}^{-\tau^0_{\mathtt{ext}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|)} + \frac{\sigma_{\mathtt{sct}}(r)}{4\pi}\int_{4\pi}P(r,\mathbf{\Omega},\mathbf{\Omega}')I(r,\mathbf{\Omega}')\,\mathrm{d}\Omega', \tag{1}$$

where $\sigma_{\mathtt{ext}}$ and $\sigma_{\mathtt{sct}}$ are the extinction and scattering coefficients, respectively, F_0 is the incident solar flux, P the scattering phase function, $\mathbf{\Omega}_0 = (-\mu_0,\varphi_0)$ with $\mu_0 > 0$ the incident solar direction, and $\tau^0_{\mathtt{ext}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|)$ the solar optical depth between a generic point $\mathbf{r}$ and the characteristic point at the top of the atmosphere $\mathbf{r}_{\mathtt{TOA}}$ in a spherical atmosphere. The formalism is pseudo-spherical, i.e. the multiple-scattering is treated in a plane-parallel atmosphere, while the solar-beam attenuation is computed in a spherical atmosphere [27]. For the phase function P, we consider the conventional expansion in terms of normalized Legendre polynomials P_n, i.e.

$$P(r,\mathbf{\Omega},\mathbf{\Omega}') = P(r,\cos\Theta) = \sum_{n=0}^{\infty} c_n\chi_n(r)\,P_n(\cos\Theta), \tag{2}$$

where $c_n = \sqrt{(2n+1)/2}$ and $\cos\Theta = \mathbf{\Omega}\cdot\mathbf{\Omega}'$. The radiative transfer equation (1) is subject to the top-of-atmosphere boundary condition ($r = r_{\mathtt{TOA}}$),

$$I(r_{\mathtt{TOA}},\mathbf{\Omega}^-) = 0, \tag{3}$$

and the surface boundary condition ($r = r_{\mathtt{s}}$),

$$I(r_{\mathtt{s}},\mathbf{\Omega}^+) = F_0\frac{A}{\pi}\mu_0\,\rho(\mathbf{\Omega}^+,\mathbf{\Omega}_0)\,\mathrm{e}^{-\tau^0_{\mathtt{ext}}(|\mathbf{r}_{\mathtt{s}}-\mathbf{r}_{\mathtt{TOA}}|)} + \frac{A}{\pi}\int_{2\pi}I(r_{\mathtt{s}},\mathbf{\Omega}^-)\,|\mu^-|\,\rho(\mathbf{\Omega}^+,\mathbf{\Omega}^-)\,\mathrm{d}\Omega^-, \tag{4}$$

where A and ρ are the surface albedo and the normalized bidirectional reflection function, respectively, and the notations $\mathbf{\Omega}^+$ and $\mathbf{\Omega}^-$ stand for upward and downward directions, respectively.

In the discrete ordinate method, we assume a cosine-azimuthal expansion of the diffuse radiance ($\varphi_0 = 0$),

$$I(r,\mathbf{\Omega}) = \sum_{m=0}^{\infty} I_m(r,\mu)\,\cos m\varphi, \tag{5}$$

and for each azimuthal component $I_m(r,\mu)$ we discretize the radiative transfer equation in the angular domain by considering a set of Gauss-Legendre quadrature points and weights $\{\mu_k, w_k\}_{k=1}^{M}$ in the interval $(0,1)$; thus, M is the number of discrete ordinates per hemisphere. The atmosphere is discretized in N levels: $r_1 = r_{\mathtt{TOA}} > r_2 > \ldots > r_N = r_{\mathtt{s}}$, and a layer j, bounded above by the level r_j and below by the level r_{j+1}, has the geometrical thickness $\triangle r_j = r_j - r_{j+1}$. The extinction and scattering coefficients as well as the phase function coefficients are assumed to be constant within each layer; their average values in layer j are $\sigma_{\mathtt{ext}j}$, $\sigma_{\mathtt{sct}j}$ and χ_{nj}, respectively. Also, we must require the intensity to be continuous across the layer interfaces. In layer j, we are led to the linear system of differential equations

$$\frac{\mathrm{d}\mathbf{i}_m}{\mathrm{d}r}(r) = \mathbf{A}_{mj}\,\mathbf{i}_m(r) + \mathrm{e}^{-\tau^0_{\mathtt{ext}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|)}\,\mathbf{b}_{mj}, \quad r_{j+1}\le r\le r_j, \tag{6}$$

where (the abbreviation "not" stands for a notation definition)

$$\mathbf{i}_m(r) = \begin{bmatrix}\mathbf{i}^+_m(r)\\ \mathbf{i}^-_m(r)\end{bmatrix} \overset{\mathrm{not}}{=} [\mathbf{i}^+_m(r)\,;\,\mathbf{i}^-_m(r)]^{\mathsf{T}} \tag{7}$$

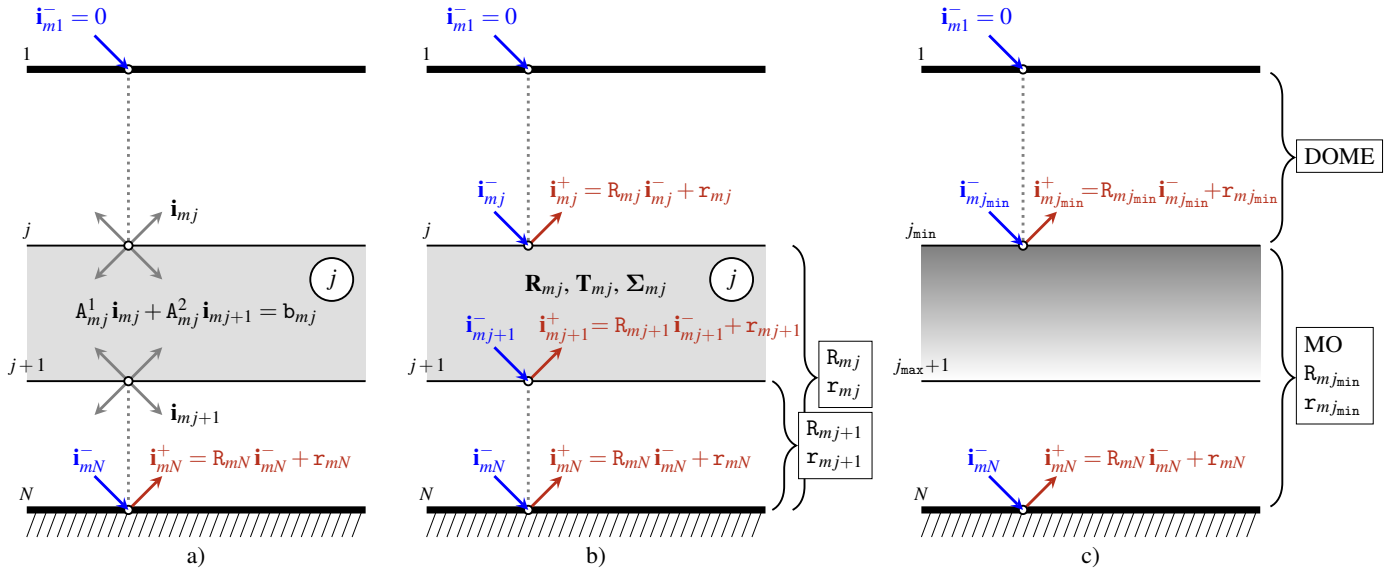


Figure 2: Illustration of DOME (a), the conventional matrix operator method (b), and MOME (c).

is the radiance vector in the discrete-ordinate space, and $[\mathbf{i}^{\pm}_m(r)]_k = I_m(r, \pm\mu_k)$, $k = 1, \ldots, M$. The layer matrix $\mathbf{A}_{mj}$ has a block structure

$$\mathbf{A}_{mj} = \begin{bmatrix} \mathbf{A}^{11}_{mj} & \mathbf{A}^{12}_{mj} \\ -\mathbf{A}^{12}_{mj} & -\mathbf{A}^{11}_{mj} \end{bmatrix}, \tag{8}$$

with entries

$$[\mathbf{A}^{11}_{mj}]_{kl} = \frac{1}{2\,\mu_k}\,[w_l\,\sigma_{\mathtt{sct}j}\,p_{mj}(\mu_k, \mu_l) - 2\,\sigma_{\mathtt{ext}j}\,\delta_{kl}], \tag{9}$$

$$[\mathbf{A}^{12}_{mj}]_{kl} = \frac{1}{2\,\mu_k}\,w_l\,\sigma_{\mathtt{sct}j}\,p_{mj}(\mu_k, -\mu_l), \tag{10}$$

while the entries of the layer vector $\mathbf{b}_{mj} = [\mathbf{b}^{+}_{mj}; \mathbf{b}^{-}_{mj}]^{\mathsf{T}}$ are given by

$$[\mathbf{b}^{\pm}_{mj}]_k = \pm\frac{1}{\mu_k}\,(2 - \delta_{m0})\,\frac{F_0}{4\pi}\,\sigma_{\mathtt{sct}j}\,p_{mj}(\pm\mu_k, -\mu_0), \tag{11}$$

where

$$p_{mj}(\mu, \mu') = \sum_{n=m}^{2M-1} \chi_{nj}\,P^m_n(\mu)\,P^m_n(\mu') \tag{12}$$

are the azimuthal expansion coefficients of the phase functions.

In the following we describe the exact and approximate radiative transfer models for the retrieval of cloud parameters.

2.1. Discrete ordinate method with matrix exponential

The Discrete Ordinate method with Matrix Exponential (DOME) [20, 28] is illustrated in Fig. 2a. The method relies on the layer equation, which relates the level values of the radiance field $\mathbf{i}_{mj} = [\mathbf{i}^{+}_{mj}; \mathbf{i}^{-}_{mj}]^{\mathsf{T}}$ with $[\mathbf{i}^{\pm}_{mj}]_k = I_m(r_j, \pm\mu_k)$, $k = 1, \ldots, M$, that is,

$$\mathtt{A}^1_{mj}\,\mathbf{i}_{mj} + \mathtt{A}^2_{mj}\,\mathbf{i}_{mj+1} = \mathtt{b}_{mj}. \tag{13}$$

The layer equation together with the boundary conditions at the top and the bottom of the atmosphere, i.e. (cf. (3))

$$\mathbf{i}^{-}_{m1} = 0, \tag{14}$$

and (cf. (4))

$$\mathbf{i}^{+}_{mN} = \mathtt{R}_{mN}\,\mathbf{i}^{-}_{mN} + \mathtt{r}_{mN}, \tag{15}$$

with

$$[\mathtt{R}_{mN}]_{kl} = 2\,A\,w_l\,\mu_l\,\rho_m(\mu_k, -\mu_l), \tag{16}$$

$$[\mathtt{r}_{mN}]_k = A\,\frac{F_0}{\pi}\,\mu_0\,\rho_m(\mu_k, -\mu_0)\,\mathrm{e}^{-\tau^0_{\mathtt{ext}}(|\mathbf{r}_{\mathtt{s}} - \mathbf{r}_{\mathtt{TOA}}|)}, \tag{17}$$

respectively, are assembled into the global matrix $\mathtt{A}_m$ of the entire atmosphere, and the solution of the resulting system of equations $\mathtt{A}_m\,\mathbf{i}_m = \mathtt{b}_m$ yields the level values of the radiance field. In (16) and (17), ρ_m are the azimuthal expansion coefficients of the normalized bi-directional reflection function, i.e.

$$\rho(\mathbf{\Omega}^+, \mathbf{\Omega}^-) = \sum_{m=0}^{2M-1} (2 - \delta_{m0})\,\rho_m(\mu, -\mu')\,\cos[m(\varphi - \varphi')], \tag{18}$$

for $\mathbf{\Omega} = (\mu, \varphi)$ and $\mathbf{\Omega}' = (-\mu', \varphi')$ with $\mu, \mu' > 0$. The matrix $\mathtt{A}_m$ of dimension $2MN \times 2MN$ has $3M-1$ sub- and super-diagonals and it may be compressed into band-storage and then inverted using, for example, the LU factorization.

The layer quantities $\mathtt{A}^1_{mj}$, $\mathtt{A}^2_{mj}$ and $\mathtt{b}_{mj}$ are expressed in terms of the exponential of the layer matrix $\mathbf{A}_{mj}$, i.e. $\exp(-\mathbf{A}_{mj}\,\triangle r_j)$. The matrix exponential can be computed by using the eigendecomposition method or the Padé approximation; the resulting expressions for $\mathtt{A}^1_{mj}$, $\mathtt{A}^2_{mj}$ and $\mathtt{b}_{mj}$ are listed in Appendix A. The Padé approximation to the matrix exponential is less time-consuming than the eigendecomposition method, but it is only applicable to optically thin layers, for which the condition $\|\mathbf{A}_{mj}\,\triangle r_j\| \le 1$ is satisfied. In practice, if $\|\mathbf{A}_{mj}\,\triangle r_j\|_1 \le 1$, the matrix exponential is computed by means of the Padé approximation; otherwise, the eigendecomposition method is used. By this procedure, the computation speed is enhanced.

2.2. Matrix operator method with matrix exponential

The Matrix Operator method with Matrix Exponential (MOME) is a combination of the conventional matrix operator method, e.g. [21], and the discrete ordinate method with matrix exponential. The method is dedicated to modeling the radiative transfer in a molecular atmosphere containing a homogeneous cloud placed between the top level $r_{j_{\min}}$ and the bottom level $r_{j_{\max}+1}$. Note that the cloud homogeneity assumption is only required by the approximate models (in particular, by the asymptotic model) to be introduced in the next subsection.

Before describing MOME, we briefly review the basic concepts of the conventional matrix operator method. The method, which is illustrated in Fig. 2b, uses the interaction principle equation

$$\begin{bmatrix} \mathbf{i}^+_{mj} \\ \mathbf{i}^-_{mj+1} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{mj} & \mathbf{T}_{mj} \\ \mathbf{T}_{mj} & \mathbf{R}_{mj} \end{bmatrix} \begin{bmatrix} \mathbf{i}^-_{mj} \\ \mathbf{i}^+_{mj+1} \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}^+_{mj} \\ \mathbf{\Sigma}^-_{mj} \end{bmatrix}, \tag{19}$$

where $\mathbf{R}_{mj}$ and $\mathbf{T}_{mj}$ are the reflection and transmission matrices of layer j, and $\mathbf{\Sigma}^{\pm}_{mj}$ is the source vector. As in DOME, $\mathbf{R}_{mj}$, $\mathbf{T}_{mj}$ and $\mathbf{\Sigma}_{mj}$ are given in terms of the exponential of the layer matrix $\mathbf{A}_{mj}$; their expressions computed by using the matrix eigendecomposition method and the Padé approximation are given in Appendix B. The computation process is an upward recurrence over the atmospheric layers, and uses the concept of a "stack" [29]. The stack $j+1$, i.e. the group of contiguous layers bounded above by the level r_{j+1} and below by the surface level r_N, is characterized by the reflection matrix $\mathtt{R}_{mj+1}$ and the reflection vector $\mathtt{r}_{mj+1}$, so that the interaction principle equation for the stack $j+1$ reads as

$$\mathbf{i}^+_{mj+1} = \mathtt{R}_{mj+1}\, \mathbf{i}^-_{mj+1} + \mathtt{r}_{mj+1}. \tag{20}$$

From (19) and (20), we obtain the interaction principle equation for stack j , i.e. the group of contiguous layers bounded above by the level r_j and below by the surface level r_N,

$$\mathbf{i}^+_{mj} = \mathtt{R}_{mj}\, \mathbf{i}^-_{mj} + \mathtt{r}_{mj}, \tag{21}$$

where $\mathtt{R}_{mj}$ and $\mathtt{r}_{mj}$ are computed by using the adding formulas

$$\mathtt{R}_{mj} = \mathbf{R}_{mj} + \mathbf{T}_{mj}\, \mathtt{R}_{mj+1}\, (\mathbf{I} - \mathbf{R}_{mj}\, \mathtt{R}_{mj+1})^{-1}\, \mathbf{T}_{mj}, \tag{22}$$

$$\mathtt{r}_{mj} = \Sigma^+_{mj} + \mathbf{T}_{mj}\, r_{mj+1} + \mathbf{T}_{mj}\, \mathtt{R}_{mj+1}\, (\mathbf{I} - \mathbf{R}_{mj}\, \mathtt{R}_{mj+1})^{-1}\, (\mathbf{R}_{mj}\, \mathtt{r}_{mj+1} + \Sigma^-_{mj}). \tag{23}$$

The procedure is initialized with (15) and it is repeated until the last layer is added to the stack. In the matrix operator method, the dimension of the problem is small, because the dimension of the matrices that have to be inverted is $2M \times 2M$. In spite of this fact, its efficiency is reduced because of the considerable number of matrix inversions and matrix multiplications required by the adding algorithm. Another disadvantage is that the computation of the radiance in an arbitrary viewing direction by the source function integration method requires the storage of the reflection and transmission matrices of all individual layers. Alternatively, and to keep the memory usage low, we may interpolate the radiance field in the discrete ordinate space by cubic splines, or we may use the method of false discrete ordinate (see Section 2.3).

The combined model MOME, which is illustrated in Fig. 2c, involves two computation steps:

1. compute the reflection matrix $\mathtt{R}_{mj_{\min}}$ and the reflection vector $\mathtt{r}_{mj_{\min}}$ of stack $j_{\min}$ (containing all layers below the cloud top height) in the framework of the conventional matrix operator method;
2. compute the level values of the radiance $\mathbf{i}_{mj}$ for $1 \le j \le j_{\min}$, in the framework of the discrete ordinate method with matrix exponential, by using as surface boundary condition the interaction principle equation

$$\mathbf{i}^+_{mj_{\min}} = \mathtt{R}_{mj_{\min}}\, \mathbf{i}^-_{mj_{\min}} + \mathtt{r}_{mj_{\min}}. \tag{24}$$

Taking this model as a starting point, several approximate and less time-consuming models have been designed. These are described in the next subsection.

2.3. Approximate models

The approximate models are based on some preliminary simplifications which we now describe. Firstly, the atmosphere below the cloud bottom height is replaced by an equivalent Lambertian surface. The resulting model corresponding to a cloud with an underlying Lambertian surface can be summarized as follows:

1. for the azimuthal mode $m = 0$, compute the reflection matrix $\mathtt{R}_{mj_{\max}+1}$ and the reflection vector $\mathtt{r}_{mj_{\max}+1}$ of stack $j_{\max}$ containing all layers below the cloud bottom height;
2. compute the spherical albedo at the cloud bottom

$$A_{\mathtt{cb}} = 2\, \mathbf{v}^{\mathsf{T}}\, \mathtt{R}_{mj_{\max}+1}\, \mathbf{1}, \tag{25}$$

   where $[\mathbf{v}]_k = w_k\, \mu_k$, $k = 1, \ldots, M$ and $\mathbf{1} = [1, \ldots, 1]^{\mathsf{T}}$;
3. use the surface boundary condition

$$\mathbf{i}^+_{mj_{\max}+1} = \mathtt{R}_{A_{\mathtt{cb}}}\, \mathbf{i}^-_{mj_{\max}+1} + \mathtt{r}_{A_{\mathtt{cb}}} \quad \text{for} \quad m = 0, \tag{26}$$

   with

$$[\mathtt{R}_{A_{\mathtt{cb}}}]_{kl} = 2\, A_{\mathtt{cb}}\, w_l\, \mu_l, \tag{27}$$

$$[\mathtt{r}_{A_{\mathtt{cb}}}]_k = A_{\mathtt{cb}}\, \frac{F_0}{\pi}\, \mu_0\, \mathrm{e}^{-\tau^0_{\mathtt{ext}}(|\mathbf{r}_{\mathtt{cb}} - \mathbf{r}_{\mathtt{TOA}}|)}, \tag{28}$$

   to initialize the adding procedure for computing the reflection matrix $\mathtt{R}_{mj_{\min}}$ and the reflection vector $\mathtt{r}_{mj_{\min}}$ of stack $j_{\min}$ containing all layers below the cloud top height;
4. compute the radiance field of the atmosphere above the cloud top height using as surface boundary condition the interaction principle equation (24).

Secondly, the atmosphere within the cloud (between the layers $j_{\min}$ and $j_{\max}$) is homogenized. The equivalent homogeneous layer $j_{\min}$ is characterized by

$$\sigma^{\mathtt{hom}}_{\mathtt{ext}\, j_{\min}} = \frac{\sum_{j=j_{\min}}^{j_{\max}} \sigma_{\mathtt{ext}\, j}\, \triangle r_j}{\sum_{j=j_{\min}}^{j_{\max}} \triangle r_j}, \quad \sigma^{\mathtt{hom}}_{\mathtt{sct}\, j_{\min}} = \frac{\sum_{j=j_{\min}}^{j_{\max}} \sigma_{\mathtt{sct}\, j}\, \triangle r_j}{\sum_{j=j_{\min}}^{j_{\max}} \triangle r_j}, \tag{29}$$

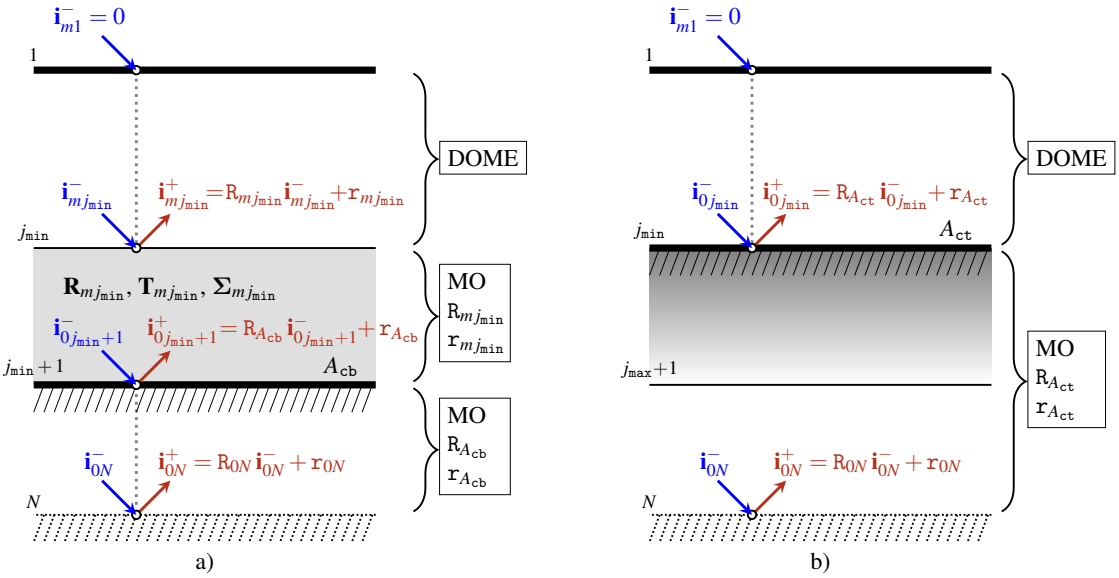


Figure 3: Illustration of a homogeneous cloud layer with an underlying Lambertian surface (a) and of an equivalent Lambertian cloud (b).

and

$$\sigma^{\mathrm{hom}}_{\mathtt{sct}\,j_{\mathtt{min}}}\chi^{\mathrm{hom}}_{n\,j_{\mathtt{min}}} = \frac{\sum_{j=j_{\mathtt{min}}}^{j_{\mathtt{max}}}\sigma_{\mathtt{sct}\,j}\chi_{nj}\,\triangle r_j}{\sum_{j=j_{\mathtt{min}}}^{j_{\mathtt{max}}}\triangle r_j}, \quad n = m,\ldots,2M-1. \quad (30)$$

The resulting model, corresponding to a homogeneous cloud layer with an underlying Lambertian surface and being illustrated in Fig. 3a, requires only one adding step. The homogenization accounts for the scattering and absorption in the cloud, the Rayleigh scattering, as well as the gas absorption between the layers $j_{\mathtt{min}}$ and $j_{\mathtt{max}}$. Strictly speaking, the equivalent homogeneous extinction coefficient $\sigma^{\mathrm{hom}}_{\mathtt{ext}\,j_{\mathtt{min}}}$ should be computed, for example, by minimizing the mismatch between the radiance corresponding to an inhomogeneous atmosphere within the cloud and the radiance corresponding to a homogeneous atmosphere. Although the homogenization formula (29) is a potential source of error (especially for strong gas absorption), it has been adopted in our analysis because of its simplicity.

