# Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO2 monitoring mission (update)

## Task 1,2 and 3 Report

Version 2.1: 2025-06-04

*Prepared by*

***RAL Space***

***Remote Sensing Group***

***Harwell Oxford***

***Chilton, OX11 0QX***

***United Kingdom***

***EUMETSAT Contract EUM/CO/20/4600002403/RL***
***EUM/CO/24/4600002860/RL***

*2025-06-04*

Science and Technology Facilities Council | RAL Space

EUMETSAT

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

## Contents

# 1 Introduction

## 1.1 Purpose and objective

This report describes work conducted in the Eumetsat project "Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO2 monitoring mission" (EUMETSAT Contract EUM/CO/20/4600002403/RL) and subsequent "Update of top-of-atmosphere test- data for CO2M -Follow-on activity" (EUM/CO/24/4600002860/RL).

The work aims to simulate level 1b (L1) spectra from the CO2M spectrometers with a quality suitable for testing the operational L1-2 processing software (currently under development via parallel activities). A fast but accurate principal component forward model (PCFM) approach is used to enable full orbits to be simulated, using rigorous radiative transfer calculations for a subset of spectral points. This follows the approach adopted for similar studies to generate test data for Sentinels 4 and 5 [RD06].

Following the statement of work [AD01], the first project was divided into 4 tasks:

- Task 1: Setup of the radiative transfer model and input data
- Task 2: Collection of input data
- Task 3: Production of the CO2IS datasets
- Task 4: Project Outreach

The follow-on project addressed three issues identified with the test data produced at the end of the first project:

1. The radiative transfer calculations mainly neglected the lower half of the bottom model layer and the upper half of the top model layer as discussed in section 6.2. The treatment of layers was changed in the follow-on project, such that the full depth of all model layers was correctly represented in the calculations.
2. An error was found in the mapping of MODIS surface reflectance parameters to the values used in each CO2M band. This resulted in the BRDF parameters for 645nm being used to simulate to the SWIR-1 and SWIR-2 bands, as described in section 5.4. This has been fixed in the follow-on work.
3. The simulations implemented a switch between using land (or ice) BRDF and the Cox and Munk model. In the follow-on work mixing between water and land or ice is now simulated. This particularly improves the realism of the simulations in sun-glint geometry near coast lines and inland water including rivers.

In addition there were some changes in L1 format and minor differences in geolocation implemented in the updated test data. Note that there was no update required to the fast "PCFM" model used to perform the simulations. Also the model input data used is unchanged compared to the previous work.

This report describes the assumptions behind the final set of simulations.

The simulations produced in the study have been supplied in the form of L1 files in the planned operational format, together with all input and intermediate output files (as described in a separate delivery note [RD26].

## 1.2 Document overview

Section 2 provides lists of applicable documents, reference documents and acronyms used.

Section 3 gives an overview of the CO2M instrument assumptions made in the project.

Section 4 describes the CO2M orbits to be simulated.

Section 5 describes the input data used.

Section 6 describes the PCFM to be used to perform the simulations

Section 7 proves some illustration of the output data.

# 2 Applicable and Reference Documents

## 2.1 Applicable Documents

[AD01] Eumetsat ITT 19-219208; Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO2 monitoring mission: Statement of Work; EUM/RSP/DOC/19/1097096; v1D, 19 September 2019

[AD02] Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO2 monitoring mission; RAL Space Technical & Management Proposal, Ref: RS028210, 2019-11-30

[AD03] Eumetsat Request for Quotation (RFQ) 24/227791 Update of top-of-atmosphere test-data for CO2M -Follow-on activity. Statement of work: EUM/COPER-CO2M/SOW/24/1406481 Issue : v1. , Date : 28 March 2024

[AD04] Update of top-of-atmosphere test-data for CO2M: Follow-on activity; RAL Technical Proposal, Ref RSRS06325, 2024-08-19

## 2.2 Reference Documents

[RD01] RTTOV V12 Users Guide, James Hocking et al. NWPSAF-MO-UD-037, Version 1.1, 31/01/2017

[RD02] RTTOV-12 Science and Validation Report, Roger Saunders et al. NWPSAF-MO-TV-41, Version 1.0, 16/02/2017

[RD03] Benedetti, A., J.-J. Morcrette, O. Boucher, A. Dethof, R. J. Engelen, M. Fisher, H. Flentjes, N. Huneeus, L. Jones, J. W. Kaiser, S. Kinne, A. Mangold, M. Razinger, A. J. Simmons, M. Suttie, and the GEMS-AER team, 2009: Aerosol analysis and forecast in the ECMWF Integrated Forecast System. Part II : Data assimilation, J. Geophys. Res., 114, D13205, doi:10.1029/2008JD011115.

[RD04] Copernicus CO2 mission requirements document - v1.2, EOP-SM/3088/YM-ym.

[RD05] Copernicus High Priority Candidate Missions: CO2M Space Segment Requirements Document (SSRD); Reference CO2M-RS-ESA-SY-0002; Issue/Revision 1.0; Date of Issue 16/09/2019

[RD06] Test Data for the EPS-SG instrument UVNS, NOV-7427-NT-4988, issue 4.

[RD07] http://adam.noveltis.com/

[RD08] Spurr R. LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In: Kokhanovsky, A, editor, Light Scattering Reviews, volume 3.Springer;2008

[RD09] Liu X, Smith WL, Zhou DK, Larar A. 2006.. Principal Component based Radiative Transfer Forward Model (PCRTM) for hyperspectral sensors, Part I: Theoretical concept, Appl. Opt. 45: 201–209

[RD10] Siddans, R., K. Weigel, H. Bovensmann, L. Vogel, H. Boesch, L. Guanter, Requirements Consolidation of the Near-Infrared Channel of the GMES-Sentinel-5 UVNS Instrument, ESA Contract 4000107856/13/NL/CT, Final Report, 2014

[RD11] CO2 Human Emissions (CHE), J.-M. Haussaire, et. al., D2.1 Model systems and simulation configurations, CHE-Project EU, https://www.che-project.eu/sites/default/files/2018-04/CHE-D2-1-V2-0.pdf.

[RD12] Agustí-Panareda, A., M. Diamantakis, V. Bayona, F. Klappenbach, and A. Butz: Improving the inter-hemispheric gradient of total column atmospheric CO2 and CH4 in simulations with the ECMWF semi-Lagrangian atmospheric global model, Geosci. Model Dev., 10, 1–18, 2017.

[RD13] Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

[RD14] CO2 Human Emissions (CHE) D2.5 Synthetic satellite datasets: Johan Strandgren (DLR); Date: 31/04/2020; Version: 0.3; Date: 31/12/2019

[RD15] CO2 Human Emissions (CHE) D2.6 Global Run V2; Anna Agusti-Panareda,,Joe McNorton, Margarita Choulga (ECMWF); Date: 20/12/2019; Version: 1.0

[RD16] CO2 Human Emissions (CHE) D2.4 Part I: Regional nature runs: Jean-Matthieu Haussaire, Dominik Brunner, Arjo Segers; Date: 23/04/2020; Version: 1

[RD17] Schaaf, C. (2019). MODIS/Terra+Aqua BRDF/Albedo Gap-Filled Snow-Free Daily L3 Global 30ArcSec CMG V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-04-10 from https://doi.org/10.5067/MODIS/MCD43GF.006

[RD18] Sathyendranath, S, Brewin, RJW, Brockmann, C, Brotas, V, Calton, B, Chuprin, A, Cipollini, P, Couto, AB, Dingle, J, Doerffer, R, Donlon, C, Dowell, M, Farman, A, Grant, M, Groom, S, Horseman, A, Jackson, T, Krasemann, H, Lavender, S, Martinez-Vicente, V, Mazeran, C, Mélin, F, Moore, TS, Müller, D, Regner, P, Roy, S, Steele, CJ, Steinmetz, F, Swinton, J, Taberner, M, Thompson, A, Valente, A, Zühlke, M, Brando, VE, Feng, H, Feldman, G, Franz, BA, Frouin, R, Gould, Jr., RW, Hooker, SB, Kahru, M, Kratzer, S, Mitchell, BG, Muller-Karger, F, Sosik, HM, Voss, KJ, Werdell, J, and Platt, T (2019) An ocean-colour time series for use in climate studies: the experience of the Ocean-Colour Climate Change Initiative (OC-CCI). Sensors: 19, 4285. doi:10.3390/s19194285

[RD19] K. F. Evans, S. J. Walter, A. J. Heymsfield, and M.N. Deeter. Modeling of submillimeter passive remote sensing of cirrus clouds. J. Appl. Met., 37:184–205, 1998.

[RD20] Wen, S., & Rose, W. I. (1994). Retrieval of sizes and total masses of particles in volcanic clouds using AVHRR bands 4 and 5. Journal of Geophysical Research, 99(D3), 5421-5431. http://dx.doi.org/10.1029/93JD03340

[RD21] Requirements Consolidation of the Near-Infrared Channel of the GMES-Sentinel-5 UVNS Instrument; ESA Contract 4000107856/13/NL/CT; Final Report; R.Siddans, K. Weigel, L. Vogel, H. Boesch, L. Guanter; Version 1.0; 26 February 2015

[RD22] Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmos. Meas. Tech., 6, 2803-2823, doi:10.5194/amt-6-2803-2013, 2013

[RD23] Butz, A, O.P. Hasekamp, C. Frankenberg, J. Vidot, I. Aben, CH4 retrievals from space based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes, JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 115, D24302, doi:10.1029/2010JD014514, 2010

[RD24] J. Joiner, P.K. Barthia, R.P. Cebula, E. Hilsenrath, R.D. Mcpeters, and H. Park. Applied Optics, 34(21), 4513-4525., 1995. Rotational Raman scattering (Ring effect) in satellite backscatter ultraviolet measurements

[RD25] K. Chance and R.J.D. Spurr. Ring effect studies: Applied Optics 36, pp. 5224-5230., 1997. Rayleigh scattering, including molecular parameters for rotational Raman scattering and the Fraunhofer spectrum.

[RD26] Siddans, R. Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO2 monitoring mission: Delivery note, EUMETSAT Contract EUM/CO/24/4600002860/RL, 2025-03-31.

[RD27] Francois-Marie Breon and Fabienne Maignan, A BRDF–BPDF database for the analysis of Earth target reflectances; Brief communication 20 Jan 2017 , Earth Syst. Sci. Data, 9, 31–45, 2017, https://doi.org/10.5194/essd-9-31-2017

[RD28] François-MarieBréon, EmilieFédèle, MarcBouvier, Polarized reflectances of natural surfaces: Spaceborne measurements and analytical modelling, Remote Sensing of Environment, Volume 113, Issue 12, 15 December 2009, Pages 2642-2650https://doi.org/10.1016/j.rse.2009.07.022

[RD29] Damien Sulla-Menashe and Mark A Friedl, User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product, May 14, 2018, https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf

## 2.3 Acronyms

Acronyms used in this report are defined in the following table.

**Table 2-1: List of acronyms used in this document**

| | |
|---|---|
| BPDF | Bi-directional polarisation distribution function |
| BRDF | Bi-directional reflectance distribution function |
| CAMS | Copernicus Atmosphere Monitoring Service |
| CHE | CO2 Human Emissions Project |
| CO2IS | CO2 NIR/SWIR instrument on CO2 Mission |
| CO2M | Copernicus CO2 Monitoring MIssion |
| ECMWF | European Centre for Medium Range Weather Forecast |
| EO | Earth Observation |
| FM | Forward model |
| FR | Final Review |
| FWHM | Full-width-half-maximum |
| GHG | Greenhouse gas |
| IGBP | International Geosphere-Biosphere Programme |
| ITT | Invitation to Tender |
| IR | Infra-Red |
| KO | Kick-off (meeting) of project |
| LIDORT | LInearized Discrete Ordinate Radiative Transfer |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MTR | Mid-term review |
| NDVI | Normalised difference vegetation index |
| NIR | Near infra-red |
| NO2IS | Visible spectrometer on CO2 Mission |
| PC | Principal Component |
| PCFM | Principal Component Forward Model |
| RAL | Rutherford Appleton Laboratory |
| RSG | Remote Sensing Group |

| | |
|---|---|
| RTM | Radiative Transfer Model |
| RTTOV | Radiative Transfer for Tiros Operational Vertical Sounder |
| SIF | Sun-induced Fluorescene |
| SoW | Statement of Work |
| ST&C | Specific Terms and Conditions |
| STFC | Science and Technology Facilities Council |
| TN | Technical Note |
| TOA | Top of atmosphere |
| VLIDORT | Vector LIDORT |
| | |

# 3 CO2M Instrument Assumptions

Instrument/system performances are based on the MRD [RD04], however detailed simulation of instrumental errors in not considered here. The following assumptions are made for the purposes of this study:

- No instrument noise is added. I.e. no random noise will be added to the simulated L1 spectra. However the estimated standard deviation (ESD) of the random noise applicable each each spectral sample may (tbc) be filled with realistic values.
- The instrument is assumed completely insensitive to polarisation (though the effects of polarisation on the radiative transfer will be simulated)
- The bands are assumed to be perfectly spatially co-registered. I.e. all bands have the same spatial sampling.
- The spectral calibration of the bands does not vary across track.
- The spectral and radiometric calibration of the instrument is perfect.

The spatial and spectral sampling of the instrument is defined by Eumetsat via L1 files with populated geolocation. In the current files there are 110 cross-track samples covering a swath of 265km. Orbit are divided into granules covering an along-track extent of ~1220km (620 scan lines). The same granularity will be used for output L1 files produced by this project.

CO2M is able to vary its pitch so as to view forward or backwards along-track, primarily to maximise viewing over the sea in sun-glint geometry (to increases the signal). Eumetsat provides orbits in both nadir and “pitched” viewing modes.

No spatial response function is simulated in the work here: simulations are carried out using geophysical parameters interpolated to the ground-pixel centre latitude and longitude (as defined in the files). The spatial response function has been considered in deriving some of the geolocation parameters in the L1 input files provided by Eumetsat (e.g. terrain height and parallax corrected ground-pixel centre latitude and longitude).

There are four CO2M channels with the spectral characteristics defined in Table 3-1. This study assumes the spectral response function to be Gaussian in shape with full-width-half-maximum (FWHM) as defined in the table. The VIS band is implemented via the N2OIS instrument on CO2M; the NIR/SWIR bands are implemented via the CO2IS instrument. It is foreseen that L1 data will be supplied in two files (one for each instrument): Band 1 data is in one (NO2IS) file and Band 2-4 in the other (CO2IS) file.

| Band | Number of spectral samples | Spectral range / nm | Spectral Resolution / nm | Spectral Sampling / nm | Oversampling ratio |
|---|---|---|---|---|---|
| Band1: VIS | 1917 | 405-490 | 0.3 | 0.04436 | 6.76 |
| Band2: NIR | 1930 | 747-773 | 0.12 | 0.01348 | 8.903 |
| Band 3: SWIR-1 | 990 | 1590-1675 | 0.3 | 0.08595 | 3.4906 |
| Band 4: SWIR-2 | 1001 | 1990-2095 | 0.35 | 0.105 | 3.333 |

**Table 3-1: CO2M spectral band definitions used in the study. Note that at the beginning of the study the VIS band was assumed to have a spectral resolution of 0.6nm (and oversampling ratio 13.52). The final settings assumed for delivery are those given in the table.**

# 4 Orbits to be simulated

The study has simulated a total of 6 orbits, simulating data from 3 CO2M platforms on two days. The orbits are referred to as follows:

- EU cent (pitch: off). Orbit in nadir mode passing over Berlin on 3 July 2025. This orbit is mainly selected to sampled detailed GHG emission features over Berlin, captured by the CHE "zoom" data (see below).
- EU west (pitch on). Orbit in glint-pointing mode to the west of "EU cent" on the same day.
- EU east (pitch: off). Orbit in nadir mode to the East of "EU cent" on the same day.
- SA cent (pitch: off). Orbit in nadir mode passing over Russia and South Africa on 23 September 2025. This orbit is chosen to maximise the fraction of scenes over land.
- SA west (pitch: on). Orbit in glint-pointing mode to the west of "SA cent" on the same day.
- SA east (pitch: off). Orbit in nadir mode to the east of "SA cent" on the same day.

The dates given above refer to the nominal date of the CHE model simulations described below. These are based on meteorological conditions for the same date in 2015.

The tracks of the orbits are illustrated in Figure 4-1.

The orbit files provided by Eumetsat define the following parameters which are used in the simulations:

- Wavelength axis of the radiance and solar irradiance spectra.
- Time/date of each observation (scanline)
- Scanline and ground-pixel (across-track) indices
- Latitude
- Longitude
- View azimuth angle
- View zenith angle
- Solar azimuth angle
- Solar zenith angle
- Terrain height
- Land fraction (i.e. fractional coverage of land rather than water)

Most geometrical parameters are given referenced to the Geoid and corrected for the intersection with the terrain, accounted for using a digital elevation model. Simulations use the terrain corrected values.

Other quantities defined in the Eumetsat L1 files are not used in the calculations but will be written to the final output L1 files produced by this project.

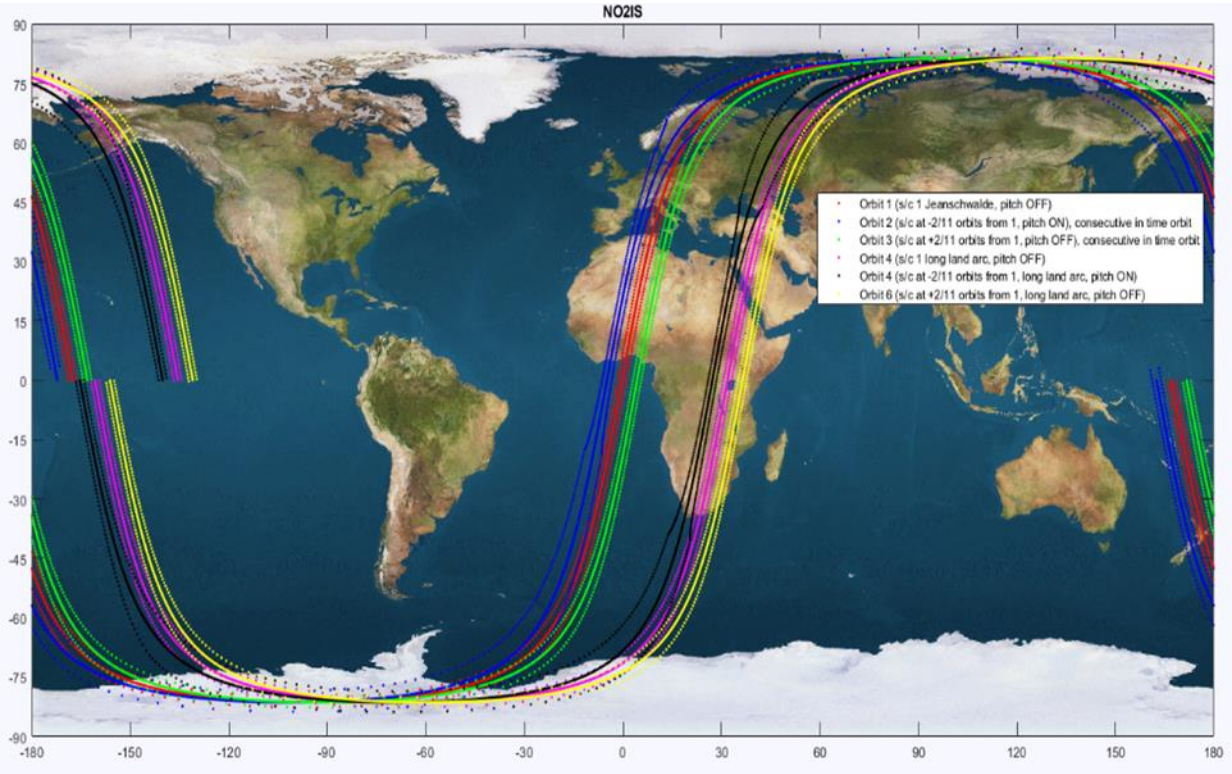



**Figure 4-1: Illustration of the orbits provided by Eumetsat (figure provided by Eumetsat)**

# 5 Geophysical parameters

## 5.1 Scenario files

"Scenario files" containing all scene specific geophysical parameters needed to run the radiative transfer calculations are constructed by interpolating data from various sources to the locations defined in the L1 orbit files provided by Eumetsat.

Scenario files are generated (and delivered to Eumetsat) in netcdf format.

The data sources used to populate the scenario files are described below.

## 5.2 CHE model fields

Simulations are based heavily on model output from the Copernicus Atmosphere Monitoring Service (CAMS) CO2 forecasting system produced in the CO2 Human emissions (CHE) project [RD11][RD12]. Fields for the dates required for this project were produced from several models. Here we use global output [RD15] (produced at ~10km spatial resolution by the ECMWF high resolution forecast model) and, for the "EU" orbits, higher spatial resolution data, restricted to domains over Europe (~5km resolution) and Berlin (~1km resolution) from the LOTOS-EUROS model [RD16]. The following parameters are taken from the CHE global model:

- The 137 level hybrid sigma grid of the global model is used as the computational grid in radiative transfer calculations (see section 6.2). All other model profiles are interpolated to this grid.
- Sea ice coverage
- Snow depth
- Geopotential height profile
- 10m wind speed components (u and v)
- Surface pressure
- Profiles of CO2, CO, CH4, temperature, water vapour. These are given as mass mixing ratio profiles, all assumed to be relative to the total mass of air.
- Profiles of cloud liquid and ice water content. These defined the cloud fields in the simulations, but are manipulated to produce higher spatial resolution cloud structures as described in section 5.8.

We use model fields at 12 noon for the "EU" orbits and 09 UT for the "SA" orbits, since these are the closest analysis times to the orbit time.

The following parameters are taken from the LOTOS Euros European and Berlin domain fields (at the same time and date):

- Surface pressure
- Profiles of CO2, CO, CH4, temperature, water vapour, ozone, NO2, formaldehyde (HCHO), SO2, HNO3

Simulations take input from the highest resolution model available at a given location. I.e. LOTOS-EUROS 1km resolution output is used within the Berlin domain; outside that but inside the European domain the LOTOS-EUROS 5km fields are used; outside that CHE global fields are used (supplemented by CAMS reanalysis, see below). This approach leads to some discontinues at the edges of the domains. These are relatively small between the Berlin and Europe domains (the two LOTOS-EUROS mode runs are consistent with each other).

The study considered using LOTOS-EUROS aerosol fields over Europe, but these were found to be difficult to reconcile with CAMS fields (large biases for some aerosol types, particular sea salt and dust, which have large sources outside that domain). The decision was made to take aerosol only from CAMS reanalysis.

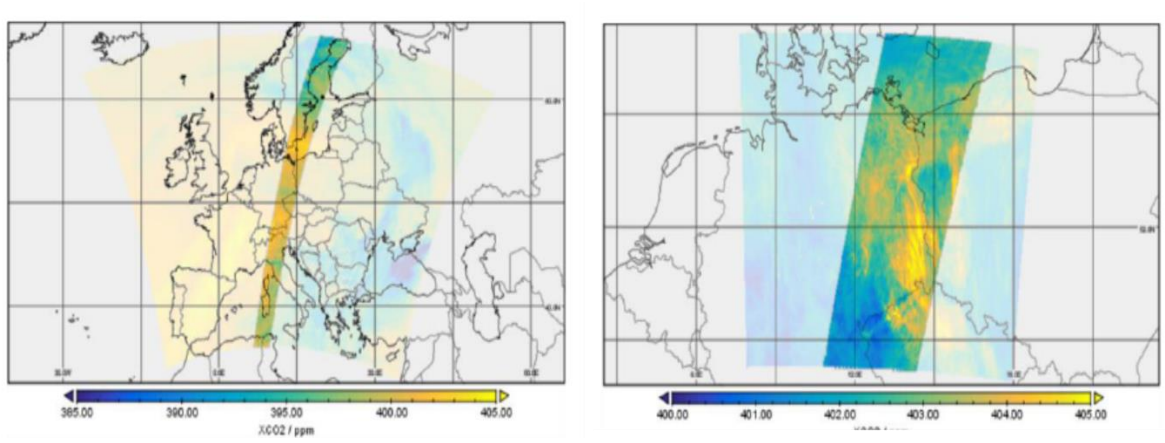



**Figure 5-1: Illustration of the CHE / LOTOS-EUROS Europe (left) and Berlin (right) domains (figure taken from [RD14]).**

## 5.3 CAMS reanalysis fields

CAMS reanalysis fields [RD13] for the relevant day are used to supplement the CHE model fields with other needed chemical species and aerosol. The following fields are used:

- Aerosol component optical depth at 550nm (ie. for black carbon, dust, organic matter, sea-salt and sulfate)
- Total aerosol optical depth at 469, 550, 670, 875 and 1240nm
- Aerosol component mass mixing ratio profiles (11 profiles for the 7 components listed above, with 3 profiles each for sea-salt and dust corresponding to different size bins).
- Profiles of CO, formaldehyde, ozone, NO2, HNO3, SO2.

## 5.4 MODIS BRDF

The land surface bi-directional reflectance distribution function (BRDF) is modelled using the MODIS MCD43GF (gap-filled) product which defines (at ~1km resolution) the coefficients of three “kernel” function (representing isotropic, geometric and volume scattering) each of which has a defined angular dependence.

However the parameters are only defined for 6 MODIS channels (centred at 469, 555, 659, 858.5, 1240, 1640 and 2130 nm), whereas we require spectrally resolved reflectance. Here we assume that the BRDF shape is constant within each spectral band, i.e. the ratio of the geometric, $k_{geo,}$ and volume coefficient, $k_{vol}$, to the isotropic coefficient, $k_{iso}$, is fixed within a band. The ratios $k_{geo}$:$k_{iso}$ and $k_{vol}$:$k_{iso}$ for each band are taken from the following MODIS bands:

- For VIS, values at 555nm are used.
- For NIR, values at 858.5nm are used.
- For SWIR-1, values at 1640nm are used
- For SWIR-2, values at 2130nm are used.