We come now to the first approximate model. When the optical thickness is sufficiently large, the reflection and transmission matrices $\mathbf{R}_{mj_{\mathtt{min}}}$ and $\mathbf{T}_{mj_{\mathtt{min}}}$ in (22) and (23) can be expressed by simple analytical expressions deriving from the asymptotic theory of thick layers. In the classical asymptotic theory, the reflection and transmission matrices for optically thick layers are given by (we omit the dependency on the layer index $j_{\mathtt{min}}$) [30]

$$\mathbb{R}_m = \mathbb{R}_{\infty m} - \delta_{m0}\,\frac{\mathtt{m}\,\mathtt{l}\,\mathrm{e}^{-2\mathtt{k}\tau}}{1-\mathtt{l}^2\,\mathrm{e}^{-2\mathtt{k}\tau}}\,\mathbf{k}_{mM}\,\mathbf{k}^{\mathsf{T}}_{mM}, \quad (31)$$

$$\mathbb{T}_m = \delta_{m0}\,\frac{\mathtt{m}\,\mathrm{e}^{-\mathtt{k}\tau}}{1-\mathtt{l}^2\,\mathrm{e}^{-2\mathtt{k}\tau}}\,\mathbf{k}_{mM}\,\mathbf{k}^{\mathsf{T}}_{mM}, \quad (32)$$

where $\mathbb{R}_m = (1/2)\,\mathbf{R}_m\,\mathbf{M}\,\mathbf{W}^{-1}$ and $\mathbb{T}_m = (1/2)\,\mathbf{T}_m\,\mathbf{M}\,\mathbf{W}^{-1}$ are the discrete-ordinate approximations of the reflection and transmission functions of the homogeneous layer $R_m(\mu,\mu')$ and $T_m(\mu,\mu')$, respectively, $\mathbb{R}_{\infty m} = (1/2)\,\mathbf{R}_{\infty m}\,\mathbf{M}\,\mathbf{W}^{-1}$ is the discrete-ordinate approximation of the reflection function of a semi-infinite atmosphere $R_{\infty m}(\mu,\mu')$, $[\mathbf{W}]_{kl} = w_k\delta_{kl}$, and $[\mathbf{M}]_{kl} = (1/\mu_k)\delta_{kl}$, $k,\,l = 1,\ldots,M$. Furthermore, in (31) and (32), $\mathtt{k} = \lambda_M$ is the diffusion exponent describing the attenuation of the radiation in the diffusion domain, λ_M is the smallest eigenvalue of the layer matrix $\mathbf{A}_m$, τ is the optical thickness of the homogeneous layer, $\mathtt{m}$ and $\mathtt{l}$ are the constants defined in [30], and $\mathbf{k}_{0M}$ is a discrete-ordinate approximation of the escape function $K(\mu)$. From (31) and (32) we see that $\mathbb{R}_m$ depends on the azimuthal index m through the reflection matrix of a semi-infinite atmosphere $\mathbb{R}_{\infty m}$, while $\mathbb{T}_m$ is azimuthally independent (it depends only on the azimuthal mode $m = 0$). A discrete-ordinate model for computing $\mathbb{R}_m$ and $\mathbb{T}_m$, and so $\mathbf{R}_m$ and $\mathbf{T}_m$, is described in [31]. In this model, the azimuthal independent parts of $\mathbb{R}_m$ and $\mathbb{T}_m$ are derived by neglecting the azimuthal modes $m > 0$, and the contributions of the terms corresponding to eigenvalues larger than the smallest eigenvalue λ_M of $\mathbf{A}_m$. A more accurate asymptotic model, the so-called P-order asymptotic model, has been introduced in [32]. This model accounts for all azimuthal modes m, and the contributions of the terms corresponding to eigenvalues

$$\lambda_{M-P} \geq \lambda_{M-(P-1)} \geq \cdots \geq \lambda_M,$$

where P is the order of approximation. The case $P = 0$ corresponds in some sense to the classical asymptotic model, as it accounts for all azimuthal modes m. It should be pointed out that, in the classical asymptotic theory, the computation of the reflection and transmission matrices can be speeded up by using parametrizations of the constants and functions appearing in (31) and (32) [32, 33]. In our numerical analysis we are mainly concerned about the accuracy of the approximate models and, for this reason, we do not use any parametrization in (31) and (32); we simulate a P-order asymptotic model by simply setting

$$\Gamma_{mj} = \mathrm{diag}[0,\ldots,0,\exp(-\lambda_{M-P}\,\triangle r_j),\ldots,\exp(-\lambda_M\,\triangle r_j)]$$

in place of (60) in Appendix B.

The second approximate model to be discussed is the Equivalent Lambertian Cloud (ELC) model (see Fig. 3b). In the ELC model, the atmosphere below the cloud top height is replaced by an equivalent Lambertian surface with a ground albedo equal to the spherical albedo of the atmosphere. The mathematical background of the ELC model is discussed in Appendix C. In the framework of MOME, this model can be simulated as follows:

1. for the azimuthal mode $m = 0$, compute the reflection matrix $\mathtt{R}_{mj_{\mathtt{min}}}$ and the reflection vector $\mathtt{r}_{mj_{\mathtt{min}}}$ of stack $j_{\mathtt{min}}$ containing all layers below the cloud top height;
2. compute the spherical albedo at the cloud top

$$A_{\mathtt{ct}} = 2\,\mathbf{v}^{\mathsf{T}}\,\mathtt{R}_{mj_{\mathtt{min}}}\,\mathbf{1};$$

3. compute the radiance field of the atmosphere above the cloud top height by using the surface boundary condition

$$\mathbf{i}^{+}_{mj_{\mathtt{min}}} = \mathtt{R}_{A_{\mathtt{ct}}}\,\mathbf{i}^{-}_{mj_{\mathtt{min}}} + \mathtt{r}_{A_{\mathtt{ct}}} \quad \text{for } m = 0.$$

The ELC model ensures the conservation of radiative flux if the plane albedo $r(\mu)$ is almost constant with respect to μ. Note that in a discrete-ordinate setting, the plane albedo $r(\mu)$ is modeled by the plane albedo vector $\mathbf{r} = \mathtt{R}_{mj_{\mathtt{min}}}\,\mathbf{1}$.

3. Acceleration techniques

In addition to the Padé approximation, the standard delta-M method [11], which expands the original phase function into a delta-function forward peak plus a new less anisotropic phase function, and the truncated-plus-single-scattering (TMS) method [34], several acceleration techniques are implemented in the radiative transfer models. These are summarized below.

3.1. The telescoping technique

The telescoping technique relies on the following result [19, 20]: For an atmosphere consisting of gas molecules (Rayleigh layers) and a group of contiguous cloud layers, the azimuthal expansion coefficients of the phase functions p_{mj} vanish for all $m > 2$, and all Rayleigh layers j. Consequently, the layer matrix $\mathbf{A}_{mj}$ becomes a diagonal matrix and the layer vector $\mathbf{b}_{mj}$ vanishes. The main idea of the telescoping technique is to solve a reduced boundary-value problem for the cloud layers, and to compute the radiances at the remaining levels recursively by means of the extinction law. The telescoping technique for DOME and MOME is outlined in Appendix D.

3.2. The method of false discrete ordinate

The method of false discrete ordinate has been discussed in connection with the matrix operator method in [21–23] and the discrete ordinate method with matrix exponential in [24]. For a single viewing direction, an additional stream (directional cosine) is introduced as an extra Gaussian quadrature point with zero weight. The upward radiance at the false discrete ordinate is exactly the upward radiance in the direction of the line of sight computed by the source function integration method. The method of false discrete ordinate eliminates the source function integration step, but increases the dimension of the layer equation. As a result, for applications involving one or two viewing zenith angles, the method of false discrete ordinate does not increase significantly the computation time of the forward model.

3.3. The correlated k-distribution method and the PCA technique

The correlated k-distribution method [10] involves grouping spectral intervals according to absorption coefficient strength, while PCA is used here as a dimensionality reduction technique of the optical properties [16–18]. Although these methods have been applied separately in the literature, they can work together. To show this, we briefly summarize their basic concepts.

Let $g(\lambda)$ be the slit function of the instrument, s the slit width, and $\{\lambda_k\}_{k=1}^{N_\lambda}$ a discrete set of N_λ equally-spaced wavelengths in the interval $[\lambda - s/2,\ \lambda + s/2]$ with

$$\lambda_1 - \frac{\triangle\lambda}{2} = \lambda - \frac{s}{2} \quad \text{and} \quad \lambda_{N_\lambda} + \frac{\triangle\lambda}{2} = \lambda + \frac{s}{2},$$

where $\triangle\lambda$ is the discretization step. The signal received by the instrument is the convolution of the slit function and the monochromatic radiance, and can be approximated by

$$\begin{aligned} I(\lambda) &= \int_{\lambda-s/2}^{\lambda+s/2} g(\lambda-\lambda')\,I(\lambda')\,\mathrm{d}\lambda' \\ &\approx \sum_{k=1}^{N_\lambda} g(\lambda-\lambda_k) \int_{\lambda_k-\triangle\lambda/2}^{\lambda_k+\triangle\lambda/2} I(\lambda')\,\mathrm{d}\lambda'. \end{aligned} \tag{33}$$

As gas absorption has greater spectral variation than molecular and particulate scattering, we may write $I(\lambda) = I(\sigma^{\mathtt{gas}}_{\mathtt{abs}}(\lambda))$. The most accurate method for computing the integral in (33) involves a detailed line-by-line (LBL) calculation of the gas absorption coefficient versus wavelength. On the other hand, the correlated k-distribution method is based on the observation that, for a homogeneous atmosphere, the transmission within a spectral interval is independent of the LBL variation of the gas absorption coefficient $\sigma^{\mathtt{gas}}_{\mathtt{abs}}$ with respect to the wavelength λ, but depends only on the distribution of $\sigma^{\mathtt{gas}}_{\mathtt{abs}}$ within the interval [35]. In this regard, let $F = F(\sigma^{\mathtt{gas}}_{\mathtt{abs}k})$ be the cumulative density function of $\sigma^{\mathtt{gas}}_{\mathtt{abs}}(\lambda)$ in the spectral interval $[\lambda_k - \triangle\lambda/2, \lambda_k + \triangle\lambda/2]$, and $\sigma^{\mathtt{gas}}_{\mathtt{abs}k}(F)$ the quantile function or the inverse distribution function. The signal received by the instrument can then be computed as

$$\begin{aligned} I(\lambda) &= \triangle\lambda \sum_{k=1}^{N_\lambda} g(\lambda-\lambda_k) \int_0^1 I(\sigma^{\mathtt{gas}}_{\mathtt{abs}k}(F))\,\mathrm{d}F \\ &= \triangle\lambda \sum_{k=1}^{N_\lambda} g(\lambda-\lambda_k) \sum_{l=1}^{N_{\mathtt{c}}} \omega_l\, I(\sigma^{\mathtt{gas}}_{\mathtt{abs}k}(F_l)), \end{aligned} \tag{34}$$

where (F_l, ω_l) are a set of $N_{\mathtt{c}}$ quadrature points and weights within the interval $[0, 1]$. The $\sigma^{\mathtt{gas}}_{\mathtt{abs}k}(F_l)$ can be computed by inverting the cumulative density functions of the LBL gas absorption coefficients or, in the case of the "exponential-sum

fitting of transmittance" method [11], by solving a nonlinear least squares problem of dimension $N_c \times M$. Let us define a new set of wavelengths $\{\overline{\lambda}_p\}_{p=1}^{N_{\overline{\lambda}}}$ and weights $\{\overline{\omega}_p\}_{p=1}^{N_{\overline{\lambda}}}$ with $N_{\overline{\lambda}} = N_\lambda N_c$, through the relations $\overline{\lambda}_p = \lambda_k$ and $\overline{\omega}_p = \triangle\lambda\,\omega_l$, where $p = l + (k-1)N_c$ for $k = 1, \ldots, N_\lambda$, $l = 1, \ldots, N_c$, and set accordingly $\sigma_{\mathtt{abs}}^{\mathtt{gas}}(\overline{\lambda}_p) = \sigma_{\mathtt{abs}k}^{\mathtt{gas}}(F_l)$. Note that $\{\overline{\lambda}_p\}_{p=1}^{N_{\overline{\lambda}}}$ contains N_λ groups of N_c identical wavelengths. By this construction, (34) becomes

$$I(\lambda) = \sum_{p=1}^{N_{\overline{\lambda}}} \overline{\omega}_p\, g(\lambda - \overline{\lambda}_p)\, I(\sigma_{\mathtt{abs}}^{\mathtt{gas}}(\overline{\lambda}_p)), \tag{35}$$

and it is apparent that (35) is a quadrature rule for (33) in the case of the correlated k-distribution method.

We come now to the PCA technique. A general approximate model for computing the radiance I at the wavelength λ is of the form

$$\ln \frac{I(\lambda)}{I_{\mathtt{a}}(\lambda)} = f_{\mathtt{I}}(\lambda), \tag{36}$$

where $I_{\mathtt{a}}$ is the radiance computed by an approximate radiative transfer model, and $f_{\mathtt{I}}$ is a correction factor. An efficient and accurate method for computing the radiance correction factor $f_{\mathtt{I}}(\lambda)$ has been given by Natraj et al. [16, 17]. This approach, which increases the computational efficiency of the radiative transfer calculations in an absorbing and scattering atmosphere, has the following attributes:

1. The exact model is the multi-stream DOME model, while the approximate model is a two-stream version of DOME, in which the eigenvalues and the eigenvectors of the layer matrix are computed analytically, and the system of equations for the entire atmosphere is solved by means of a pentadiagonal linear algebra solver.
2. The principal component analysis (PCA) is used to reduce the dimensionality of the optical parameters of the atmospheric system. In particular, PCA is performed on the logarithms of the layer values of the gas absorption coefficient $\ln \sigma_{\mathtt{abs}}^{\mathtt{gas}}(\lambda)$ and the molecular scattering coefficient $\ln \sigma_{\mathtt{sct}}^{\mathtt{mol}}(\lambda)$; for each λ_k, $k = 1, \ldots N_\lambda$, we define the $(2N-2)$-dimensional vector

$$\mathbf{x}(\lambda_k) \overset{\text{not}}{=} \mathbf{x}_k = [\ln \sigma_{\mathtt{abs}\,j}^{\mathtt{gas}}(\lambda_k); \ln \sigma_{\mathtt{sct}\,j}^{\mathtt{mol}}(\lambda_k)]^{\mathsf{T}},$$

   $j = 1, \ldots, N-1$, so that the wavelength variability of the optical parameters is encapsulated in $\mathbf{x}_k$.
3. 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.

Parenthetically, we note that in [18] the PCA-based radiative transfer model of Natraj et al. [16, 17] has been generalized to include several dimensionality reduction techniques as, for example, linear embedding methods and discrete orthogonal transforms. The PCA formalism remains valid if instead of the set $\{\lambda_k\}_{k=1}^{N_\lambda}$ we consider the set $\{\overline{\lambda}_p\}_{p=1}^{N_{\overline{\lambda}}}$ of the correlated k-distribution method (the wavelength variability is encapsulated in $\mathbf{x}(\overline{\lambda}_p)$). This claim is checked in the next section.

4. Numerical simulations

In this section we analyze the accuracy and efficiency of the exact and approximate radiative transfer models in the oxygen (O_2) A-band absorption channel at 764 nm (channel 9 of the EPIC instrument). For this purpose, simulations are performed for a water-cloud model with a Gamma size distribution

$$P(a) \propto a^\alpha \exp\left[-\alpha\left(\frac{a}{a_{\mathtt{mod}}}\right)\right] \tag{37}$$

with parameters $a_{\mathtt{mod}} = 8$ μm and $\alpha = 6$. The droplet size ranges between 0.02 and 50.0 μm. The cloud top height is $h_{\mathtt{t}} = 4$ km, and the cloud geometrical thickness $\triangle h$ is chosen as

$$\triangle h = \begin{cases} 1.0\,\text{km}, & \tau_c \le 4 \\ 1.5\,\text{km}, & 4 < \tau_c \le 8 \\ 2.0\,\text{km}, & 8 < \tau_c \le 14 \\ 2.5\,\text{km}, & 14 < \tau_c \le 20 \\ 3.0\,\text{km}, & \tau_c > 20 \end{cases},$$

where τ_c is the cloud optical thickness. The atmosphere is discretized with a step of 0.5 km between 0 and 16 km, a step of 2 km between 16 km and 20 km, a step of 5 km between 20 km and 30 km, and, finally, a step of 10 km between 30 km and 50 km. The ground surface is Lambertian with albedo $A = 0.06$, and, as in [3], the solar and viewing zenith angles are taken to be equal, i.e. $\theta = \theta_0$, while the relative azimuthal angle between the solar and viewing directions is chosen as $\triangle\varphi = 176°$. The azimuthal convergence test is the standard DISORT double convergence test [36] with a tolerance of 10^{-6}. The O_2 absorption cross sections are computed by using LBL calculations [37] with optimized rational approximations for the Voigt line profile [38]. The wavenumber grid point spacing is chosen as a fraction (e.g. 1/4) of the minimum half-width of the Voigt lines taken from HITRAN database [39]. 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], and this atmosphere is considered free of aerosols. The radiances are solar-flux normalized, and are computed by means of the delta-M approximation in conjunction with the TMS correction. If not stated otherwise the simulations are based on LBL calculations. The instrument spectral response functions (ISRF) for the different EPIC channels are available from NASA public servers.[1]

4.1. Exact models

In Fig. 4 we show the DOME radiances as functions of the cloud optical thickness τ_c and for different values of the viewing zenith angle θ. The simulations are performed by using a large number of discrete ordinates, namely $M = 128$. The plots show that, for a fixed value of θ, the radiance is an increasing function of τ_c. The reason is that the multiple-scattering contribution, which is an increasing function of τ_c, dominates the

[1] https://avdc.gsfc.nasa.gov/pub/DSCOVR/EPIC_Filter_Data/EPIC_Filters_Original_Data.xlsx.

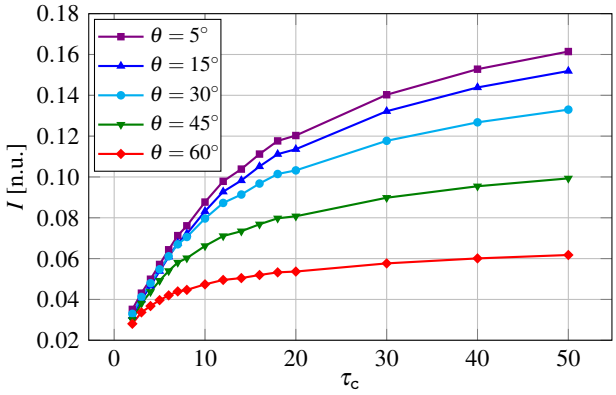


Figure 4: Reference DOME radiances as functions of τ_c for $M = 128$. In the y-axis, [n.u.] stands for normalized units.

single-scattering contribution, which is a decreasing function of τ_c. Moreover, for a fixed value of τ_c, the radiances decrease as the viewing zenith angle θ increases (if θ increases, the path through the atmosphere increases, the amount of oxygen absorption increases, and so, the single-scattering contribution decreases). These results are taken as the "exact" reference values for the rest of the simulations that follow.

In Fig. 5 we illustrate the relative errors in DOME radiances for different values of the number of discrete ordinates M. The plots show that, for a fixed value of θ, the absolute values of the relative errors decrease with τ_c. The reason is that, now, the relative error in the single-scattering contribution dominates the relative error in the multiple-scattering contribution. The results in Fig. 5 can be summarized as follows:

1. for $M = 8$ and $M = 16$, the relative errors are smaller than 0.015 for $\tau_c \geq 10$, and lie roughly between 0.015 and 0.03 for $\tau_c < 10$;
2. for $M = 24$ and $M = 32$, the relative errors are smaller than 0.008 for $\tau_c \geq 10$, and lie between 0.008 and 0.015 for $\tau_c < 10$;
3. for $M = 48$, the relative errors are smaller than 0.01 over the entire range of τ_c;
4. for $M = 64$, the relative errors are smaller than 0.003 over the entire range of τ_c.

The computation times for these simulations are given in Table 2. Note that the simulations were performed on a server Intel(R) Xeon(R) CPU E5-2695 v3 @ 2.30 GHz using up to 56 threads, and that the computation times in Table 2 correspond to the accumulated time over all threads. Thus, even though the delta-M scaling is used, an accurate description of the backscattering region requires $M \geq 48$, although a compromise between speed and accuracy can be reached by the choice $24 \leq M \leq 32$. Note also that for $M \geq 24$ more than 30 azimuthal modes are required to reach convergence with an accuracy of 10^{-6}.

Fig. 6 illustrates the relative errors in MOME radiances using DOME as a reference. Taking into account that the relative errors are smaller than $5 \cdot 10^{-4}$, and that the MOME computation user time is 1660 minutes for $M = 32$, we infer that DOME and MOME yield similar results from the point of view of accuracy and efficiency.

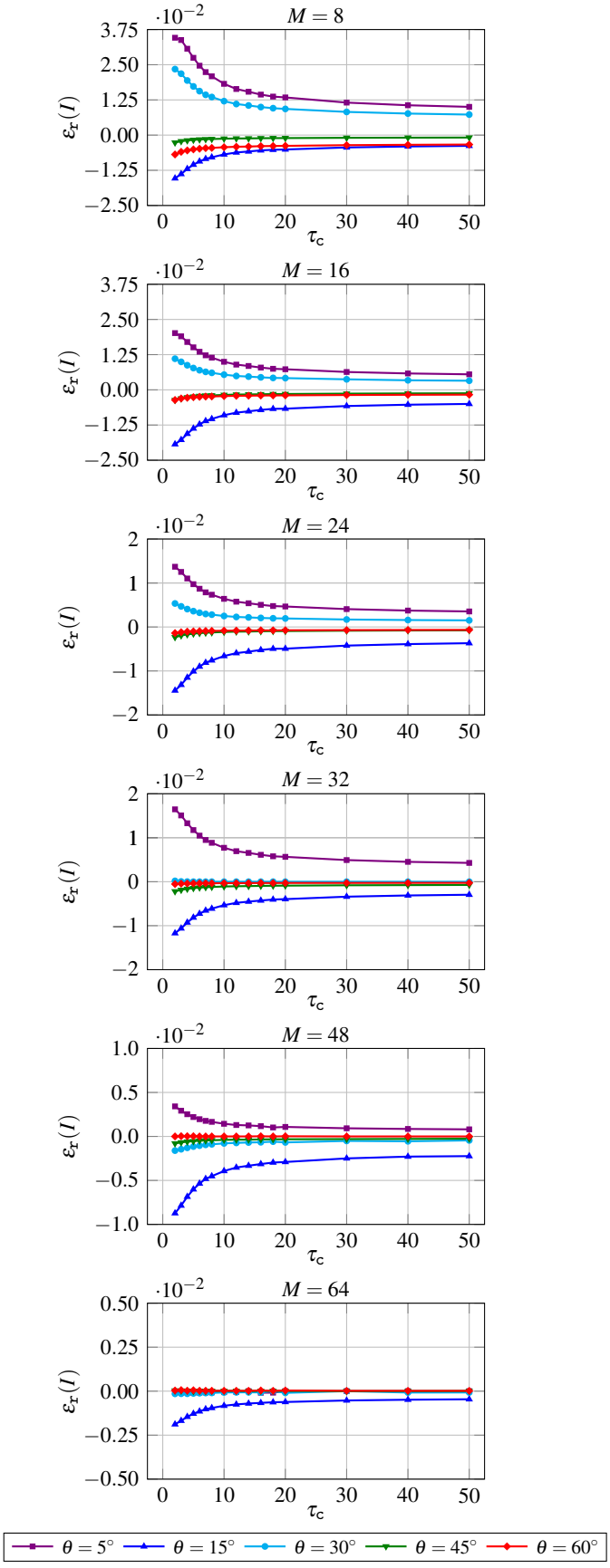


Figure 5: Relative errors in DOME radiances as functions of τ_c and for different values of M.

Table 2: Computation user times in minutes for the results in Fig.5, and speed-up factors with respect to the case $M = 128$.

M	t / min	Speed-up factor
8	67	2389.8
16	284	563.8
24	760	210.7
32	1711	93.6
48	5698	28.1
64	14548	11.0
128	160116	–

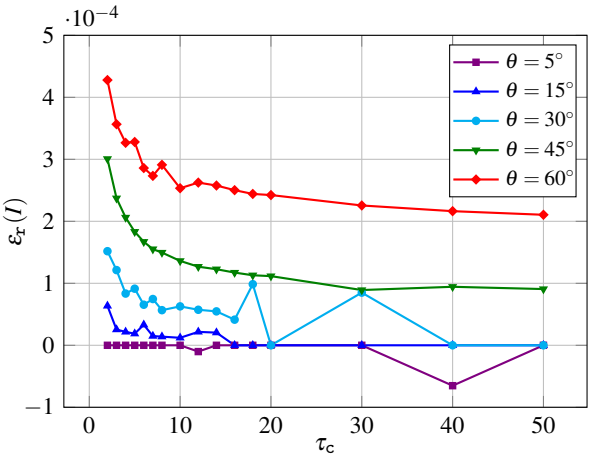


Figure 6: Relative errors in MOME radiances with respect to DOME radiances for $M = 32$.

4.2. *Approximate models*

The relative errors of the asymptotic (fourth- and zeroth-order) and ELC models are shown in Fig.7. A general finding is that an increase of the number of discrete ordinates does not significantly improve the accuracy of the results; a real improvement is apparent only for the fourth-order asymptotic model. The relative errors of the fourth-order asymptotic model are smaller than 0.02 for $\tau_c \geq 10$, and smaller than 0.04 for $\tau_c < 10$. For the zeroth-order asymptotic model, the relative errors have a similar behaviour when $\tau_c \geq 10$, but they can reach 0.24 when $\tau_c < 10$.

The relative errors of the ELC model are excessively large; they can reach 0.4 for small τ_c, and are 0.1 for large τ_c. The main reason for this is that the plane albedo $r(\mu)$ varies significantly with μ, and so, the radiative flux is not conserved (Fig.8). Note that, even in the limit of large τ_c, the flux conservation does not imply an accurate description of the radiance field (the radiative flux description involves only the azimuthal mode $m = 0$). Also note that the spherical albedo A_{ct} changes with τ_c, a fact which contradicts the common assumption of a constant cloud albedo [9]. The following result is a consequence of the low accuracy of the ELC model. Let $I(A_{ct0}, h_{t0})$ be the radiance computed by an exact model corresponding to a cloud with top height h_{t0} and a spherical albedo $A_{ct0} = A_{ct0}(\tau_{c0}, h_{t0})$, where τ_{c0} is the optical thickness. Further, let $h_t^\star$ solve the minimization problem

$$h_t^\star = \arg\min_{h_t} [I_{ELC}(A_{ct0}, h_t) - I(A_{ct0}, h_{t0})]^2,$$

where $I_{ELC}(A_{ct0}, h_t)$ is the radiance computed by the ELC model. The minimizer $h_t^\star$ that yields a small residual, and so, an acceptable accuracy, is different from h_{t0}. Thus, the physical significance of the cloud top height is lost.

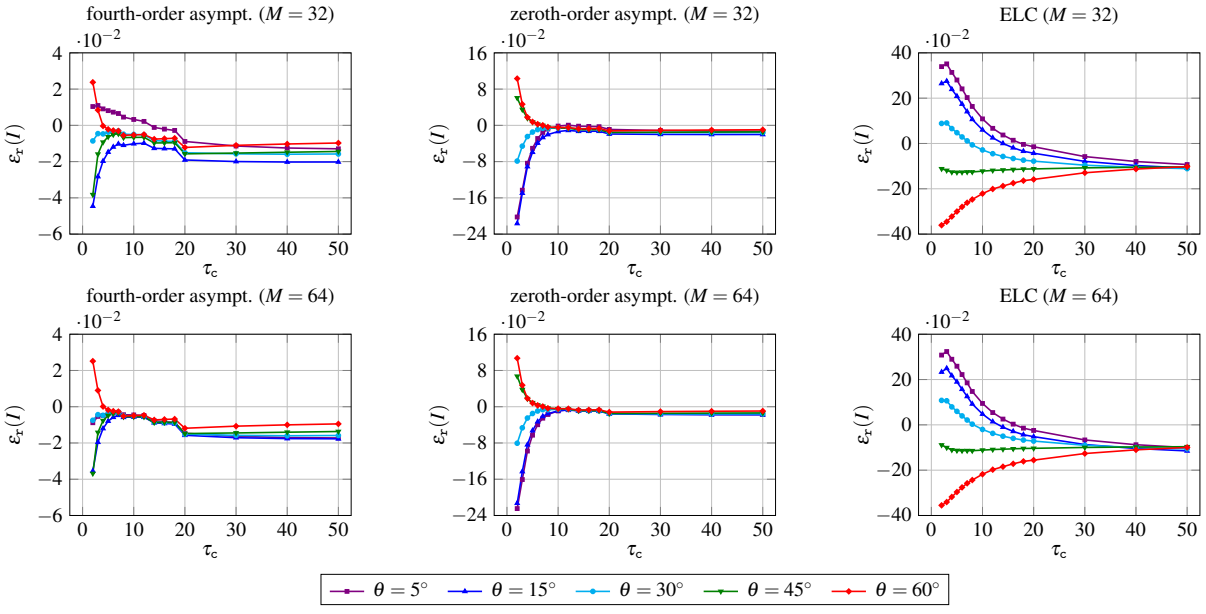


Figure 7: Relative errors in radiances as functions of τ_c when using the approximate models. The results correspond to $M = 32$ (top panels) and $M = 64$ (bottom panels), and are computed by using the fourth-order asymptotic model, the zeroth-order asymptotic model and the ELC model.

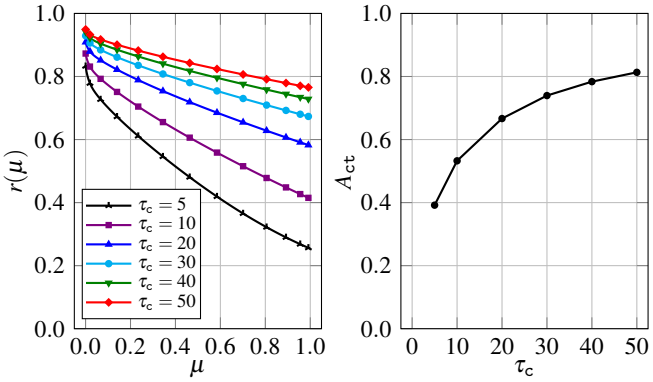


Figure 8: The plane albedo $r(\mu)$ for different values of τ_c (left panel), and the spherical albedo A_{ct} as a function of τ_c (right panel).

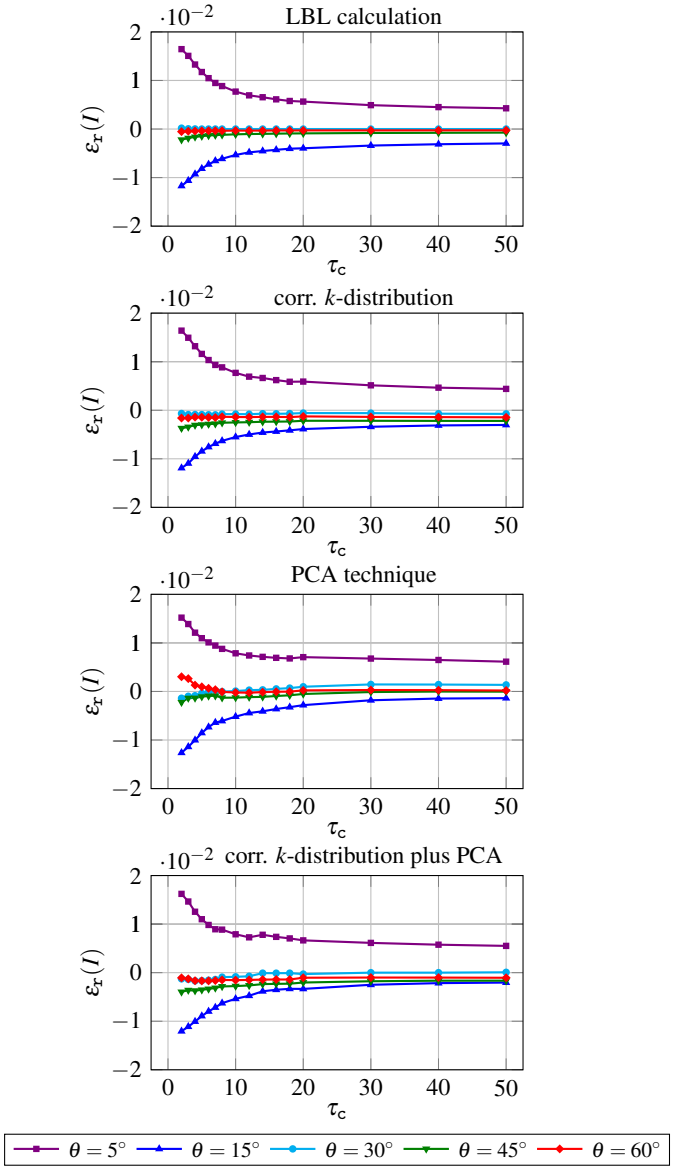


Figure 9: Relative errors in DOME radiances as functions of $\tau_c$when the acceleration techniques are used. The results correspond to $M = 32$, and are computed by using LBL calculations, correlated k-distribution, PCA technique and correlated k-distribution plus PCA. Note that the first plot is the same one as the fourth plot of Fig. 5.

Table 3: Computation user times in minutes for the results in Fig. 9, and speed-up factors with respect to the case of LBL calculation.

Acceleration technique	**t / min**	**Speed-up factor**
LBL calculation	1711	–
Correlated k-distribution	159	10.8
PCA technique	22	77.8
Corr. k-distribution plus PCA technique	5	342.2

Table 4: Computation times in seconds for the approximate models using the combined correlated k-distribution plus PCA technique. The simulations correspond to the same scenario as in Fig. 9.

Forward model	**t / s**	**Speed-up factor**
DOME	285.0	–
Fourth-order asymptotic	221.8	1.3
Zeroth-order asymptotic	222.2	1.3
ELC	188.7	1.5

4.3. Acceleration techniques

In Fig. 9 we illustrate the relative errors in DOME radiances when LBL calculations are replaced with the correlated k-distribution method, the PCA technique, and the combined correlated k-distribution plus PCA method. The corresponding computation times are given in Table 3. Note that almost identical relative errors and computation times are obtained for MOME (not shown here). The general observation is that the relative error curves are practically unchanged when these acceleration techniques are used. The computation time using the combined correlated k-distribution plus PCA method is 5 times smaller than that for the PCA technique alone, and 40 times smaller than that for the correlated k-distribution method alone. Taking into account that the results in Table 3 correspond to 16 values of τ_c, the average time to simultaneously compute the radiances at 5 viewing zenith angles is 17.8 seconds.

In Table 4 we show the computation times of the approximate models using the combined correlated k-distribution plus PCA method. As in Fig. 9, the radiances are computed for 16 values of τ_c and at 5 viewing zenith angles. Note that the relative error curves are very similar to those plotted in the top panels of Fig. 7 (for $M = 32$). The conclusion is that the computation times of the approximate models are lower, but not significantly lower than those of DOME.

5. Conclusions

We analyzed several radiative transfer models which can be used for the retrieval of cloud parameters from EPIC measurements. The radiative transfer models are the exact DOME and MOME models, and the approximate asymptotic and equivalent Lambertian cloud models. MOME is a combination of the conventional matrix operator method and DOME. Essentially, the matrix exponential formalism enabled us to combine the two methods in a simple way and to introduce the higher-order

asymptotic and equivalent Lambertian cloud models in a natural manner. The radiative transfer models used several acceleration techniques such as the Padé approximation, the telescoping technique, the method of false discrete ordinate, the correlated k-distribution method and the PCA technique. We computed the radiances also by combining the correlated k-distribution method with the PCA technique.

We analyzed the accuracy and efficiency of the radiative transfer models when simulating radiances for the channel 9 of the EPIC instrument. In summary, the following conclusions can be drawn:

1. The exact DOME and MOME models using correlated k-distribution in conjunction with PCA yield an accuracy better than 0.015 and a computation time of approximately 18 seconds for radiance calculations at 5 viewing zenith angles.
2. The approximate models are slightly more efficient than the exact models but their accuracy leaves much to be desired. In particular, the relative errors of the zeroth-order asymptotic model are large for cloud optical thicknesses smaller than 10, while the relative errors of the ELC model are large for all cloud optical thicknesses.

Thus, it appears that the exact DOME and MOME models could fulfill the accuracy and efficiency requirements of an offline processor for retrieval of cloud parameters from EPIC measurements. For online (i.e. operational) retrieval algorithms, the inversion is frequently performed by using look-up table (LUT) approaches [4, 5, 42, 43]. The size of such a LUT is extremely large, and so the time for computing a LUT by using exact radiative transfer models is normally too high. Even though the inversion is performed by neural network techniques [7, 8, 44], the time for neural network training by using exact radiative transfer models is high. In this regard, the above radiative transfer models are an efficient tool for LUTs computation and neural network training.

6. Acknowledgements

This work was funded by the programme DLR/DAAD Research Fellowships 2015 (57186656), with reference numbers 91613528 and 91627488, organized by the German Academic Exchange Service (DAAD) and the German Aerospace Center (DLR).

Appendix A

In this appendix we give the expressions for the layer quantities $\mathtt{A}^1_{mj}$, $\mathtt{A}^2_{mj}$ and $\mathtt{b}_{mj}$ in (13) computed by the eigendecomposition method and the Padé approximation.

The eigendecomposition method for computing the matrix exponential is based on a spectral decomposition of the matrix $\mathbf{A}_{mj}$ given by (8). This can be obtained by one of the following methods: direct decomposition of an asymmetric matrix [45], square-root decomposition [46], and Cholesky decomposition [47]. Here, we use the direct decomposition method. Exploiting the block symmetry of $\mathbf{A}_{mj}$, we find

$$\mathbf{A}_{mj} = \mathbf{V}_{mj} \begin{bmatrix} \mathbf{\Lambda}_{mj} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda}_{mj} \end{bmatrix} \mathbf{V}^{-1}_{mj}, \tag{38}$$

with

$$\mathbf{V}_{mj} = \begin{bmatrix} \mathbf{V}^{+}_{mj} & \mathbf{V}^{-}_{mj} \\ \mathbf{V}^{-}_{mj} & \mathbf{V}^{+}_{mj} \end{bmatrix}, \tag{39}$$

$$\mathbf{V}^{\pm}_{mj} = [\mathbf{v}^{\pm}_1, \ldots, \mathbf{v}^{\pm}_M], \tag{40}$$

and

$$\Lambda_{mj} = \mathrm{diag}[\lambda_1, \ldots, \lambda_M] \overset{\text{not}}{=} \mathrm{diag}[\lambda_k]. \tag{41}$$

Note that $\begin{bmatrix} \mathbf{v}^{+}_k \\ \mathbf{v}^{-}_k \end{bmatrix}$ are the right eigenvectors of $\mathbf{A}_{mj}$ corresponding to the eigenvalue λ_k, and $\begin{bmatrix} \mathbf{v}^{-}_k \\ \mathbf{v}^{+}_k \end{bmatrix}$ are the right eigenvectors of $\mathbf{A}_{mj}$ corresponding to the eigenvalue $-\lambda_k$. As a result, the layer matrices can be expressed in compact form as

$$\mathtt{A}^1_{mj} = \mathbf{D}^1_{mj} \mathbf{V}^{-1}_{mj}, \tag{42}$$

$$\mathtt{A}^2_{mj} = -\mathbf{D}^2_{mj} \mathbf{V}^{-1}_{mj}, \tag{43}$$

where the diagonal scaling matrices $\mathbf{D}^1_{mj}$ and $\mathbf{D}^2_{mj}$ are given by

$$\mathbf{D}^1_{mj} = \mathrm{diag}[a_0(\lambda_k \triangle r_j); 1], \tag{44}$$

$$\mathbf{D}^2_{mj} = \mathrm{diag}[1; a_0(\lambda_k \triangle r_j)], \tag{45}$$

with $a_0(x) = \mathrm{e}^{-x}$. The layer vector is computed as

$$\mathtt{b}_{mj} = \mathbf{B}_{mj} \mathbf{V}^{-1}_{mj} (\triangle r_j \, \mathbf{b}_{mj}), \tag{46}$$

where, in the secant approximation of the solar beam, the diagonal scaling matrix $\mathbf{B}_{mj}$ is given by

$$\mathbf{B}_{mj} = \mathrm{diag}[b_1(\lambda_k \triangle r_j); b_2(\lambda_k \triangle r_j)], \tag{47}$$

with

$$b_1(x) = \frac{\mathrm{e}^{-(\tau^0_{\mathtt{ext}j} + x)} - \mathrm{e}^{-\tau^0_{\mathtt{ext}j+1}}}{\tau^0_{\mathtt{ext}j+1} - \tau^0_{\mathtt{ext}j} - x}, \tag{48}$$

$$b_2(x) = \frac{\mathrm{e}^{-(\tau^0_{\mathtt{ext}j+1} + x)} - \mathrm{e}^{-\tau^0_{\mathtt{ext}j}}}{\tau^0_{\mathtt{ext}j} - \tau^0_{\mathtt{ext}j+1} - x}. \tag{49}$$

In (48) and (49), $\tau^0_{\mathtt{ext}j}$ and $\tau^0_{\mathtt{ext}j+1}$ are the solar optical depths at the boundary levels j and $j+1$, respectively.

In the first-order Padé approximation to the matrix exponential, the layer equation reads as

$$\mathtt{A}^1_{mj} = \mathbf{I} - \frac{\triangle r_j}{2} \mathbf{A}_{mj}, \tag{50}$$

$$\mathtt{A}^2_{mj} = -\Big(\mathbf{I} + \frac{\triangle r_j}{2} \mathbf{A}_{mj}\Big), \tag{51}$$

$$\mathtt{b}_{mj} = \mathbf{B}_{mj} (\triangle r_j \, \mathbf{b}_{mj}) \, \mathrm{e}^{-\tau^0_{\mathtt{ext}j}}, \tag{52}$$

where $\mathbf{B}_{mj}$ is now given by

$$\mathbf{B}_{mj} = I_0(\tau_0)\,\mathbf{I} + \Big[\frac{1}{2}\,I_0(\tau_0) - I_1(\tau_0)\Big]\,\triangle r_j\,\mathbf{A}_{mj}, \tag{53}$$

with $\tau_0 = \tau^0_{\mathtt{ext}\,j+1} - \tau^0_{\mathtt{ext}\,j}$ and

$$I_0(x) = \frac{1-\mathrm{e}^{-x}}{x}, \tag{54}$$

$$I_1(x) = \frac{1}{x}\Big[1 - \frac{\mathrm{e}^{-x}}{x}\,(1-\mathrm{e}^{-x})\Big]. \tag{55}$$

Appendix B

In this appendix we give the expressions of $\mathbf{R}_{mj}$, $\mathbf{T}_{mj}$ and $\mathbf{\Sigma}_{mj}$ in (19) computed by the eigendecomposition method and the Padé approximation.

The reflection and transmission matrices of layer j can be derived by representing the layer equation (13) as

$$\begin{bmatrix} \mathbf{i}^+_{mj} \\ \mathbf{i}^-_{mj+1} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{mj} & \mathbf{T}_{mj} \\ \mathbf{T}_{mj} & \mathbf{R}_{mj} \end{bmatrix} \begin{bmatrix} \mathbf{i}^-_{mj} \\ \mathbf{i}^+_{mj+1} \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}^+_{mj} \\ \mathbf{\Sigma}^-_{mj} \end{bmatrix}, \tag{56}$$

from which we find that

$$\begin{bmatrix} \mathbf{R}_{mj} & \mathbf{T}_{mj} \\ \mathbf{T}_{mj} & \mathbf{R}_{mj} \end{bmatrix} = -\begin{bmatrix} \Gamma_{mj}\,\mathbf{V}^{11}_{mj} & -\mathbf{V}^{12}_{mj} \\ \mathbf{V}^{21}_{mj} & -\Gamma_{mj}\,\mathbf{V}^{22}_{mj} \end{bmatrix}^{-1} \times \begin{bmatrix} \Gamma_{mj}\,\mathbf{V}^{12}_{mj} & -\mathbf{V}^{11}_{mj} \\ \mathbf{V}^{22}_{mj} & -\Gamma_{mj}\,\mathbf{V}^{21}_{mj} \end{bmatrix}, \tag{57}$$

$$\begin{bmatrix} \mathbf{\Sigma}^+_{mj} \\ \mathbf{\Sigma}^-_{mj} \end{bmatrix} = \begin{bmatrix} \Gamma_{mj}\,\mathbf{V}^{11}_{mj} & -\mathbf{V}^{12}_{mj} \\ \mathbf{V}^{21}_{mj} & -\Gamma_{mj}\,\mathbf{V}^{22}_{mj} \end{bmatrix}^{-1} \mathtt{b}_{mj}, \tag{58}$$

with

$$\mathbf{V}^{-1}_{mj} = \begin{bmatrix} \mathbf{V}^{11}_{mj} & \mathbf{V}^{12}_{mj} \\ \mathbf{V}^{22}_{mj} & \mathbf{V}^{22}_{mj} \end{bmatrix}, \tag{59}$$

and

$$\Gamma_{mj} = \mathrm{diag}[\exp(-\lambda_k\,\triangle r_j)]. \tag{60}$$

By inspection of (57) and (58) it is apparent that the computation of $\mathbf{R}_{mj}$, $\mathbf{T}_{mj}$ and $\mathbf{\Sigma}_{mj}$ requires an inversion and a multiplication of matrices of dimension $2M \times 2M$. Other methods in which the computation of $\mathbf{R}_{mj}$ and $\mathbf{T}_{mj}$ can be halved in order are discussed in [32].