Spectral variation of the isotropic coefficient (and hence total BRDF via the ratios) could be modelled using the approach developed in the ESA Adam study and illustrated in Figure 5-2 to Figure 5-4: Over land, spectral albedo is modelled using the four most significant principal components of the variation in spectral albedo contained in a large database of reference measurements. To obtain the spectral albedo for a specific scene the four PCs are fitted to best match the values in the 6 MODIS bands. Over ice, Adam implements a physical model of ice reflectance which is optimised to fit the MODIS reflectance. Here we use an adapted version of the code provided by the Adam project (from http://adam.noveltis.com/), to model both the land and ice albedo.

This approach to model within band, spectrally dependent albedo was implemented in the previous work for Sentinel 4 and 5, and tested for CO2M. However, it was found that including spectral variability of surface albedo within the 2 micron CO2M band, increased the spectral variability to a degree which significantly degraded the PCFM accuracy. This behaviour was attributed to some (possibly spurious) variation in spectral surface albedo predicted by the Adam tool for a few scenarios (highlighted by the random sampling of the independent set). Note that the example surface reflectances illustrated in Figure 5-2 to Figure 5-4 show that Adam predicts some strong spectral features within the SWIR-2 band.)

It was therefore decided to reduce the complexity of the simulations in all bands and assume spectrally independent BRDF within each NO2I/CO2I band. For each NO2I/CO2I band the parameters for the nearest MODIS band are assumed (as defined above for the BRDF shape). It is recommended that the potential for spectral dependence of albedo to impact L2 quality, especially in the 2 micron range, should be studied further in future.

*It should be noted however, than in performing the simulations, an error was present in the code, which has meant that the final simulations for the SWIR channels actually use BRDF parameters from MODIS band 1 (645nm) instead of the intended channels (indicated in the table above).*

We use the MODIS product defined for the relevant day, however sometimes this contains gaps particularly in the Southern hemisphere (due to high solar zenith / high cloud coverage). We therefore supplement this with data from a second MODIS data which samples a complementary range of latitudes (missing values are filled with data from this “backup” day). For the EU orbits 20 Dec 2015 is used. For the SA orbits, 20 Dec 2015 is also used in the Southern hemisphere; 20 March 2015 is used in the Northern hemisphere.

In addition, we extend the MODIS coverage beyond the coast into the sea. This is done to avoid possible missing data around the coast arising in scenes defined as land by the L1 orbit file, as follows:

- The original fields (at 1km resolution) are convolved with a square kernel of 3x3 cells. The convolution takes account of missing data within each 3x3 area, returning the average of only the valid samples. Cells which are not defined in the original grid are filled with any valid results from the convolution (i.e. where at least one of the cells involved in the convolution was valid). This has the effect of extending valid data out to neighbouring cells (but retaining the original data where it is valid.
- This procedure is iterated with larger square kernels of dimension 5, 9 and 17 cells. Each iteration further extends the valid coverage, without replacing data filled on a previous iteration.

## 5.5 Land surface polarised reflectance

The MODIS kernels provide the scalar BRDF shape. The approach has been extended to model the bi-directional polarised distribution function (BPDF), using an approach described in [RD27] and [RD28], using code provided by Eumetsat (originally provided by the authors [RD28] via an earlier Eumetsat study). This involves adding the polarised reflection to MODIS BRDF, assuming specular reflection to define the Muller matrix. The approach is

implemented in VLIDORT using the built-in “BPDF2009” kernel, in addition to the MODIS BRDF kernels. There are two parameters required to define the polarised reflectance:

- The normalised difference vegetation index (NDVI) of the scene: This is obtained from the MODIS isotropic reflectances as described in section 5.9, below.
- The “calibration factor”, C, which is parameterised as a function of IGBP land classification index, as defined in table Table 5-1.The IGBP land classes are defined for each scene simulated here from the annual MODIS land cover type product (MCD12Q1) [RD29] for 2015. This gives the IGBP classifications are a spatial resolution of approximately 500m. The IGBP indices are nearest-neighbour sampled to the centre latitude/longitude of each CO2M footprint.

In general, the polarised term can add up to a few % to the direct reflectance computed using only the MODIS BRDF kernels. In order to more closely match the reflectance obtained from scalar-only and vector calculations, in the current work we scale down the MODIS BRDF (applying the same factor to the weights for each kernel) such that the total direct surface reflection (for the scene specific view/solar geometry) is unchanged by introducing polarisation.

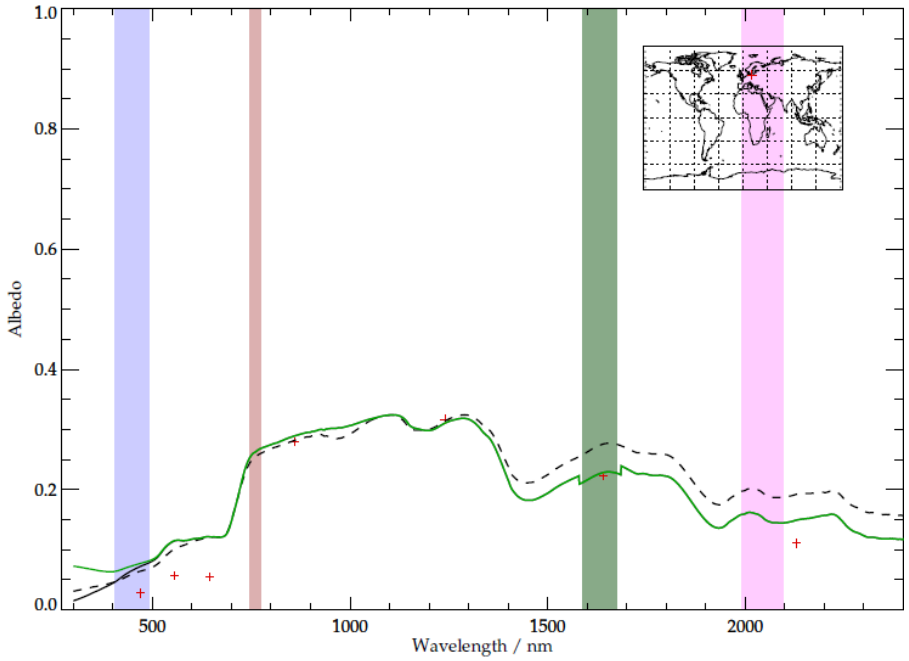



**Figure 5-2: Example spectral albedo (over Europe) generated from the MODIS isotropic reflectances (channels indicated by red +), using the Adam algorithm. The black dashed line is the Adam global mean albedo spectrum. The green solid line shows the original Adam spectrum. The black curve (usually underlying the green) shows the spectrum after corrections are applied. Coloured vertical bars mark the spectral ranges of CO2M.**

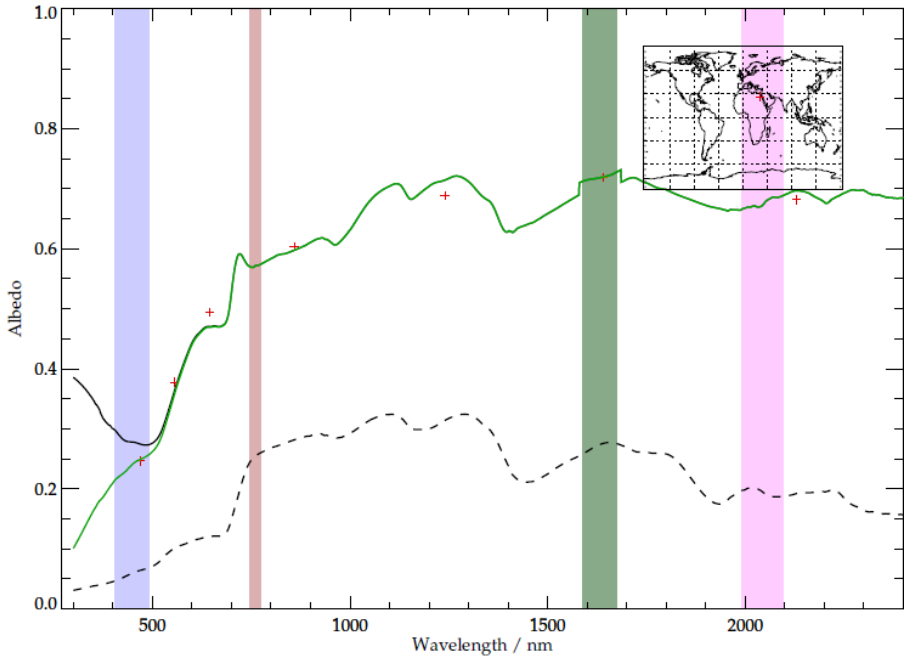



**Figure 5-3: Example spectral albedo (over Egypt) generated from the MODIS isotropic reflectances (channels indicated by red +), using the Adam algorithm. The black dashed line is the Adam global mean albedo spectrum. The green solid line shows the original Adam spectrum. The black curve (usually underlying the green) shows the spectrum after corrections are applied. Coloured vertical bars mark the spectral ranges of CO2M.**

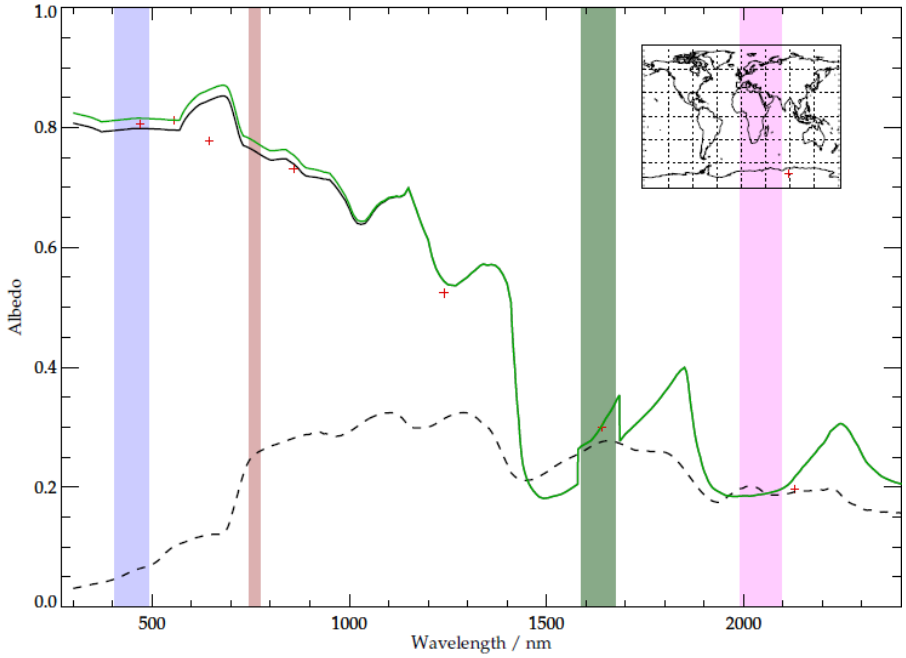



**Figure 5-4: Example spectral albedo (over Antarctica) generated from the MODIS isotropic reflectances (channels indicated by red +), using the Adam algorithm. The black dashed line is the Adam global mean albedo spectrum. The green solid line shows the original Adam spectrum. The black curve (usually underlying the green) shows the spectrum after corrections are applied. Coloured vertical bars mark the spectral ranges of CO2m.**

| IGBP land classification name | Index | C |
|---|---|---|
| Evergreen Needleleaf Forests | 1 | 4.3 |
| Evergreen Broadleaf Forests | 2 | 5.2 |
| Deciduous Needleleaf Forests | 3 | 5.9 |
| Deciduous Broadleaf Forests | 4 | 5.9 |
| Mixed Forests | 5 | 5.9 |
| Closed Shrublands | 6 | 4.7 |
| Open Shrublands | 7 | 5.4 |
| Woody Savannas | 8 | 5.8 |
| Savannas | 9 | 5.9 |
| Grasslands | 10 | 5.8 |
| Permanent Wetlands | 11 | 6.5 |
| Croplands | 12 | 6.8 |
| Urban and Built-up Lands | 13 | 6.7 |
| Cropland/Natural Vegetation Mosaics | 14 | 7.3 |
| Permanent Snow and Ice | 15 | 6.5 |
| Barren | 16 | 7.9 |
| Water Bodies | 17 | N/A |
| Unclassified | 18 | N/A |

**Table 5-1: IGBP Land classification indices and the corresponding BPDF calibration factor used to scale the specular polarised reflectance.**

## 5.6 Ocean Colour CCI data

BRDF over ocean is simulated as the sum of the isotropic reflectance taken from Ocean Colour CCI data [RD18] and the Cox and Munk model implemented in VLIDORT using the near surface (10m) wind speed from CHE. The VLIDORT Cox and Munk models includes polarisation.

We use v4.2, monthly, 5km CCI data, which provides reflectance at 412,443,490,510,555 and 670nm. Here we associated the CCI wavelengths with the MODIS channels, so as to obtain values for each MODIS channel which can be interpolated in wavelength using the same Adam based approach: Channels are used as follows

- CCI 670nm value assigned to MODIS 659 and 858 nm channels
- CCI 490nm value assigned to MODIS 470nm channel
- CCI 555nm value assigned to MODIS 555nm channel
- Ocean reflectance is assumed 0 in the 1240, 1640 and 2130 nm MODIS channels (though this is superceded by a minimum value as described in the following section).

There are significant gaps in the CCI data, which are filled as follows:

- The convolution approach described for MODIS above is applied to fill cells around the coast and any small gaps. The same convolution dimensions are used, though applied to the coarser resolution input grid (so valid data is propagated further in terms of distance on the ground).
- Any cells which are still not filled by this procedure are filled with the 2-degree latitude binned zonal mean of the (valid) reflectance data (extrapolated at fixed value to high southern latitudes where there is no ocean at any longitude).

At the end of this process, a complete global field of ocean reflectance (at each wavelength) is defined.

Ultimately, whether the filled land or ocean values are used in the simulations depends on selection rules defined below.

## 5.7 Selection rules for land vs ocean reflectance data

The simulations ultimately use either the MODIS land and ocean colour CCI data depending on the land fraction defined in the orbit L1 files. In addition, CHE sea ice data is used to include sea ice over ocean. Scenario files are produced containing MODIS BRDF parameters which combine the original MODIS, Ocean CCI and CHE sea-ice data as follows:

- The land fraction defined in the L1 orbit file is used to determine if a scene is land or water (based on threshold of 0.5). Scenes over land use the original MODIS BRDF values.
- For scenes over sea with zero sea ice fraction and zero snow depth according to CHE (global), the Ocean Colour CCI values are used to define the MODIS isotropic BRDF parameter (with the MODIS/CCI channel mapping described above). Geometric and volume BRDF parameters are assumed to be zero (Cox and Munk will also be used in the radiative transfer itself).
- For scenes over sea with sea ice, the Ocean Colour CCI derived isotropic reflectance is linearly combined with fixed isotropic reflectance values assumed for sea ice (weighted by the sea ice fraction). The following values are assumed for isotropic sea ice reflectance, based on typical values found in the MODIS BRDF data:
    - 0.7888 at 470nm

  - 0.7256 at 555nm
  - 0.6749 at 659nm
  - 0.6022 at 858nm
  - 0.427 at 1250nm
  - 0.2217 at 1640nm
  - 0.1144 at 2130nm
- Finally, if the resulting isotropic reflectance in any channel is found to be less than 0.005 (over land or sea), then it is set equal to 0.005. I.e. isotropic reflectance is not allowed to be less than 0.005 in the simulations.

## 5.8 Modelling of cloud and aerosol

### 5.8.1 Conversion of mass mixing ratio profiles to optical properties

In order to perform the RT calculations, optical properties need to be defined based on the mass mixing ratio profiles defined in the model fields (and written to the scenario files). For aerosol these are given in terms of the CAMS aerosol components, together with their total optical depths. The vertical distribution of cloud is given by CHE cloud water content profiles.. The RTM requires profiles of optical properties (extinction optical depth, single scatter albedo, phase matrix).

For aerosol the optical properties (including polarised phase function parameters) have been obtained from the 3MI/MetImage study. Four aerosol models are defined which are associated to CAMS aerosol types (or components) as indicated in Table 5-2. The aerosol optical properties used are illustrated in Figure 5-5.

The scenario files define (from CAMS) the mass mixing ratio profiles of various aerosol components, some of which are divided into size modes. The total optical depth of the components is also provided, though not for each mode. To determine the aerosol extinction coefficient profile the following approach is adopted:

- Mass mixing ratios are converted to mass density (scale by density of air).
- An effective radius is associated to each component, with appropriate values for each mode defined based on the ECMWF documentation [RD03].
- Mass densities are normalised by the associated effective radius. Extinction coefficient is assumed proportional to these profiles.
- The total optical depth of each component is used to scale the effective radius normalised mass density profiles such that the total optical depth of each component is correct.
- The total extinction coefficient is obtained.
- The single scattering albedo and phase function (applicable to all levels in the profile) is obtained from the 3MI/MetImage file, choosing the class which has largest individual optical depth.

Figure 5-6 illustrates this process for a single aerosol profile.

A similar approach is adopted to compute the extinction profile for ice and liquid cloud, which are both simulated separately, using appropriate spectral optical properties. Here we use the optical properties supplied with RTTOV12 [RD01][RD02]. These are defined, separately for ice and liquid, as a function of effective radius. For these simulations we assumed the ice cloud effective radius is given be an assumed linear correlation with mass density [RD19]. Liquid cloud is always assumed to have an effective radius of 10 microns. The optical depth of liquid and ice cloud is derived from the integrated water path, $L$ (in g/m$^2$), using the following formula [RD20]:

$$\tau = \frac{3\ L\ Q_{ext}}{4\ R_{eff}\ \rho}$$

**Equation 1**

Where

- $Q_{ext}$ is the extinction efficiency factor (assumed to be 2)
- $R_{eff}$ is the effective radius (m)
- $\rho$ is the density of water (1e-6 g/m$^3$)

Figure 5-7 illustrates the process for an example profile.

| **3MI study class** | **MACC class(es)** |
|---|---|
| OPAC continental average | Black carbon |
| Desert dust (Dubovik et al, 2002) | Dust |
| OPAC Urban | Organic matter + sulphate |
| OPAC Maritime clean | Sea Salt aerosol |

**Table 5-2: Relationship between aerosol classes used in the study.**

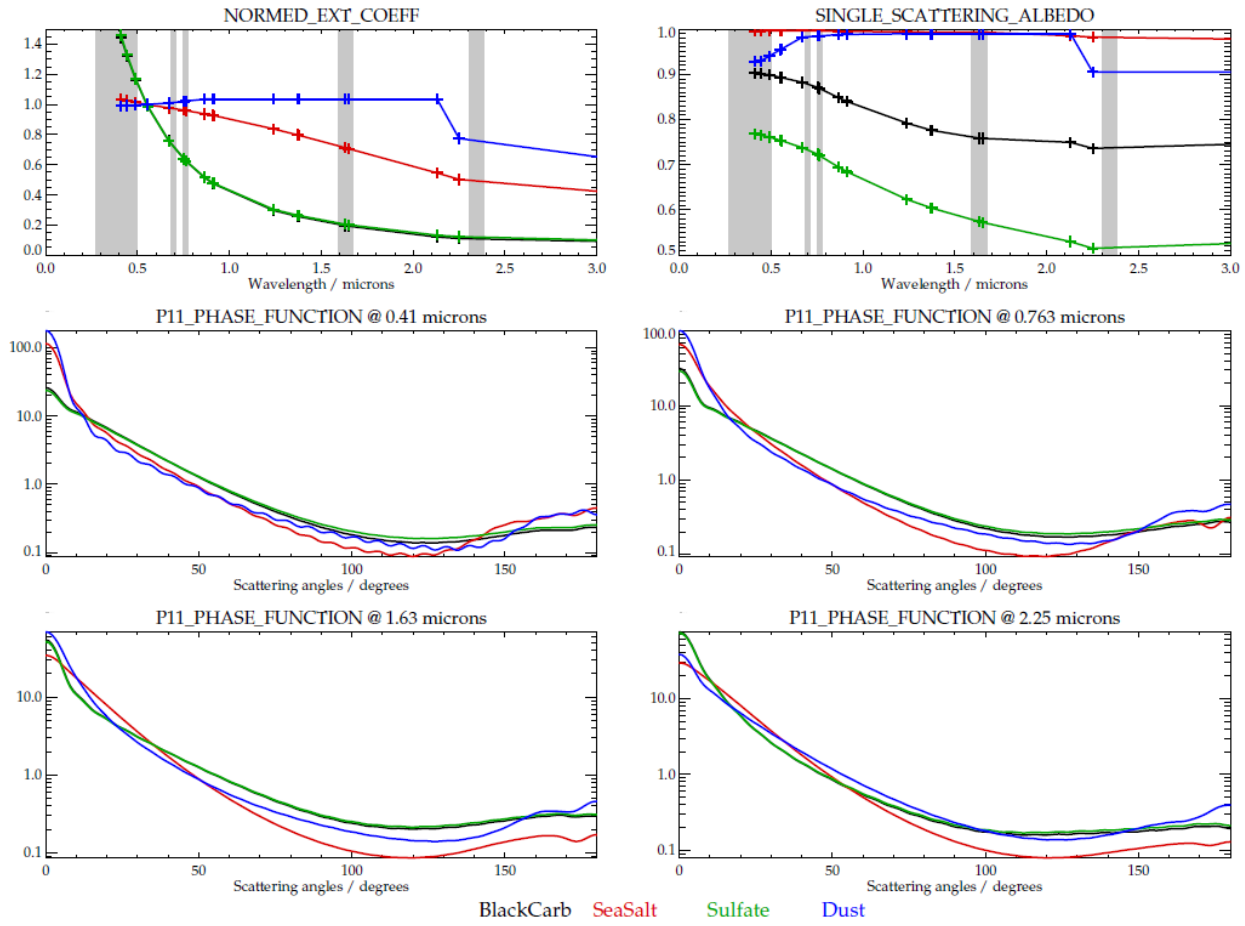



**Figure 5-5: Aerosol optical properties used (from [RD06])**

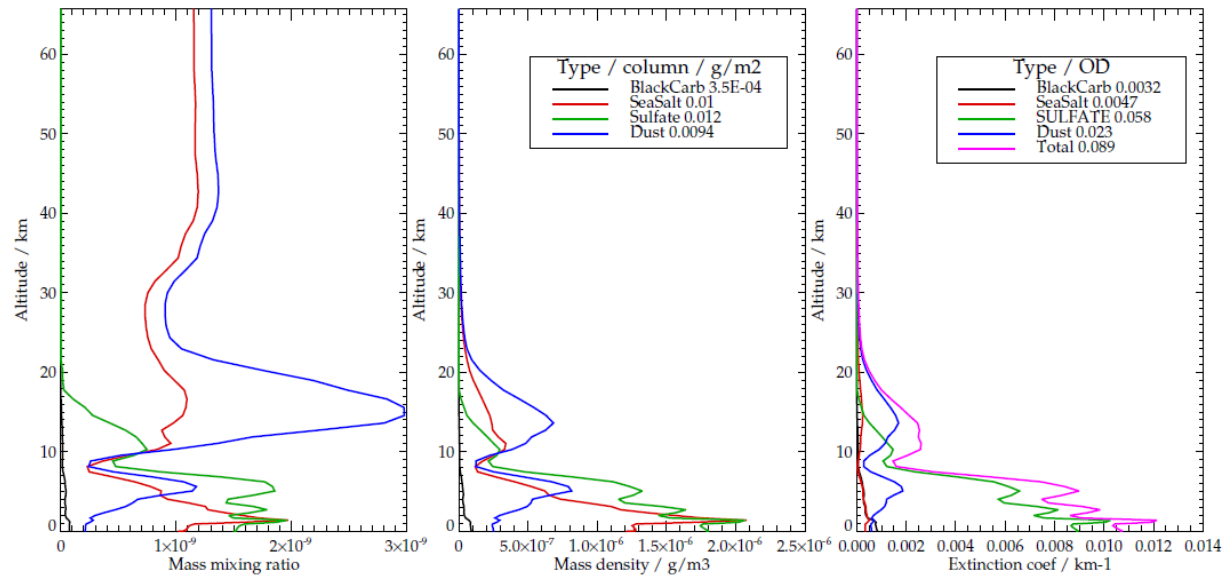



**Figure 5-6: Example aerosol profiles. Left hand panel shows the component mass mixing ratios from CAMS defined in the scenario file. Middle panel shows the profiles in mass density units. Right hand panel shows inferred extinction coefficient profiles. Values the legend give total column amount. The dominant aerosol class, is indicated in upper case in the right-hand panel legend. (From [RD06])**

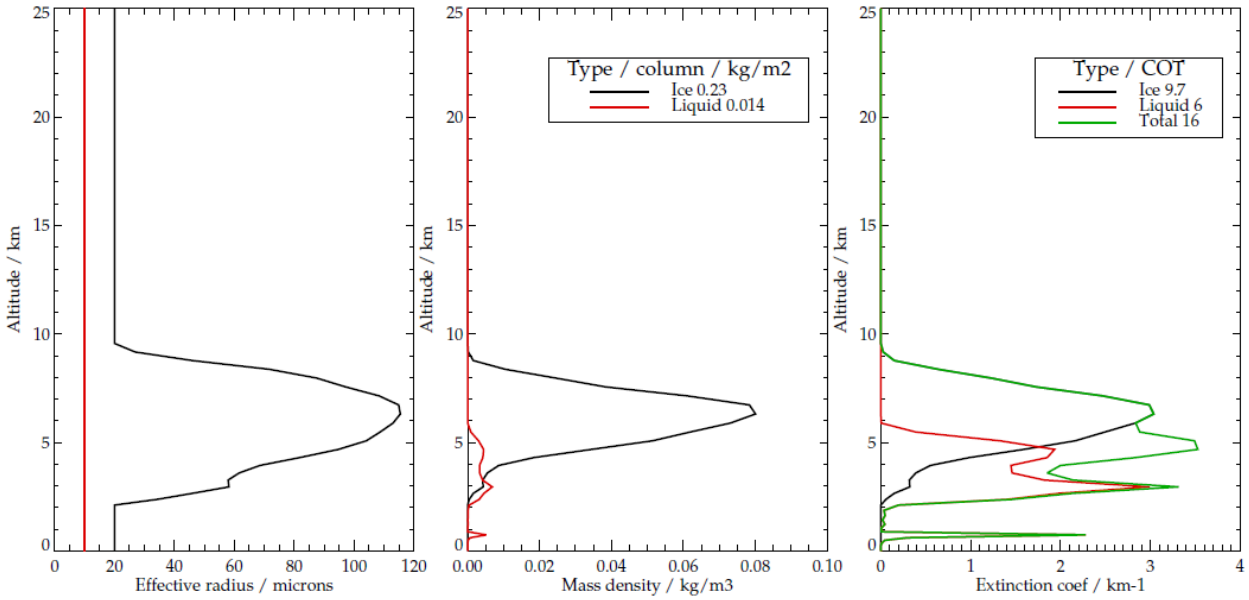



**Figure 5-7: Derivation of cloud extinction profiles. Left hand panel shows effective radius assigned to liquid and ice cloud. Middle panel shows the mass density profiles. Right-hand panel shows the derived extinction coefficient. (From [RD06])**

### 5.8.2 Simulation of high-spatial resolution cloud fields

Cloud fields at ~1km spatial resolution are defined for the simulations from the 10km resolution CHE global data as follows:

1. The (10km) resolution (ice+liquid) cloud optical depth field is determined from the CHE profiles as described in the previous section.
2. As described further below, pseudo random cloud field is generated at 1km spatial resolution which has (almost) the same optical depth when degraded to theoriginal 10km resolution.
3. The ice and liquid water content profiles defined at 10km resolution are interpolated to each location defined in the L1 orbit files. The high-resolution optical depth is also interpolated to each location.
4. These interpolated ice and liquid water content profiles are both scaled by an altitude independent factor (the same for both liquid and water at a given location), such that the total optical depth matches the interpolated high-resolution total optical depth.