In the first-order Padé approximation to the matrix exponential, we have

$$\begin{bmatrix} \mathbf{R}_{mj} & \mathbf{T}_{mj} \\ \mathbf{T}_{mj} & \mathbf{R}_{mj} \end{bmatrix} = -\begin{bmatrix} \mathbf{I} - \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj} & -\frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} \\ \frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} & -(\mathbf{I} - \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj}) \end{bmatrix}^{-1} \times \begin{bmatrix} -\frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} & -(\mathbf{I} + \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj}) \\ \mathbf{I} + \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj} & \frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} \end{bmatrix}, \tag{61}$$

$$\begin{bmatrix} \mathbf{\Sigma}^+_{mj} \\ \mathbf{\Sigma}^-_{mj} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj} & -\frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} \\ \frac{1}{2}\triangle r_j\mathbf{A}^{12}_{mj} & -(\mathbf{I} - \frac{1}{2}\triangle r_j\mathbf{A}^{11}_{mj}) \end{bmatrix}^{-1} \mathtt{b}_{mj}. \tag{62}$$

Appendix C

In this appendix we give a justification of the ELC model.

For the azimuthal mode m, the reflection law at the cloud top height is (we omit the $r_{j_{\mathtt{min}}}$ dependency)

$$I_m(\mu) = 2\int_0^1 R_m(\mu,\mu')\,I_{\mathtt{t}m}(-\mu')\,\mu'\,\mathrm{d}\mu', \quad m \geq 0, \tag{63}$$

where $I_m(\mu)$ is the upward diffuse radiance in the direction μ, $R_m(\mu,\mu')$ is the reflection function of the atmosphere below the cloud, and $I_{\mathtt{t}m}(-\mu')$ is the downward total radiance. The upward flux is

$$E^{\uparrow} = \int_0^{2\pi}\!\!\int_0^1 I(\mu,\varphi)\,\mu\,\mathrm{d}\mu\,\mathrm{d}\varphi = 2\pi\int_0^1 r(\mu')\,I_{\mathtt{t}0}(-\mu')\,\mu'\,\mathrm{d}\mu', \tag{64}$$

where

$$r(\mu) = 2\int_0^1 R_0(\mu,\mu')\,\mu'\,\mathrm{d}\mu' = 2\int_0^1 R_0(\mu',\mu)\,\mu'\,\mathrm{d}\mu' \tag{65}$$

is the plane albedo of the atmosphere below the cloud. Let us assume that the plane albedo $r(\mu)$ is almost constant with respect to μ, so that we can write $r(\mu) = r_0$. Then, the spherical albedo of the atmosphere below the cloud is

$$A_{\mathtt{ct}} = 2\int_0^1 r(\mu)\,\mu\,\mathrm{d}\mu = r_0, \tag{66}$$

and the upward flux is

$$E^{\uparrow} = 2\pi\,A_{\mathtt{ct}}\int_0^1 I_{\mathtt{t}0}(-\mu')\,\mu'\,\mathrm{d}\mu'. \tag{67}$$

The upward flux given by (67) is that of a Lambertian surface with a ground albedo equal to the spherical albedo $A_{\mathtt{ct}}$. By replacing the atmosphere below the cloud by a Lambertian surface with the reflection function $R_m(\mu,\mu') = A_{\mathtt{ct}}\,\delta_{m0}$, we try to ensure the conservation of radiative flux. However, this is possible only when the plane albedo $r(\mu)$ does not vary significantly with μ.

Appendix D

In this appendix we describe the telescoping technique for DOME and MOME.

Let us make the assumption that for all layers j, with $j < j_{\mathtt{min}}$ and $j > j_{\mathtt{max}}$, and all azimuthal modes $m > 2$, the azimuthal expansion coefficients of the phase functions vanish. In this case, $\mathbf{A}_{mj}$ is a diagonal matrix,

$$\mathbf{A}_{mj} = \mathrm{diag}\Big[-\frac{\sigma_{\mathtt{ext}\,j}}{\mu_k}; \frac{\sigma_{\mathtt{ext}\,j}}{\mu_k}\Big], \tag{68}$$

and the solar layer vector $\mathbf{b}_{mj}$ vanishes. The layer equation simplifies to

$$\mathbf{i}_{mj+1} = \mathrm{e}^{-\mathbf{A}_{mj}\triangle r_j}\,\mathbf{i}_{mj}, \tag{69}$$

which further implies that

$$\mathbf{i}^{+}_{mj} = \mathrm{diag}\Big[\exp\Big(-\frac{\sigma_{\mathtt{ext}j}\,\triangle r_j}{\mu_k}\Big)\Big]\,\mathbf{i}^{+}_{mj+1}, \tag{70}$$

$$\mathbf{i}^{-}_{mj+1} = \mathrm{diag}\Big[\exp\Big(-\frac{\sigma_{\mathtt{ext}j}\,\triangle r_j}{\mu_k}\Big)\Big]\,\mathbf{i}^{-}_{mj}. \tag{71}$$

In the matrix operator method, we have

$$\mathbf{R}_{mj} = 0, \quad \mathbf{T}_{mj} = \mathrm{diag}\Big[\exp\Big(-\frac{\sigma_{\mathtt{ext}j}\,\triangle r_j}{\mu_k}\Big)\Big], \quad \mathbf{\Sigma}^{\pm}_{mj} = 0, \tag{72}$$

and the adding formula becomes

$$\mathtt{R}_{mj} = \mathbf{T}_{mj}\,\mathtt{R}_{mj+1}\,\mathbf{T}_{mj}, \tag{73}$$

$$\mathtt{r}_{mj} = \mathbf{T}_{mj}\,\mathtt{r}_{mj+1}. \tag{74}$$

Let

$$\mathtt{T} = \mathrm{diag}\Big[\exp\Big(-\frac{1}{\mu_k}\sum_{j=j_{\mathtt{max}}+1}^{N-1}\sigma_{\mathtt{ext}j}\,\triangle r_j\Big)\Big]$$

be the transmission matrix of the layers below the cloud. The DOME telescoping technique can be summarized as follows. For $m > 2$,

1. solve the radiative transfer equation for the cloud layers with the top and bottom boundary conditions

$$\mathbf{i}^{-}_{m,j_{\mathtt{min}}} = 0, \tag{75}$$

and

$$\mathbf{i}^{+}_{mj_{\mathtt{max}}+1} = \mathtt{R}_{mj_{\mathtt{max}}+1}\,\mathbf{i}^{-}_{mj_{\mathtt{max}}+1} + \mathtt{r}_{mj_{\mathtt{max}}+1}, \tag{76}$$

$$\mathtt{R}_{mj_{\mathtt{max}}+1} = \mathtt{T}\,\mathtt{R}_{mN}\,\mathtt{T}, \tag{77}$$

$$\mathtt{r}_{mj_{\mathtt{max}}+1} = \mathtt{T}\,\mathtt{r}_{mN}, \tag{78}$$

respectively;
2. set $\mathbf{i}^{-}_{mj} = 0$ for $j_{\mathtt{min}} - 1 \le j \le 1$, and compute the radiances above the cloud and in the upward directions $\mathbf{i}^{+}_{mj}$ for $j_{\mathtt{min}} - 1 \le j \le 1$, by using recurrence (70);
3. compute the radiances below the cloud and in the downward directions $\mathbf{i}^{-}_{mj+1}$ for $j_{\mathtt{max}} + 1 \le j \le N - 1$, by using recurrence (71);
4. compute the upward radiance at the surface $\mathbf{i}^{+}_{mN}$ from the boundary condition (15);
5. compute the radiances below the cloud and in the upward directions $\mathbf{i}^{+}_{mj}$ for $j_{\mathtt{max}} + 2 \le j \le N - 1$, by using recurrence (70).

The MOME telescoping technique involves the following steps. For $m > 2$,

1. compute the reflection matrix $\mathtt{R}_{mj_{\mathtt{max}}+1}$ and the reflection vector $\mathtt{r}_{mj_{\mathtt{max}}+1}$ at the cloud bottom by means of (77) and (78), respectively;
2. compute $\mathtt{R}_{mj_{\mathtt{min}}}$ and $\mathtt{r}_{mj_{\mathtt{min}}}$ by using the adding algorithm;
3. set $\mathbf{i}^{-}_{mj} = 0$ for $j_{\mathtt{min}} \le j \le 1$, and $\mathbf{i}^{+}_{mj_{\mathtt{min}}} = \mathtt{r}_{mj_{\mathtt{min}}}$(boundary condition at the cloud top);
4. compute the radiances above the cloud and in the upward directions $\mathbf{i}^{+}_{mj}$ for $j_{\mathtt{min}} - 1 \le j \le 1$, by using recurrence (70).

References

[1] A. Marshak, Y. Knyazikhin, The spectral invariant approximation within canopy radiative transfer to support the use of the EPIC/DSCOVR oxygen B-band for monitoring vegetation, Journal of Quantitative Spectroscopy and Radiative Transfer 191 (2017) 7–12. doi:10.1016/j.jqsrt.2017.01.015

[2] X. Xu, J. Wang, Y. Wang, J. Zeng, O. Torres, Y. Yang, A. Marshak, J. Reid, S. Miller, Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point, Geophysical Research Letters 44 (14) (2017) 7544–7554. doi:10.1002/2017gl073939

[3] Y. Yang, A. Marshak, J. Mao, A. Lyapustin, J. Herman, A method of retrieving cloud top height and cloud geometrical thickness with oxygen A and B bands for the Deep Space Climate Observatory (DSCOVR) mission: Radiative transfer simulations, Journal of Quantitative Spectroscopy and Radiative Transfer 122 (2013) 141–149. doi:10.1016/j.jqsrt.2012.09.017

[4] S. Platnick, M. King, S. Ackerman, W. Menzel, B. Baum, J. Rièdi, R. Frey, The MODIS cloud products: Algorithms and examples from Terra, IEEE Transactions on Geoscience and Remote Sensing 41 (2003) 459–473. doi:10.1109/tgrs.2002.808301

[5] M. King, S.-C. Tsay, S. Platnick, M. Wang, K.-N. Liou, Cloud retrieval algorithms for MODIS: Optical thickness, effective particle radius, and thermodynamic phase, Tech. rep., National Aeronautics and Space Administration, United States, available at https://eospso.nasa.gov/sites/default/files/atbd/atbd_mod05.pdf (1997).

[6] A. Kokhanovsky, V. Rozanov, P. Zege, H. Bovensmann, J. Burrows, A semianalytical cloud retrieval algorithm using backscattered radiation in 0.4-2.4 μm spectral region, Journal of Geophysical Research 108 (2003) AAC 4–1–AAC 4–19. doi:10.1029/2001jd001543

[7] D. Loyola, W. Thomas, Y. Livschitz, T. Ruppert, P. Albert, R. Hollmann, Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, IEEE Transactions on Geoscience and Remote Sensing 45 (9) (2007) 2747–2758. doi:10.1109/tgrs.2007.901043

[8] D. Loyola R., W. Thomas, R. Spurr, B. Mayer, Global patterns in daytime cloud properties as derived from GOME backscatter UV-VIS measurements, International Journal of Remote Sensing 31 (16) (2010) 4295–4318. doi:10.1080/01431160903246741

[9] R. B. A. Koelemeijer, P. Stammes, J. W. Hovenier, J. F. de Haan, A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, Journal of Geophysical Research 106 (D4) (2001) 3475. doi:10.1029/2000jd900657

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

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

[12] S. A. Tjemkes, J. Schmetz, Synthetic satellite radiances using the radiance sampling method, Journal of Geophysical Research: Atmospheres 102 (D2) (1997) 1807–1818. doi:10.1029/96jd02684

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

[14] X. Liu, W. L. Smith, D. K. Zhou, A. Larar, Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept, Applied Optics 45 (1) (2006) 201. doi:10.1364/ao.45.000201

[15] P. Eriksson, C. Jiménez, S. Bühler, D. Murtagh, A Hotelling transformation approach for rapid inversion of atmospheric spectra, Journal of Quantitative Spectroscopy and Radiative Transfer 73 (6) (2002) 529–543. doi:10.1016/s0022-4073(01)00175-3

[16] V. Natraj, X. Jiang, R. Shia, X. Huang, J. Margolis, Y. Yung, 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) (2005) 539–556. doi:10.1016/j.jqsrt.2004.12.024

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

[18] D. Efremenko, A. Doicu, D. Loyola, 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 (2014) 128–135. doi:10.1016/j.jqsrt.2013.07.023

[19] R. Spurr, LIDORT and VLIDORT. Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: A. Kokhanovsky (Ed.), Light scattering reviews, Vol. 3, 2008, pp. 229–275. doi:10.1007/978-3-540-48546-9_7

[20] A. Doicu, 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) (2009) 146–158. doi:10.1016/j.jqsrt.2008.09.014

[21] Q. Liu, F. Weng, Advanced doubling-adding method for radiative transfer in planetary atmospheres, Journal of the Atmospheric Sciences 63 (12) (2006) 3459–3465. doi:10.1175/jas3808.1

[22] K. Evans, G. Stephens, A new polarized atmospheric radiative transfer model, Journal of Quantitative Spectroscopy and Radiative Transfer 46 (5) (1991) 413–423. doi:10.1016/0022-4073(91)90043-p

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

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

[25] B. Mayer, M. Schröder, R. Preusker, L. Schüller, Remote sensing of water cloud droplet size distributions using the backscatter glory: A case study, Atmospheric Chemistry and Physics 4 (5) (2004) 1255–1263. doi:10.5194/acp-4-1255-2004

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

[27] R. Spurr, Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, Journal of Quantitative Spectroscopy and Radiative Transfer 75 (2) (2002) 129–175. doi:10.1016/s0022-4073(01)00245-x

[28] A. Doicu, T. Trautmann, Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Vector case, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (1-2) (2009) 159–172. doi:10.1016/j.jqsrt.2008.09.013

[29] R. Spurr, M. Christi, Linearization of the interaction principle: Analytic jacobians in the "radiant model", Journal of Quantitative Spectroscopy and Radiative Transfer 103 (3) (2007) 431–446. doi:10.1016/j.jqsrt.2006.05.001

[30] E. Yanovitskij, Light scattering in inhomogeneous atmospheres, Springer-Verlag, 1997.

[31] T. Nakajima, M. D. King, Asymptotic theory for optically thick layers: Application to the discrete ordinates method, Applied Optics 31 (36) (1992) 7669–7683. doi:10.1364/ao.31.007669

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

[33] M. D. King, Determination of the scaled optical thickness of clouds from reflected solar radiation measurements, Journal of the Atmospheric Sciences 44 (13) (1987) 1734–1751. doi:10.1175/1520-0469(1987)044<1734:dotsot>2.0.co;2

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

[35] 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 (1936) 7–18.

[36] K. Stamnes, S.-C. Tsay, I. Laszlo, DISORT, a general-purpose Fortran program for discrete-ordinate-method radiative transfer in scattering and emitting layered media: Documentation of methodology, Tech. rep., Dept. of Physics and Engineering Physics, Stevens Institute of Technology, Hoboken, NJ 07030, available at http://www.libradtran.org/lib/exe/fetch.php?media=disortreport1.1.pdf (2000).

[37] F. Schreier, S. Gimeno García, P. Hedelt, M. Hess, J. Mendrok, M. Vasquez, J. Xu, GARLIC — A general purpose atmospheric radiative transfer line-by-line infrared-microwave code: Implementation and evaluation, Journal of Quantitative Spectroscopy and Radiative Transfer 137 (2014) 29–50. doi:10.1016/j.jqsrt.2013.11.018

[38] F. Schreier, Optimized implementations of rational approximations for the Voigt and complex error function, Journal of Quantitative Spectroscopy and Radiative Transfer 112 (6) (2011) 1010–1025. doi:10.1016/j.jqsrt.2010.12.010

[39] L. Rothman, I. Gordon, A. Barbe, D. C. Benner, P. Bernath, M. Birk, V. Boudon, L. Brown, A. Campargue, J.-P. Champion, K. Chance, L. Coudert, V. Dana, V. Devi, S. Fally, J.-M. Flaud, R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. Lafferty, J.-Y. Mandin, S. Massie, S. Mikhailenko, C. Miller, N. Moazzen-Ahmadi, O. Naumenko, A. Nikitin, J. Orphal, V. Perevalov, A. Perrin, A. Predoi-Cross, C. Rinsland, M. Rotger, M. Šimečková, M. Smith, K. Sung, S. Tashkun, J. Tennyson, R. Toth, A. Vandaele, J. Vander Auwera, The HITRAN 2008 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (9-10) (2009) 533–572. doi:10.1016/j.jqsrt.2009.02.013

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

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

[42] T. Y. Nakajima, T. Nakajma, Wide-area determination of cloud microphysical properties from NOAA AVHRR measurements for FIRE and ASTEX regions, Journal of the Atmospheric Sciences 52 (23) (1995) 4043–4059. doi:10.1175/1520-0469(1995)052<4043:wadocm>2.0.co;2

[43] K. Kawamoto, T. Nakajima, T. Y. Nakajima, A global determination of cloud microphysics with AVHRR remote sensing, Journal of Climate 14 (9) (2001) 2054–2068. doi:10.1175/1520-0442(2001)014<2054:agdocm>2.0.co;2

[44] D. S. Efremenko, D. G. Loyola R., P. Hedelt, R. J. D. Spurr, Volcanic SO_2 plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm, International Journal of Remote Sensing 38 (sup1) (2017) 1–27. doi:10.1080/01431161.2017.1348644

[45] K. Stamnes, R. A. Swanson, A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres, Journal of the Atmospheric Sciences 38 (2) (1981) 387–399. doi:10.1175/1520-0469(1981)038<0387:anlatd>2.0.co;2

[46] T. Nakajima, 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) (1986) 13–21. doi:10.1016/0022-4073(86)90088-9

[47] K. Stamnes, S.-C. Tsay, W. Wiscombe, K. Jayaweera, Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, Applied Optics 27 (12) (1988) 2502. doi:10.1364/ao.27.002502

The paper was published in Journal of Quantitative Spectroscopy and Radiative Transfer **213**(2018)241–251
DOI: 10.1016/j.jqsrt.2018.03.008

Linearized radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements

Víctor Molina García*, Sruthy Sasi, Dmitry S. Efremenko, Adrian Doicu, Diego Loyola

Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung (IMF), 82234 Oberpfaffenhofen, Germany

Abstract

In this paper, we describe several linearized radiative transfer models which can be used for the retrieval of cloud parameters from EPIC (Earth Polychromatic Imaging Camera) measurements. The approaches under examination are (1) the linearized forward approach, represented in this paper by the linearized discrete ordinate and matrix operator methods with matrix exponential, and (2) the forward-adjoint approach based on the discrete ordinate method with matrix exponential. To enhance the performance of the radiative transfer computations, the correlated k-distribution method and the Principal Component Analysis (PCA) technique are used. We provide a compact description of the proposed methods, as well as a numerical analysis of their accuracy and efficiency when simulating EPIC measurements in the oxygen A-band channel at 764 nm. We found that the computation time of the forward-adjoint approach using the correlated k-distribution method in conjunction with PCA is approximately 13 s for simultaneously computing the derivatives with respect to cloud optical thickness and cloud top height.

Keywords: inversion, linearized models, adjoint radiative transfer, EPIC

1. Introduction

EPIC (Earth Polychromatic Imaging Camera) is a 10-channel spectroradiometer (317–780 nm) onboard the spacecraft DSCOVR (Deep Space Climate Observatory), and was designed to measure the atmosphere and surface properties over the whole sunlit half of the Earth from the Lagrange point L_1. DSCOVR provides observations of the Earth at near backward directions with scattering angles ranging from 168° to 176°. EPIC has two pairs of reference and absorption channels in the oxygen A-band (780 and 764 nm) and B-band (680 and 688 nm), which are used for monitoring the vegetation condition [1], the aerosol layer height and optical depth [2], as well as the cloud height and optical depth [3]. A further description of EPIC/DSCOVR geometry and their channels can be found in [4].

In [4], we analyzed exact and approximate radiative transfer models regarding their applicability to the retrieval of cloud parameters from EPIC measurements. It has been shown that the exact Discrete Ordinate method with Matrix Exponential (DOME) and the Matrix Operator method with Matrix Exponential (MOME) using the correlated k-distribution method [5] in conjunction with the Principal Component Analysis (PCA) technique [6–9] fulfill the accuracy and efficiency requirements for such kind of application.

However, the retrieval of atmospheric constituents from satellite measurements also requires the knowledge of weighting functions, i.e. the partial derivatives of the measured radiance with respect to the atmospheric parameters being retrieved. The process of obtaining this set of partial derivatives, which constitute the matrix of weighting functions or Jacobian, is commonly referred to as linearization analysis. There are two common linearization approaches: the linearized forward approach and the forward-adjoint approach.

1. In the linearized forward approach, the partial derivatives are computed analytically. Such linearized radiative transfer models based on the conventional discrete ordinate method and the matrix operator method have been developed by Spurr [10–13].
2. In the forward-adjoint approach, the measured radiance is expressed as the scalar product of the solution of the adjoint problem and the source function of the forward problem. Using the linearization technique to the forward and adjoint problems (i.e. using the differentiation operator by means of the chain rule), analytical expressions for the weighting functions have been derived in [14–20] and also compared to other methods in [21, 22]. The forward-adjoint approach is extremely efficient because only two radiative transfer calculations are required for computing the derivatives.

In this paper, we discuss the linearization of these radiative transfer models for EPIC retrieval purposes. More precisely, we will apply the linearized forward approach to DOME and MOME, and design a forward-adjoint approach based on DOME. As for the radiance calculations, the performances of the derivative calculations will be enhanced using acceleration techniques. The accuracy and efficiency of the linearized radiative transfer models will be analyzed by performing simulations which are pertinent to the retrieval of cloud parameters from EPIC oxygen A- and B-band measurements.

*Corresponding author
Email address: Victor.MolinaGarcia@dlr.de (Víctor Molina García)

The paper is organized as follows. In Section 2 we review the radiative transfer models DOME and MOME, in Section 3 we discuss the linearized radiative transfer models, and in Section 4 we present runtime and error analyses for these models. The results are summarized in Section 5. Some specific features of the linearization procedure are outlined in the appendices.