The high resolution optical depth field is derived as follows (based on that adopted for the previous S4/5 studies [RD06]):

1. A pseudo random image (with floating point values ranging from 0-1) is generated at the ~1km resolution. This is defined exactly be oversampling the grid of the CHE global model by a factor of 10. The image is generated to follow a spatial power law distribution proportional to $1/v^2$, where v is spatial wavenumber. This distribution, which gives more variability at larger spatial scales, is typical of cloud fields. The intention of the following is to use this image to define structure at sub-CHE spatial resolution, while maintaining the larger scale structure in the CHE field.
2. A series of monotonically increasing threshold values, $t_L$, are defined spanning the range of optical depths. 250 values which span the optical depth range from 0.01 to 200 equally spaced in log optical depth are used.

3. For each $t_L$ (starting at the lowest) a binary mask, $M_L$, is defined (at CHE resolution) with value 0 or 1 depending on whether the CHE value is larger than the threshold or not.
4. The fraction, $f$, of the 100 FCI resolution pixels within each CHE pixel which exceed the threshold is estimated. This is done by determining the fraction of the CHE data with $M_L$=1 in areas of 3x3 and 5x5 CHE pixels about each CHE pixel. $f$ is estimated by extrapolating the fractions at 3x and 5x over-sampling linearly to 1 pixel sampling.
5. Within each CHE pixel, a threshold $t_H$, is applied to the pseudo-random image such that the fraction of high-resolution pixels which exceed $t_H$ matches $f$. This is repeated for each CHE pixel, considering the corresponding 3x3 high-resolution pixels from the pseudo-random image, choosing a different threshold each time to always match the CHE pixel specific value of $f$. Thus, a binary mask, $M_H$, is generated from the high-resolution image which has the same structure as $M_L$ at CHE resolution, but has realistic structure at finer scales from the pseudo-random image. The high-scale structure tends to be continuous across CHE resolution pixels (as the pseudo-random image is continuous), though there will be discontinuities where there are strong changes in $f$.
6. A high-resolution image is constructed by filling an image at FCI resolution by setting all pixels where $M_H$=1 with the mid value between the current value of $t_L$ and the next higher threshold value.
7. The process repeats from step 3 for each $t_L$ in ascending order until the complete image is formed. The process results in a high-resolution image with realistic structure, but digitised to only have values which are the mid-points between the chosen $t_L$. The levels are chosen to be sufficiently finely spaced that this does not affect the realism of the data for the intended purpose. In any case these will subsequently be interpolated spatially to the locations in the orbit files, which will largely remove any digitisation of the values.

An implementation detail is that the approach would be too computationally demanding to implement using a full global 1km field. Instead the globe was divided into 15x15 degree segments and the approach above applied separately in each. This does not generate any discontinuous structure at the segment boundaries because the same pseudo-random field is used in each segment and this generated (using a fast-fourier transform) to be cyclically symmetric, i.e. the structures near one edge of the image are continuous with the opposite edge. The approach will mean that the fine scale structures repeat at 15 degree intervals, but this is not expected to have any practical consequence for the realism of these simulations (especially as the structures will be adapted to different optical depth levels in different segments).

The derived cloud fields are illustrated for an example orbit granule in Figure 5-8. This is and RGB (red-green-blue) false colour derived from the isotropic BRDF and cloud optical depth values in the scenario file as follows:

- $R_{clear}$ (red channel) is set to the MODIS 1600nm isotropic BRDF
- $G_{clear}$ (green channel) is set to the MODIS 858nm isotropic BRDF
- $B_{clear}$ (blue channel) is set to the MODIS 555nm isotropic BRDF
- The high resolution cloud optical depth,$\tau$, is converted to an effective transmission, $T$:

$$T = e^{-\tau/3}$$

**Equation 2**

- Cloudy effective reflectances in the red channel are obtained via

$$R_{cloud} = R_{clear} * T + \ (1 - T)$$

**Equation 3**

Similar equations are used for the green and blue channels.

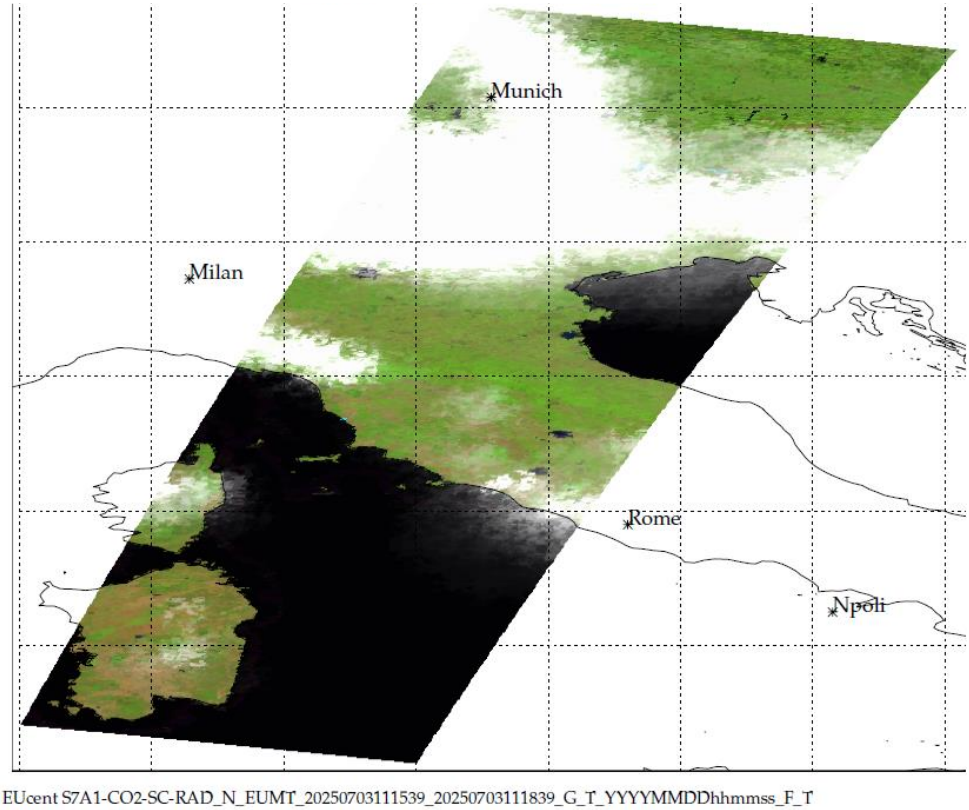



**Figure 5-8: False colour image defined from data in the scenario file for a single granule, illustrating spatial structures in the modelled surface reflectance and cloud fields.**

## 5.9 Solar induced fluorescence (SIF)

The approach to model solar-induced chlorophyll fluorescence (SIF) in this study is based on that adopted for an ESA S5 NIR band simulation study [RD21]. This made use of a set of geophysical scenarios (adapted from those of [RD23]). The same scenarios were used to train the PCFM in the S4/5 Eumetsat simulation studies and have been adapted here for the purposes of this study (see section 6.6):

- Global average maps of Solar-induced chlorophyll fluorescence (SIF) at 740nm were constructed for each of the SWIR study geophysical scenarios using mean data provided by L. Guanter (FUB), based on [RD22]. For this study we require plausible estimates of SIF at the spatial resolution of CO2M. To achieve this we have correlated the S5 NIR study SIF values with MODIS normalised difference vegetation index (NDVI), derived from the isotropic BRDF parameters at 858 (NIR) and 555nm (VIR) as follows:

$$NDVI = \frac{NIR - VIS}{NIR + VIR}$$

**Equation 4**

- The correlation between NDVI and SIF is illustrated in Figure 5-9. The derived linear relationship (illustrated in the right-hand panel) between SIF and NDVI is then used to prescribe SIF for each CO2M scene (by applying the linear fit to the MODIS parameters stored for each location in the scenario file). I.e. SIF in mW/m2/nm/sr is assumed to be

$$SIF = 2.15\ NDVI - 0.280$$

**Equation 5**

(Where NDVI is such that this results in negative SIF, SIF is assumed to be 0.)

- The spectral dependence of fluorescence is modelled using two Gaussian functions which are fit to observed data provided by FUB for the S5 SWIR study, illustrated in Figure 5-10. The same spectral dependence is assumed at all locations.
- SIF is only modelled in the CO2M NIR band.
- Note that the core scattering model used in this study (VLIDORT, see section 6) is able to simulate surface emission directly (i.e. it is possible to add a diffuse emission term at the surface).

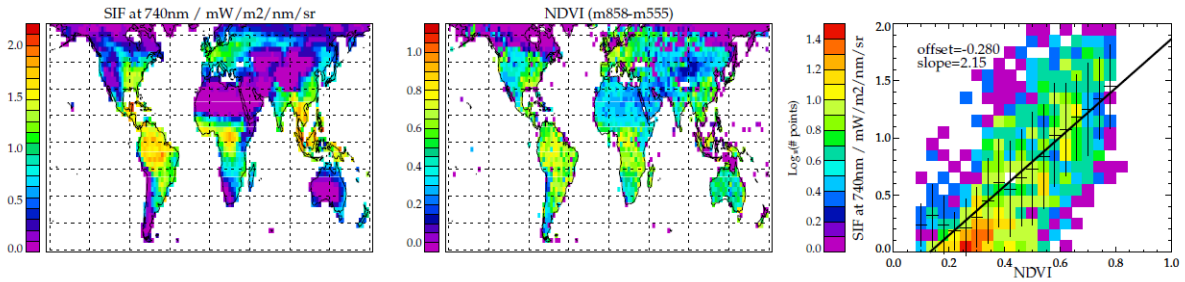



**Figure 5-9: Relationship between SIF and NDVI used to model SIF in the CO2M simulations. Left-hand panel shows the mean fluorescence intensity at 740 nm, derived for the S5 SWIR scenarios by L. Guanter. Centre panel shows MODIS NDVI for the same locations, derived from the MODIS BRDF at 858 and 555nm. Right-hand panel shows the scatter density plot of the relationship between SIF and NDVI (where SIF is positive, NDVI>0.1, in the latitude range 60S to 65N). The linear fit is shown by the solid line; the fit parameters are indicated in the lop-left of the panel.**

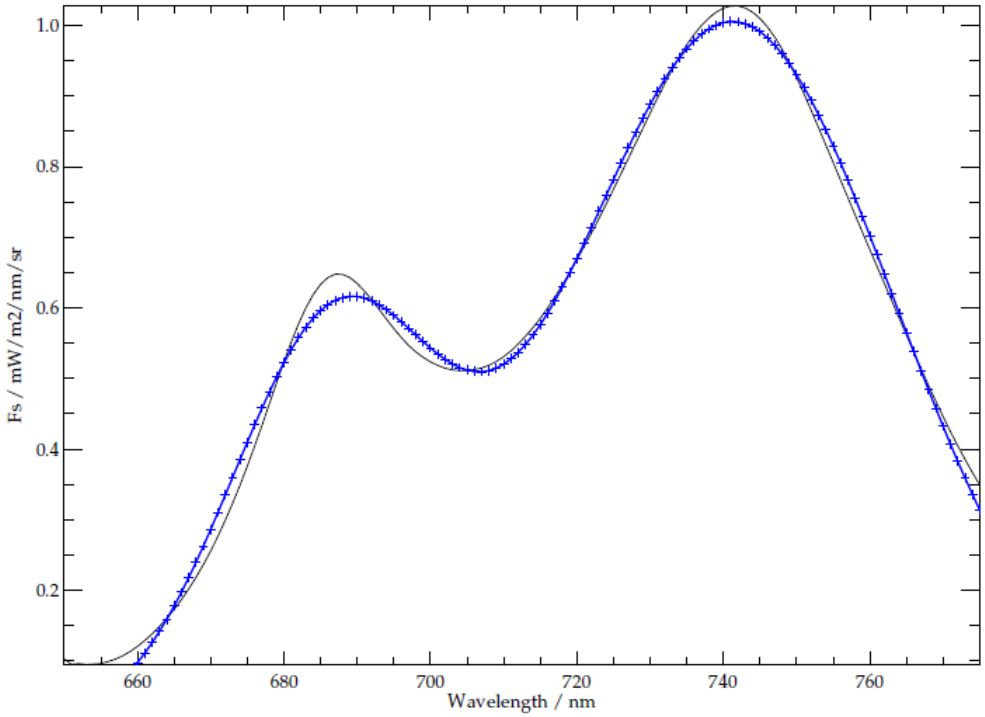



**Figure 5-10: Fluorescence source spectrum provide by L.Guanter (black line), and fit (based on two Gaussian functions) used in the height-resolved aerosol retrieval simulations (blue line).**

## 5.10 Spectroscopic parameters

### 5.10.1 Line-by-line model

For line absorbing species ($O_2$, $H_2O$, CO, $CH_4$, $N_2O$, $CO_2$), Hitran 2016 spectroscopic parameters are used, and the line-by-line calculations are performed by the Oxford RFM (version 5.1). Voigt/CKD line shape, only considering self and air broadening. The file containing the spectroscopic parameters was downloaded from hitran online (http://hitran.org) on 29 September 2020, and is provided with the final delivery of data. Line mixing not modelled. Continuum absorption is defined by the CKD model (see http://eodg.atm.ox.ac.uk/RFM/).

### 5.10.2 Cross-sections

In the UV-VIS, the cross-sections for relevant species are taken, based on those used in the S5P operational processors:

- Ozone ($O_3$): Brion-Malicet-Daumont. Temperature dependence is modelled using a 2nd order polynomial fit.
- Oxygen dimer ($O_4$): Thalman+Volkamer 2013. Temperature dependence is modelled as linear between the spectra give at 203 and 293K.
- Nitrogen dioxide ($NO_2$): Vandaele, 2002. Temperature dependence is modelled as linear between the spectra give at 220 and 294K.

### 5.10.3 Solar spectra

The monochromatic RT calculations are performed using high-resolution solar reference spectra to define the incoming solar irradiance. For the UV-VIS the spectra of [Dobber, 2008] used. For NIR and SWIR bands the spectra computed by R.L. Kurucz are used:

http://kurucz.harvard.edu/sun/irradiance/irrcmasun.asc

### 5.10.4 Conversion of mass-mixing ratio profiles to volume mixing ratio and number density

Model gas profiles are generally defined in mass mixing ratio. In the radiative transfer model these are assumed to be with respect to the total mass of air (including water vapour) and are converted to volume mixing ratio by scaling by the ratio of the molecular mass of air to that of the specific gas. For this purpose, the molecular mass of air is to be a constant value of 28.964 g/mol. The assumed molecular mass of each gas is as follows (g/mol):

- CH4 = 16.04275
- CO= 28.01055
- CO2 = 44.00995
- H2O = 18.01528
- HCHO = 30.02635
- NO2 = 46.0055
- O3 = 47.9982
- SO2 = 64.0628

The radiative transfer model also assumes the total air molecular number density, $N_{air}$, to be given by the ideal gas equation:

$$N_{air} = \frac{N_a p}{RT}$$

Where p in pressure (Pa), T is temperature (K), R is the gas constant = 8.314472 J/K/mole and Na is the Avogadro number = 6.02214199x$10^{23}$ $mole^{-1}$. The number density of a given gas is obtained by multiplying the air number density by the corresponding volume mixing ratio.

# 6 PCFM

## 6.1 Overview of the PCFM

The principal component forward model (PCFM) [RD09] aims to compute instrument resolution spectra from a limited number of rigorous RTM calculations performed for a sub-set of the monochromatic spectral points. With the PCFM, for each ground scene, only a few (10s-100s) rigorous monochromatic calculations need to be performed (compared to many thousands which would otherwise be needed). The reconstruction of full spectra from the reduced number of points is achieved exploiting the relationship between the few monochromatic values and the principal components of the spectral variation of the (instrument resolution) spectrum. These principal components are determined from a "training set" of geophysical scenarios which are simulated in a "brute-force" manner using full-up RTM calculations performed at all monochromatic points.

The approach works well because there is a large degree of correlated variability in the monochromatic radiances which means it can be expected that most of the information needed to construct the convolved radiance are contained in a limited number of (carefully selected) monochromatic radiances.

The steps to implement the PCFM are summarised as follows:

1. A representative set of $N_{tr}$ geophysical conditions is defined, spanning the expected variability of the eventual dataset to be simulated. Several thousand tests scenes are needed, in the previous study these were taken from a set of profiles developed in an earlier S5 study. Here we would select an appropriate sub-set from the atmospheric and surface data used for the orbit simulation (but considering global extent to ensure applicability beyond the specific test orbits)

2. Monochromatic spectra are computed via full RTM calculations, for each profile/geometry in the training set. These spectra are stored in matrix $\boldsymbol{R_{rtm}}$ (number of full-grid monochromatic points, $N_{rtm}$ x $N_{tr}$). Here VLIDORT [RD08] (version 2.5) is used as the "core" RTM. This is called from within over-arching forward model (FM) which computes optical properties from the atmospheric state / surface state parameters in the scenario files for input to VLIDORT. In the training of the PCFM, VLIDORT is run purely in scalar mode (i.e. polarisation is not included in the radiative transfer). It is assumed that polarisation does not itself introduce further spectral patterns which cannot be captured by the principal components of the training set based on scalar calculations. Polarisation is included only in the *application* of the PCFM, when simulating the sub-set of monochromatic points.

**3.** Principal components of the full (convolved) spectra are determined, $\boldsymbol{U}$ ($N_{rtm}$ x number of significant principal components, $N_{pc}$), such that a given radiance vector can be accurately computed using

$$\boldsymbol{r_{ch}} = \boldsymbol{U}\,\boldsymbol{y}$$

**Equation 6**

where $\boldsymbol{y}$ is vector of "scores" for each principle component. These are known for every spectrum in the training set from

$$\boldsymbol{y} = \boldsymbol{U^T r_{ch}}$$

**Equation 7**

**4.** The basis of the PCFM is to suppose that the principle component weights for a new spectrum can be estimated quickly from radiances computed from a sub-set of monochromatic points by computing:

$$\boldsymbol{y} = \boldsymbol{W}\,\boldsymbol{r_{mo}}$$

**Equation 8**

Where $\boldsymbol{W}$ is a matrix of fixed weights ($N_{pc}$ x number selected monochromatic points, $N_{mo}$) and $\boldsymbol{r_{mo}}$ is the vector containing the reduced number of monochromatic radiances. The matrix $\boldsymbol{W}$ is determined by performing a linear regression of the known $\boldsymbol{y}$ against the corresponding values of $\boldsymbol{r_{mo}}$ from the training set. This is carried out by simple linear least squares retrieval of the values in $\boldsymbol{W}$ which minimise the difference between the predicted and "true" $\boldsymbol{y}$, considering all profiles in the training set, for each principal component in turn. The "measurement" being fit has $N_{tr}$ elements and $N_{mo}$ weights are retrieved, for each principal component. Having obtained the predicted PC scores, $\boldsymbol{y}$, a fully sampled radiance spectrum is obtained by applying Equation 6.

Key to making the approach work effectively is the selection of the sub-set of monochromatic points. In principle, any set could be chosen, but unless the set is chosen carefully, a large $N_{mo}$ will be required to achieve a given level of accuracy. An information content based approach was developed for S4/5 and will be applied again here.

The accuracy of the PCFM can be assessed (using independent test data) as a function of the number, $N_v$, of monochromatic points selected from the ordered list generated above. A suitable number/level of accuracy is chosen and this define the points which need to be simulated in full-orbit calculations.

The process used to train and verify the accuracy of the PCFM is illustrated in Figure 6-1.

Note that the PCFM approach is not overly sensitive to the completeness / representativity of the training set: The process is always grounded by the detailed sub-set of calculations which are performed for the sub-set of points. The PCFM is a purely linear regression (exploiting the fact that there is a definite linear relationship between monochromatic and convolved spectra). This greatly reduces the scope for the reconstruction to generate unrealistic features (in contrast to a more complex non-linear regression e.g. neural network).

While some aspects of the PCFM are fixed by assumptions made in the training data, it retains a larger degree of flexibility e.g.:

1) The approach relies mainly on monochromatic simulations, which are independent of the detailed instrument spectral response. The full monochromatic calculations for the training set are retained; the selection of the sub-set of monochromatic points is independent of the spectral convolution and the sub-set of monochromatic calculations for each ground pixel in the simulated orbits are retained. (All this data can be delivered in addition to the final L1 data). This means that it is straightforward to regenerate output files for a modified spectral response (no new radiative transfer calculations are needed).
2) Potentially spectra could be generated at finer-than instrument resolution such that the effect of changing instrument spectral response could also be simulated (within limits) by convolving L1 output spectra.
3) To some extent, the sub-set monochromatic orbit output can be used also to extend the spectral ranges of the instrument, provided the sub-set of monochromatic points are good enough predictors of the extended range (only RTM calculations of the training set are needed).
4) The cloud fraction can trivially be varied given the cloudy/cloud-free spectra for each scene which will be provided.
5) Simulations for different orbits or changed atmospheric state (other than cloud) require new calculations for those scenes for the sub-set of monochromatic points, however it is not expected to be necessary to repeat the training (we intend to adequately span a suitable range of variability in the training set).

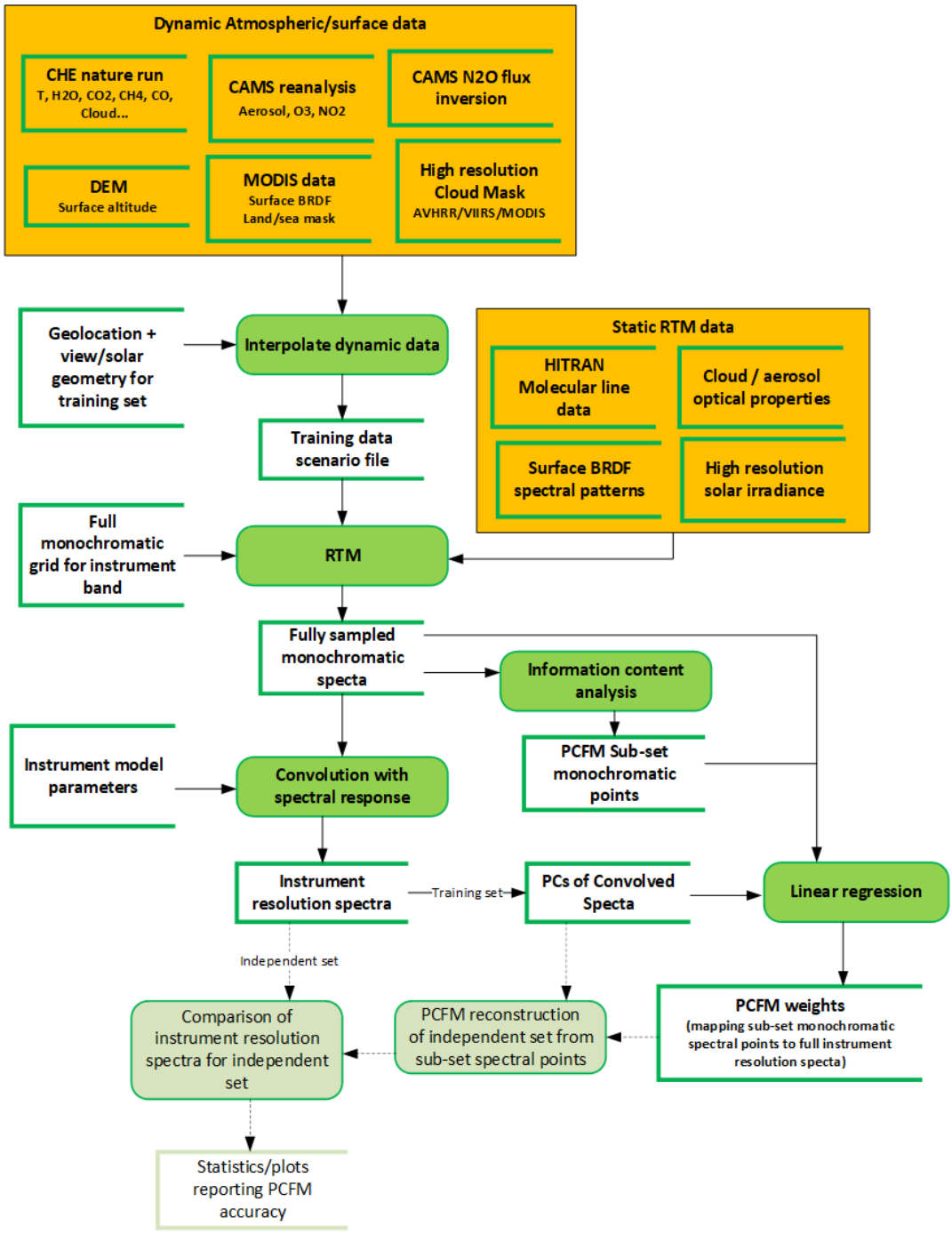



**Figure 6-1: Overview of the approach used to train and test the PCFM. Open 3-sided boxes indicate data; solid green shapes indicate processing modules. Dashed arrows and light-green shapes indicate steps involved in verifying the accuracy of the PCFM.**

## 6.2 RTM vertical grids

The core radiative transfer calculations performed to produce both the PCFM training spectra and the sub-set of monochromatic points for individual scenes with the orbits, are performed by calling RFM to compute absorption coefficient profiles which are then input (together with surface and atmospheric scattering parameters) to VLIDORT. The vertical grid of the CHE global model output is used in the VLIDORT calculations. The hybrid sigma coordinate defines the pressure, $\boldsymbol{p_i}$, of model level *i* via the following equation:

$$p_i = A_i + B_i p_{surf}$$

**Equation 9**

Where $A_i$ and $B_i$ are parameters for a specific level, defined in table **Table 6-1**. The values of $A_i$ tend to zero near the surface and $B_i$ tend to unity, such that model levels follow the surface pressure near the ground. Note that l*evels* are defined as the boundaries between model *layers*, together with the top of atmosphere and the surface. There are 137 model layers and 138 levels in this case.