2. Radiative transfer models

The radiative transfer equation for the total radiance $I(r, \mathbf{\Omega})$ at point r and in the direction $\mathbf{\Omega} = (\mu, \varphi)$ reads as

$$\mu \frac{\mathrm{d}I}{\mathrm{d}r}(r, \mathbf{\Omega}) = -\sigma_{\mathtt{ext}}(r)\, I(r, \mathbf{\Omega}) + \frac{\sigma_{\mathtt{sct}}(r)}{4\pi} \int_{4\pi} P(r, \mathbf{\Omega}, \mathbf{\Omega}')\, I(r, \mathbf{\Omega}')\, \mathrm{d}\Omega', \quad (1)$$

where $\sigma_{\mathtt{ext}}$ and $\sigma_{\mathtt{sct}}$ are the extinction and scattering coefficients, respectively, and P is the scattering phase function. The phase function P is expanded in terms of normalized Legendre polynomials P_n, i.e.

$$P(r, \mathbf{\Omega}, \mathbf{\Omega}') = P(r, \cos\Theta) = \sum_{n=0}^{\infty} c_n\, \chi_n(r)\, P_n(\cos\Theta), \quad (2)$$

where $c_n = \sqrt{(2n+1)\ /\ 2}$ and $\cos\Theta = \mathbf{\Omega} \cdot \mathbf{\Omega}'$. The differential equation (1) is endowed with the top-of-atmosphere (TOA) boundary condition ($r = r_{\mathtt{TOA}}$),

$$I(r_{\mathtt{TOA}}, \mathbf{\Omega}^-) = F_0\, \delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0), \quad (3)$$

and the surface boundary condition ($r = r_{\mathtt{s}}$),

$$I(r_{\mathtt{s}}, \mathbf{\Omega}^+) = \frac{A}{\pi} \int_{2\pi} \rho(\mathbf{\Omega}^+, \mathbf{\Omega}^-)\, |\mu^-|\, I(r_{\mathtt{s}}, \mathbf{\Omega}^-)\, \mathrm{d}\Omega^-, \quad (4)$$

where F_0 is the incident solar flux, $\mathbf{\Omega}_0 = (-\mu_0, \varphi_0)$ with $\mu_0 > 0$ is the incident solar direction, and A and ρ are the surface albedo and the normalized bi-directional reflection function, respectively. The notations $\mathbf{\Omega}^+$ and $\mathbf{\Omega}^-$ stand for an upward and a downward direction, respectively. The total radiance is decomposed into the diffuse radiance $I_{\mathtt{d}}(r, \mathbf{\Omega})$ and the direct solar beam $I_{\odot}(r, \mathbf{\Omega})$, i.e.

$$I(r, \mathbf{\Omega}) = I_{\mathtt{d}}(r, \mathbf{\Omega}) + I_{\odot}(r, \mathbf{\Omega}), \quad (5)$$

where

$$I_{\odot}(r, \mathbf{\Omega}) = F_0\, \delta(\mathbf{\Omega} - \mathbf{\Omega}_0)\, T(r)\,, \quad (6)$$

$T(r) = \exp[-\tau_{\mathtt{ext}}(|\mathbf{r} - \mathbf{r}_{\mathtt{TOA}}|, \mathbf{\Omega}_0)]$ is the solar transmission, and $\tau_{\mathtt{ext}}(|\mathbf{r} - \mathbf{r}_{\mathtt{TOA}}|, \mathbf{\Omega}_0)$ is the solar optical depth between a general point $\mathbf{r}$ and the specific point at the top of the atmosphere $\mathbf{r}_{\mathtt{TOA}}$ in a spherical atmosphere.

Assuming cosine-azimuthal expansions for the diffuse radiance, the scattering phase function and the normalized bi-directional reflection function, we formulate a boundary-value problem for each Fourier component of the diffuse radiance $I_{\mathtt{d}m}(r, \mu)$ by means of the discrete ordinate method. The angular variation of the phase function and radiance is discretized in M discrete-ordinate directions per hemisphere, and the set of Gauss-Legendre quadrature points and weights in the interval $(0, 1)$ is denoted by $\{\mu_k, w_k\}_{k=1}^{M}$. The atmosphere is discretized in N levels: $r_1 = r_{\mathtt{TOA}} > r_2 > \ldots > r_N = r_{\mathtt{s}}$, and a layer j, bounded above by the level r_j and below by the level r_{j+1}, has the geometrical thickness $\triangle r_j = r_j - r_{j+1}$. The extinction and scattering coefficients, as well as the phase function coefficients, are assumed to be constant within each layer; their average values on layer j are $\sigma_{\mathtt{ext}j}$, $\sigma_{\mathtt{sct}j}$ and χ_{nj}, respectively.

The radiative transfer models analyzed in [4] are DOME and MOME. In summary, these models can be characterized as follows:

1. DOME [23, 24] relies on the layer equation, which relates the level values of the radiance field $\mathbf{i}_{mj} = [\mathbf{i}^+_{mj}; \mathbf{i}^-_{mj}]^{\mathsf{T}}$ with $[\mathbf{i}^{\pm}_{mj}]_k = I_{\mathtt{d}m}(r_j, \pm\mu_k)$, $k = 1, \ldots, M$, that is,

$$\mathtt{A}^1_{mj}\, \mathbf{i}_{mj} + \mathtt{A}^2_{mj}\, \mathbf{i}_{mj+1} = \mathtt{b}_{mj}. \quad (7)$$

   The layer equation together with the boundary conditions at the top and the bottom of the atmosphere, as well as the continuity of the radiance across layers, are assembled into the global matrix of the entire atmosphere, and the solution of the resulting system of equations yields the level values of the diffuse radiance field.

2. MOME applies to a molecular atmosphere containing a homogeneous cloud placed between the top level $r_{j_{\mathtt{min}}}$ and the bottom level $r_{j_{\mathtt{max}}+1}$. The method is a combination of the conventional matrix operator method and the discrete ordinate method with matrix exponential, in the sense that the reflection matrix and vector of the atmosphere below the cloud top are computed in the framework of the conventional matrix operator method, while the level values of the diffuse radiance field above the cloud top are computed in the framework of the discrete ordinate method with matrix exponential. The conventional matrix operator method, e.g. [25], uses the interaction principle equation

$$\begin{bmatrix} \mathbf{i}^+_{mj} \\ \mathbf{i}^-_{mj+1} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{mj} & \mathbf{T}_{mj} \\ \mathbf{T}_{mj} & \mathbf{R}_{mj} \end{bmatrix} \begin{bmatrix} \mathbf{i}^-_{mj} \\ \mathbf{i}^+_{mj+1} \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}^+_{mj} \\ \mathbf{\Sigma}^-_{mj} \end{bmatrix}, \quad (8)$$

   where $\mathbf{R}_{mj}$ and $\mathbf{T}_{mj}$ are the reflection and transmission matrices of layer j, respectively, and $\mathbf{\Sigma}^{\pm}_{mj}$ is the source vector. The computation process is an upward recurrence over the atmospheric layers [13]: If $\mathtt{R}_{mj+1}$ and $\mathtt{r}_{mj+1}$ are respectively the reflection matrix and the reflection vector of stack $j+1$ (the group of layers bounded above by the level r_{j+1} and below by the surface level r_N), the interaction principle equation for stack j (the group of layers bounded above by the level r_j and below by the surface level r_N) reads as

$$\mathbf{i}^+_{mj} = \mathtt{R}_{mj}\, \mathbf{i}^-_{mj} + \mathtt{r}_{mj}, \quad (9)$$

   where $\mathtt{R}_{mj}$ and $\mathtt{r}_{mj}$ are computed with the adding formulas

$$\mathtt{R}_{mj} = \mathbf{R}_{mj} + \mathbf{T}_{mj}\, \mathtt{R}_{mj+1}\, (\mathbf{I} - \mathbf{R}_{mj}\, \mathtt{R}_{mj+1})^{-1}\, \mathbf{T}_{mj}, \quad (10)$$

$$\mathtt{r}_{mj} = \Sigma^+_{mj} + \mathbf{T}_{mj}\, \mathtt{r}_{mj+1} + \mathbf{T}_{mj}\, \mathtt{R}_{mj+1} (\mathbf{I} - \mathbf{R}_{mj}\, \mathtt{R}_{mj+1})^{-1} (\mathbf{R}_{mj}\, \mathtt{r}_{mj+1} + \Sigma^-_{mj}). \quad (11)$$

The two methods are combined by considering the interaction principle equation at the cloud top, i.e.

$$\mathbf{i}^{+}_{mj_{\mathtt{min}}} = \mathtt{R}_{mj_{\mathtt{min}}}\,\mathbf{i}^{-}_{mj_{\mathtt{min}}} + \mathtt{r}_{mj_{\mathtt{min}}}, \tag{12}$$

as the surface boundary condition for the discrete ordinate method with matrix exponential.

The layer quantities $\mathtt{A}^1_{mj}$, $\mathtt{A}^2_{mj}$ and $\mathtt{b}_{mj}$ in (7), as well as $\mathbf{R}_{mj}$, $\mathbf{T}_{mj}$ and $\mathbf{\Sigma}^{\pm}_{mj}$ in (8), are expressed in terms of the exponential of the layer matrix

$$\mathbf{A}_{mj} = \begin{bmatrix} \mathbf{A}^{11}_{mj} & \mathbf{A}^{12}_{mj} \\ -\mathbf{A}^{12}_{mj} & -\mathbf{A}^{11}_{mj} \end{bmatrix}, \tag{13}$$

with entries

$$[\mathbf{A}^{11}_{mj}]_{kl} = \frac{1}{2\,\mu_k}\,[w_l\,\sigma_{\mathtt{sct}j}\,p_{mj}(\mu_k,\mu_l) - 2\,\sigma_{\mathtt{ext}j}\,\delta_{kl}], \tag{14}$$

$$[\mathbf{A}^{12}_{mj}]_{kl} = \frac{1}{2\,\mu_k}\,w_l\,\sigma_{\mathtt{sct}j}\,p_{mj}(\mu_k,-\mu_l). \tag{15}$$

Note that the product $\sigma_{\mathtt{sct}j}\,p_{mj}(\mu,\mu')$ in (14) and (15) is computed as

$$\sigma_{\mathtt{sct}j}\,p_{mj}(\mu,\mu') = \sum_{n=m}^{2M-1} \xi_{nj}\,P^m_n(\mu)\,P^m_n(\mu'), \tag{16}$$

$$\xi_{nj} = \sigma_{\mathtt{sct}j}\chi_{nj}, \tag{17}$$

so that the optical input parameters of the radiative transfer models are $\sigma_{\mathtt{ext}j}$ and ξ_{nj} for $j = 1,\ldots,N-1$ and $n = 0,\ldots,2M-1$. DOME and MOME rely on the same matrix exponential formalism. The matrix exponential can be computed by employing the eigendecomposition method or the Padé approximation. In the first case, the layer quantities depend on the inverse of the eigenvector matrix $\mathbf{V}_{mj}$ and the eigenvalue matrix $\mathbf{\Lambda}_{mj} = \mathrm{diag}[\lambda_1,\ldots,\lambda_M]$ of $\mathbf{A}_{mj}$, while, in the second case, the layer quantities are expressed in terms of the layer matrix $\mathbf{A}_{mj}$ [4]. In the layer j, the matrix exponential is computed by means of the Padé approximation if $\|\mathbf{A}_{mj}\,\triangle r_j\| \le 1$, and by the eigendecomposition method if this is not the case. For radiance calculations in the EPIC oxygen A-band absorption channel, DOME and MOME have been used in [4] together with the following acceleration techniques:

1. the delta-M [26] and the truncated-plus-single-scattering (TMS) method [27];
2. the telescoping technique [12, 23], which consists of the solution of a (reduced) boundary-value problem for the cloud layers and azimuthal modes $m > 2$;
3. the method of false discrete ordinate [25, 28–30], which eliminates the source integration in the post-processing step of the discrete ordinate method;
4. the correlated k-distribution method [5] and the PCA technique [6–9] (independently and together).

3. Linearized radiative transfer models

The radiance measured by the instrument is a function of various atmospheric and surface parameters of interest. In the first category we include, for example, the layer values of the trace gas extinction coefficient and the cloud parameters, while in the second category we include the surface albedo and parameters characterizing the normalized bi-directional reflection function. In the following, we consider the computation of the partial derivatives of the measured radiance with respect to a set of atmospheric parameters ς_i, $i = 1,\ldots,N_{\mathrm{p}}$, since the basic concepts are fully represented in this case. Before proceeding, we mention that the optical input parameters of the linearized radiative transfer models are $\partial\sigma_{\mathtt{ext}j}/\partial\varsigma_i$ and $\partial\xi_{nj}/\partial\varsigma_i$ for $i = 1,\ldots,N_{\mathrm{p}}$, $j = 1,\ldots,N-1$, and $n = 0,\ldots,2M-1$.

3.1. Linearized forward approach

A detailed description of the linearized forward approach can be found in [10–12]. In our case, the linearized forward approach is applied to DOME and MOME. To compute the partial derivative with respect to the atmospheric property ς_i, we proceed as follows:

1. In DOME, we linearize the layer equation (7) and obtain [22]

$$\mathtt{A}^1_{mj}\frac{\partial\mathbf{i}_{mj}}{\partial\varsigma_i} + \mathtt{A}^2_{mj}\frac{\partial\mathbf{i}_{mj+1}}{\partial\varsigma_i} = \frac{\partial\mathtt{b}_{mj}}{\partial\varsigma_i} - \frac{\partial\mathtt{A}^1_{mj}}{\partial\varsigma_i}\mathbf{i}_{mj} - \frac{\partial\mathtt{A}^2_{mj}}{\partial\varsigma_i}\mathbf{i}_{mj+1}. \tag{18}$$

   As for radiance calculations, the linearized layer equations (18) are assembled into a global system of equations for the entire atmosphere. It is worth noticing that the system matrix for derivative calculations coincides with the system matrix for radiance calculations; only the right-hand sides are different. In order to increase the efficiency of the method we compute the partial derivatives with respect to all atmospheric parameters ς_i, $i = 1,\ldots,N_{\mathrm{p}}$, that is, we solve a system of equations with multiple right-hand sides.
2. In MOME, we linearize the interaction principle equation in (12), and compute the partial derivatives $\partial\mathtt{R}_{mj_{\mathtt{min}}}/\partial\varsigma_i$ and $\partial\mathtt{r}_{mj_{\mathtt{min}}}/\partial\varsigma_i$ recursively by linearizing the adding formulas (10) and (11). From (10) and (11), it is apparent that the recursion relation involves the partial derivatives $\partial\mathbf{R}_{mj}/\partial\varsigma_i$, $\partial\mathbf{T}_{mj}/\partial\varsigma_i$ and $\partial\mathbf{\Sigma}^{\pm}_{mj}/\partial\varsigma_i$.

To compute $\partial\mathtt{A}^1_{mj}/\partial\varsigma_i$, $\partial\mathtt{A}^2_{mj}/\partial\varsigma_i$ and $\partial\mathtt{b}_{mj}/\partial\varsigma_i$ as well as $\partial\mathbf{R}_{mj}/\partial\varsigma_i$, $\partial\mathbf{T}_{mj}/\partial\varsigma_i$ and $\partial\mathbf{\Sigma}^{\pm}_{mj}/\partial\varsigma_i$, we apply the chain rule. In the case of the Padé approximation, the derivative calculations are trivial, but in the case of the eigendecomposition method we face the calculation of the partial derivatives of the inverse of the eigenvector matrix $\mathbf{V}^{-1}_{mj}$ and of the eigenvalues λ_k. This computation step is outlined in Appendix A.

3.2. Forward-adjoint approach

The steps of the adjoint radiative transfer approach consist of the formulation of the boundary-value problem for radiative transfer as an operator equation with homogeneous boundary conditions, the derivation of the adjoint radiative transfer operator, and the representation of the measured radiance in terms

of the solution of the adjoint radiative transfer problem and the forward source function. In our analysis, the forward and the adjoint radiative transfer problem are solved by using DOME.

Essentially, the forward-adjoint approach relies on the following basic result: If the (total) radiance field $I(r,\mathbf{\Omega})$ solves the forward problem

$$\mathcal{L}I(r,\mathbf{\Omega}) = Q(r,\mathbf{\Omega}), \tag{19}$$
$$I(r_{\mathtt{TOA}},\mathbf{\Omega}^-) = I(r_{\mathtt{s}},\mathbf{\Omega}^+) = 0, \tag{20}$$

and the radiance field $I^\dagger(r,\mathbf{\Omega})$ solves the adjoint problem

$$\mathcal{L}^\dagger I^\dagger(r,\mathbf{\Omega}) = Q^\dagger(r,\mathbf{\Omega}), \tag{21}$$
$$I^\dagger(r_{\mathtt{TOA}},\mathbf{\Omega}^+) = I^\dagger(r_{\mathtt{s}},\mathbf{\Omega}^-) = 0, \tag{22}$$

then the measured radiance at the top of the atmosphere and in the direction $\mathbf{\Omega}_{\mathtt{m}} = (\mu_{\mathtt{m}},\varphi_{\mathtt{m}})$ with $\mu_{\mathtt{m}} > 0$, can be computed as

$$I_{\mathtt{m}} = \langle Q^\dagger, I\rangle = \langle I^\dagger, Q\rangle. \tag{23}$$

The forward radiative transfer operator $\mathcal{L}$ and the forward source function Q are given by

$$\begin{aligned}\mathcal{L}I(r,\mathbf{\Omega}) &= \mu\,\frac{\mathrm{d}I}{\mathrm{d}r}(r,\mathbf{\Omega}) + \sigma_{\mathtt{ext}}(r)\,I(r,\mathbf{\Omega})\\ &-\frac{\sigma_{\mathtt{sct}}(r)}{4\pi}\int_{4\pi} P(r,\mathbf{\Omega},\mathbf{\Omega}')\,I(r,\mathbf{\Omega}')\,\mathrm{d}\Omega'\\ &-\frac{A}{\pi}\,\delta(r-r_{\mathtt{s}})\,H(\mu)\,\mu\int_{4\pi}\rho(\mathbf{\Omega},\mathbf{\Omega}')\,H(-\mu')\,|\mu'|\,I(r,\mathbf{\Omega}')\,\mathrm{d}\Omega'\end{aligned} \tag{24}$$

and

$$Q(r,\mathbf{\Omega}) = F_0\,\mu_0\,\delta(r-r_{\mathtt{TOA}})\,\delta(\mathbf{\Omega}-\mathbf{\Omega}_0), \tag{25}$$

respectively, where δ is the Dirac delta function and H is the Heaviside step function. In this context, the forward problem consisting of the operator equation (19) and the homogeneous boundary conditions (20) is equivalent to the radiative transfer equation (1) with the boundary conditions (3) and (4). The adjoint radiative transfer operator $\mathcal{L}^\dagger$ is defined through the Lagrange identity

$$\langle \mathcal{L}I, I^\dagger\rangle = \langle I, \mathcal{L}^\dagger I^\dagger\rangle, \tag{26}$$

where the scalar product of fields I_1 and I_2 is given by

$$\langle I_1, I_2\rangle = \int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\int_{4\pi} I_1(r,\mathbf{\Omega})\,I_2(r,\mathbf{\Omega})\,\mathrm{d}\Omega\,\mathrm{d}r.$$

For the homogeneous boundary conditions (20) and (22), the expressions of the adjoint operator $\mathcal{L}^\dagger$ and the adjoint source function $Q^\dagger$ are

$$\begin{aligned}\mathcal{L}^\dagger I^\dagger(r,\mathbf{\Omega}) &= -\mu\,\frac{\mathrm{d}I^\dagger}{\mathrm{d}r}(r,\mathbf{\Omega}) + \sigma_{\mathtt{ext}}(r)\;I^\dagger(r,\mathbf{\Omega})\\ &-\frac{\sigma_{\mathtt{sct}}(r)}{4\pi}\int_{4\pi} P(r,\mathbf{\Omega}',\mathbf{\Omega})\,I^\dagger(r,\mathbf{\Omega}')\,\mathrm{d}\Omega'\\ &-\frac{A}{\pi}\,\delta(r-r_{\mathtt{s}})\,H(-\mu)\,|\mu|\int_{4\pi}\rho(\mathbf{\Omega}',\mathbf{\Omega})\,H(\mu')\,|\mu'|\,I^\dagger(r,\mathbf{\Omega}')\,\mathrm{d}\Omega'\end{aligned} \tag{27}$$

and

$$Q^\dagger(r,\mathbf{\Omega}) = \delta(r-r_{\mathtt{TOA}})\,\delta(\mathbf{\Omega}-\mathbf{\Omega}_{\mathtt{m}}), \tag{28}$$

respectively. The solution of the adjoint radiative transfer problem can be found by using the same solution method as for the forward problem with a modified source function. Actually, it can be shown that the conjugate adjoint radiance $\widehat{I}^\dagger$ defined by

$$\widehat{I}^\dagger(r,\mathbf{\Omega}) = I^\dagger(r,-\mathbf{\Omega})$$

solves the conjugate adjoint problem

$$\mathcal{L}\widehat{I}^\dagger(r,\mathbf{\Omega}) = \widehat{Q}^\dagger(r,\mathbf{\Omega}), \tag{29}$$
$$\widehat{I}^\dagger(r_{\mathtt{TOA}},\mathbf{\Omega}^-) = \widehat{I}^\dagger(r_{\mathtt{s}},\mathbf{\Omega}^+) = 0, \tag{30}$$

where

$$\widehat{Q}^\dagger(r,\mathbf{\Omega}) = \widehat{F}_0\,\mu_{\mathtt{m}}\,\delta(r-r_{\mathtt{TOA}})\,\delta(\mathbf{\Omega}-\widehat{\mathbf{\Omega}}_{\mathtt{m}}) \tag{31}$$

is the conjugate adjoint source function, $\widehat{\mathbf{\Omega}}_{\mathtt{m}} = -\mathbf{\Omega}_{\mathtt{m}} = (\widehat{\mu}_{\mathtt{m}},\widehat{\varphi}_{\mathtt{m}})$ is the conjugate adjoint direction ($\widehat{\mu}_{\mathtt{m}} = -\mu_{\mathtt{m}}$ and $\widehat{\varphi}_{\mathtt{m}} = \varphi_{\mathtt{m}} + \pi$), and $\widehat{F}_0 = 1/\mu_{\mathtt{m}}$. The boundary value problems (19)-(20) and (29)-(30) are identical excepting the source functions (25) and (31), which, however, are of similar forms. As a result, in the discrete ordinate method with matrix exponential, the system matrices for the forward and conjugate adjoint problems coincide, and solving a system of equations with two right-hand sides yields the level values of the forward and (conjugate) adjoint radiance fields.

We come now to the derivative calculations. Taking the variation of the measured radiance with respect to variations of atmospheric parameters, and using (21) and (26) yields ($\delta Q^\dagger = 0$)

$$\delta I_{\mathtt{m}} = \langle Q^\dagger, \delta I\rangle = \langle I^\dagger, \mathcal{L}\delta I\rangle. \tag{32}$$

The variation of the forward operator equation (19) gives

$$\mathcal{L}\delta I = \delta Q - \delta\mathcal{L}I, \tag{33}$$

and, since $\delta Q = 0$, we obtain

$$\delta I_{\mathtt{m}} = -\langle I^\dagger, \delta\mathcal{L}I\rangle. \tag{34}$$

In the first step of the forward-adjoint approach, we separate the total radiance $I(r,\mathbf{\Omega})$ into a diffuse and a direct component $I_{\mathtt{d}}(r,\mathbf{\Omega})$ and $I_\odot(r,\mathbf{\Omega})$, respectively, (cf. (5)), and do the same for the conjugate adjoint total radiance $\widehat{I}^\dagger(r,\mathbf{\Omega})$, i.e.

$$\widehat{I}^\dagger(r,\mathbf{\Omega}) = \widehat{I}^\dagger_{\mathtt{d}}(r,\mathbf{\Omega}) + \widehat{I}^\dagger_\odot(r,\mathbf{\Omega}), \tag{35}$$

with

$$\widehat{I}^\dagger_\odot(r,\mathbf{\Omega}) = \widehat{F}_0\,\delta(\mathbf{\Omega}-\widehat{\mathbf{\Omega}}_{\mathtt{m}})\,\widehat{T}^\dagger(r)$$

and $\widehat{T}^\dagger(r) = \exp[-\tau_{\mathtt{ext}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|,\widehat{\mathbf{\Omega}}_{\mathtt{m}})]$. Inserting the resulting expressions in (34) it gives

$$\frac{\partial I_{\mathtt{m}}}{\partial\varsigma_i} = T_1 + T_2, \tag{36}$$

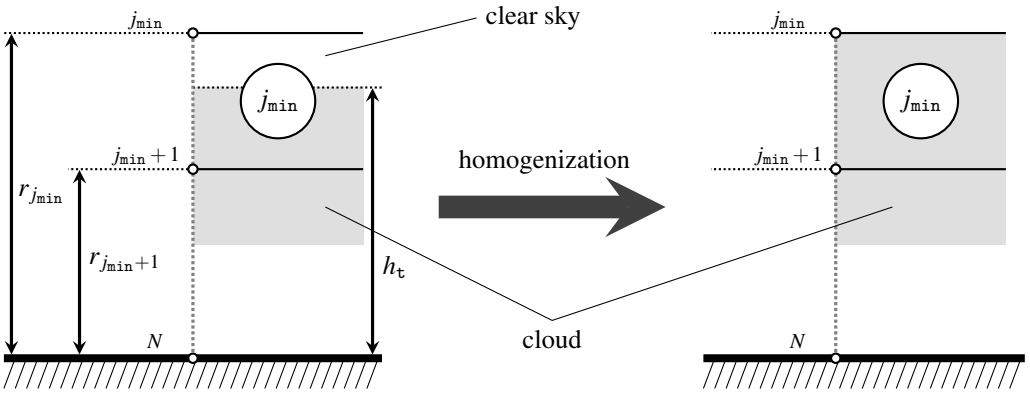


Figure 1: Homogenization of the atmospheric layer $j_{\min}$ containing the cloud top height h_{t}. The layer $j_{\min}$, with the boundary levels $j_{\min}$ and $j_{\min}+1$, is indicated by a circle.

where the first term T_1 involves integrals of the conjugate adjoint diffuse radiance $\widehat{I}_{\mathrm{d}}^{\dagger}$,

$$
\begin{aligned}
T_1 = &- \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \int_{4\pi} \frac{\partial \sigma_{\mathrm{ext}}}{\partial \varsigma_i}(r)\, \widehat{I}_{\mathrm{d}}^{\dagger}(r,-\boldsymbol{\Omega})\, I_{\mathrm{d}}(r,\boldsymbol{\Omega})\, \mathrm{d}r\, \mathrm{d}\Omega \\
&- F_0 \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \frac{\partial \sigma_{\mathrm{ext}}}{\partial \varsigma_i}(r)\, \widehat{I}_{\mathrm{d}}^{\dagger}(r,-\boldsymbol{\Omega}_0)\, T(r)\, \mathrm{d}r \\
&+ \frac{1}{4\pi} \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \int_{4\pi} \widehat{I}_{\mathrm{d}}^{\dagger}(r,-\boldsymbol{\Omega})\, \mathrm{d}r\, \mathrm{d}\Omega \int_{4\pi} \frac{\partial}{\partial \varsigma_i} [\sigma_{\mathtt{sct}}(r) P(r,\boldsymbol{\Omega},\boldsymbol{\Omega}')]\, I_{\mathrm{d}}(r,\boldsymbol{\Omega}')\, \mathrm{d}\Omega' \\
&+ \frac{F_0}{4\pi} \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \int_{4\pi} \frac{\partial}{\partial \varsigma_i} [\sigma_{\mathtt{sct}}(r) P(r,\boldsymbol{\Omega},\boldsymbol{\Omega}_0)]\, \widehat{I}_{\mathrm{d}}^{\dagger}(r,-\boldsymbol{\Omega})\, T(r)\, \mathrm{d}r\, \mathrm{d}\Omega, \qquad (37)
\end{aligned}
$$

while the second term T_2 involves integrals of the (conjugate) adjoint transmission $\widehat{T}^{\dagger}$,

$$
\begin{aligned}
T_2 = &- \widehat{F}_0 \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \frac{\partial \sigma_{\mathrm{ext}}}{\partial \varsigma_i}(r)\, I_{\mathrm{d}}(r,\boldsymbol{\Omega}_{\mathrm{m}})\, \widehat{T}^{\dagger}(r)\, \mathrm{d}r \\
&+ \frac{\widehat{F}_0}{4\pi} \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \widehat{T}^{\dagger}(r)\, \mathrm{d}r \int_{4\pi} \frac{\partial}{\partial \varsigma_i} [\sigma_{\mathtt{sct}}(r)\, P(r,\boldsymbol{\Omega}_{\mathrm{m}},\boldsymbol{\Omega}')]\, I_{\mathrm{d}}(r,\boldsymbol{\Omega}')\, \mathrm{d}\Omega' \\
&+ \frac{\widehat{F}_0 F_0}{4\pi} \int_{r_{\mathrm{s}}}^{r_{\mathrm{TOA}}} \frac{\partial}{\partial \varsigma_i} [\sigma_{\mathtt{sct}}(r)\, P(r,\boldsymbol{\Omega}_{\mathrm{m}},\boldsymbol{\Omega}_0)]\, T(r)\, \widehat{T}^{\dagger}(r)\, \mathrm{d}r. \qquad (38)
\end{aligned}
$$

As a next step, the integration over the azimuthal angle in (37) and (38) is performed. This computation step together with the integration over the radial coordinate is described in Appendix B.

3.3. Derivatives with respect to cloud geometrical parameters

We must be cautious when computing the derivatives with respect to the cloud geometrical parameters (cloud top height h_{t} and cloud bottom height h_{b}). In fact, this process is intimately connected with the discretization of the atmosphere containing an homogeneous cloud. We have two options:

1. the atmosphere above, below and in the cloud is discretized, so that the cloud top and bottom heights are atmospheric levels;
2. the atmosphere is discretized into a fixed grid, and the layers containing the cloud top and bottom heights are homogenized.

The first discretization method, in which the atmosphere is discretized each time when a new cloud position is considered, can be used in conjunction with the forward-adjoint approach and is described in Appendix C. The second discretization method, in which a fixed altitude grid is used, can be applied for both the linearized forward and forward-adjoint approaches and is summarized below.

Let the optical properties of the homogeneous cloud be described by the cloud extinction coefficient $\sigma_{\mathtt{ext}}^{\mathtt{c}}$ and the expansion coefficients $\xi_n^{\mathtt{c}}$. Assume that the cloud top height h_{t} lies between the levels $r_{j_{\min}+1}$ and $r_{j_{\min}}$, i.e. $r_{j_{\min}+1} < h_{\mathrm{t}} < r_{j_{\min}}$ as shown in Fig. 1. In layer $j_{\min}$ containing the cloud top height h_{t}, we perform an homogenization which consists of the computation of the optical parameters $\sigma_{\mathtt{ext}\,j_{\min}}$ and $\xi_{n\,j_{\min}}$ by using

$$
\sigma_{\mathtt{ext}\,j_{\min}} = \sigma^{0}_{\mathtt{ext}\,j_{\min}} + w(h_{\mathrm{t}})\, \sigma^{\mathtt{c}}_{\mathtt{ext}}, \qquad (39)
$$

$$
\xi_{n\,j_{\min}} = \xi^{0}_{n\,j_{\min}} + w(h_{\mathrm{t}})\, \xi^{\mathtt{c}}_{n}, \qquad (40)
$$

where the superscript "0" refers to the clear sky atmosphere, and the weighting factor w is given by

$$
w(h_{\mathrm{t}}) = \frac{h_{\mathrm{t}} - r_{j_{\min}+1}}{\triangle r_j}. \qquad (41)
$$

By this procedure, $\sigma_{\mathtt{ext}\,j_{\min}}$ and $\xi_{n\,j_{\min}}$ are functions of h_{t}, i.e.

$$
\sigma_{\mathtt{ext}\,j_{\min}} = \sigma_{\mathtt{ext}\,j_{\min}}(\sigma^{\mathtt{c}}_{\mathtt{ext}}, h_{\mathrm{t}}) \quad \text{and} \quad \xi_{n\,j_{\min}} = \xi_{n\,j_{\min}}(\sigma^{\mathtt{c}}_{\mathtt{ext}}, h_{\mathrm{t}}),
$$

so that the partial derivatives $\partial \sigma_{\mathtt{ext}\,j_{\min}}/\partial h_{\mathrm{t}}$ and $\partial \xi_{n\,j_{\min}}/\partial h_{\mathrm{t}}$ can be readily computed. The derivatives with respect to the cloud bottom height h_{b} are computed in an analogous manner. Note that, for an accurate homogenization, the discretization step $\triangle r_j$ should not be too large.

3.4. PCA-based methods for derivative calculations

As for radiance calculations, acceleration techniques can be used to increase the computation speed. These include the Padé approximation, the delta-M scaling and TMS correction methods, the telescoping technique, the method of false discrete ordinate, the correlated k-distribution method, and the PCA technique. For derivative calculations, two PCA-based methods can be designed. To explain these methods, we give a short overview of dimensionality reduction techniques for optical parameters [9].

An approximate model for computing the radiance I at the wavelength λ reads as

$$\ln \frac{I(\lambda)}{I_{\mathtt{a}}(\lambda)} = f_{\mathtt{I}}(\lambda), \tag{42}$$

where $I_{\mathtt{a}}$ is the radiance computed by an approximate radiative transfer model, and $f_{\mathtt{I}}$ is a correction factor.

The optical property dimensionality reduction is performed on the layer values of the gas absorption coefficient $\ln \sigma^{\mathtt{gas}}_{\mathtt{abs}}(\lambda)$ and the molecular scattering coefficient $\ln \sigma^{\mathtt{mol}}_{\mathtt{sct}}(\lambda)$. For each λ_k, $k = 1, \ldots N_\lambda$, where N_λ is the number of discrete wavelengths, we define an $\mathcal{N}$-dimensional vector (the abbreviation "not" stands for notation)

$$\mathbf{x}(\lambda_k) \overset{\text{not}}{=} \mathbf{x}_k = [\ln \sigma^{\mathtt{gas}}_{\mathtt{abs}_j}(\lambda_k); \ln \sigma^{\mathtt{mol}}_{\mathtt{sct}_j}(\lambda_k)]^{\mathsf{T}}, \;\; j = 1, \ldots, N-1, \tag{43}$$

with $\mathcal{N} = 2N-2$, so that the wavelength variability of the optical parameters is contained in $\mathbf{x}_k$. High-dimensional real data often lie on or near a lower-dimensional manifold. The fundamental issues in dimensionality reduction are the modeling of the geometry structure of the manifold, and the design of an appropriate embedding for data projection. For the $\mathcal{N}$-dimensional data set $\{\mathbf{x}_k\}_{k=1}^{N_\lambda}$, where $\mathbf{x}_k \in \mathbb{R}^{\mathcal{N}}$, let $\overline{\mathbf{x}} = (1/N_\lambda) \sum_{k=1}^{N_\lambda} \mathbf{x}_k$ be the sample mean of the data. The goal of a linear embedding method is to find an $\mathcal{M}$-dimensional subspace ($\mathcal{M} < \mathcal{N}$) spanned by a set of linear independent vectors $\{\mathbf{a}_l\}_{l=1}^{\mathcal{M}}$, such that the centered data $\mathbf{x}_k - \overline{\mathbf{x}}$ lie mainly on this subspace (manifold), i.e. $\mathbf{x}_k \approx \overline{\mathbf{x}} + \sum_{l=1}^{\mathcal{M}} y_{kl} \mathbf{a}_l = \overline{\mathbf{x}} + \mathbf{A}\mathbf{y}_k$, $k = 1, \ldots, N_\lambda$. Here, $\mathbf{A} \overset{\text{not}}{=} [\mathbf{a}_l]_{l=1}^{\mathcal{M}}$ is an $\mathcal{N} \times \mathcal{M}$ matrix comprising the column vectors $\mathbf{a}_l$, and y_{kl} is the lth component of the vector of parameters $\mathbf{y}_k \in \mathbb{R}^{\mathcal{M}}$. The vector of parameters $\mathbf{y}_k$ is given by the forward mapping from the high-dimensional space to the low-dimensional space, i.e. $\mathbf{y}_k = \mathbf{A}^\dagger(\mathbf{x}_k - \overline{\mathbf{x}})$, where $\mathbf{A}^\dagger = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}}$ is the pseudoinverse of $\mathbf{A}$. Now, let $f(\mathbf{x}_k)$ be a scalar function, which by assumption is not too nonlinear in $\mathbf{x}_k$. Setting

$$\triangle \mathbf{x}_k = \sum_{l=1}^{\mathcal{M}} y_{kl}\, \mathbf{a}_l, \tag{44}$$

we approximate $f(\mathbf{x}_k)$ by a second-order Taylor expansion, expressed in finite-difference form by

$$\begin{aligned} f(\mathbf{x}_k) \approx f(\overline{\mathbf{x}}) &+ \frac{1}{2} \sum_{l=1}^{\mathcal{M}} [f(\overline{\mathbf{x}} + \mathbf{a}_l) - f(\overline{\mathbf{x}} - \mathbf{a}_l)]\, y_{kl} \\ &+ \frac{1}{2} \sum_{l=1}^{\mathcal{M}} [f(\overline{\mathbf{x}} + \mathbf{a}_l) - 2\, f(\overline{\mathbf{x}}) + f(\overline{\mathbf{x}} - \mathbf{a}_l)]\, y_{kl}^2. \end{aligned} \tag{45}$$

To compute the radiance correction factor we identify

$$f(\mathbf{x}_k) = f_{\mathtt{I}}(\lambda_k) = \ln[I(\lambda_k)/I_{\mathtt{a}}(\lambda_k)], \tag{46}$$

and from (45) it is apparent that the computation of the correction factor requires $2\mathcal{M} + 1$ calls of the exact and approximate models. As a result and taking into account that $\mathcal{M} \ll N_\lambda$, we are led to a substantial reduction of the computational time. A short remark is though needed. PCA produces a global linear model of the data and is appropriate when the manifold is embedded linearly or almost linearly in the data space. The method preserves only the global structure of the data, and may fail to preserve the local structure if the data lies on a nonlinear manifold. In contrast, the linear embedding methods presented in [9] optimally preserve local neighbourhood information (the local structure of the data) in a certain sense. In fact, the dimensionality reduction approach used in [6, 7] is also a local linear model, which combines PCA with the clustering of the data space. In general, a local model implementation of PCA involves a two-step procedure: (1) a clustering of the data space into disjoint regions by using, for instance, the Lloyd algorithm with Euclidean distances as the distortion function, and (2) the estimation of the linear mappings within each region by PCA. Then, each region (bin of wavelengths) is characterized by its own orthogonal basis, and so, by its own set of correction factors. If P is the number of the disjoint regions, then $(2\mathcal{M}+1)P$ calls of the exact and the approximate models are required to compute all correction factors.

Two methods can be used for computing the derivatives with respect to the atmospheric parameter ς_i.

1. The first method is similar to (46), and uses the identification
$$f(\mathbf{x}_k) = f_{\varsigma_i}(\lambda_k) = \ln\Big[\frac{\partial I}{\partial \varsigma_i}(\lambda_k) / \frac{\partial I_{\mathtt{a}}}{\partial \varsigma_i}(\lambda_k)\Big]. \tag{47}$$
2. The second method is based on the linearization of the restoration equation (45) for the radiance correction factor $f_{\mathtt{I}}$, and involves the computation of the derivatives $\partial y_{kl}/\partial \varsigma_i$ [8]. In our framework, this technique can be summarized as follows. Express the restoration equation as
$$I(\lambda_k) = I_{\mathtt{a}}(\lambda_k)\, \mathrm{e}^{f_{\mathtt{I}}(\lambda_k)}, \tag{48}$$
and take the derivative to obtain
$$\frac{\partial I}{\partial \varsigma_i}(\lambda_k) = \frac{\partial I_{\mathtt{a}}}{\partial \varsigma_i}(\lambda_k)\, \mathrm{e}^{f_{\mathtt{I}}(\lambda_k)} + I_{\mathtt{a}}(\lambda_k)\, \mathrm{e}^{f_{\mathtt{I}}(\lambda_k)} \frac{\partial f_{\mathtt{I}}}{\partial \varsigma_i}(\lambda_k), \tag{49}$$
with
$$\begin{aligned} \frac{\partial f_{\mathtt{I}}}{\partial \varsigma_i}(\lambda_k) &\equiv \frac{\partial f}{\partial \varsigma_i}(\mathbf{x}_k) \\ &\approx \frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}}) + \frac{1}{2} \sum_{l=1}^{\mathcal{M}} [\frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}} + \mathbf{a}_l) - \frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}} - \mathbf{a}_l)]\, y_{kl} \\ &+ \frac{1}{2} \sum_{l=1}^{\mathcal{M}} [f(\overline{\mathbf{x}} + \mathbf{a}_l) - f(\overline{\mathbf{x}} - \mathbf{a}_l)]\, \frac{\partial y_{kl}}{\partial \varsigma_i} \\ &+ \frac{1}{2} \sum_{l=1}^{\mathcal{M}} [\frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}} + \mathbf{a}_l) - 2\, \frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}}) + \frac{\partial f}{\partial \varsigma_i}(\overline{\mathbf{x}} - \mathbf{a}_l)]\, y_{kl}^2 \\ &+ \sum_{l=1}^{\mathcal{M}} [f(\overline{\mathbf{x}} + \mathbf{a}_l) - 2\, f(\overline{\mathbf{x}}) + f(\overline{\mathbf{x}} - \mathbf{a}_l)]\, y_{kl}\, \frac{\partial y_{kl}}{\partial \varsigma_i}. \end{aligned} \tag{50}$$

In general, from a vector $\mathbf{x}$ (which can be identified with $\overline{\mathbf{x}}$ or $\overline{\mathbf{x}} \pm \mathbf{a}_l$) we extract $\{\sigma^{\mathtt{gas}}_{\mathtt{abs}j}\}_{j=1}^{N-1}$ and $\{\sigma^{\mathtt{mol}}_{\mathtt{sct}j}\}_{j=1}^{N-1}$ according to the representation $\mathbf{x} = [\ln \sigma^{\mathtt{gas}}_{\mathtt{abs}j}; \ln \sigma^{\mathtt{mol}}_{\mathtt{sct}j}]^{\intercal}$ and, for these optical parameters, we compute the exact and approximate radiances $I(\mathbf{x})$ and $I_{\mathtt{a}}(\mathbf{x})$, respectively. Then, we have

$$\frac{\partial f}{\partial \varsigma_i}(\mathbf{x}) = \frac{1}{I(\mathbf{x})\, I_{\mathtt{a}}(\mathbf{x})} \Big[\frac{\partial I}{\partial \varsigma_i}(\mathbf{x})\, I_{\mathtt{a}}(\mathbf{x}) - \frac{\partial I_{\mathtt{a}}}{\partial \varsigma_i}(\mathbf{x})\, I(\mathbf{x}) \Big], \tag{51}$$

and what is left is the computation of $\partial y_{kl} \,/\, \partial \varsigma_i$. For doing this, we need to specify the mapping from the high-dimensional space to the low-dimensional space. Let PCA be chosen as the dimensionality reduction technique, and let us assume that all centered data $\mathbf{x}_k - \overline{\mathbf{x}}$, $k = 1, \ldots, N_\lambda$, are stacked into the columns of an $\mathcal{N} \times N_\lambda$ matrix $\mathbf{X}$, i.e. $\mathbf{X} = [\mathbf{x}_k - \overline{\mathbf{x}}]_{k=1}^{N_\lambda}$. Essentially, PCA performs a dimensionality reduction by projecting the original $\mathcal{N}$-dimensional data on the $\mathcal{M}$-dimensional subspace spanned by the dominant singular vectors of the data's covariance matrix. Therefore, assuming the singular value decomposition $\mathbf{C}_{\mathbf{x}} = (1/N_\lambda)\, \mathbf{X}\, \mathbf{X}^{\intercal} = \mathbf{U}\, \mathbf{\Sigma}\, \mathbf{U}^{\intercal}$, where $\mathbf{\Sigma} \overset{\text{not}}{=} \mathrm{diag}[\sigma_l]_{l=1}^{\mathcal{N}}$ is the $\mathcal{N} \times \mathcal{N}$ diagonal matrix of the singular values appearing in decreasing order $\sigma_1 > \sigma_2 > \cdots > \sigma_{\mathcal{N}} > 0$, and $\mathbf{U} = [\mathbf{u}_l]_{l=1}^{\mathcal{N}}$ is the $\mathcal{N} \times \mathcal{N}$ orthogonal (or orthonormal) matrix of the singular vectors $\mathbf{u}_l$, we take $\mathbf{A} = \mathbf{U}_{\mathcal{M}} = [\mathbf{u}_l]_{l=1}^{\mathcal{M}}$, yielding $\mathbf{A}^{\dagger} = \mathbf{U}_{\mathcal{M}}^{\intercal}$. As in [6, 7], we consider the scaled orthogonal vectors $\overline{\mathbf{u}}_l = \sigma_l\, \mathbf{u}_l$, in which case we have $\mathbf{A} = \overline{\mathbf{U}}_{\mathcal{M}} = \mathbf{U}_{\mathcal{M}} \mathbf{\Sigma}_{\mathcal{M}}$ and $\mathbf{A}^{\dagger} = \overline{\mathbf{U}}_{\mathcal{M}}^{\dagger} = \mathbf{\Sigma}_{\mathcal{M}}^{-1} \mathbf{U}_{\mathcal{M}}^{\intercal}$, with $\mathbf{\Sigma}_{\mathcal{M}} = \mathrm{diag}[\sigma_l]_{l=1}^{\mathcal{M}}$. We then get

$$y_{kl} = \frac{1}{\sigma_l}\, \mathbf{u}_l^{\intercal} (\mathbf{x}_k - \overline{\mathbf{x}}), \tag{52}$$

and, further,

$$\frac{\partial y_{kl}}{\partial \varsigma_i} = -\frac{1}{\sigma_l^2}\, \mathbf{u}_l^{\intercal} (\mathbf{x}_k - \overline{\mathbf{x}})\, \frac{\partial \sigma_l}{\partial \varsigma_i} + \frac{1}{\sigma_l} \Big(\frac{\partial \mathbf{u}_l}{\partial \varsigma_i} \Big)^{\intercal} (\mathbf{x}_k - \overline{\mathbf{x}}) + \frac{1}{\sigma_l}\, \mathbf{u}_l^{\intercal} \Big(\frac{\partial \mathbf{x}_k}{\partial \varsigma_i} - \frac{\partial \overline{\mathbf{x}}}{\partial \varsigma_i} \Big). \tag{53}$$

To compute $\partial y_{kl} \,/\, \partial \varsigma_i$ we need $\partial \sigma_l \,/\, \partial \varsigma_i$ and $\partial \mathbf{u}_l \,/\, \partial \varsigma_i$. As $\mathbf{C}_{\mathbf{x}}\, \mathbf{u}_l = \sigma_l\, \mathbf{u}_l$ and $\mathbf{u}_l^{\intercal}\, \mathbf{u}_l = 1$, these quantities can be computed by solving the $(\mathcal{N}+1) \times (\mathcal{N}+1)$ system of equations (as in Appendix A)

$$\begin{bmatrix} \mathbf{u}_l & \sigma_l\, \mathbf{I}_{\mathcal{N}} - \mathbf{C}_{\mathbf{x}} \\ 0 & \mathbf{u}_l^{\intercal} \end{bmatrix} \begin{bmatrix} \frac{\partial \sigma_l}{\partial \varsigma_i} \\ \frac{\partial \mathbf{u}_l}{\partial \varsigma_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{C}_{\mathbf{x}}}{\partial \varsigma_i}\, \mathbf{u}_l \\ 0 \end{bmatrix}. \tag{54}$$

The disadvantage of the first method is that the second-order Taylor approximation should be valid for both $f_{\mathtt{I}}(\lambda_k)$ and $f_{\varsigma_i}(\lambda_k)$, while the disadvantage of the second method lies in an increase of the computation time. However, in the second method, if ς_i stands for the cloud optical thickness or the cloud top height, then $\mathbf{x}_k$ does not depend on ς_i, and so, $\partial y_{kl} / \partial \varsigma_i = 0$. Consequently, the system of equations (54) needs not to be solved, and the computation time of the second method is comparable to that of the first method.