The model defines temperature and mixing ratio at the 137 layer mid-points. It is implicitly assumed that temperature and mixing ratios are constant across the layer. On the other hand, pressure is defined at the 138 levels / layer boundaries. Altitude above the surface is also calculated at the 138 levels (assuming hydrostatic equilibrium).

While VLIDORT requires layer optical depth to be input, The RFM line-by-line model used to simulate the gas absorption requires quantities to be defined on monotonically decreasing altitude *levels* (from space down to the surface). It is not trivial to faithfully represent the model profile in this way – the model defines fixed mixing ratios across discrete layers while RFM implicitly assumes continuous variation of values between the defined levels. The approach we adopt is to define profiles for RFM (only) on a new grid with twice the number of levels as there are model layers (274 levels). We then represent the discrete layers of the model on this grid, as illustrated in Figure 6-2. For mixing ratio and temperature, level i (0 to 273) in the new grid is populated as follows:

- For i is even (including 0): Model value for layer i/2.
- For i is odd: Model value for layer (i-1)/2.

I.e mixing ratios and temperature in consecutive pairs of levels are the same, Pressure and altitude at new-grid level i are set to capture the layer bounds as follows:

- For i is 0: 0198hPa (considered top-of-atmosphere; i.e. slightly lower than the pressure on the 2$^{nd}$ model level; altitude is set consistent with this assumed top-of-atmosphere pressure.
- For i is even (excluding 0): Model value (pressure or altitude) at level i/2 .
- For i is odd, excluding the last level: The model value at level (i+1)/2, reduced by 0.1% of the pressuredifference of the layer (i.e. between model pressure at level (i+2)/2 and level (i-1)/2). The small reduction is to avoid having identical altitudes/pressures in the grid.
- For i=237 (the last level): The surface pressure or zero for altitude.

For i is even, the bounded by levels i to i+1 is therefore very close to the ideal representation of the model layer i/2. In these “main” layers, only 0.1% of the model layer pressure thickness is omitted (nothing is omitted from the near-surface layer itself), For i is odd, the layers bounded by i to i+1 span the remaining 0.1% of the model layer pressure difference (completing the previous “main” layer). These thin “transition” layers implicitly define (within RFM) a linear transition in mixing ratio and temperature across this very thin layer.

RFM is run with profiles defined on this modified grid. From the output of this run, optical depths are computed for the full, original 137 model layers, by summing the optical depths in the consecutive pairs of “main” and associated thin “transition” layers. The last near-surface layer has no transition layer – in this case the “main” layer defines its full model layer.

VLIDORT is the run using the optical depths of the 137 layers (in combination with the corresponding layer scattering properties).

| Index | A (level) / Pa | B (level) | A (layer) / Pa | B (layer) | Index | A (level) / Pa | B (level) | A (layer) / Pa | B (layer) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.00000E+00 | 0 | 1.00018E+00 | 0 | 70 | 1.49756E+04 | 0.009261 | 1.52419E+04 | 0.010534 |
| 2 | 2.00037E+00 | 0 | 2.55130E+00 | 0 | 71 | 1.55083E+04 | 0.011806 | 1.57672E+04 | 0.013311 |
| 3 | 3.10224E+00 | 0 | 3.88416E+00 | 0 | 72 | 1.60261E+04 | 0.014816 | 1.62767E+04 | 0.016567 |
| 4 | 4.66608E+00 | 0 | 5.74703E+00 | 0 | 73 | 1.65273E+04 | 0.018318 | 1.67681E+04 | 0.020336 |
| 5 | 6.82798E+00 | 0 | 8.28747E+00 | 0 | 74 | 1.70088E+04 | 0.022355 | 1.72382E+04 | 0.024659 |
| 6 | 9.74697E+00 | 0 | 1.16762E+01 | 0 | 75 | 1.74676E+04 | 0.026964 | 1.76846E+04 | 0.02957 |
| 7 | 1.36054E+01 | 0 | 1.61072E+01 | 0 | 76 | 1.79016E+04 | 0.032176 | 1.81050E+04 | 0.035101 |
| 8 | 1.86089E+01 | 0 | 2.17973E+01 | 0 | 77 | 1.83084E+04 | 0.038026 | 1.84971E+04 | 0.041287 |
| 9 | 2.49857E+01 | 0 | 2.89857E+01 | 0 | 78 | 1.86857E+04 | 0.044548 | 1.88585E+04 | 0.048161 |
| 10 | 3.29857E+01 | 0 | 3.79325E+01 | 0 | 79 | 1.90313E+04 | 0.051773 | 1.91874E+04 | 0.055751 |
| 11 | 4.28792E+01 | 0 | 4.89174E+01 | 0 | 80 | 1.93435E+04 | 0.059728 | 1.94818E+04 | 0.064088 |
| 12 | 5.49555E+01 | 0 | 6.22380E+01 | 0 | 81 | 1.96200E+04 | 0.068448 | 1.97397E+04 | 0.073203 |
| 13 | 6.95206E+01 | 0 | 7.82082E+01 | 0 | 82 | 1.98594E+04 | 0.077958 | 1.99597E+04 | 0.083122 |
| 14 | 8.68959E+01 | 0 | 9.71558E+01 | 0 | 83 | 2.00599E+04 | 0.088286 | 2.01398E+04 | 0.093874 |
| 15 | 1.07416E+02 | 0 | 1.19421E+02 | 0 | 84 | 2.02197E+04 | 0.099462 | 2.02788E+04 | 0.105484 |
| 16 | 1.31426E+02 | 0 | 1.45352E+02 | 0 | 85 | 2.03379E+04 | 0.111505 | 2.03751E+04 | 0.117977 |
| 17 | 1.59279E+02 | 0 | 1.75309E+02 | 0 | 86 | 2.04123E+04 | 0.124448 | 2.04272E+04 | 0.13138 |
| 18 | 1.91339E+02 | 0 | 2.09654E+02 | 0 | 87 | 2.04421E+04 | 0.138313 | 2.04339E+04 | 0.145719 |
| 19 | 2.27969E+02 | 0 | 2.48754E+02 | 0 | 88 | 2.04257E+04 | 0.153125 | 2.03938E+04 | 0.161018 |
| 20 | 2.69540E+02 | 0 | 2.92980E+02 | 0 | 89 | 2.03618E+04 | 0.16891 | 2.03057E+04 | 0.177299 |
| 21 | 3.16421E+02 | 0 | 3.42702E+02 | 0 | 90 | 2.02495E+04 | 0.185689 | 2.01683E+04 | 0.19459 |
| 22 | 3.68982E+02 | 0 | 3.98287E+02 | 0 | 91 | 2.00871E+04 | 0.203491 | 1.99806E+04 | 0.212912 |
| 23 | 4.27592E+02 | 0 | 4.60104E+02 | 0 | 92 | 1.98740E+04 | 0.222333 | 1.97413E+04 | 0.232289 |
| 24 | 4.92616E+02 | 0 | 5.28515E+02 | 0 | 93 | 1.96086E+04 | 0.242244 | 1.94494E+04 | 0.252743 |
| 25 | 5.64413E+02 | 0 | 6.03877E+02 | 0 | 94 | 1.92902E+04 | 0.263242 | 1.91038E+04 | 0.274298 |
| 26 | 6.43340E+02 | 0 | 6.86542E+02 | 0 | 95 | 1.89175E+04 | 0.285354 | 1.87036E+04 | 0.296976 |
| 27 | 7.29744E+02 | 0 | 7.76856E+02 | 0 | 96 | 1.84897E+04 | 0.308598 | 1.82483E+04 | 0.320769 |
| 28 | 8.23968E+02 | 0 | 8.75156E+02 | 0 | 97 | 1.80069E+04 | 0.332939 | 1.77394E+04 | 0.345596 |
| 29 | 9.26345E+02 | 0 | 9.81773E+02 | 0 | 98 | 1.74718E+04 | 0.358254 | 1.71803E+04 | 0.371309 |
| 30 | 1.03720E+03 | 0 | 1.09703E+03 | 0 | 99 | 1.68887E+04 | 0.384363 | 1.65754E+04 | 0.397744 |
| 31 | 1.15685E+03 | 0 | 1.22123E+03 | 0 | 100 | 1.62620E+04 | 0.411125 | 1.59294E+04 | 0.424758 |
| 32 | 1.28561E+03 | 0 | 1.35469E+03 | 0 | 101 | 1.55967E+04 | 0.438391 | 1.52476E+04 | 0.452197 |
| 33 | 1.42377E+03 | 0 | 1.49770E+03 | 0 | 102 | 1.48985E+04 | 0.466003 | 1.45359E+04 | 0.479901 |
| 34 | 1.57162E+03 | 0 | 1.65054E+03 | 0 | 103 | 1.41733E+04 | 0.4938 | 1.38005E+04 | 0.50771 |
| 35 | 1.72945E+03 | 0 | 1.81348E+03 | 0 | 104 | 1.34278E+04 | 0.521619 | 1.30480E+04 | 0.53546 |
| 36 | 1.89752E+03 | 0 | 1.98681E+03 | 0 | 105 | 1.26683E+04 | 0.549301 | 1.22848E+04 | 0.562997 |
| 37 | 2.07610E+03 | 0 | 2.17076E+03 | 0 | 106 | 1.19013E+04 | 0.576692 | 1.15173E+04 | 0.59017 |
| 38 | 2.26543E+03 | 0 | 2.36560E+03 | 0 | 107 | 1.11333E+04 | 0.603648 | 1.07517E+04 | 0.616842 |
| 39 | 2.46577E+03 | 0 | 2.57156E+03 | 0 | 108 | 1.03702E+04 | 0.630036 | 9.99385E+03 | 0.642886 |
| 40 | 2.67735E+03 | 0 | 2.78887E+03 | 0 | 109 | 9.61752E+03 | 0.655736 | 9.24898E+03 | 0.66819 |
| 41 | 2.90039E+03 | 0 | 3.01776E+03 | 0 | 110 | 8.88045E+03 | 0.680643 | 8.52191E+03 | 0.692656 |
| 42 | 3.13512E+03 | 0 | 3.25843E+03 | 0 | 111 | 8.16338E+03 | 0.704669 | 7.81686E+03 | 0.716204 |
| 43 | 3.38174E+03 | 0 | 3.51111E+03 | 0 | 112 | 7.47034E+03 | 0.727739 | 7.13738E+03 | 0.738768 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 44 | 3.64047E+03 | 0 | 3.77598E+03 | 0 | 113 | 6.80442E+03 | 0.749797 | 6.48648E+03 | 0.760298 |
| 45 | 3.91149E+03 | 0 | 4.05321E+03 | 0 | 114 | 6.16853E+03 | 0.770798 | 5.86646E+03 | 0.780758 |
| 46 | 4.19493E+03 | 0 | 4.34287E+03 | 0 | 115 | 5.56438E+03 | 0.790717 | 5.27909E+03 | 0.800126 |
| 47 | 4.49082E+03 | 0 | 4.64498E+03 | 0 | 116 | 4.99380E+03 | 0.809536 | 4.72559E+03 | 0.818396 |
| 48 | 4.79915E+03 | 0 | 4.95952E+03 | 0 | 117 | 4.45738E+03 | 0.827256 | 4.20667E+03 | 0.835569 |
| 49 | 5.11990E+03 | 0 | 5.28644E+03 | 0 | 118 | 3.95596E+03 | 0.843881 | 3.72260E+03 | 0.851656 |
| 50 | 5.45299E+03 | 0 | 5.62567E+03 | 0 | 119 | 3.48923E+03 | 0.859432 | 3.27325E+03 | 0.866681 |
| 51 | 5.79834E+03 | 0 | 5.97721E+03 | 0 | 120 | 3.05727E+03 | 0.873929 | 2.85820E+03 | 0.880669 |
| 52 | 6.15607E+03 | 0 | 6.34151E+03 | 0 | 121 | 2.65914E+03 | 0.887408 | 2.47669E+03 | 0.893654 |
| 53 | 6.52695E+03 | 0 | 6.71941E+03 | 0 | 122 | 2.29424E+03 | 0.8999 | 2.12787E+03 | 0.905674 |
| 54 | 6.91187E+03 | 0 | 7.11187E+03 | 0 | 123 | 1.96150E+03 | 0.911448 | 1.81049E+03 | 0.916772 |
| 55 | 7.31187E+03 | 0 | 7.51964E+03 | 3.5E-06 | 124 | 1.65948E+03 | 0.922096 | 1.52351E+03 | 0.926988 |
| 56 | 7.72741E+03 | 0.000007 | 7.94338E+03 | 1.55E-05 | 125 | 1.38755E+03 | 0.931881 | 1.26540E+03 | 0.93637 |
| 57 | 8.15935E+03 | 0.000024 | 8.38394E+03 | 4.15E-05 | 126 | 1.14325E+03 | 0.94086 | 1.03488E+03 | 0.944962 |
| 58 | 8.60853E+03 | 0.000059 | 8.84246E+03 | 8.55E-05 | 127 | 9.26508E+02 | 0.949064 | 8.30750E+02 | 0.952807 |
| 59 | 9.07640E+03 | 0.000112 | 9.31954E+03 | 0.000156 | 128 | 7.34992E+02 | 0.95655 | 6.51527E+02 | 0.959951 |
| 60 | 9.56268E+03 | 0.000199 | 9.81433E+03 | 0.00027 | 129 | 5.68063E+02 | 0.963352 | 4.96238E+02 | 0.966433 |
| 61 | 1.00660E+04 | 0.00034 | 1.03253E+04 | 0.000451 | 130 | 4.24414E+02 | 0.969513 | 3.63445E+02 | 0.972296 |
| 62 | 1.05846E+04 | 0.000562 | 1.08506E+04 | 0.000726 | 131 | 3.02477E+02 | 0.975078 | 2.52480E+02 | 0.977575 |
| 63 | 1.11167E+04 | 0.00089 | 1.13884E+04 | 0.001122 | 132 | 2.02484E+02 | 0.980072 | 1.62293E+02 | 0.982307 |
| 64 | 1.16601E+04 | 0.001353 | 1.19358E+04 | 0.001673 | 133 | 1.22102E+02 | 0.984542 | 9.24414E+01 | 0.986521 |
| 65 | 1.22115E+04 | 0.001992 | 1.24892E+04 | 0.002425 | 134 | 6.27813E+01 | 0.9885 | 4.28086E+01 | 0.990242 |
| 66 | 1.27669E+04 | 0.002857 | 1.30458E+04 | 0.003414 | 135 | 2.28359E+01 | 0.991984 | 1.32969E+01 | 0.993493 |
| 67 | 1.33247E+04 | 0.003971 | 1.36030E+04 | 0.004674 | 136 | 3.75781E+00 | 0.995003 | 1.87891E+00 | 0.996316 |
| 68 | 1.38813E+04 | 0.005378 | 1.41567E+04 | 0.006256 | 137 | 0.00000E+00 | 0.99763 | 0.00000E+00 | 0.998815 |
| 69 | 1.44321E+04 | 0.007133 | 1.47039E+04 | 0.008197 | Surface | 0.00000E+00 | 1 | | |

**Table 6-1: Hybrid sigma levels used by CHE global. There are 137 model layers and hence 138 model levels, which delimit the contiguous layers. The bottom level is at the Earth surface. Each row in the table gives a layer index, together with the hybrid sigma A and B parameters for the level at the top of that layer and the values for the mid-pressure of the layer. The parameters for the level lower boundary are those for the upper level of the next layer down.**

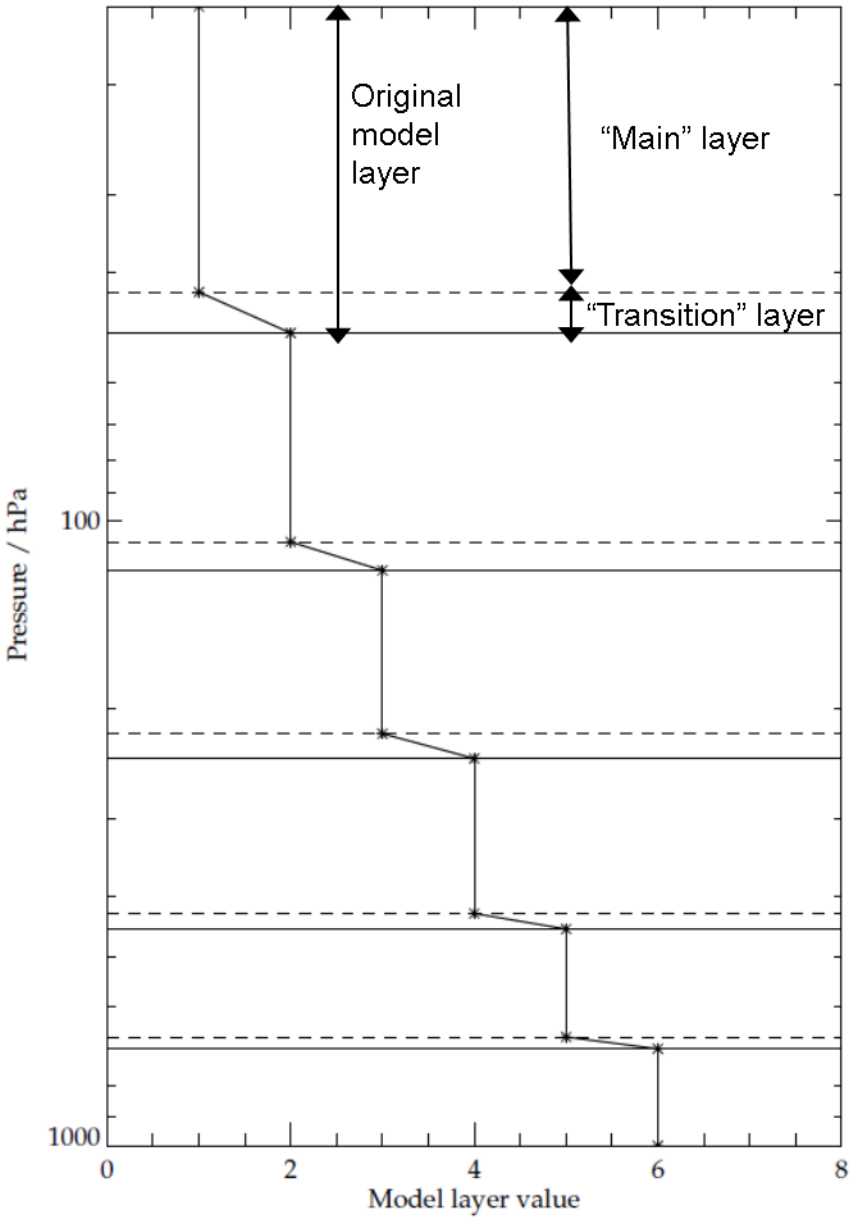



**Figure 6-2: Schematic illustration of the approach to represent model layers in the RFM calculations. The pressure thickness of the model layers and relative thickness of the transition layers are exaggerated for clarity.**

## 6.3 Modelling surface altitude variations

The model profiles represent the effects of surface topography at the spatial resolution of the given model (CHE global, regional or CAMS). However, the L1 orbit files from Eumetsat include surface altitude for each CO2M observation derived from a high resolution digital elevation model which has relatively high spatial resolution, commensurate with the CO2M sampling. The following equation is used to adapt the assumed surface pressure (in hPa) to the more highly resolved terrain information:

$$p_{surf}^{HR} = p_{surf}^{CHE} \frac{p_*^{HR}}{p_*^{CHE}}$$

Equation 10

Where

- $p_{surf}^{HR}$ is the adapted, high resolution surface pressure (eventually used in the radiative transfer).
- $p_{surf}^{CHE}$ is the surface pressure from CHE global (interpolated to the location of the CO2M observation).
- $p_*^{HR}$ is an estimate of the mean surface pressure based simply on an assumed fixed pressure/altitude scale-height and the surface altitude in the Eumetsat orbit files, $z_{surf}^{HR}$, (in km) using the following equation:

$$p_*^{HR} = 10^{(3-\frac{z_{surf}^{HR}}{16})}$$

Equation 11

- $p_*^{CHE}$ is computed in the same way as $p_*^{HR}$, from the relatively coarsely resolved surface altitudefrom the CHE model, $z_{surf}^{CHE}$.

The modified surface pressure, $p_{surf}^{HR}$, is used as the surface pressure in expanding the hybrid sigma vertical coordinate of the model profiles. This includes the model profiles from the CHE "zoom" model fields over Europe and Berlin and the CAMS global fields, to enable these to be interpolated, in pressure, onto the CHE global vertical grid, without any further complication around matching the surface altitude assumed by the different models. *This is the only adaptation to the high-resolution surface altitude performed here.*

This approach is reasonable in making the implicit assumption that mixing ratio is conserved along the hybrid sigma coordinates, following the surface. However the approach is not very realistic for the temperature, aerosol and cloud profiles (or humidity where saturation may occur). For current purposes more sophistication is not required: In any case the assumed "true" profiles, with this adaptation, are defined in the scenario files. It is, however, noted that L2 algorithms may need to use a more sophisticated approach in this respect to obtain accurate results in the vicinity of mountains[1].

[1] Although not done in this work, when the $z_{surf}^{HR} > z_{surf}^{CHE}$, it may be more appropriate to simply "cut" the model profiles at the high-resolution surface altitude. This would reflect e.g. lapse rate decrease in temperature over a mountain and tendency for aerosol to stay in a valley. However, since the model reflects the lower resolution mean altitude, it is often the case around mountains that $z_{surf}^{HR} < z_{surf}^{CHE}$ and in this case it is less clear how to extrapolate below the model surface.

Note that the radiative transfer calculations themselves do not take into account detailed variations in surface altitude (other than via the adjustment of the pressure profile described above). These are performed assuming an altitude grid with origin at the surface, with geometric layer thicknesses calculated using the hydrostatic equation from the model pressure and temperature. In evaluating the hydrostatic equation, the following equation is used to account for the variation of the acceleration due to gravity, $g(\phi, z)$, with latitude, $\phi$, and altitude, $z$:

$$g(\phi,z) = GM\left(\frac{1}{(R_e(\phi)+z)^2} - 3J_2R_{eq}^2\frac{3\sin^2\phi - 1}{2(R_e(\phi)+z)^4}\right) - \Omega^2(R_e(\phi)+z)(1-\sin^2\phi)$$

Equation 12

Where

- G: Gravitational constant = 6.670× $10^{-11}$ Nm$^2$/kg$^2$.
- M: Mass of the Earth = 5.976× $10^{24}$ kg
- $J_2$: Empirical constant = 1082.64× $10^{-6}$
- $\Omega$ : Earth angular velocity= 7.2722052× $10^{-5}$ radians/second
- $R_{eq}$: Earth radius at equator = =6378.388 km
- $R_e(\phi)$: Earth radius at latitude, $\phi$, which is given by

$$R_e(\phi) = \frac{R_{eq}(1-\alpha)}{\sqrt{\sin^2\phi + (1-\alpha)^2\cos^2\phi}}$$

Equation 13

Where $\alpha$ is the Earth ellipticity = 0.003367.

When calculating the altitude of each model level, the hydrostatic equation is integrated from the surface upwards, assuming $z = 0$ at the surface. For the lowest layer (with lower bound at the surface), $g$ is evaluated using $z = 0$ and this is used in the hydrostatic equation to compute the layer geometric thickness and hence altitude of the upper boundary. For successive layers, the integrated altitude of the previous layers upper boundary is used to evaluate $g$ and hence the layer thickness. In principle the surface altitude above the geoid should be used to evaluate $g$ for the lowest layer, however this is not expected to have any significant impact on the derived layer geometric thicknesses for current purposes.

## 6.4 Simulation of polarised radiances

CO2M is assumed to be polarisation insensitive and the incoming solar irradiance is unpolarised. It is nevertheless important to model the effects of polarisation in simulating the CO2M observed radiances, since both the surface and atmospheric scattering polarises the light, and the degree of polarisation affects subsequent scattering events. Effects of scattering on the total intensity are negligible where only single scattering contributes to the TOA radiance (e.g. in the far-UV) or the order of scattering is very high. The error introduced by neglecting scattering tends to be largest where (on average) photons undergo around two Rayleigh scattering events (which is most likely to occur in the CO2M VIS band).

The PCFM can simulate vector or scalar RT simply by using vector or scalar RT to compute the monochromatic radiances at the selected wavelengths. The vector radiative transfer calculations needed to simulate polarisation effects are ~16 times more computationally expensive than scalar equivalents. Even with the PCFM approach it

proved to be unfeasible to perform simulations for this study using vector RT to simulate the (up to) 1000 monochromatic points chosen to drive the PCFM for each CO2M band. A faster approach is implemented as follows:

- Scalar radiances, $I_{\lambda:s1000}$, are simulated using Ns monochromatic points (per band)
- Vector and scalar radiances are simulated using a smaller number of Nv=100 monochromatic points, $I_{\lambda:v100}$ and $I_{\lambda:s100}$.
- Accurate vector radiance are simulated using the ratio of the vector:scalar radiance simulated based on Nv points:

$$I_{\lambda:v} = I_{\lambda:s1000} \frac{I_{\lambda:v100}}{I_{\lambda:s100}}$$

**Equation 14**

The approach was tested during the Sentinel 4/5 work (for a sub-set of geophysical conditions) radiances simulated applying the PCFM to vector calculations performed for all monochromatic points. Errors were found to be substantially smaller than the error of the PCFM itself.

Polarisation is included in all CO2M bands, since it was found to be non-negligible in all bands. See results illustrating the effect of including/excluding polarisation in section 7.

## 6.5 Ring effect

Rayleigh scattering by air molecules is not a purely elastic process: Photons are absorbed and subsequently re-emitted. Most of the photons are re-emitted at the same wavelength, but a fraction are re-emitted at a different wavelength, having gained or lost energy. The process is quantised so that a scattered monochromatic source will exhibit a line spectrum composed of

- The central Cabannes line, itself consisting of the Gross line (corresponding to elastic scattering) and the Brillouin lines (associated with translational energy transitions).
- The rotational Raman spectrum, which consists of a set of spectral lines with a range of displacements from the source wavelength.

This process leads to absorption features in a measured spectrum being “filled-in” by photons from brighter parts of the spectral range, referred to as the Ring effect.

Simulation of Ring effect here is performed (in an approximate manner), as a correction to the simulated (instrument resolution) spectra produced by the PCFM, largely following the approach of [RD24], using the Raman spectroscopic parameters of [RD25]. The approach is similar to that used in the RAL ozone profile scheme (applied in the ESA Ozone CCI project), where Ring effect is computed by convolving simulated radiances with the Raman spectrum:

$$\mathrm{I}_{\lambda}^{\mathrm{R}} \;=\; \mathrm{f}\;\mathrm{I}_{\lambda} \;+\; (1 - \mathrm{f})\,(\mathrm{I}_{\lambda} * \mathrm{R})$$

**Equation 15**

Where

- $I_{\lambda}^{R}$ is the simulated radiance including Ring effect
- $I_{\lambda}$ is the modelled radiance (at instrument resolution and sampling)
- $* R$ denotes convolution with the (normalised) Raman line spectrum.
- $f$ is the fraction of signal which is (inelastically) Raman scattered

Here the assumption is made that the contribution from photons which are Raman scattered more than once can be neglected. It is also assumed that Raman scattering takes place after absorption in the atmosphere. This leads to an accurate model of the filling-in of solar absorption features, but tends to over-estimate the filling-in of the absorption features from atmospheric molecules.

For single scattering in a purely Rayleigh atmosphere (and no surface reflection), $f = f_0$, which can be calculated based on laboratory measurements, and is a function of the scattering geometry (and temperature). In a multiple-scattering atmosphere, over a reflecting surface, $f$ may be larger than $f_0$ in the case of multiple Rayleigh scattering, or smaller, where the relative contribution of Rayleigh scattering is small (e.g. over high-altitude optically thick cloud).

In the RAL retrieval scheme, $f$ is derived from the observations themselves (by a scale factor for the single-scatter value being included in the retrieval state vector).

For the CO2M simulation we calculate $f$ by:

$$f = f_0 \frac{\partial I_\lambda}{\partial \omega_R}$$

**Equation 16**

Where $\omega_R$ is the Rayleigh single scattering albedo. The derivative is calculated by a seperate run of the PCFM (in scalar mode) in which the Rayleigh single scatter albedo is set to 0.99 (instead of 1 in the nominal case). In a single run of VLIDORT, both the single and multiple scattering contributions to the TOA radiance are provided. These are used to compute factor $f$ separately for the two contributions, in order to take into account the scattering angle dependence of $f_0$. In the single-scattering case, $f_0$ is evaluated for the actual scattering angle. For the multiple scattering contribution and effective scattering angle of 135° is assumed. Single and multiple-scattering contributions to the total Ring-corrected scalar radiance are then computed separately using Equation 15, and summed to obtain the single-scatter Ring-corrected radiance, $I^R_{\lambda:s}$.

To save computational effort, The Ring effect correction is applied using the Ring-filling in spectrum derived from the scalar simulation computed for the first 100 monochromatic points only:

$$k_\lambda = \frac{I^R_{\lambda:s100}}{I_{\lambda:s100}} - 1$$

**Equation 17**

This is used to estimate the vector result for all monochromatic points:

$$I^R_{\lambda:v} = (k_\lambda + 1) I_{\lambda:v}$$

**Equation 18**

Testing of the approach during the S4/5 projects showed that the predicted filling-in has similar variation with view / solar geometry, cloud-cover and surface properties compared to the variation inferred from GOME observations via the scaling parameter retrieved in the RAL ozone scheme.

## 6.6 Training scenarios

For this study, training scenarios have been constructed based on the scenarios used in the previous S4/5 work, which in turn was based on the scenarios used in the ESA S5 NIR study.

As in the previous work, 2723 geophysical locations are defined, each simulated with and without cloud (5446 spectra, per band in total). The following aspects remain identical to the S4/5 work:

- Location of each scene (see Figure 6-3).
- Cloud fields: These were defined in the S5 NIR study to (reasonably comprehensively) sample variability of cloud height and optical depth (see [RD21] and Figure 6-4).
- View/solar geometry: This was based on sampling the globe along S5 orbits (this covers the range of view zenith angles sampled by CO2M even when pitched viewing is enabled). See Figure 6-5 and Figure 6-6.

Other fields have been redefined for this study by sampling the same input data sets used for the main simulations to the training scenario locations (for 3 July 2015, as used for the main orbit simulations).

For most species the input fields are simply interpolated to the training scenario location. However, for f CO2, CH4 and CO, we adopt the following approach to improve sampling of the variability of each gas:

- All CHE global profiles within the lat/lon box associated with each training scenario location are identified (these span 2.8 degrees of latitude/longitude and so contain 784 CHE global profiles).
- In each box, the minimum, median and maximum total column amount is found from all theseprofiles.
- A random number generator is used to decide whether to use the profile associated with the minimum, medium or maximum total column for this particular training scenario location.

Some examples of the training scenario fields are illustrated Figure 6-7 and Figure 6-8. Results from the radiative transfer calculations performed for the training scenarios are illustrated in Figure 6-9 and Figure 6-10.

The training scenario data is written to a netcdf file in the same format as the main scenario files generated from the L1 orbit files.

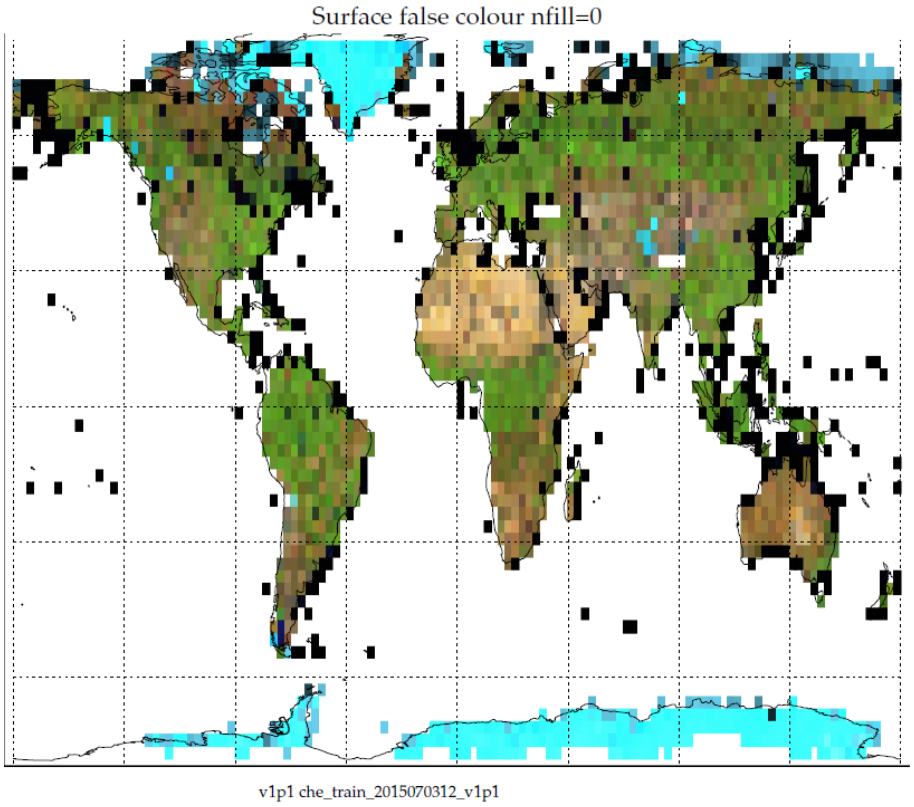



**Figure 6-3: False colour image illustrating MODIS isotropic surface reflectance spectral variations in the PCFM training set.**

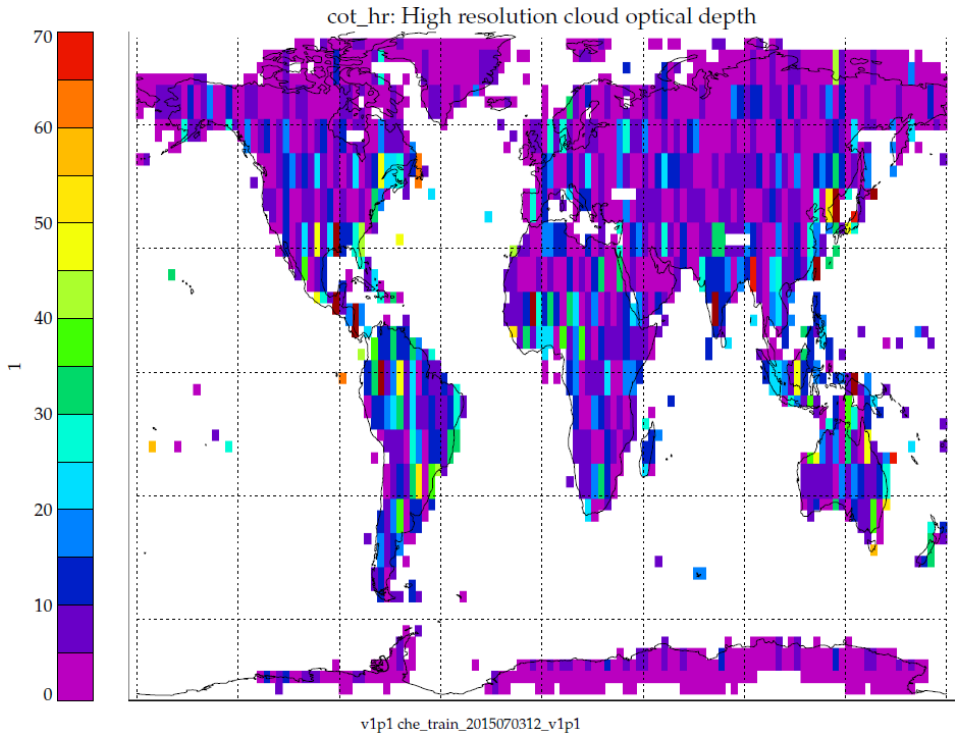



**Figure 6-4: Cloud optical depth from the PCFM training set (from S5 NIR study scenarios). The same optical depth is used in 3 neighbouring cells in the N-S direction. In each of these 3 cases the cloud height is varied (set to the 10, 50 and 90$^{th}$ percentile height from a CALIPSO based climatology at for the given latitude).**

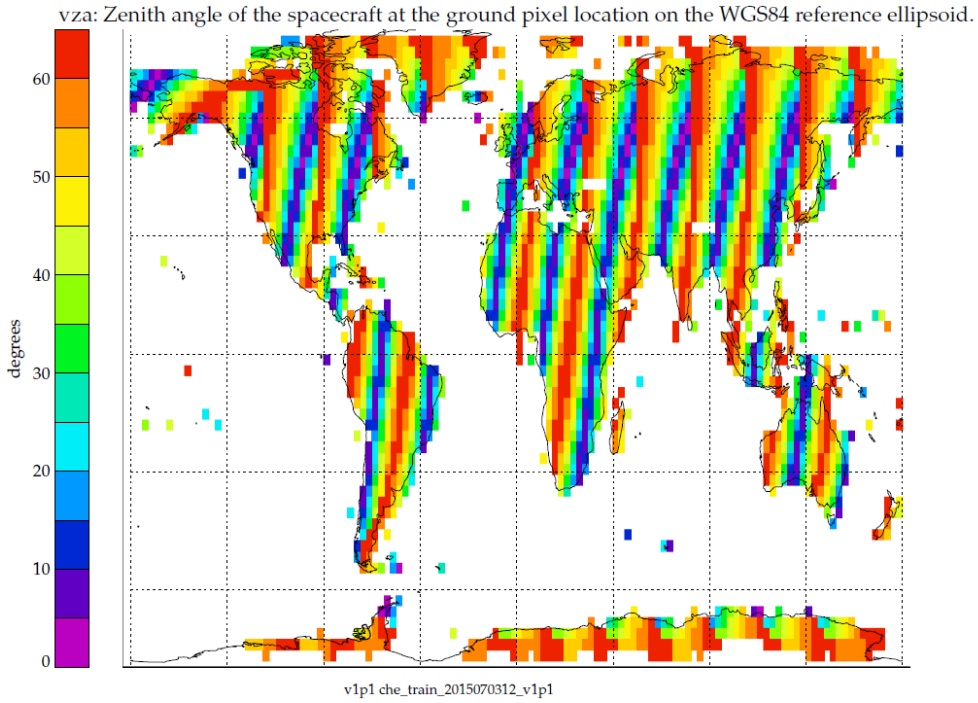



**Figure 6-5: View zenith angle assumed in the training cases.**

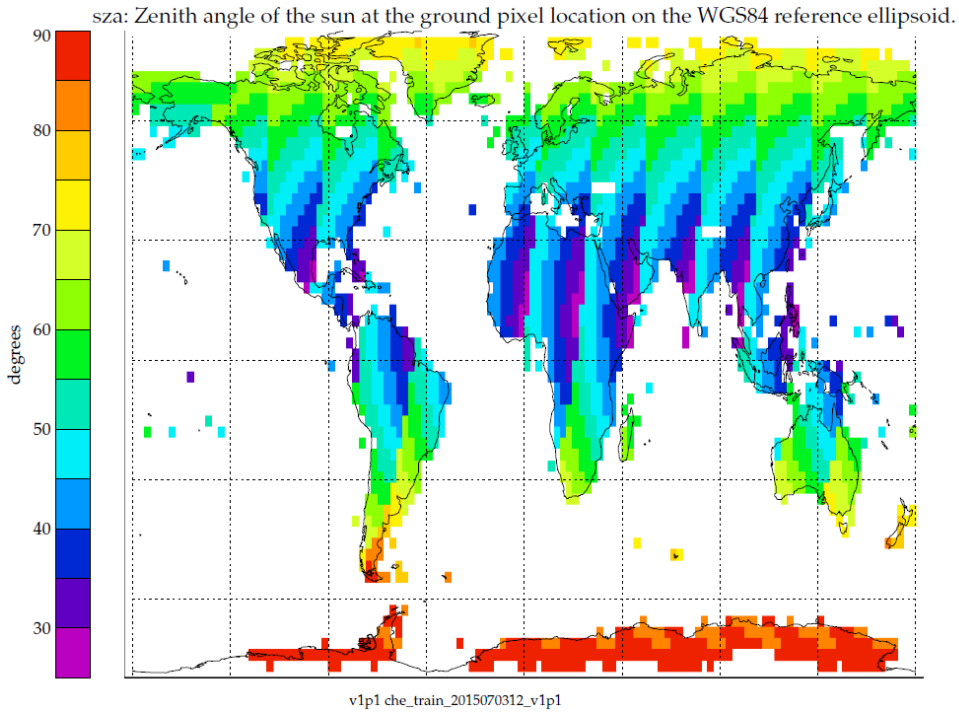



**Figure 6-6: Solar zenith angle assumed in the training cases.**

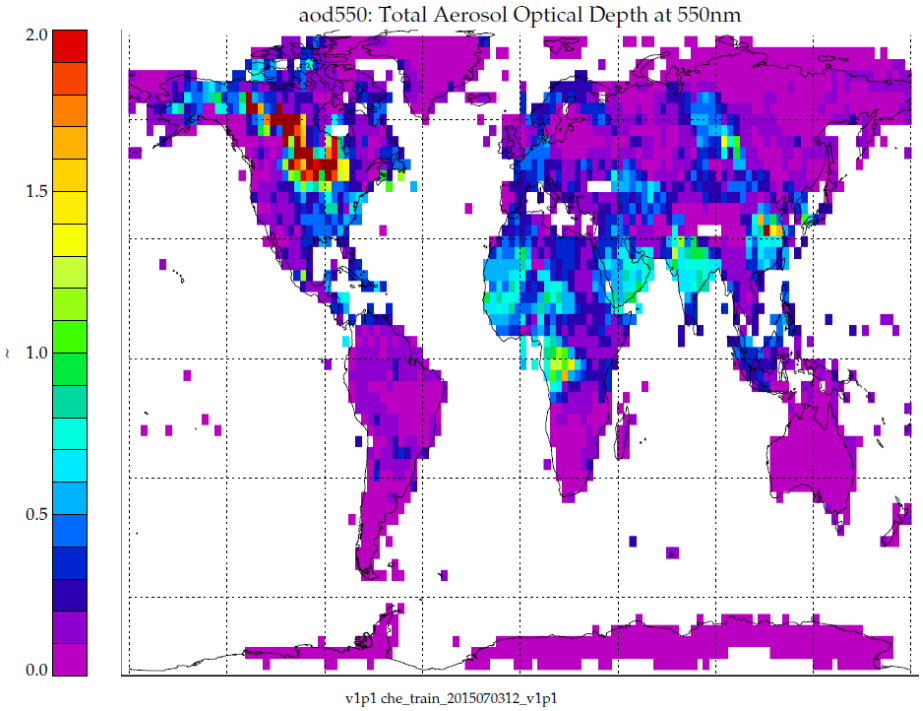



**Figure 6-7: Total aerosol optical depth in the PCFM training set (based on CAMS)**

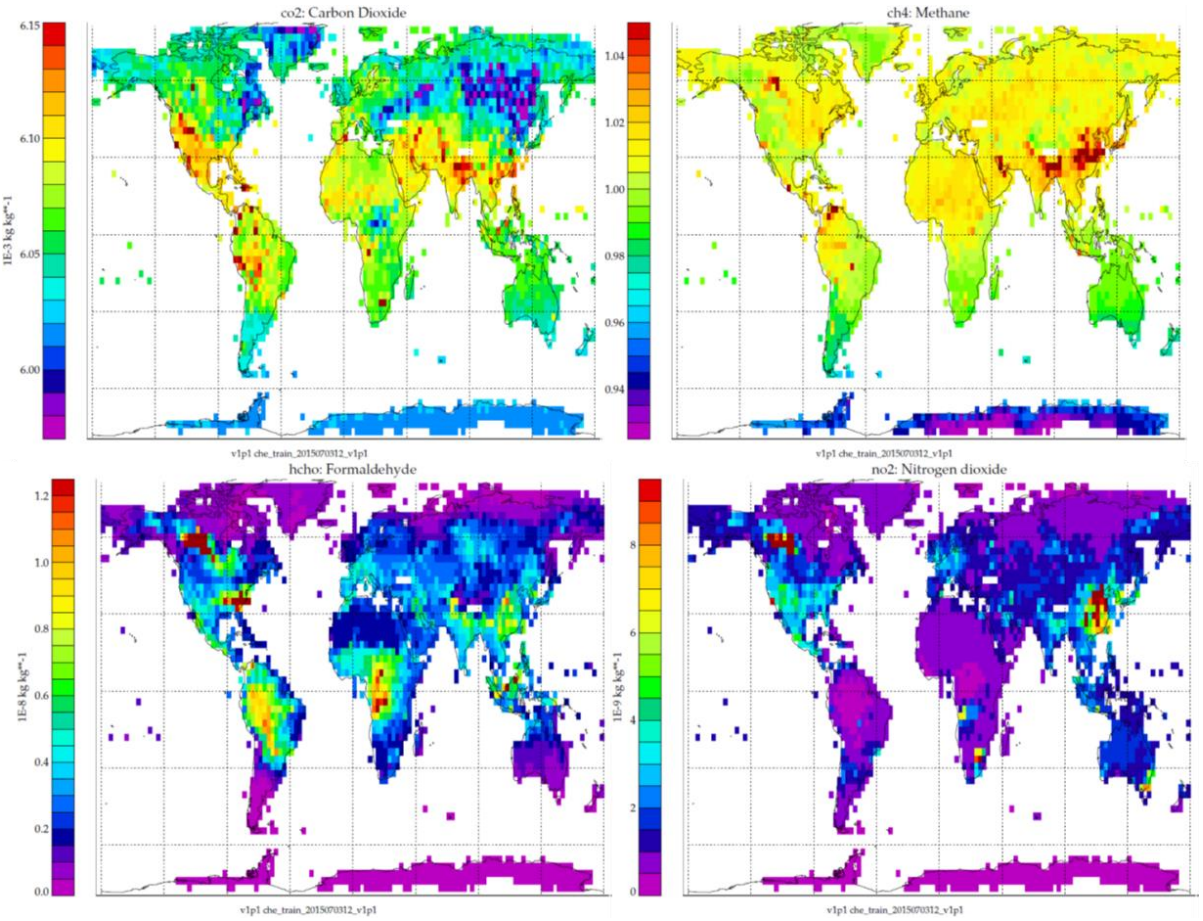



**Figure 6-8: Total column average mass mixing ratios of main gases used in the simulations (based on CHE global for CO2 and CH4 and CAMS for formaldehyde and NO2).**

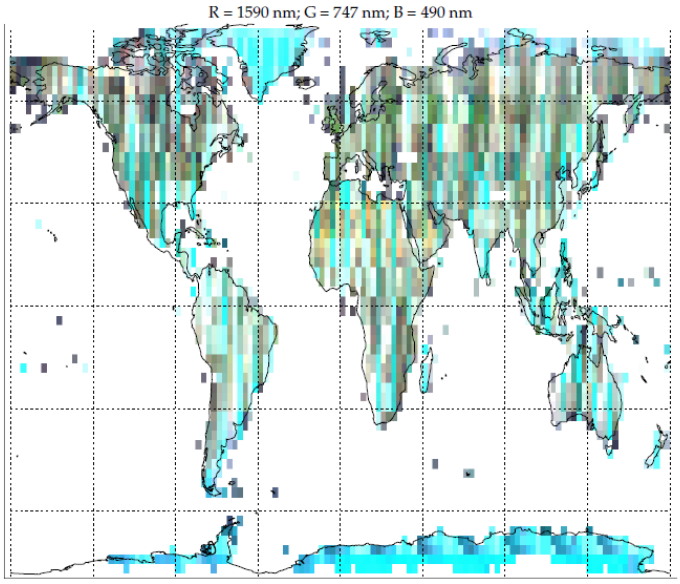



**Figure 6-9: False colour image from simulated window channel reflectances sampled from the radiative transfer simulations or the VIS, NIR and SWIR1 bands.**

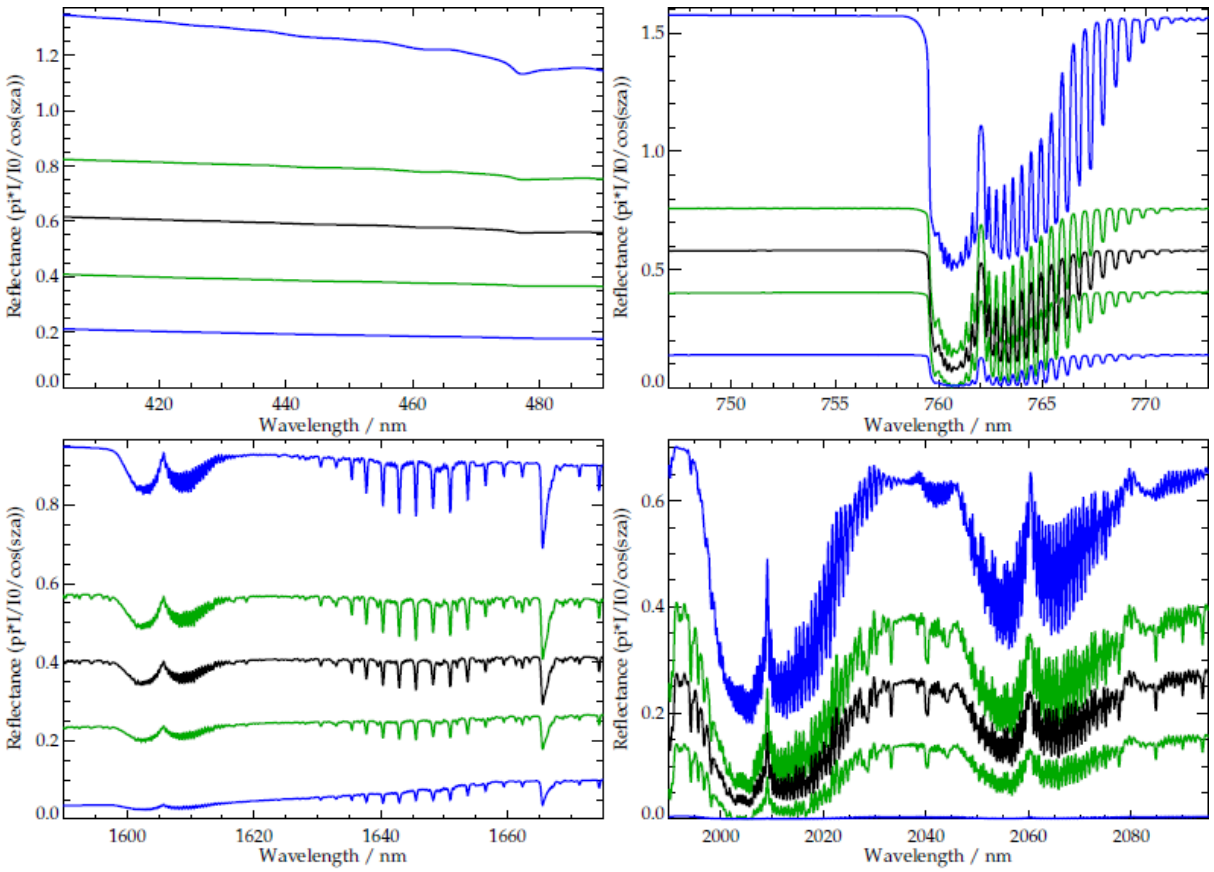



**Figure 6-10: Summary of simulated reflectance spectra in all 4 bands for all training scenarios. Black lines show the mean spectra; Green lines show the mean +/- the standard deviation; blue lines show the minmum and maximum spectra.**

## 6.7 PCFM performance for CO2M channels

The PCFM is trained using a randomly selected subset of 80% of the training scenarios. The remaining 20% are used as independent cases against which to confirm the accuracy of the PCFM. Furthermore the PCFM is trained to reproduce the ratio of the convolved radiance in all channels to the monochromatic radiance in the first selected monochromatic wavelength. This wavelength always corresponds to a window channel, so this means the PCFM training optimises the ratio of the spectra to its spectral background level. I.e. the relative accuracy of spectral features is optimised, irrespective of the amplitude of the overall spectrum. In this way, spectra for low albedo or high solar zenith angle are weighted similarly to those for bright scenes. This normalisation is used as it retains the perfectly linear relationship between monochromatic spectra and the convolved result.