In [4], the correlated k-distribution method has been combined with the PCA technique to speed-up the radiance calculations. The same technique is used here for the derivative calculations.

4. Numerical simulations

In this section, we analyze the accuracy and efficiency of the linearized radiative transfer models in computing the derivatives of the measured radiance with respect to the cloud optical thickness $\tau_{\mathtt{c}}$ and the cloud top height $h_{\mathtt{t}}$. The models to be analyzed are the Linearized Discrete Ordinate method with Matrix Exponential (LDOME), the Linearized Matrix Operator method with Matrix Exponential (LMOME), and the Forward-Adjoint approach using the Discrete Ordinate method with Matrix Exponential (FADOME). The derivatives with respect to the cloud geometrical parameters are computed by homogenizing the layers containing the cloud top and bottom heights. The simulations were performed for channel 9 of the EPIC instrument, which is an oxygen A-band absorption channel at 764 nm with a bandwidth of 1.0 nm. As in [4], we consider a water-cloud model with a Gamma size distribution

$$P(a) \propto a^{\alpha} \exp\left[-\alpha \left(\frac{a}{a_{\mathtt{mod}}} \right) \right] \tag{55}$$

of parameters $a_{\mathtt{mod}} = 8$ μm and $\alpha = 6$. The droplet size ranges between 0.02 and 50.0 μm, and the cloud geometrical thickness is $\triangle h = h_{\mathtt{t}} - h_{\mathtt{b}} = 1.5$ km. The atmosphere is discretized with a step of 0.5 km between 0 and 16 km, a step of 2 km between 16 and 20 km, a step of 5 km between 20 and 30 km, and, finally, a step of 10 km between 30 and 50 km. The ground surface is Lambertian with albedo $A = 0.2$, the solar and viewing zenith angles are $\theta_{\mathtt{m}} = \theta_0 = 30°$, and the relative azimuthal angle is $\triangle\varphi = \varphi_{\mathtt{m}} - \varphi_0 = 176°$. In [4], it was found that a compromise between an accurate description of the scattering in the backward direction and computation time can be reached for a number of discrete ordinates M in the range $24 \leq M \leq 32$. For this reason, we fix the number of discrete ordinates to 32. The radiances are solar-flux normalized, and the delta-M scaling together with the TMS correction is used. The simulations were performed on a server Intel(R) Xeon(R) CPU E5-2695 v3 @ 2.30 GHz using up to 56 threads.

In Fig. 2, we plot the derivatives with respect to $\tau_{\mathtt{c}}$ for $h_{\mathtt{t}} = 4$ km, and with respect to $h_{\mathtt{t}}$ for $\tau_{\mathtt{c}} = 5$. The results are computed by using line-by-line (LBL) calculations for the oxygen absorption cross sections as in [31, 32]. The agreement between LDOME and LMOME is perfect. The results corresponding to FADOME show small deviations from LDOME results; in general, the relative errors are larger when the derivative values are smaller, but they are below 10^{-4} for $\partial I_{\mathtt{m}} / \partial \tau_{\mathtt{c}}$ and below 10^{-3} for $\partial I_{\mathtt{m}} / \partial h_{\mathtt{t}}$. Note that, in atmospheric remote sensing, due to the general nature of the Gauss-Newton iterative method commonly used to solve the nonlinear inverse problem, the accuracies of the weighting functions up to a few percent do not deteriorate the convergence rate of the solution, and the application of the adjoint approach is not critical. Therefore, an accuracy better than 10^{-3} in the derivatives when using FADOME in conjunction with LBL calculations is more than satisfactory.

In Fig. 3, we show the relative errors when LBL calculations are replaced with the correlated k-distribution method, the PCA technique, and the correlated k-distribution plus PCA method.

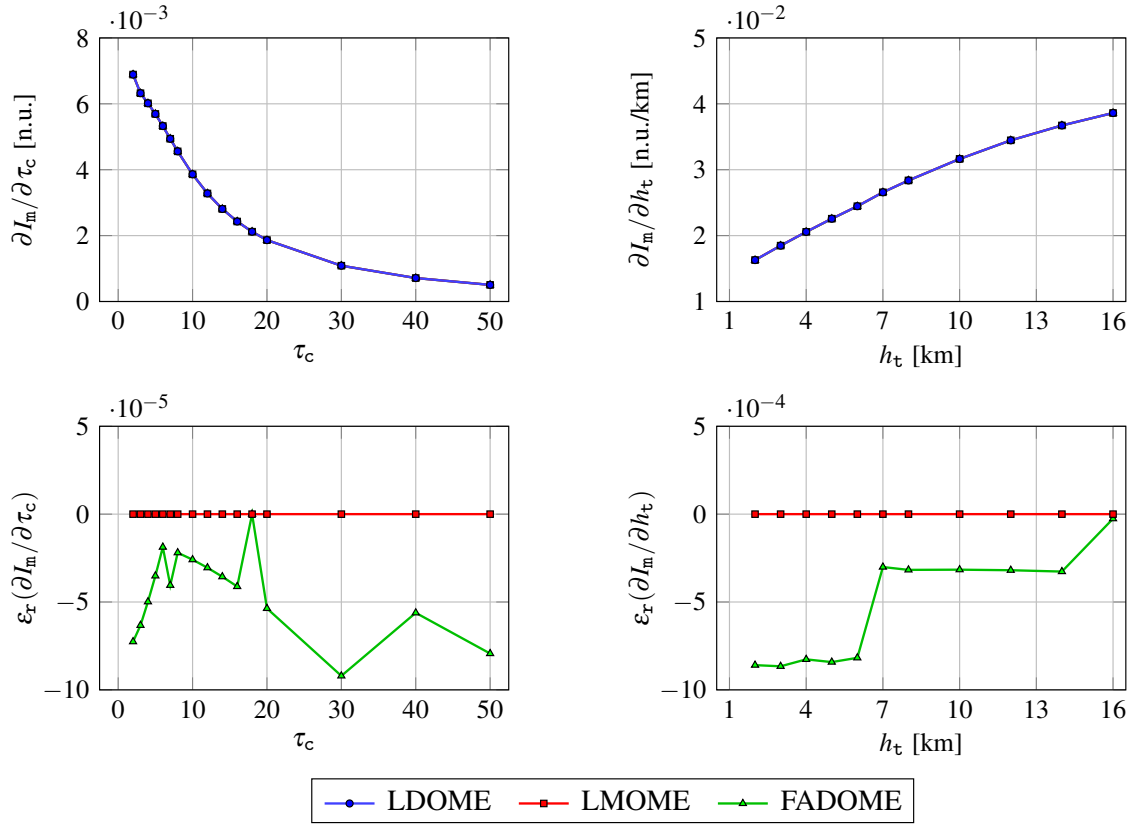


Figure 2: Upper panels: Radiance derivatives (n.u. denotes Sun-normalized units) with respect to τ_c for $h_t = 4$ km (upper-left panel) and with respect to h_t for $\tau_c = 5$ (upper-right panel). Lower panels: Relative errors in the derivatives when using the linearized DOME approach as a reference.

PCA is used with two principal components and one wavelength bin, while the method based on the linearization of the restoration equation (45) is used for derivative calculations. The general observation is that the relative errors in $\partial I_m / \partial \tau_c$ are smaller than $5 \cdot 10^{-3}$, and that the relative errors in $\partial I_m / \partial h_t$ are smaller than 10^{-2}. Indeed, the relative errors of the correlated k-distribution plus PCA method are smaller than $4 \cdot 10^{-3}$ for $\partial I_m/\partial \tau_c$ in the range $2 \leq \tau_c \leq 50$, and smaller than $7 \cdot 10^{-3}$ for $\partial I_m/\partial h_t$ in the range $2 \leq h_t \leq 16$ km.

In Table 1 and Table 2, we show the computation times (i.e. the accumulated time over all threads) for the simulations with the linearized radiative transfer models and various acceleration techniques. It is apparent that FADOME is the fastest method, followed by LDOME, equally whether LBL calculations or acceleration techniques are used. The best time performance is obtained when FADOME is combined with the correlated k-distribution plus PCA method. In this case, and taking into account that the results in Table 1 correspond to 16 values of τ_c, and the results in Table 2 correspond to 11 values of h_t, the average time to simultaneously compute $\partial I_m/\partial \tau_c$ and $\partial I_m/\partial h_t$ at (τ_c, h_t) is 13 s.

5. Conclusions

Several linearized radiative transfer models have been analyzed with respect to their applicability to the retrieval of cloud parameters from EPIC measurements. The models under consideration are the linearized discrete ordinate method with matrix exponential, the linearized matrix operator method with matrix exponential, and a forward-adjoint approach based on the discrete ordinate method with matrix exponential.

The numerical simulations show that these three models provide similar results when computing the derivatives with respect to the cloud optical thickness and the cloud top height. Furthermore, the forward-adjoint approach based on the discrete ordinate method with matrix exponential, and using the correlated k-distribution method in conjunction with the PCA technique, is an accurate and efficient tool for the offline retrieval of cloud optical thickness and cloud top height from EPIC measurements, with a speed-up factor of 2 when compared to the conventional linearization approaches. This forward-adjoint radiative transfer model can be combined with a linearized Mie or T-matrix code [33] to retrieve microphysical properties of clouds and aerosols.

6. Acknowledgements

This work was funded by the programme DLR/DAAD Research Fellowships 2015 (57186656), with reference numbers 91613528 and 91627488, organized by the German Academic Exchange Service (DAAD) and the German Aerospace Center (DLR).

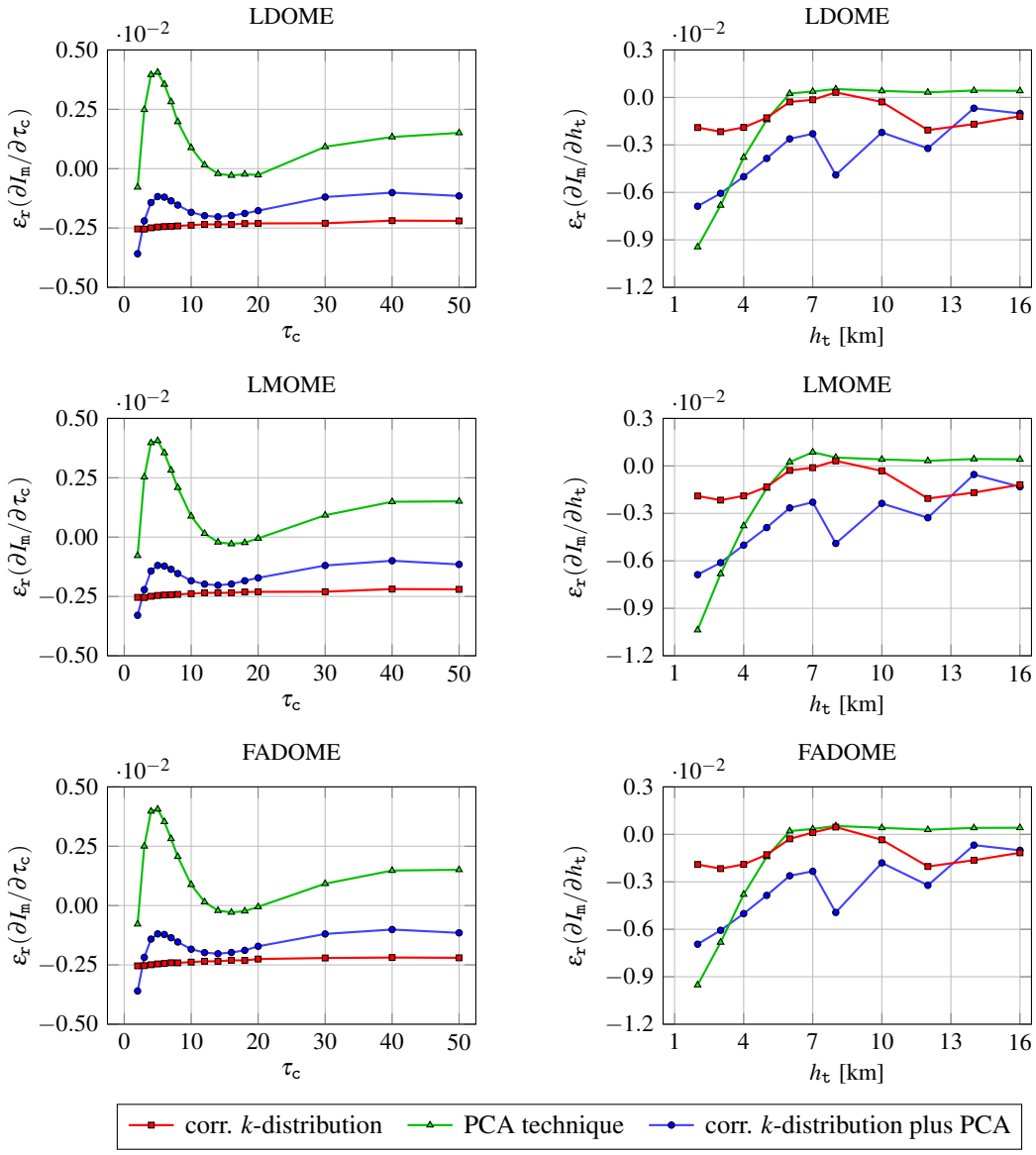


Figure 3: Relative errors in $\partial I_m/\partial \tau_c$ (left panels) and $\partial I_m/\partial h_t$ (right panels) when comparing LBL-based simulations with those for correlated k-distribution, PCA technique, and correlated k-distribution plus PCA.

Table 1: Computation times in minutes to simulate the derivatives with respect to τ_c for different models (LDOME, LMOME, FADOME) and acceleration techniques. The speed-up factor corresponds to the ratio between LDOME and FADOME.

Acceleration techniques	**Linearized models**			**Speed-up factor**
	LDOME	LMOME	FADOME	
LBL calculation	3625	4602	1804	2.0
Correlated k-distribution	678	869	335	2.0
PCA technique	8	9.5	6	1.3
Corr. k-distribution plus PCA technique	6	7.5	3.5	1.7

Table 2: Computation times in minutes to simulate the derivatives with respect to h_t for different models (LDOME, LMOME, FADOME) and acceleration techniques. The speed-up factor corresponds to the ratio between LDOME and FADOME.

Acceleration techniques	**Linearized models**			**Speed-up factor**
	LDOME	LMOME	FADOME	
LBL calculation	2451	3079	1240	2.0
Correlated k-distribution	456	574	224	2.0
PCA technique	5.5	6.5	4.5	1.2
Corr. k-distribution plus PCA technique	4	5	2.5	1.6

Appendix A

In this appendix we compute the partial derivatives of the inverse of the eigenvector matrix $\mathbf{V}_{mj}^{-1}$ and the eigenvalues λ_k.

The steps for computing an eigensystem of the matrix $\mathbf{A}_{mj}$ can be summarized as follows:

1. Compute
$$\mathcal{A}_+ = \mathbf{A}_-\mathbf{A}_+, \tag{56}$$
where $\mathbf{A}_+ = \mathbf{A}_{mj}^{11}+\mathbf{A}_{mj}^{12}$ and $\mathbf{A}_- = \mathbf{A}_{mj}^{11}-\mathbf{A}_{mj}^{12}$, and determine an eigensystem $\{\mu_k, \mathbf{w}_k^+\}_{k=1}^M$ of the matrix $\mathcal{A}_+$.
2. Normalize the vectors $\mathbf{w}_k^+$ for $k = 1, ..., M$.
3. Compute the eigenvectors of the matrix $\mathcal{A}_- = \mathbf{A}_+\mathbf{A}_-$,
$$\mathbf{w}_k^- = \frac{1}{\lambda_k}\mathbf{A}_+\mathbf{w}_k^+, \quad k = 1, \ldots, M, \tag{57}$$
where $\lambda_k = \sqrt{\mu_k}$ are the positive eigenvalues of the matrix $\mathbf{A}_{mj}$.
4. Set $\mathbf{v}_k^+ = (\mathbf{w}_k^+ + \mathbf{w}_k^-)\ /\ 2$ and $\mathbf{v}_k^- = (\mathbf{w}_k^+ - \mathbf{w}_k^-)\ /\ 2$ for $k = 1, \ldots, M$.
5. Construct the eigenvectors of $\mathbf{A}_{mj}$ as
$$\overline{\mathbf{v}}_k^+ = \begin{bmatrix} \mathbf{v}_k^+ \\ \mathbf{v}_k^- \end{bmatrix}, \quad \overline{\mathbf{v}}_k^- = \begin{bmatrix} \mathbf{v}_k^- \\ \mathbf{v}_k^+ \end{bmatrix}, \quad k = 1, \ldots, M. \tag{58}$$

The spectral decomposition of the matrix $\mathbf{A}_{mj}$ is then

$$\mathbf{A}_{mj} = \mathbf{V}_{mj}\begin{bmatrix} \mathbf{\Lambda}_{mj} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda}_{mj} \end{bmatrix}\mathbf{V}_{mj}^{-1}, \tag{59}$$

with

$$\mathbf{V}_{m,j} = [\,\overline{\mathbf{v}}_1^+, \ldots, \overline{\mathbf{v}}_M^+, \overline{\mathbf{v}}_1^-, \ldots, \overline{\mathbf{v}}_M^-\,] \tag{60}$$

and

$$\Lambda_{mj} = \mathrm{diag}[\lambda_1, \ldots, \lambda_M]. \tag{61}$$

To compute $\partial\mathbf{V}_{mj}^{-1}/\partial\varsigma_i$ and $\partial\lambda_k/\partial\varsigma_i$, we follow the exposition given in [11]. Considering the eigenvalue problem for the matrix $\mathcal{A}_+$, i.e.

$$\mathcal{A}_+\mathbf{w}_k^+ = \mu_k\mathbf{w}_k^+, \tag{62}$$

and taking the derivative with respect to ς_i, we obtain

$$\frac{\partial\mathcal{A}_+}{\partial\varsigma_i}\,\mathbf{w}_k^+ + \mathcal{A}_+\,\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i} = \frac{\partial\mu_k}{\partial\varsigma_i}\,\mathbf{w}_k^+ + \mu_k\,\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i}. \tag{63}$$

Equation (63) is a system of M equations with $M+1$ unknowns: the scalar $\partial\mu_k/\partial\varsigma_i$ and the vector $\partial\mathbf{w}_k^+/\partial\varsigma_i$. Since the eigenvectors $\mathbf{w}_k^+$ are normalized, we derive an additional equation

$$\mathbf{w}_k^{+\mathsf{T}}\,\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i} = 0, \tag{64}$$

which yields the compatibility of the system of equations. By (63) and (64), the resulting system of equations can be written in matrix form as

$$\begin{bmatrix} \mathbf{w}_k^+ & \mu_k\,\mathbf{I} - \mathcal{A}_+ \\ 0 & \mathbf{w}_k^{+\mathsf{T}} \end{bmatrix}\begin{bmatrix} \frac{\partial\mu_k}{\partial\varsigma_i} \\ \frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial\mathcal{A}_+}{\partial\varsigma_i}\,\mathbf{w}_k^+ \\ 0 \end{bmatrix}. \tag{65}$$

It is important to observe that the we can solve the above system of equations for all atmospheric parameters ς_i, $i = 1, \ldots, N_\mathrm{p}$, that is, we can solve the matrix equation

$$\begin{bmatrix} \mathbf{w}_k^+ & \mu_k\,\mathbf{I} - \mathcal{A}_+ \\ 0 & \mathbf{w}_k^{+\mathsf{T}} \end{bmatrix}\begin{bmatrix} \frac{\partial\mu_k}{\partial\varsigma_1} & \cdots & \frac{\partial\mu_k}{\partial\varsigma_{N_\mathrm{p}}} \\ \frac{\partial\mathbf{w}_k^+}{\partial\varsigma_1} & \cdots & \frac{\partial\mathbf{w}_k^+}{\partial\varsigma_{N_\mathrm{p}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial\mathcal{A}_+}{\partial\varsigma_1}\,\mathbf{w}_k^+ & \cdots & \frac{\partial\mathcal{A}_+}{\partial\varsigma_{N_\mathrm{p}}}\,\mathbf{w}_k^+ \\ 0 & \cdots & 0 \end{bmatrix}. \tag{66}$$

If $\partial\mu_k/\partial\varsigma_i$ is known, the partial derivative of $\lambda_k = \sqrt{\mu_k}$ with respect to ς_i follows immediately as

$$\frac{\partial\lambda_k}{\partial\varsigma_i} = \frac{1}{2\,\lambda_k}\,\frac{\partial\mu_k}{\partial\varsigma_i}. \tag{67}$$

To compute the partial derivative of $\mathbf{w}_k^-$, we use definition (57) and apply the chain rule to obtain

$$\frac{\partial\mathbf{w}_k^-}{\partial\varsigma_i} = \Big(-\frac{1}{\lambda_k^2}\,\frac{\partial\lambda_k}{\partial\varsigma_i}\Big)\,\mathbf{A}_+\,\mathbf{w}_k^+ + \frac{1}{\lambda_k}\,\frac{\partial\mathbf{A}_+}{\partial\varsigma_i}\,\mathbf{w}_k^+ + \frac{1}{\lambda_k}\,\mathbf{A}_+\,\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i}. \tag{68}$$

Further calculations give

$$\frac{\partial\mathbf{v}_k^+}{\partial\varsigma_i} = \frac{1}{2}\Big(\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i} + \frac{\partial\mathbf{w}_k^-}{\partial\varsigma_i}\Big), \quad \frac{\partial\mathbf{v}_k^-}{\partial\varsigma_i} = \frac{1}{2}\Big(\frac{\partial\mathbf{w}_k^+}{\partial\varsigma_i} - \frac{\partial\mathbf{w}_k^-}{\partial\varsigma_i}\Big), \tag{69}$$

and

$$\frac{\partial\overline{\mathbf{v}}_k^+}{\partial\varsigma_i} = \begin{bmatrix} \frac{\partial\mathbf{v}_k^+}{\partial\varsigma_i} \\ \frac{\partial\mathbf{v}_k^-}{\partial\varsigma_i} \end{bmatrix}, \quad \frac{\partial\overline{\mathbf{v}}_k^-}{\partial\varsigma_i} = \begin{bmatrix} \frac{\partial\mathbf{v}_k^-}{\partial\varsigma_i} \\ \frac{\partial\mathbf{v}_k^+}{\partial\varsigma_i} \end{bmatrix}. \tag{70}$$

Using now the definition of $\mathbf{V}_{mj}$ we obtain

$$\frac{\partial\mathbf{V}_{mj}}{\partial\varsigma_i} = \Big[\frac{\partial\overline{\mathbf{v}}_1^+}{\partial\varsigma_i}, \ldots, \frac{\partial\overline{\mathbf{v}}_M^+}{\partial\varsigma_i}, \frac{\partial\overline{\mathbf{v}}_1^-}{\partial\varsigma_i}, \ldots, \frac{\partial\overline{\mathbf{v}}_M^-}{\partial\varsigma_i}\Big]; \tag{71}$$

whence, taking into account that $\mathbf{V}_{mj}\,\mathbf{V}_{mj}^{-1} = \mathbf{I}$, we end up with

$$\frac{\partial\mathbf{V}_{mj}^{-1}}{\partial\varsigma_i} = -\mathbf{V}_{mj}^{-1}\,\frac{\partial\mathbf{V}_{mj}}{\partial\varsigma_i}\,\mathbf{V}_{mj}^{-1}. \tag{72}$$

Appendix B

In this appendix, we perform the integration over the azimuthal angle in (37) and (38), and discuss the integration over the radial coordinate.