The accuracy of the PCFM is assessed by examining the difference between PCFM reconstructed spectra and the reference spectra (in the normalised radiance unit) for the corresponding scenario. This is done both for the training and independent sets of spectra. The regression is performed for different numbers of monochromatic points, with a view to choosing the smallest number of points which provides acceptable accuracy.

Results are summarised for each band in Figure 6-11 to Figure 6-15. Different coloured lines show the standard deviation in the difference between PCFM and reference radiance, for a different number of monochromatic points used in the PCFM. The PCFM behaves well for all bands. Accuracy is comparable for training and independent sets and an accuracy of a few times $10^{-4}$ (in the normalised radiance unit) is achieved with a few 100 to 1000 points. These results are for the final settings which assume spectrally independent surface reflectance within band (varying only between bands). As discussed in section 5.4, results were significantly worse in SWIR-2 when reflectance was allowed to vary following the Adam database. In that case, the standard deviation against the independent set remained at a level of a few times $10^{-3}$ even when 2000 monochromatic points are used. Not that for the NIR band, results are shown for a PCFM with (Figure 6-13) and without (Figure 6-12) included. Results are comparable in both cases, so we include the effect of SIF directly in the PCFM.

Based on these results it was decided to use the following number of monochromatic points in the PCFM for the full orbit simulations:

- VIS: 700 points
- NIR: 700 points
- SWIR-1:200 points
- SWIR-2:1000 points

Figure 6-16 to Figure 6-20 show maps of the standard deviation over all spectral samples in the difference between PCFM and reference calculations, split between cloudy and clear cases, for the number of monochromatic points which were finally selected. These generally emphasise that in most scenes the PCFM accuracy is better than indicated by the spectral plots of standard deviation over all locations, i.e. larger errors occur in some isolated scenes, typically where albedo is very low and/or solar zenith angle is very high.

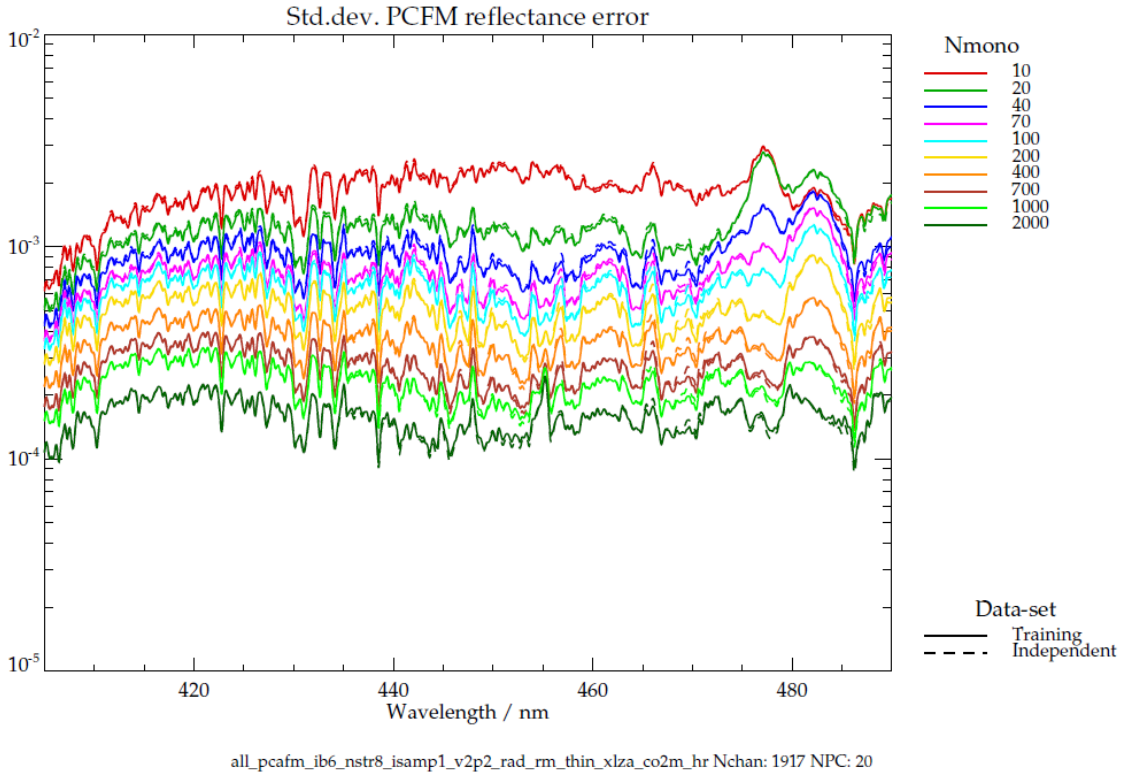



**Figure 6-11: Spectral standard deviation over all scenarios for the VIS, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

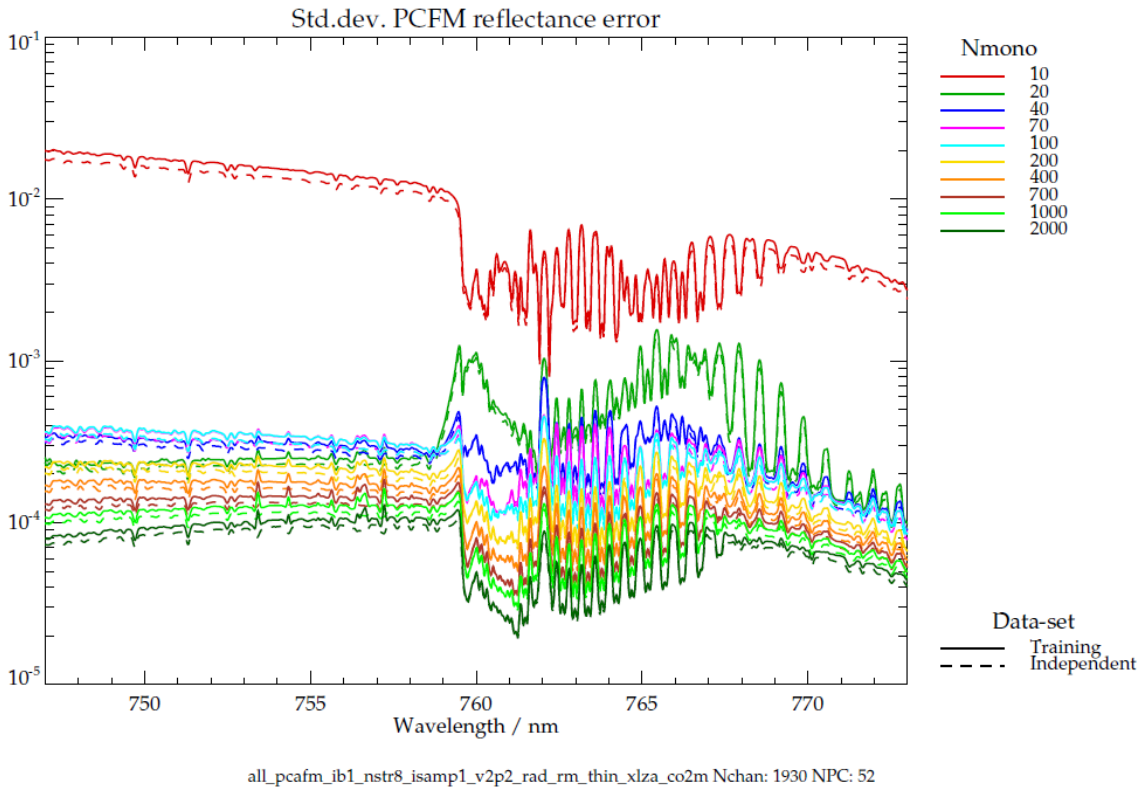



**Figure 6-12: Spectral standard deviation over all scenarios for the NIR (without SIF included), considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

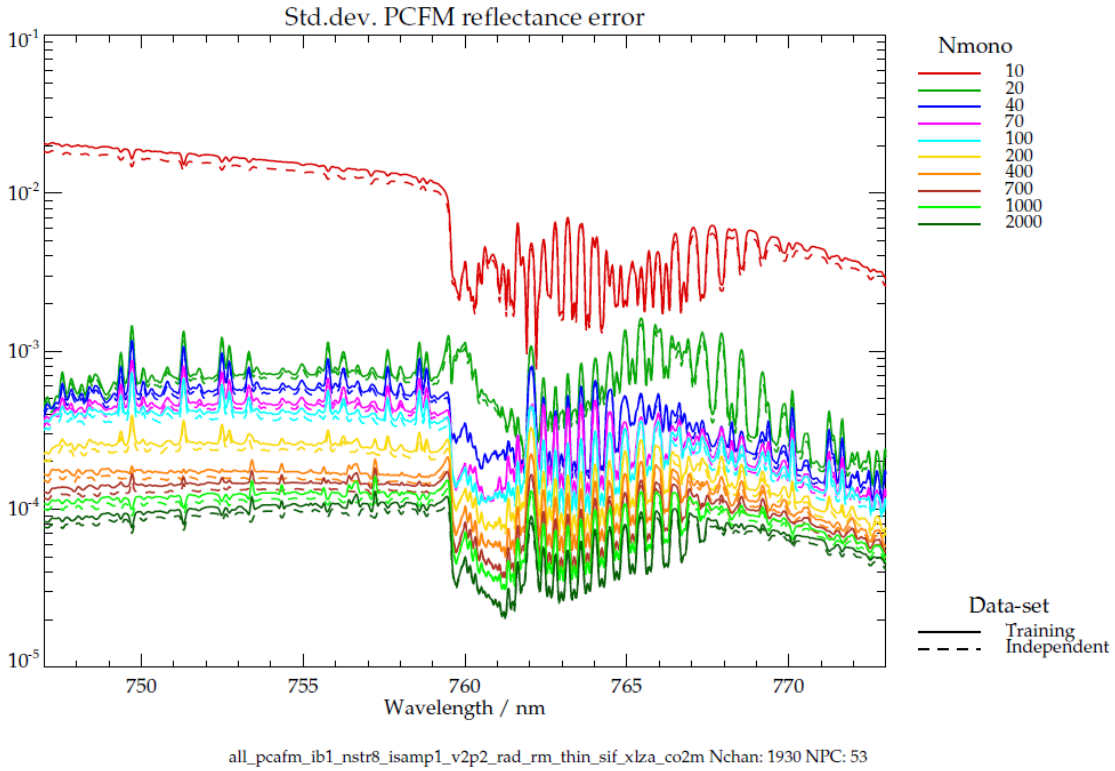



**Figure 6-13: Spectral standard deviation over all scenarios for the NIR band (with SIF included), considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

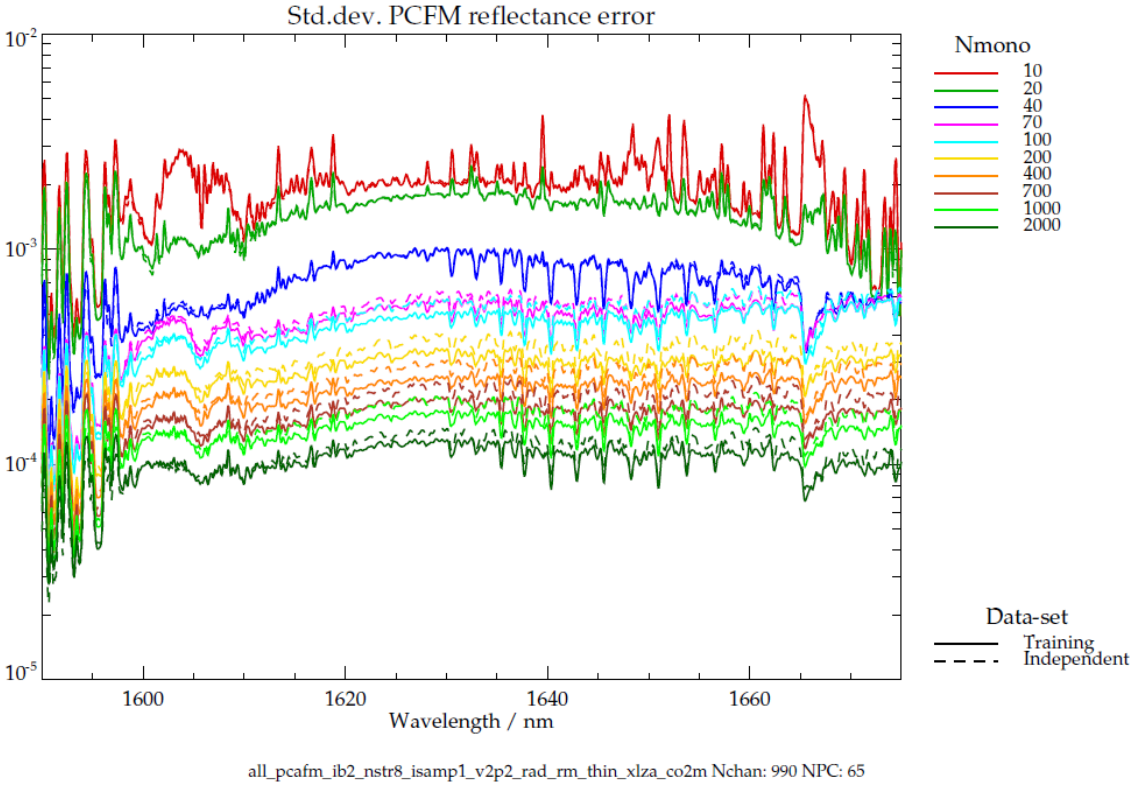



**Figure 6-14: Spectral standard deviation over all scenarios for the SWIR-1 band, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

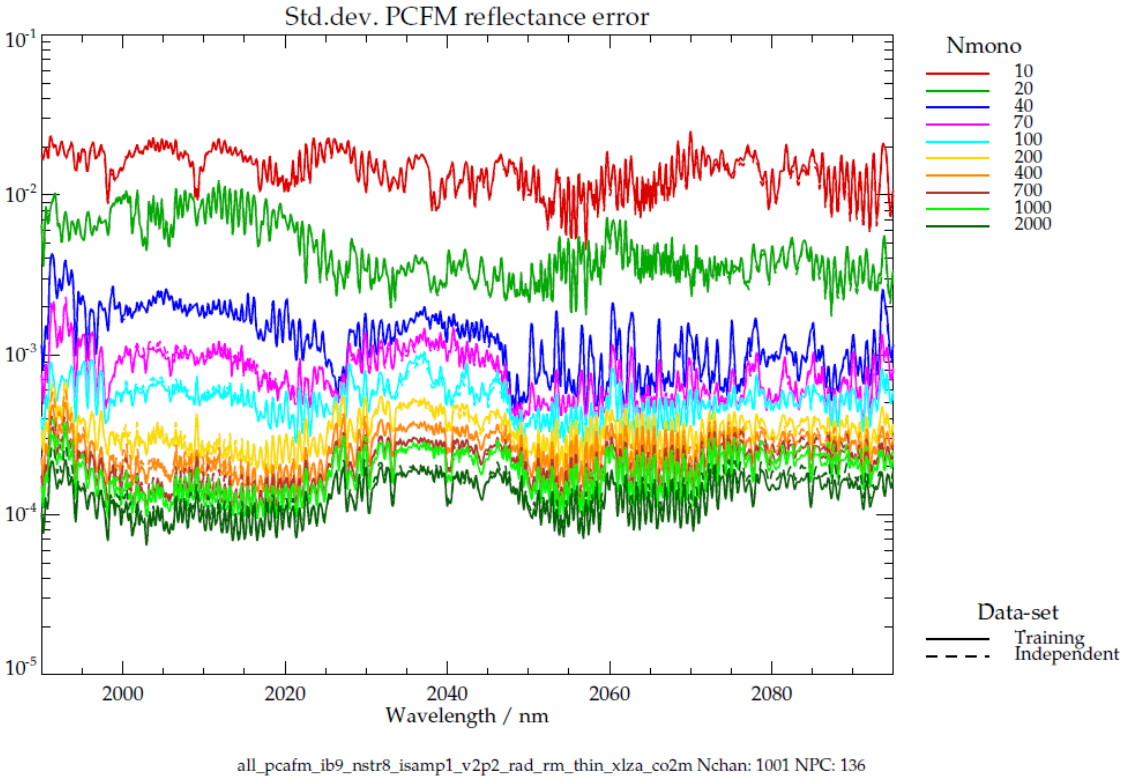



**Figure 6-15: Spectral standard deviation over all scenarios for the SWIR-2 band, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

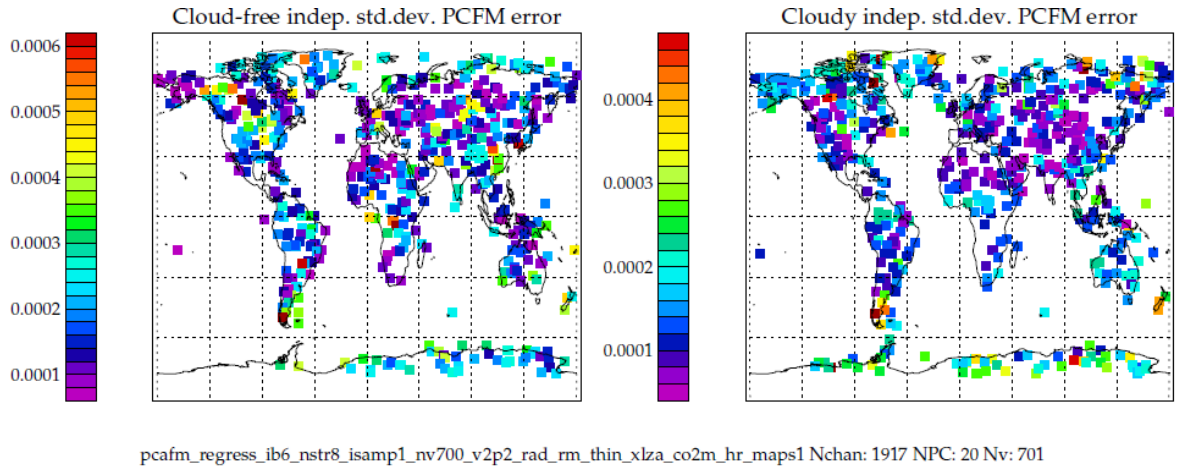



**Figure 6-16: Maps, for the VIS band, of the standard deviation (over all spectral samples) in differences between PCFM (with 700 monochromatic points) and reference simulations for the independent scenes from the training set (split into cloud and clear cases).**

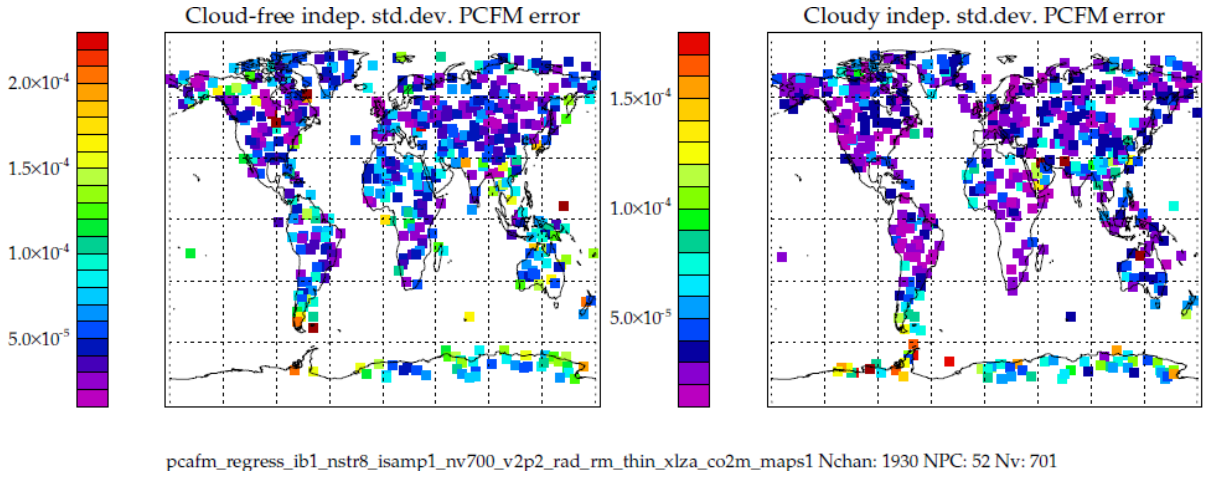



**Figure 6-17: Maps, for the NIR band (without SIF), of the standard deviation (over all spectral samples) in differences between PCFM (with 700 monochromatic points) and reference simulations for the independent scenes from the training set (split into cloud and clear cases).**

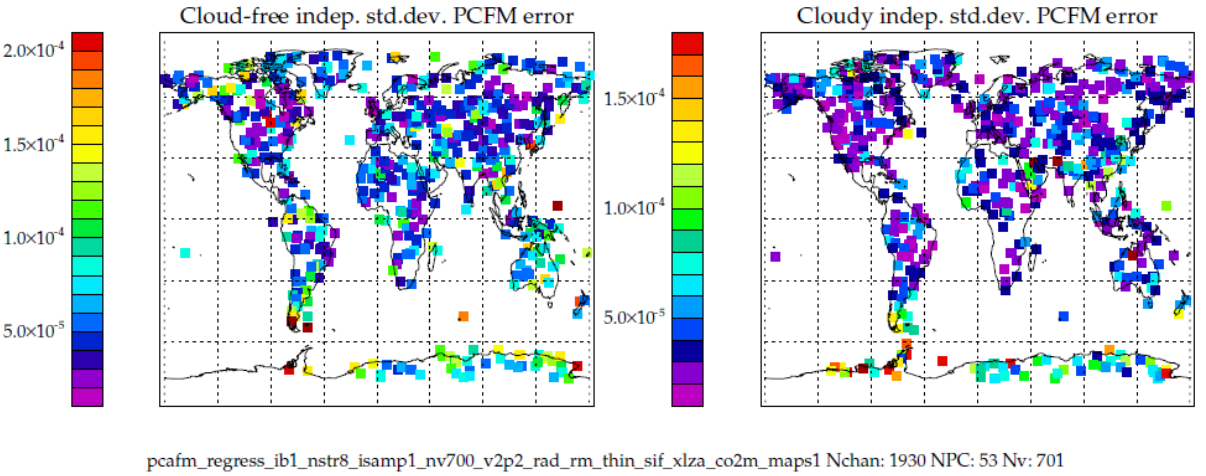



**Figure 6-18: Maps, for the NIR band (with SIF), of the standard deviation (over all spectral samples) in differences between PCFM (with 700 monochromatic points) and reference simulations for the independent scenes from the training set (split into cloud and clear cases).**

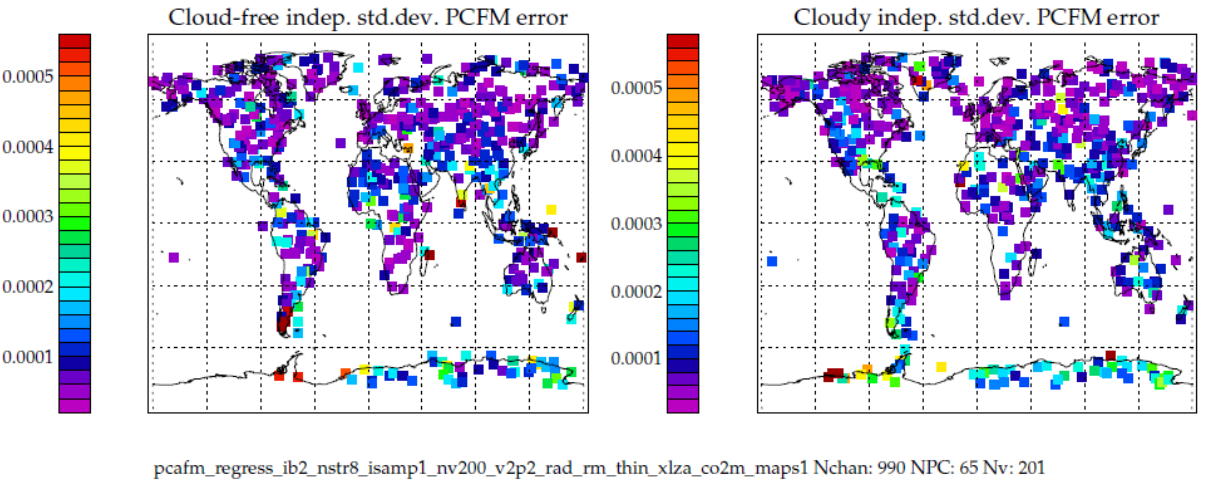



**Figure 6-19: Maps, for the SWIR-1 band, of the standard deviation (over all spectral samples) in differences between PCFM (with 200 monochromatic points) and reference simulations for the independent scenes from the training set (split into cloud and clear cases).**

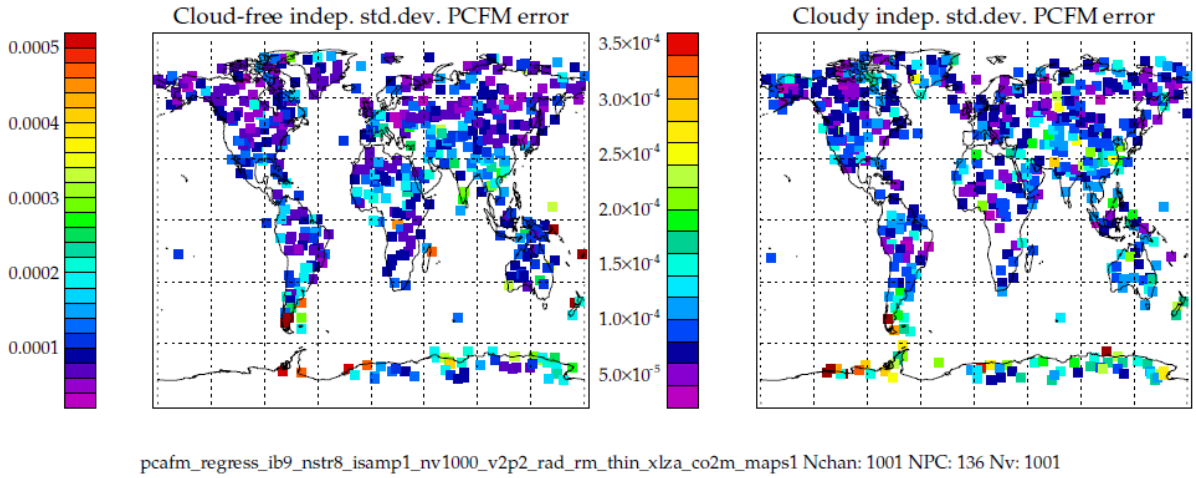



**Figure 6-20: Maps, for the SWIR-2 band, of the standard deviation (over all spectral samples) in differences between PCFM (with 1000 monochromatic points) and reference simulations for the independent scenes from the training set (split into cloud and clear cases).**

## 6.8 Approach to running full orbit simulations

Even with the PCFM approach, the computational demands of this project are extremely large. Processing of orbits requires large number of jobs to be executed on a high performance computing cluster (in this case the Jasmin facility at RAL, https://jasmin.ac.uk). Most of the computational effort is expended in the monochromatic radiative transfer calculations performed for each scene. For a given granule these calculations, carried out by many individual jobs on the cluster, can take weeks to perform. Once all calculations are complete, then the L1 file can be produced, using the PCFM to reconstruct full spectra from the sub-set of monochromatic points. It is important to ensure that all the individual jobs are complete and correct at the time the L1 file is produced. Finding missing and/or incorrect data only at the point the L1 file is made can lead to long delays in the production of the final dataset, particularly if errors are found which necessitate re-running the radiative transfer calculations.

The processing logic to generate the L1 orbit files is illustrated in Figure 6-21. It is preferable for processing efficiency (and simplicity) on Jasmin to ensure that individual batch jobs run in less than 24 hours (on a single core). This could be easily achieved by running jobs for individual ground-pixels (which would typically take a few minutes), however this would result in a huge number of individual output files, which leads to relatively inefficient post-processing e.g. when aggregating results for L1 file production or when reading output to check completeness of results. In practise, the choice was made to run individual jobs which process all (110) ground pixels in a specific scanline (within a specific orbit granule). Each of these jobs processes a specific "batch" of monochromatic points from a single NO2I/COI band, running the core radiative transfer model (RTM) in a specific "mode", as follows.

- Batch 1 (used for all bands): Processed the first 100 monochromatic points as they are ranked in descending order of information content (see section 6.1). In this case four RTM modes are used (only one of which is used in a given individual job):
    - Scalar mode
    - Scalar mode with SIF (only for the NIR band)
    - Scalar mode with perturbed Rayleigh single-scattering albedo (to simulated Ring effect, see section 6.5)
    - Vector mode (to simulated polarised radiances, see section 6.3)

- Batch 2 (used for all bands): Processes monochromatic points ranked 100-199. Only scalar RTM mode is used.
- Batch 3 (used for VIS, NIR and SWIR-2): Processes monochromatic points ranked 200-399. Only scalar RTM mode is used.
- Batch 4 (used for VIS, NIR and SWIR-2): Processes monochromatic points ranked 400-799. Only scalar RTM mode
- Batch 5 (used for SWIR-2): Processes monochromatic points ranked 700-1000. Only scalar RTM mode is used.

E.g. in order to fully process a scanline for the NIR band (which is based on a specific scan line 7 individual jobs are run: Batch 1 in all 4 modes, together with batches 2,3 and 4 (in scalar mode). Some individual jobs take significantly longer than others. E.g. batches 3-5 process more monochromatic points and will take proportionately longer than batch 1 or 2 scalar mode. The vector mode batch 1 calculations take approximately 10 times longer than the scalar mode calculations.

Scripts are used to (a) initiate all batch jobs needed to process each orbit granule; (b) check that all expected output has been produced and (c) in case any output is missing, re-run the missing jobs. (It is common for a small fraction of submitted jobs to fail due to contention for memory or disk resources during the run.)

In principle, full quality L1 output can only be generated once all individual jobs have been run for all bands and scanlines. In order to comprehensively test most aspects of the full processing chain, before expending a large computational effort, the following step-by-step approach was taken:

1. All of the batch 1 scalar calculations are run. L1 products are produced from these limited calculations. Plots and statistics are produced (like those shown in section 7 for the full output) from the resulting L1 files to confirm that the simulations are complete and without spurious features. These calculations can run for a full orbit in a few days of real time on the Jasmin cluster, allowing any issues found to be diagnosed quickly.
2. Full simulations, including all spectral points and vector/Ring effect, were then carried out for the granule over Berlin, and plots produced to confirm that the changes introduced by including more spectral points, polarisation and Ring effect are in line with expectations.
3. Full orbits are processed to L1 including Ring effect (but scalar only).
4. Batch 1 vector calculations are run so that the final output can be produced.

In total, the calculations for all 6 orbits, took of order 100,000 CPU days on the jasmin cluster.

As in the S4/5 study, the code used is maintained under version control (SVN).

With the final output it is possible to produce L1 files based on radiances calculated with and without the inclusion of cloud, Ring effect, SIF and polarisation (in any combination). For the purpose of testing the L2 processors it may be desirable to omit some of these effects, e.g. to apply the processor to more cloud-free scenes or to check if Ring-effect, SIF or polarisation cause specific issues for the L2 scheme. To allow flexibility to perform such tests in future, multiple L1 files have been produced for each orbit granule. In all cases completely cloud-free versions of the orbits have been produced. L1 files for all combinations of the 4 effects have been produced for the granule over Berlin. For all other granules a limited set of combinations has been produced (see the delivery note , [RD26]).

The delivered output also includes the monochromatic output of all individual jobs run. With these files it would be possible to produce, at a later date, L1 files for NO2I/CO2I bands with modified spectral response function (e.g. after this has been measured on ground). This would require re-training of the PCFM, but no new radiative transfer calculations. Monochromatic output needed to train the PCFM is also provided in the delivery.

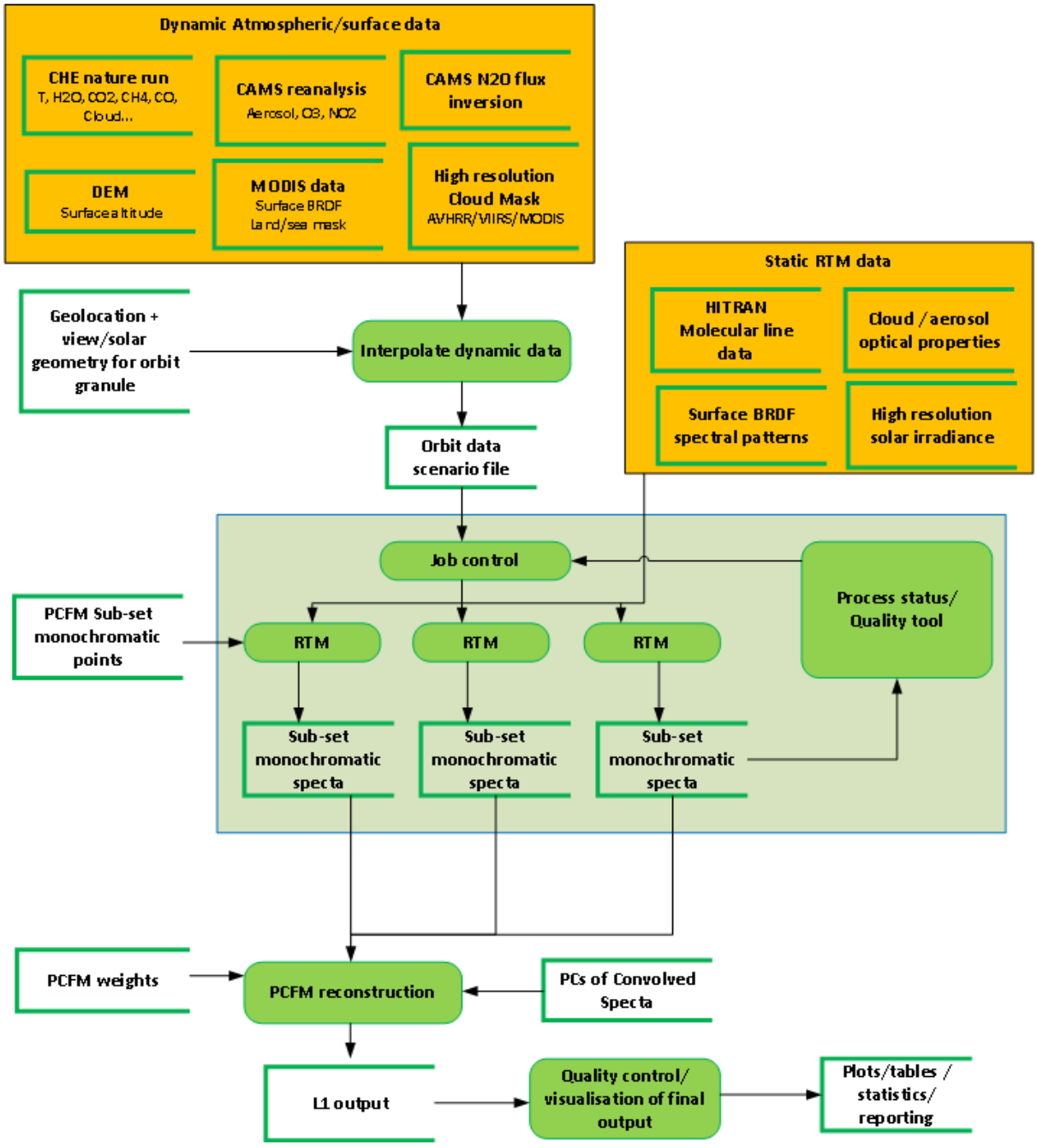



**Figure 6-21: Overview of the simulation of orbit data using the PCFM. Processes in the light green rectangle indicate the main computationally intensive step to produce radiances for the sub-set of spectral points, which requires (carefully monitored) batch processing. The diagram indicates 3 parallel RTM jobs, in reality there would be 100s-1000s running in parallel (millions in total), one job for each scanline in an orbit granule, and RTM mode.**

## 6.9 Use of alternative instrument spectral response function.

The simulations described above are based on an assumed Gaussian instrument spectral response function (ISRF). The same function is assumed for all spectral channels in a given spectral band. In is assumed that the L2 schemes will eventual use more realistic (measured) spectral response functions representative of the real instrument.

It is noted that the output of this study can be used to generate L1 files for different instrument spectral response without the need to redo any of the time consuming radiative transfer calculations. This is because the PCFM approach is based on selecting monochromatic points to represent the band (and performing radiative transfer for monochromatic points). The selection of monochromatic points is not dependent on the assumed ISRF. The accuracy of the model for a given number of monochromatic points is somewhat dependent on the assumed ISRF (since the training regression is based on convolved spectra), but there is no reason to expect a strong dependence of the PCFM accuracy on the detailed shape of the ISRF, provided the width of the function is not significantly narrower than that assumed. Furthermore, all monochromatic output for both the PCFM training and the orbit simulations are delivered (and therefore available for re-use at a later date).

Simulation of spectra with modified ISRF would involve the following steps:

- Monochromatic spectra for the training set should be convolved with the new ISRF. Variation of the ISRF within each band could be modelled if required (this does not complicate the approach significantly).
- PCFM training should be repeated with the modified spectra. It should be confirmed that the same number of monochromatic points are sufficient (this is expected to be the case).
- The new PCFM coefficients are then used in combination with the existing set of monochromatic orbit simulations to re-generate modified L1 files.

In this way it would require only a few days CPU time to re-generate the 6 orbits (compared to ~100 000 days for the full simulations). An additional set of test data has now been simulated in this way using a more realistic ISRF (see section 7.2).

While this would be the most straightforward and robust approach to produce a new L1 dataset for an arbitrary specific new ISRF, the following points are noted:

- Certain kinds of perturbations to the assumed ISRF could be simulated by convolving the existing results. E.g. a perturbation in the width of the ISRF could be simulated by convolving with a 2$^{nd}$ Gaussian function. This will be accurate because of the spectral over-sampling of CO2M, The “true” ISRF to assume in a L2 scheme applied to such data would be convolution of the original ISRF with the 2$^{nd}$ function applied to smooth the L1 simulations.
- It would also be possible to make a PCFM which simulates CO2M spectra at significantly higher spectral resolution and sampling than nominal and use this to produce higher resolution L1 data. This would result in a data set which could then be convolved a 2$^{nd}$ time to realistically simulate a range of spectral response functions without needing to specifically re-train and re-apply the PCFM for each modified ISRF (or perform the time consuming monochromatic radiative transfer). However this approach would require some further testing and development of the PCFM: In general more spectral points would be needed for a given level of accuracy at the higher spectral resolution. To avoid needing more points (and hence new radiative transfer simulations), the regression could be adapted to optimise for accuracy at CO2M nominal resolution (i.e. include a 2$^{nd}$ convolution to CO2M nominal resolution in the regression). This could lead to a more robust simulation of CO2M resolution spectra with modified ISRFs, provided these have spectral width comparable to that assumed in the regression.

# 7 Results

## 7.1 Results assuming Gaussian ISRF

Results for all six orbits are visualised as false-colour images in Figure 7-1. The false colour images are formed as an RGB composite in which red, green and blue channels shows the reflectance at 1590nm, 747nm and 405nm, respectively. In this RGB composite liquid cloud will tend to appear white, while ice cloud will appear light blue or cyan (due to the relatively low ice reflectance in the SWIR). Vegetated land shows as green, while desert surfaces are more yellow/brown and ice surface are cyan. Reflectance here is the sun-normalised radiance divided by the cosine of the solar zenith angle. The images are displayed in columns. Within each column, results images from the west, central and east orbit are shown alongside each other. The left hand column shows the first 1242 rows from each orbit. This corresponds to the first 2 granules of the output L1 files, since there are 621 scanline in each granule (except the last granule of each orbit). Columns from left-right show images from successive scan lines along the orbit. Maps under each column show the three orbits, with the green area indicating the area covered by the images above. The set of upper panels show the “EU” orbits and the lower panels show the “SA” orbits.

Figure 7-2 shows the same orbits, but for L1 data simulated without the presence of cloud.

These figures clearly show the realistic way in which surface and cloud features are represented in the simulations. The Western orbits (in glint-pointing mode) also clearly show sun-glint, particularly towards the ends of the orbit. The mixing of water and land reflectance models within each scene means leads to more realistic results near coastlines than was the case for the previously delivered simulations from the original study. Rivers are also bright in the glint orbit (especially in the first part of the SAwest orbit).

Similar format plots in Figure 7-3 to Figure 7-8, illustrate some of the key atmospheric parameters in the simulations: Total columns of $CO_2$, $CH_4$, CO, $NO_2$, water vapour and aerosol optical depth. The particularly high spatial resolution of some of the input datasets over Europe is evident.

Figure 7-9 to Figure 7-12, illustrate spectra from the granule over Berlin. Each of the figures shows results for one of the four NO2I/CO2I bands. In each case, the top-left panel shows the mean simulated reflectance spectrum in black (considering all ground pixels in the granule). Green lines show the mean +/- one standard deviation. Blue lines show the minimum and maximum reflectance spectrum in the granule. The top-right panel shows similar plots of the relative differences (in percent) between the full simulation and one which neglects polarisation. In each case, the relative difference is computed with respect to the full simulation for the same scene, which will tend to emphasise cases where the radiances in particularly low (either spectrally or spatially). The bottom-left panel shows similar plots of the relative differences (in percent) between the full simulation and one which neglects Ring effect. The bottom-right panel shows similar plots of the relative differences (in percent) between the full simulation and one which neglects solar induced fluorescence (SIF). SIF is only included in the NIR simulations, so this panel shows zero difference for the other bands.

Figure 7-13 to Figure 7-15, show for each CO2M band, maps of the root-mean-square (RMS) relative differences (considering all spectral points of a given spectrum) between the full simulation and one which neglects polarisation (Figure 7-13), page 5), Ring effect (Figure 7-14) and SIF (Figure 7-15). Being relative differences, these will tend to emphasise differences in strong absorption bands.

**Figure 7-1: False colour images illustrating simulated radiances from all six orbits. The top set of panels shows the "EU" orbits and the bottom set of panels show the "SA" orbits.**