We begin by considering the cosine-azimuthal expansions

$$I_\mathrm{d}(r, \mathbf{\Omega}) = \sum_{m=0}^{2M-1} I_m(r, \mu)\cos[m(\varphi - \varphi_0)], \tag{73}$$

and

$$\widehat{I}_\mathrm{d}^{\uparrow}(r, \mathbf{\Omega}) = \sum_{m=0}^{2M-1} \widehat{I}_m^{\uparrow}(r, \mu)\cos[m(\varphi - \widehat{\varphi}_\mathtt{m})], \tag{74}$$

yielding

$$\widehat{I}_\mathrm{d}^{\uparrow}(r, -\mathbf{\Omega}) = \sum_{m=0}^{2M-1} \widehat{I}_m^{\uparrow}(r, -\mu)\cos[m(\varphi - \varphi_\mathtt{m})]. \tag{75}$$

Further, we write

$$\sigma_{\mathtt{sct}}(r)\,P(r,\mathbf{\Omega},\mathbf{\Omega}')=\sum_{m=0}^{2M-1}(2-\delta_{m0})s_m(r,\mu,\mu')\cos[m(\varphi-\varphi')], \quad (76)$$

where (cf. (16) and (17))

$$s_m(r,\mu,\mu')=\sigma_{\mathtt{sct}}(r)\,p_m(r,\mu,\mu')=\sum_{n=m}^{2M-1}\xi_n(r)\,P_n^m(\mu)\,P_n^m(\mu'). \quad (77)$$

Inserting (73), (75), and (76) in (36)-(38), and using the orthogonality relations of the Fourier cosine basis functions, we obtain

$$\frac{\partial I_{\mathtt{m}}}{\partial\varsigma_i}=\sum_{m=0}^{2M-1}(T_{1m}+T_{2m})\cos[m(\varphi_{\mathtt{m}}-\varphi_0)], \quad (78)$$

where

$$\begin{aligned}T_{1m}=&-(1+\delta_{m0})\,\pi\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\int_{-1}^{1}\frac{\partial\sigma_{\mathtt{ext}}}{\partial\varsigma_i}(r)\,\widehat{I}_m^{\dagger}(r,-\mu)\,I_m(r,\mu)\,\mathrm{d}r\mathrm{d}\mu\Big]\\&-F_0\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\frac{\partial\sigma_{\mathtt{ext}}}{\partial\varsigma_i}(r)\,\widehat{I}_m^{\dagger}(r,\mu_0)\,T(r)\;\mathrm{d}r\Big]\\&+\frac{1}{2}(1+\delta_{m0})\,\pi\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\int_{-1}^{1}\widehat{I}_m^{\dagger}(r,-\mu)\,\mathrm{d}r\mathrm{d}\mu\int_{-1}^{1}\frac{\partial s_m}{\partial\varsigma_i}(r,\mu,\mu')\,I_m(r,\mu')\,\mathrm{d}\mu'\Big]\\&+\frac{F_0}{2}\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\int_{-1}^{1}\frac{\partial s_m}{\partial\varsigma_i}(r,\mu,-\mu_0)\,\widehat{I}_m^{\dagger}(r,-\mu)\,T(r)\;\mathrm{d}r\mathrm{d}\mu\Big] \quad (79)\end{aligned}$$

and

$$\begin{aligned}T_{2m}=&-\widehat{F}_0\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\frac{\partial\sigma_{\mathtt{ext}}}{\partial\varsigma_i}(r)\,I_m(r,\mu_{\mathtt{m}})\,\widehat{T}^{\dagger}(r)\;\mathrm{d}r\Big]\\&+\frac{\widehat{F}_0}{2}\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\widehat{T}^{\dagger}(r)\;\mathrm{d}r\int_{-1}^{1}\frac{\partial s_m}{\partial\varsigma_i}(r,\mu_{\mathtt{m}},\mu')\,I_m(r,\mu')\,\mathrm{d}\mu'\Big]\\&+\frac{\widehat{F}_0F_0}{4\pi}(2-\delta_{m0})\Big[\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\frac{\partial s_m}{\partial\varsigma_i}(r,\mu_{\mathtt{m}},-\mu_0)\,T(r)\,\widehat{T}^{\dagger}(r)\;\mathrm{d}r\Big]. \quad (80)\end{aligned}$$

To perform the integration over the radial coordinate in (79) and (80), we assume for simplicity that the partial derivatives $\partial\sigma_{\mathtt{ext}\,j}/\partial\varsigma_i$ and $\partial\xi_{n,j}/\partial\varsigma_i$ are nonzero for all layers $j\in D_i$, where D_i is a subset of $\{1,\ldots,N-1\}$. In this case, the integration with respect to the radial coordinate reduces to an integration over all layers in D_i; we have

$$\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}}\mathrm{d}r=\sum_{j\in D_i}\int_{r_{j+1}}^{r_j}\mathrm{d}r.$$

For the azimuthal mode m, let $I_{mj}(\rho,\pm\mu_k)$ and $\widehat{I}_{mj}^{\dagger}(\rho,\pm\mu_k)$ be, respectively, the forward and conjugate adjoint radiances at an internal point ρ in the layer j ($0\le\rho\le\triangle r_j$). Using the analytic representation of the radiance at an internal layer point as given in [23], the integrals

$$\begin{aligned}\mathcal{I}_{mj}(\mu)&=\int_0^{\triangle r_j}I_{mj}(\rho,\mu)\,\widehat{T}^{\dagger}(\rho)\,\mathrm{d}\rho\\&=\int_0^{\triangle r_j}\mathrm{e}^{-[\frac{\rho}{\triangle r_j}\tau^{\mathtt{m}}_{\mathtt{ext}\,j}+(1-\frac{\rho}{\triangle r_j})\tau^{\mathtt{m}}_{\mathtt{ext}\,j+1}]}\,I_{mj}(\rho,\mu)\,\mathrm{d}\rho, \quad (81)\\\widehat{\mathcal{I}}_{mj}(\mu)&=\int_0^{\triangle r_j}\widehat{I}_{mj}^{\dagger}(\rho,\mu)\,T(\rho)\,\mathrm{d}\rho\\&=\int_0^{\triangle r_j}\mathrm{e}^{-[\frac{\rho}{\triangle r_j}\tau^{0}_{\mathtt{ext}\,j}+(1-\frac{\rho}{\triangle r_j})\tau^{0}_{\mathtt{ext}\,j+1}]}\,\widehat{I}_{mj}^{\dagger}(\rho,\mu)\,\mathrm{d}\rho, \quad (82)\\\mathcal{J}_{mj}(\mu,\mu')&=\int_0^{\triangle r_j}\widehat{I}_{mj}^{\dagger}(\rho,\mu)\,I_{mj}(\rho,\mu')\,\mathrm{d}\rho, \quad (83)\end{aligned}$$

for $\mu=\pm\mu_k$, $\mu'=\pm\mu_l$, $k,l=1,\ldots,M$, as well as the integral

$$\begin{aligned}\mathcal{T}_j&=\int_0^{\triangle r_j}T(\rho)\,\widehat{T}^{\dagger}(\rho)\;\mathrm{d}\rho\\&=\int_0^{\triangle r_j}\mathrm{e}^{-[\frac{\rho}{\triangle r_j}(\tau^{0}_{\mathtt{ext}\,j}+\tau^{\mathtt{m}}_{\mathtt{ext}\,j})+(1-\frac{\rho}{\triangle r_j})(\tau^{0}_{\mathtt{ext}\,j+1}+\tau^{\mathtt{m}}_{\mathtt{ext}\,j+1})]}\,\mathrm{d}\rho, \quad (84)\end{aligned}$$

can be computed analytically. Here, $\tau^0_{\mathtt{ext}\,j}$ and $\tau^0_{\mathtt{ext}\,j+1}$ are the optical depths along the geometrical characteristic $\mathbf{\Omega}_0=(-\mu_0,\varphi_0)$ at the boundary levels j and $j+1$, respectively, while $\tau^{\mathtt{m}}_{\mathtt{ext}\,j}$ and $\tau^{\mathtt{m}}_{\mathtt{ext}\,j+1}$ are the optical depths along the geometrical characteristic $\widehat{\mathbf{\Omega}}_{\mathtt{m}}=(\widehat{\mu}_{\mathtt{m}},\widehat{\varphi}_{\mathtt{m}})=(-\mu_{\mathtt{m}},\varphi_{\mathtt{m}}+\pi)$ at the boundary levels j and $j+1$, respectively. We end up with

$$\begin{aligned}T_{1m}=&-(1+\delta_{m0})\,\pi\sum_{j\in D_i}T^a_{1mj}-F_0\sum_{j\in D_i}T^b_{1mj}\\&-\frac{1}{2}(1+\delta_{m0})\,\pi\sum_{j\in D_i}T^c_{1mj}+\frac{F_0}{2}\sum_{j\in D_i}T^d_{1mj}, \quad (85)\\T_{2m}=&-\widehat{F}_0\sum_{j\in D_i}T^a_{2mj}+\frac{\widehat{F}_0}{2}\sum_{j\in D_i}T^b_{2mj}\\&+\frac{\widehat{F}_0F_0}{4\pi}(2-\delta_{m0})\sum_{j\in D_i}T^c_{2mj}, \quad (86)\end{aligned}$$

where

$$\begin{aligned}T^a_{1mj}&=\frac{\partial\sigma_{\mathtt{ext}\,j}}{\partial\varsigma_i}\sum_k w_k\,[\mathcal{J}_{mj}(-\mu_k,\mu_k)+\mathcal{J}_{mj}(\mu_k,-\mu_k)], \quad (87)\\T^b_{1mj}&=\frac{\partial\sigma_{\mathtt{ext}\,j}}{\partial\varsigma_i}\widehat{\mathcal{I}}_{mj}(\mu_0), \quad (88)\\T^c_{1mj}&=\sum_l\sum_k w_l\,w_k\Big[\frac{\partial s_{mj}}{\partial\varsigma_i}(\mu_l,\mu_k)\,\mathcal{J}_{mj}(-\mu_l,\mu_k)\\&\quad+\frac{\partial s_{mj}}{\partial\varsigma_i}(\mu_l,-\mu_k)\,\mathcal{J}_{mj}(-\mu_l,-\mu_k)+\frac{\partial s_{mj}}{\partial\varsigma_i}(-\mu_l,\mu_k)\,\mathcal{J}_{mj}(\mu_l,\mu_k)\\&\quad+\frac{\partial s_{mj}}{\partial\varsigma_i}(-\mu_l,-\mu_k)\,\mathcal{J}_{mj}(\mu_l,-\mu_k)\Big], \quad (89)\\T^d_{1mj}&=\sum_k w_k\Big[\frac{\partial s_{mj}}{\partial\varsigma_i}(\mu_k,-\mu_0)\,\widehat{\mathcal{I}}_{mj}(-\mu_k)\\&\quad+\frac{\partial s_{mj}}{\partial\varsigma_i}(-\mu_k,-\mu_0)\,\widehat{\mathcal{I}}_{mj}(\mu_k)\Big], \quad (90)\end{aligned}$$

and

$$T^a_{2mj} = \frac{\partial \sigma_{\mathtt{ext}\,j}}{\partial \varsigma_i}\, \mathcal{I}_{mj}(\mu_{\mathtt{m}}), \tag{91}$$

$$T^b_{2mj} = \sum_k w_k \Big[\frac{\partial s_{mj}}{\partial \varsigma_i}(\mu_{\mathtt{m}}, \mu_k)\, \mathcal{I}_{mj}(\mu_k) + \frac{\partial s_{mj}}{\partial \varsigma_i}(\mu_{\mathtt{m}}, -\mu_k)\, \mathcal{I}_{mj}(-\mu_k) \Big], \tag{92}$$

$$T^c_{2mj} = \frac{\partial s_{mj}}{\partial \varsigma_i}(\mu_{\mathtt{m}}, -\mu_0)\, \mathcal{T}_j. \tag{93}$$

The integrals $\widehat{\mathcal{I}}_{mj}(\mu_0)$ and $\mathcal{I}_{mj}(\mu_{\mathtt{m}})$, which enter the expressions of T^b_{1mj} and T^a_{2mj}, respectively, can be computed by using the method of false discrete ordinate (i.e. the false discrete ordinates μ_0 and $\mu_{\mathtt{m}}$ with zero weights are added to the set $\{\mu_k\}_{k=1}^M$).

Appendix C

In this appendix, we present the first discretization method of Section 3.3, which consists of the discretization of the atmosphere above, below and in the cloud. The linearization method pertinent to this discretization scheme is the forward-adjoint approach. Let us consider an homogeneous cloud with top height $h_{\mathtt{t}}$ and bottom height $h_{\mathtt{b}}$, and let the optical properties of the homogeneous cloud be described by the cloud extinction coefficient $\sigma^{\mathtt{c}}_{\mathtt{ext}}$ and the expansion coefficients $\xi^{\mathtt{c}}_n$, so that $s^{\mathtt{c}}_m(\mu,\mu') = \sum_{n=m}^{2M-1} \xi^{\mathtt{c}}_n\, P^m_n(\mu)\, P^m_n(\mu')$. Set

$$\sigma_{\mathtt{ext}}(r) = \sigma^0_{\mathtt{ext}}(r) + H(h_{\mathtt{t}} - r)\, H(r - h_{\mathtt{b}})\, \sigma^{\mathtt{c}}_{\mathtt{ext}}, \tag{94}$$

$$s_m(r,\mu,\mu') = s^0_m(r,\mu,\mu') + H(h_{\mathtt{t}} - r)\, H(r - h_{\mathtt{b}})\, s^{\mathtt{c}}_m(\mu,\mu'), \tag{95}$$

where the superscript "0" refers to the clear sky atmosphere, so that from

$$\frac{\partial H(h_{\mathtt{t}} - r)}{\partial h_{\mathtt{t}}} = \delta(h_{\mathtt{t}} - r) \tag{96}$$

we get

$$\frac{\partial \sigma_{\mathtt{ext}}}{\partial h_{\mathtt{t}}}(r) = \delta(h_{\mathtt{t}} - r)\, H(r - h_{\mathtt{b}})\, \sigma^{\mathtt{c}}_{\mathtt{ext}}, \tag{97}$$

$$\frac{\partial s_m}{\partial h_{\mathtt{t}}}(r,\mu,\mu') = \delta(h_{\mathtt{t}} - r)\, H(r - h_{\mathtt{b}})\, s^{\mathtt{c}}_m(\mu,\mu'). \tag{98}$$

Inserting (97) and (98) in (79) and (80), we find, for example, that the first integral in (79) is given by

$$\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}} \int_{-1}^{1} \frac{\partial \sigma_{\mathtt{ext}}(r)}{\partial \varsigma_i}\, \widehat{I}^{\dagger}_m(r,-\mu)\, I_m(r,\mu)\, \mathrm{d}r\mathrm{d}\mu = \int_{-1}^{1} \sigma^{\mathtt{c}}_{\mathtt{ext}}\, \widehat{I}^{\dagger}_m(h_{\mathtt{t}},-\mu)\, I_m(h_{\mathtt{t}},\mu)\, \mathrm{d}\mu \tag{99}$$

and that the third integral in (79) is given by

$$\int_{r_{\mathtt{s}}}^{r_{\mathtt{TOA}}} \int_{-1}^{1} \widehat{I}^{\dagger}_m(r,-\mu)\, \mathrm{d}r\mathrm{d}\mu \int_{-1}^{1} \frac{\partial s_m}{\partial \varsigma_i}(r,\mu,\mu')\, I_m(r,\mu')\, \mathrm{d}\mu' = \int_{-1}^{1} \widehat{I}^{\dagger}_m(h_{\mathtt{t}},-\mu)\, \mathrm{d}\mu \int_{-1}^{1} s^{\mathtt{c}}_m(\mu,\mu')\, I_m(h_{\mathtt{t}},\mu')\, \mathrm{d}\mu'. \tag{100}$$

For the derivatives with respect to the cloud bottom height $h_{\mathtt{b}}$, we use

$$\frac{\partial H(r - h_{\mathtt{b}})}{\partial h_{\mathtt{b}}} = -\delta(r - h_{\mathtt{b}}) \tag{101}$$

and proceed analogously.

References

[1] A. Marshak, Y. Knyazikhin, The spectral invariant approximation within canopy radiative transfer to support the use of the EPIC/DSCOVR oxygen B-band for monitoring vegetation, Journal of Quantitative Spectroscopy and Radiative Transfer 191 (2017) 7–12. doi:10.1016/j.jqsrt.2017.01.015

[2] X. Xu, J. Wang, Y. Wang, J. Zeng, O. Torres, Y. Yang, A. Marshak, J. Reid, S. Miller, Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point, Geophysical Research Letters 44 (14) (2017) 7544–7554. doi:10.1002/2017gl073939

[3] Y. Yang, A. Marshak, J. Mao, A. Lyapustin, J. Herman, A method of retrieving cloud top height and cloud geometrical thickness with oxygen A and B bands for the Deep Space Climate Observatory (DSCOVR) mission: Radiative transfer simulations, Journal of Quantitative Spectroscopy and Radiative Transfer 122 (2013) 141–149. doi:10.1016/j.jqsrt.2012.09.017

[4] V. Molina García, S. Sasi, D. Efremenko, A. Doicu, D. Loyola, Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements, Journal of Quantitative Spectroscopy and Radiative Transfer 123 (2018) 228–240. doi:10.1016/j.jqsrt.2018.03.014

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

[6] V. Natraj, X. Jiang, R. Shia, X. Huang, J. Margolis, Y. Yung, 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) (2005) 539–556. doi:10.1016/j.jqsrt.2004.12.024

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

[8] R. Spurr, V. Natraj, C. Lerot, V. Roozendael, 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 (2013) 1–17. doi:10.1016/j.jqsrt.2013.04.002

[9] D. Efremenko, A. Doicu, D. Loyola, 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 (2014) 128–135. doi:10.1016/j.jqsrt.2013.07.023

[10] R. Spurr, T. Kurosu, K. Chance, A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, Journal of Quantitative Spectroscopy and Radiative Transfer 68 (6) (2001) 689–735. doi:10.1016/s0022-4073(00)00055-8

[11] R. Spurr, Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, Journal of Quantitative Spectroscopy and Radiative Transfer 75 (2) (2002) 129–175. doi:10.1016/s0022-4073(01)00245-x

[12] 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. Kokhanovsky (Ed.), Light scattering reviews, Vol. 3, 2008, pp. 229–275. doi:10.1007/978-3-540-48546-9_7

[13] R. Spurr, M. Christi, Linearization of the interaction principle: Analytic jacobians in the "radiant model", Journal of Quantitative Spectroscopy and Radiative Transfer 103 (3) (2007) 431–446. doi:10.1016/j.jqsrt.2006.05.001

[14] G. Marchuk, Equation for the value of information from weather satellites and formulation of inverse problems, Cosmic Research (3) (1964) 462–477, [in Russian].

[15] G. Marchuk, Adjoint equations and analysis of complex systems, Springer Nature, 1995. doi:10.1007/978-94-017-0621-6

[16] M. Box, Radiative perturbation theory: A review, Environmental Modelling & Software 17 (1) (2002) 95–106. doi:10.1016/s1364-8152(01)00056-1

[17] E. Ustinov, Adjoint sensitivity analysis of radiative transfer equation: temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR, Journal of Quantitative Spectroscopy and Radiative Transfer 68 (2) (2001) 195–211. doi:10.1016/s0022-4073(00)00022-4

[18] E. Ustinov, Atmospheric weighting functions and surface partial derivatives for remote sensing of scattering planetary atmospheres in thermal spectral region: General adjoint approach, Journal of Quantitative Spectroscopy and Radiative Transfer 92 (3) (2005) 351–371. doi:10.1016/j.jqsrt.2004.08.003

[19] V. Rozanov, A. Rozanov, Relationship between different approaches to derive weighting functions related to atmospheric remote sensing problems, Journal of Quantitative Spectroscopy and Radiative Transfer 105 (2) (2007) 217–242. doi:10.1016/j.jqsrt.2006.12.006

[20] A. Doicu, T. Trautmann, Adjoint problem of radiative transfer for a pseudo-spherical atmosphere and general viewing geometries, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (8) (2009) 464–476. doi:10.1016/j.jqsrt.2009.01.027

[21] J. Landgraf, O. Hasekamp, M. Box, T. Trautmann, A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory approach, Journal of Geophysical Research: Atmospheres 106 (D21) (2001) 27291–27305. doi:10.1029/2001jd000636

[22] A. Doicu, T. Trautmann, Two linearization methods for atmospheric remote sensing, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (8) (2009) 477–490. doi:10.1016/j.jqsrt.2009.02.001

[23] A. Doicu, 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) (2009) 146–158. doi:10.1016/j.jqsrt.2008.09.014

[24] A. Doicu, T. Trautmann, Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Vector case, Journal of Quantitative Spectroscopy and Radiative Transfer 110 (1-2) (2009) 159–172. doi:10.1016/j.jqsrt.2008.09.013

[25] Q. Liu, F. Weng, Advanced doubling-adding method for radiative transfer in planetary atmospheres, Journal of the Atmospheric Sciences 63 (12) (2006) 3459–3465. doi:10.1175/jas3808.1

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

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

[28] K. Evans, G. Stephens, A new polarized atmospheric radiative transfer model, Journal of Quantitative Spectroscopy and Radiative Transfer 46 (5) (1991) 413–423. doi:10.1016/0022-4073(91)90043-p

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

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

[31] F. Schreier, S. Gimeno García, P. Hedelt, M. Hess, J. Mendrok, M. Vasquez, J. Xu, GARLIC — A general purpose atmospheric radiative transfer line-by-line infrared-microwave code: Implementation and evaluation, Journal of Quantitative Spectroscopy and Radiative Transfer 137 (2014) 29–50. doi:10.1016/j.jqsrt.2013.11.018

[32] F. Schreier, Optimized implementations of rational approximations for the Voigt and complex error function, Journal of Quantitative Spectroscopy and Radiative Transfer 112 (6) (2011) 1010–1025. doi:10.1016/j.jqsrt.2010.12.010

[33] R. Spurr, J. Wang, J. Zeng, M. Mishchenko, Linearized T-matrix and Mie scattering computations, Journal of Quantitative Spectroscopy and Radiative Transfer 113 (6) (2012) 425–439. doi:10.1016/j.jqsrt.2011.11.014