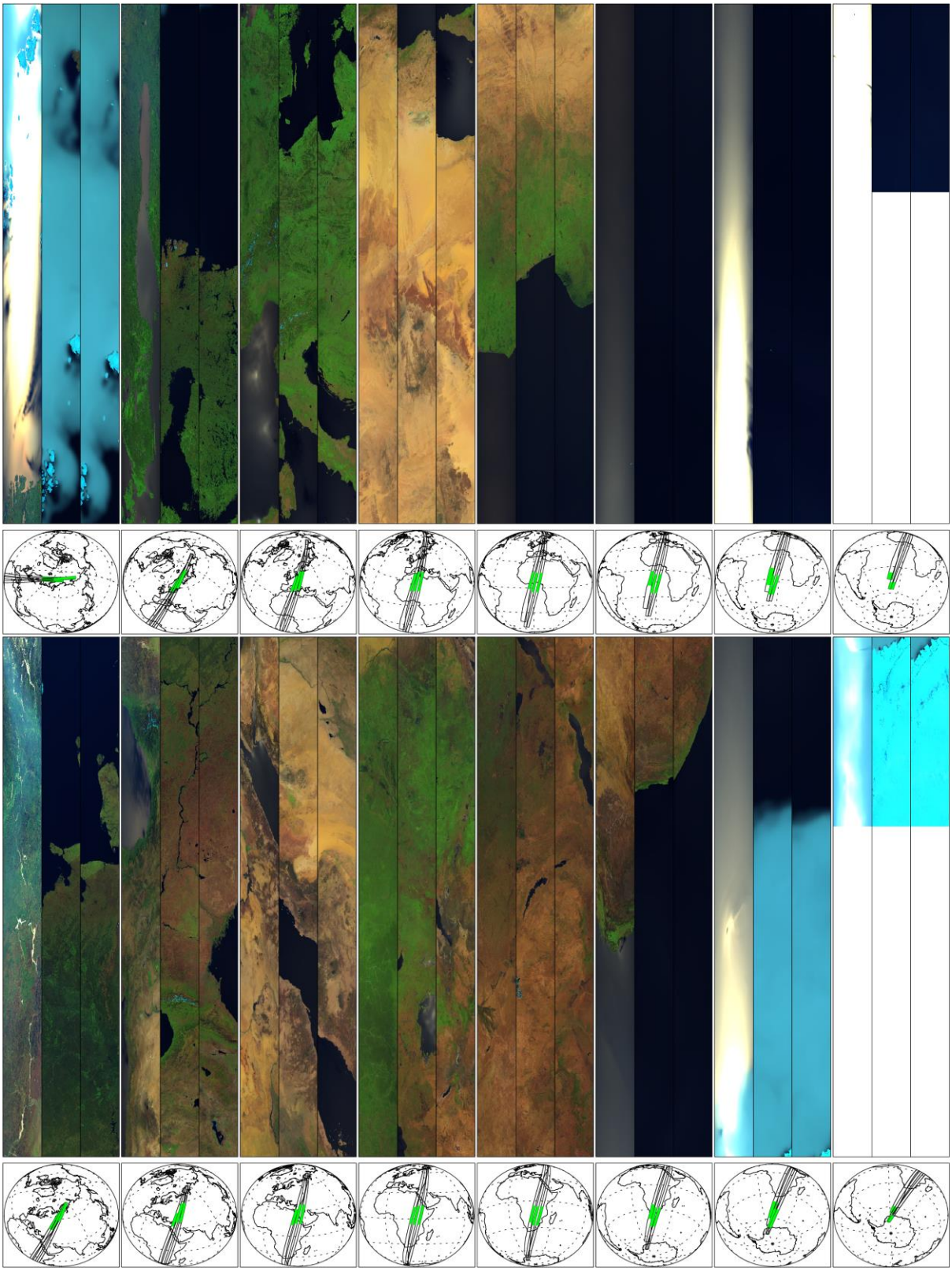


**Figure 7-2: False colour images illustrating simulated radiances from all six orbits (neglecting cloud).**

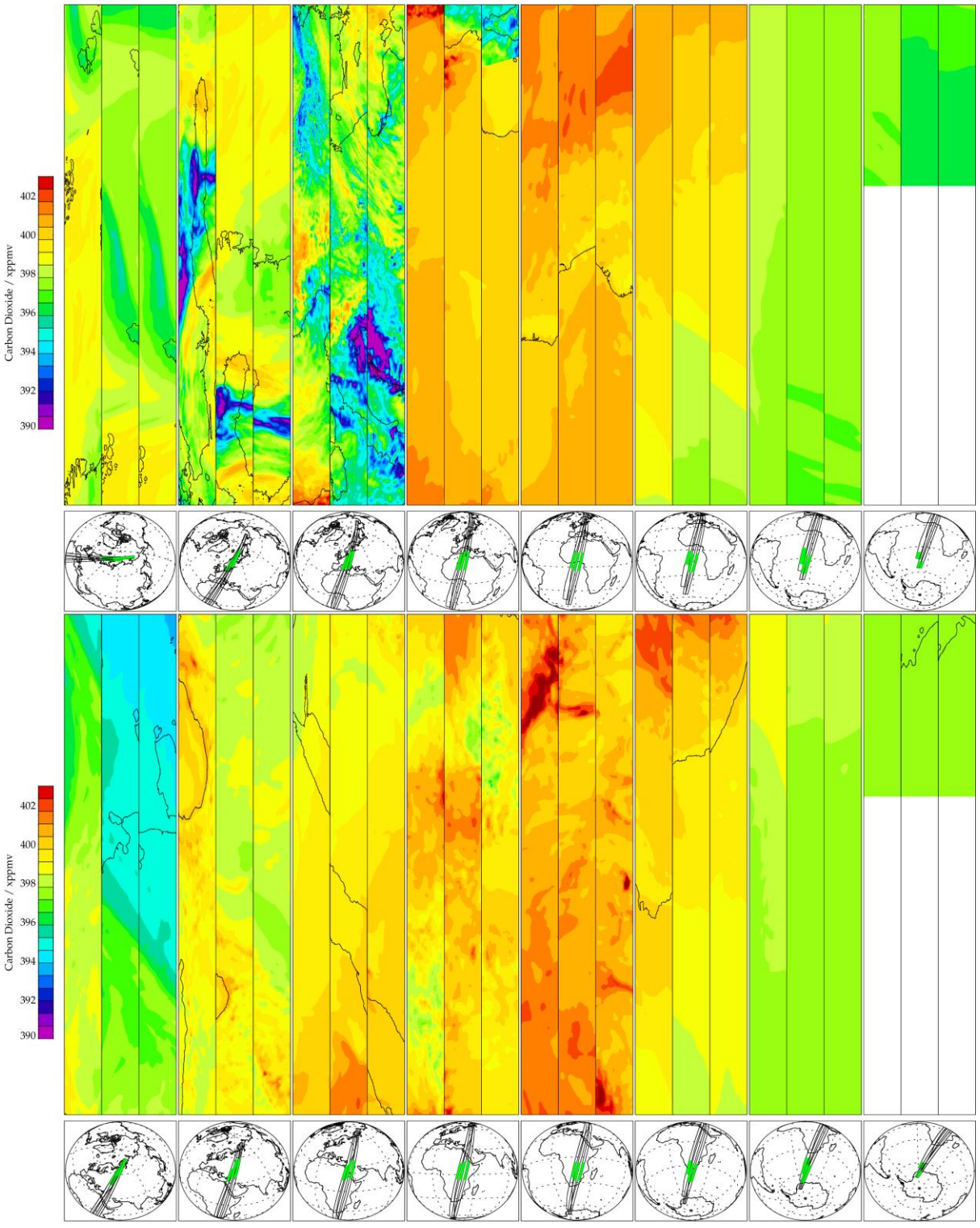



**Figure 7-3: Total column average mixing ratio of CO2 as simulated in all orbits.**

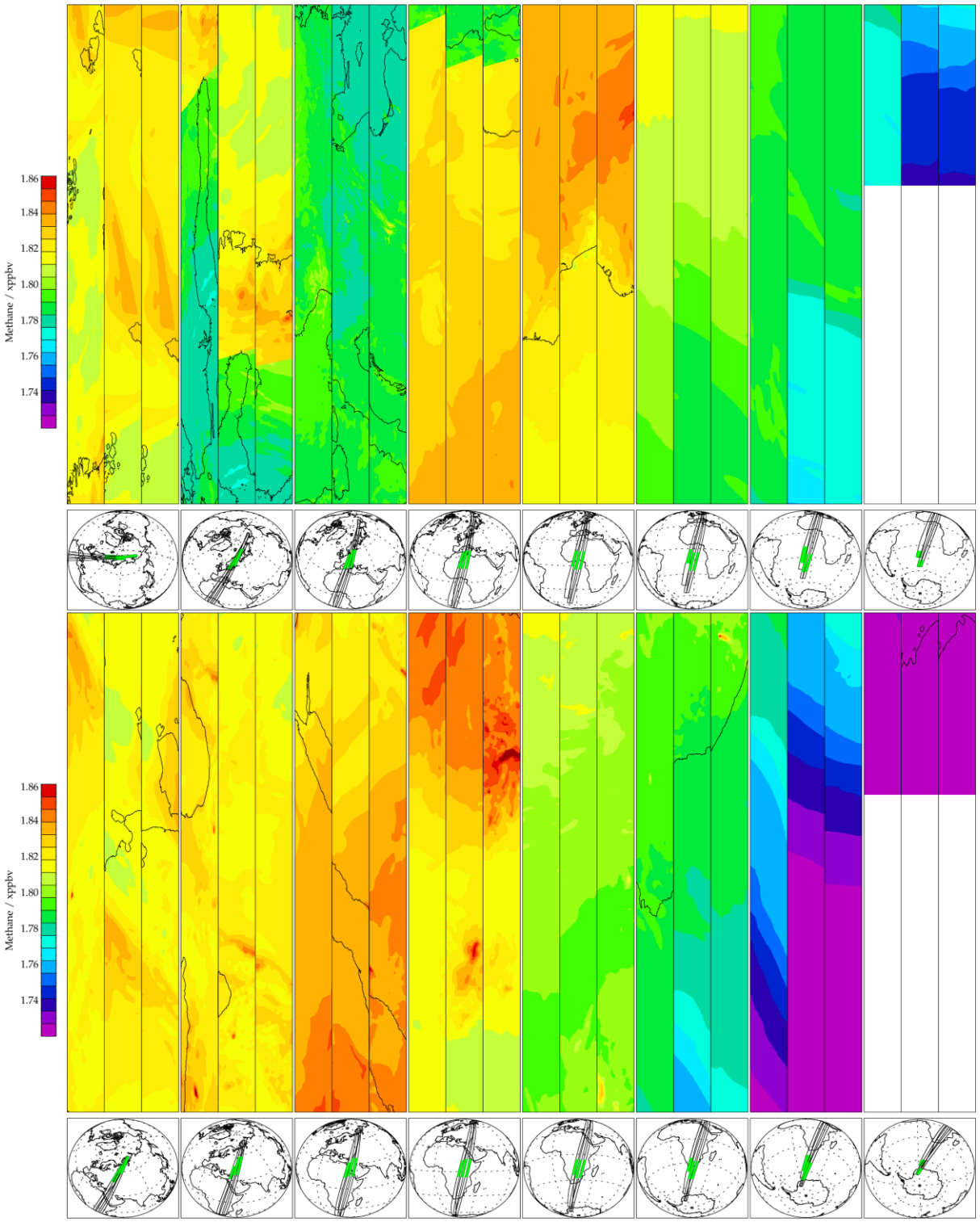



**Figure 7-4: Total column average mixing ratio of CH4 as simulated in all orbits.**

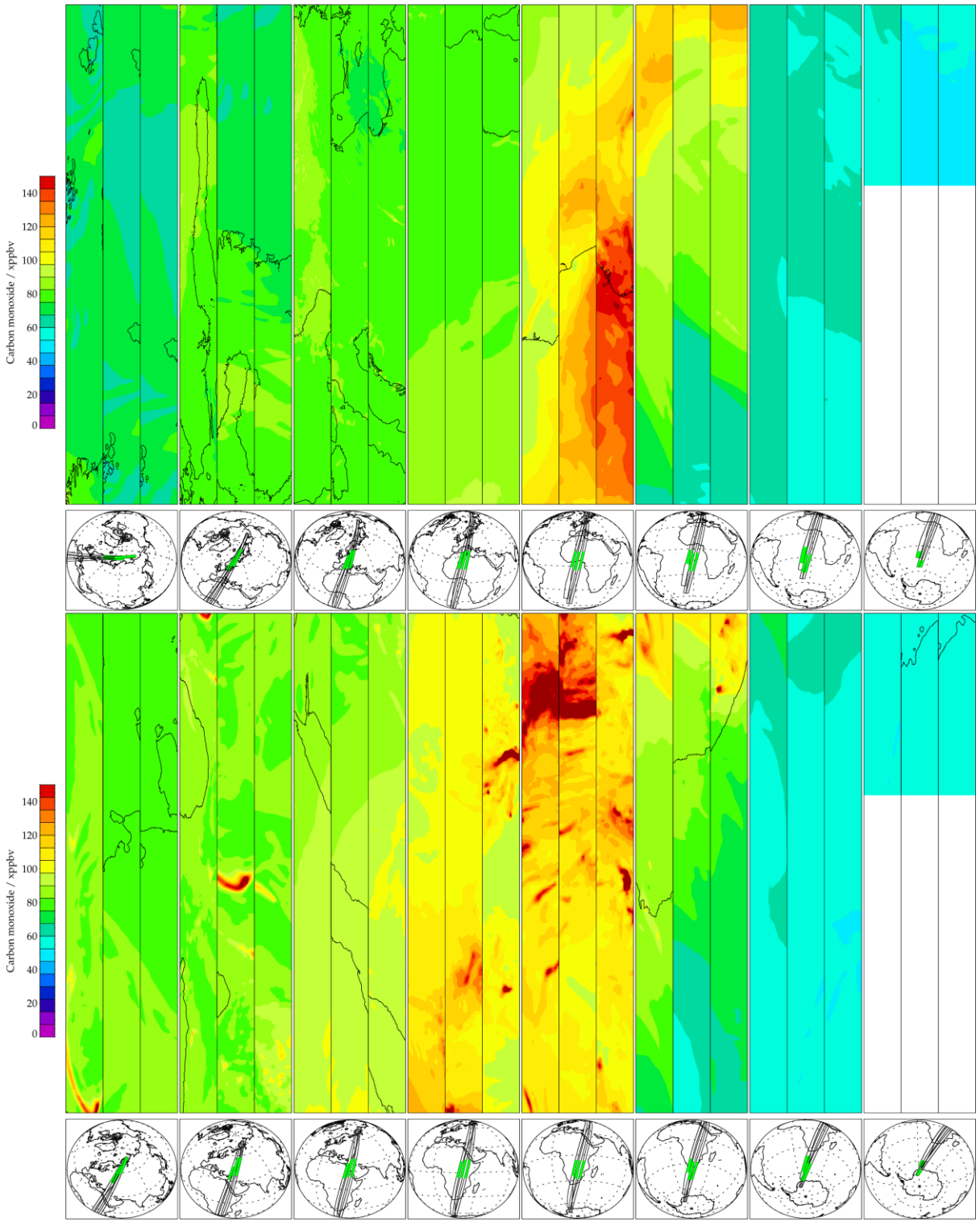



**Figure 7-5: Total column average mixing ratio of CO as simulated in all orbits.**

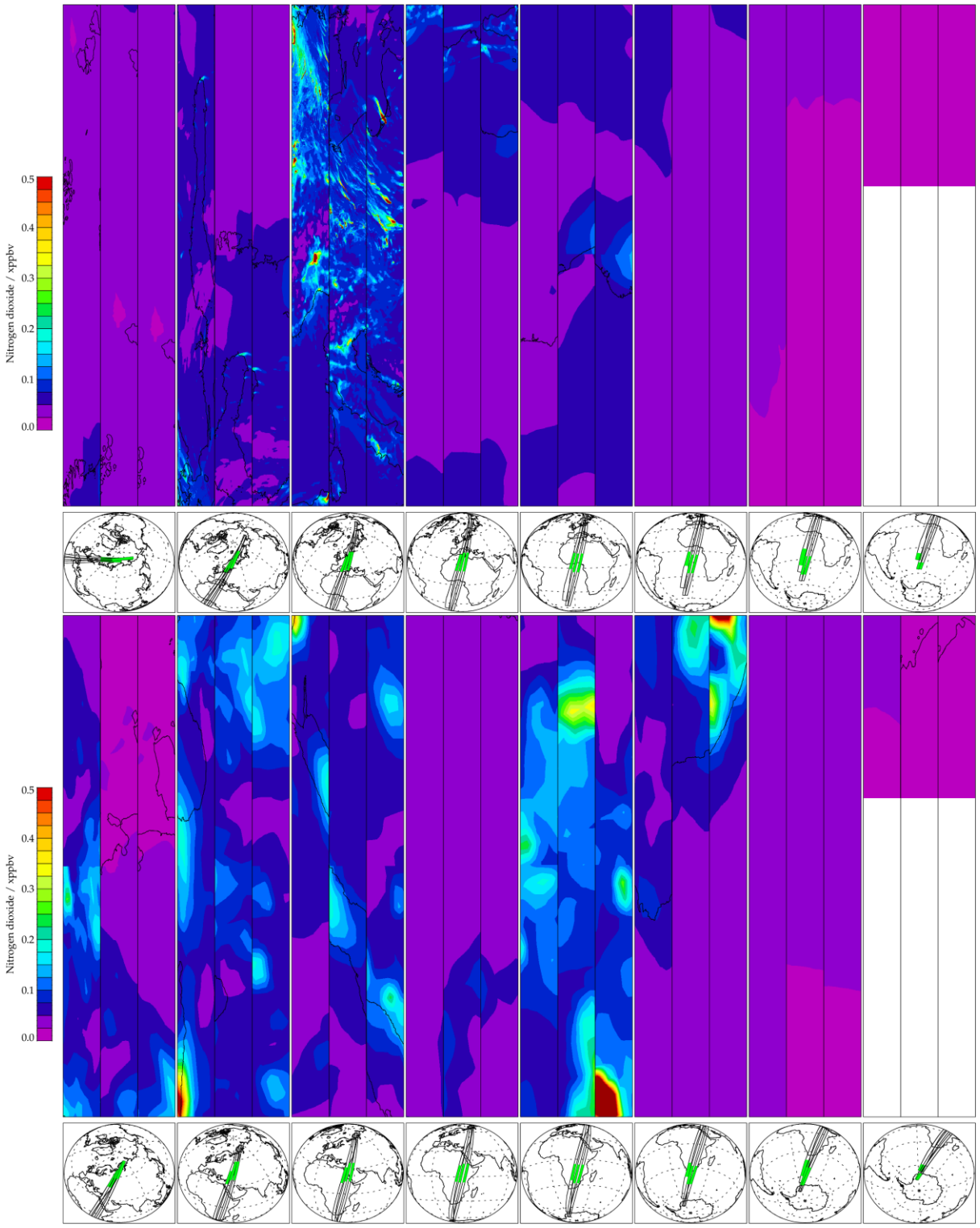



**Figure 7-6: Total column average mixing ratio of NO2 as simulated in all orbits.**

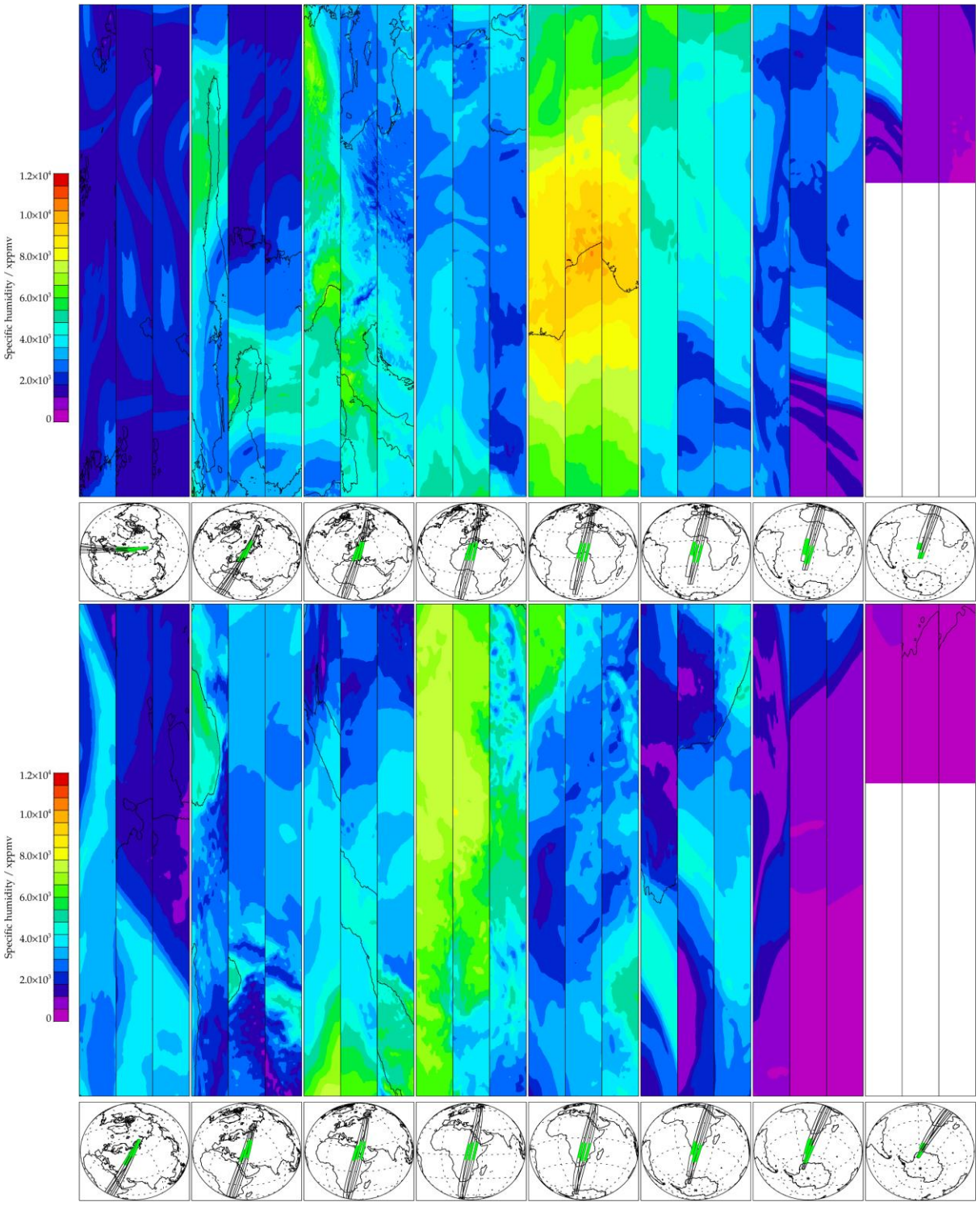



**Figure 7-7: Total column average mixing ratio of water vapour as simulated in all orbits.**

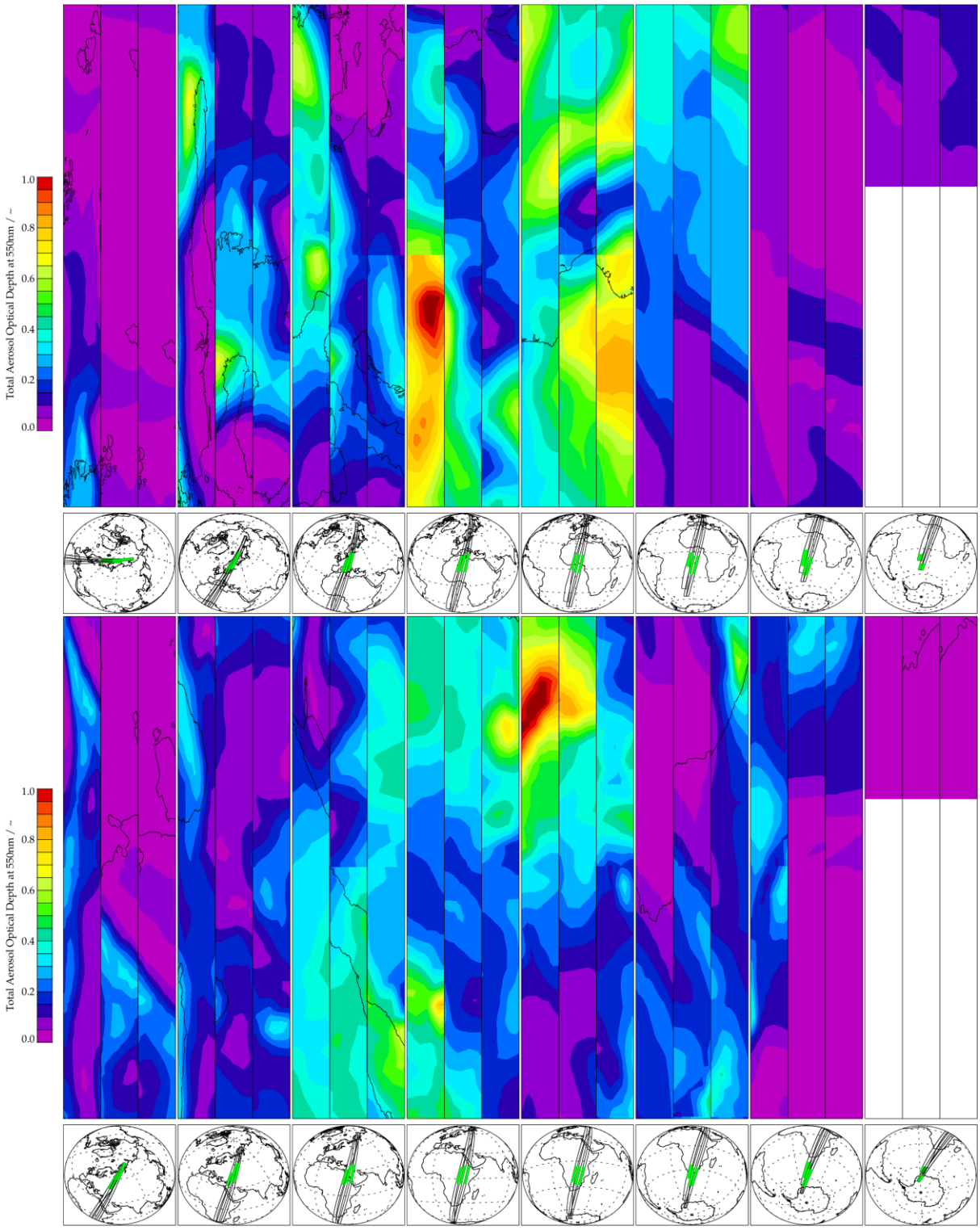
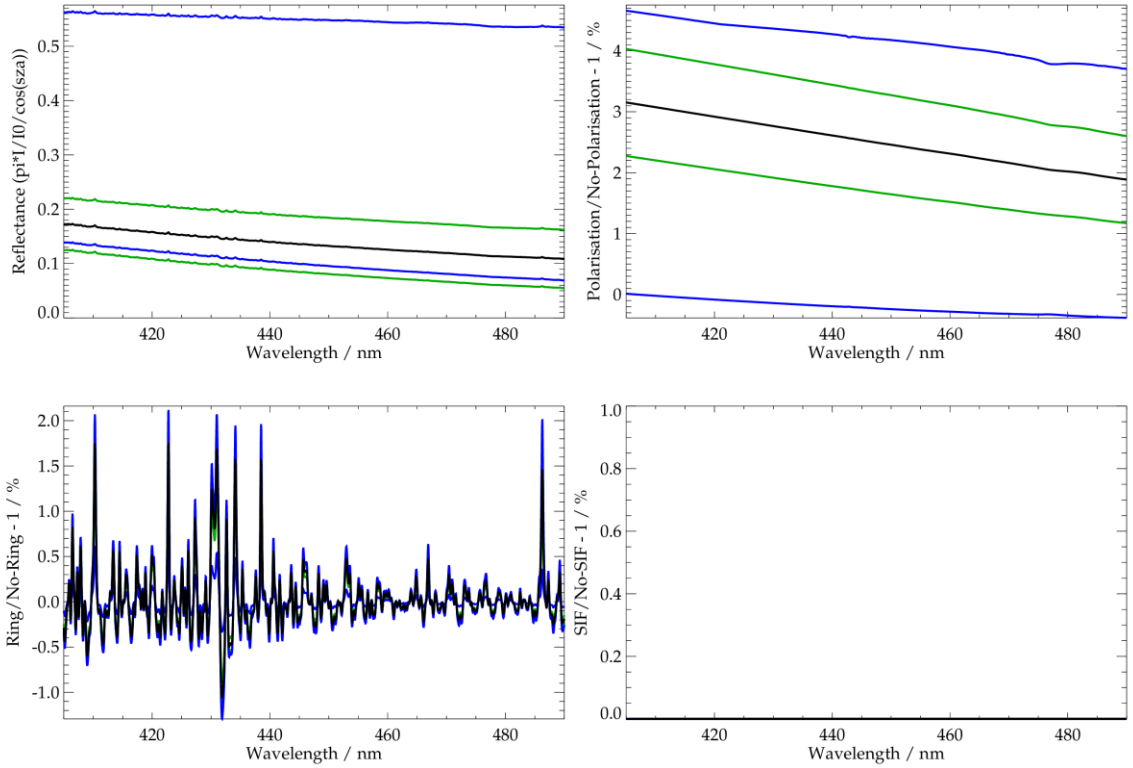
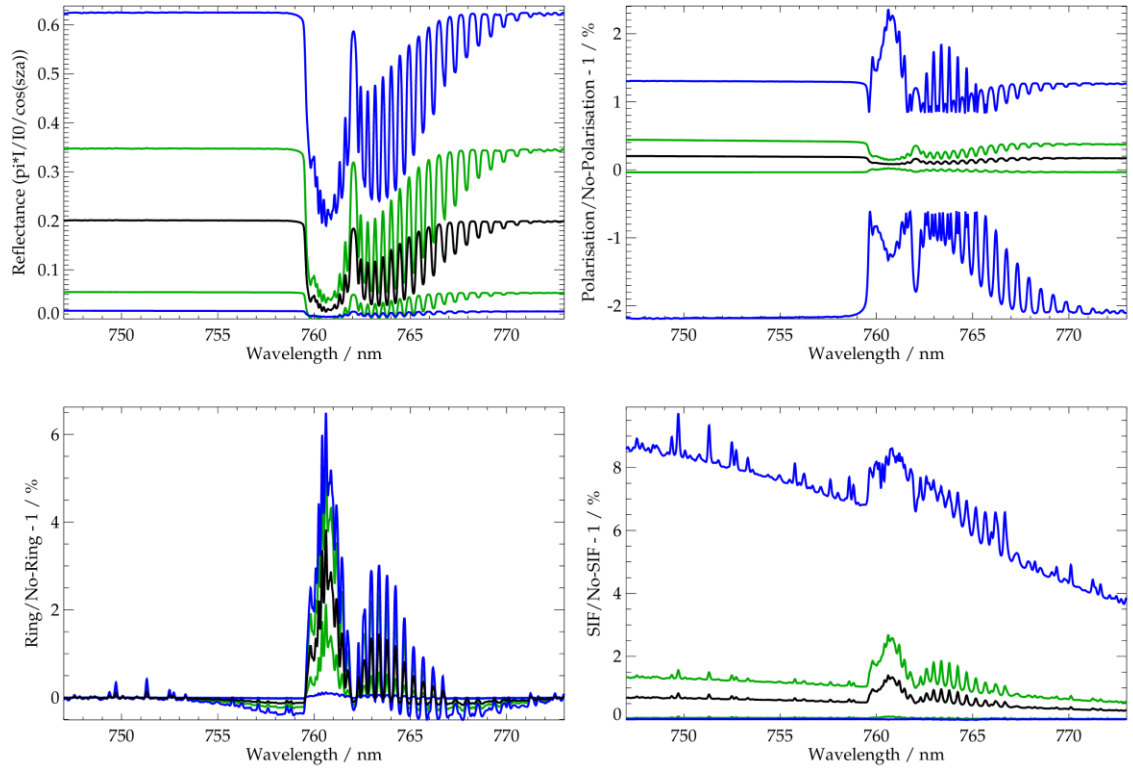



**Figure 7-8: Total aerosol optical depth (at 550nm) as simulated in all orbits.**



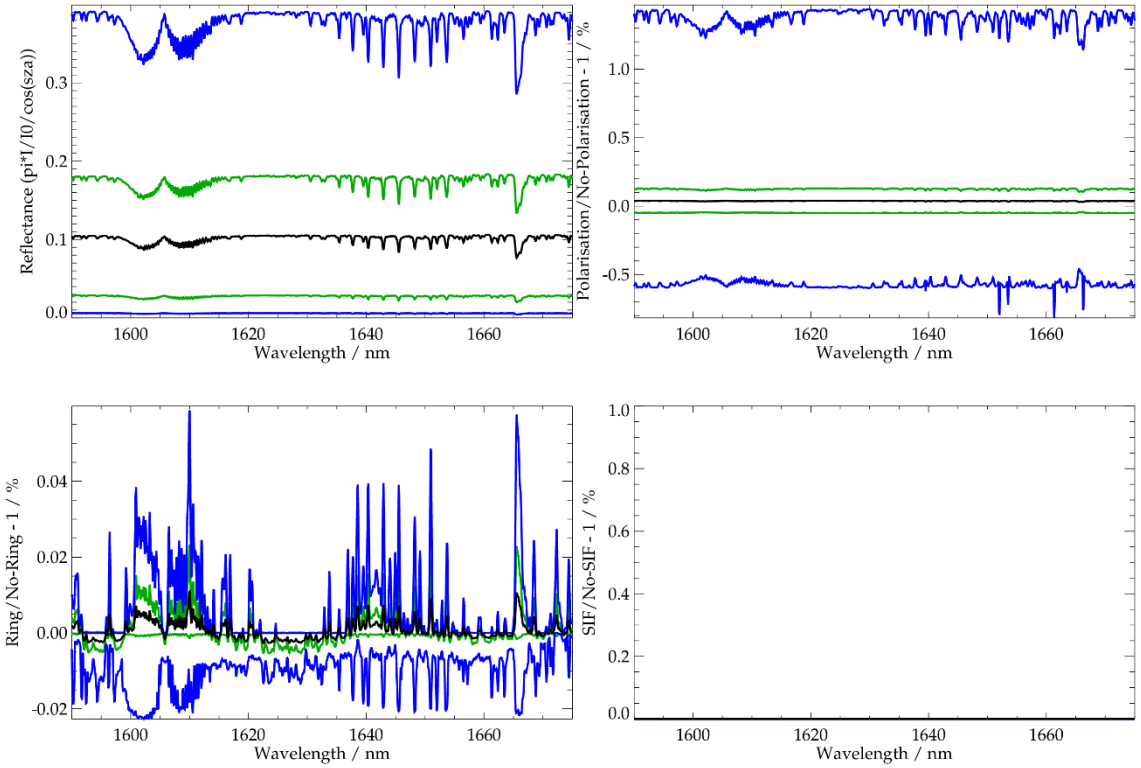

2p0/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol

**Figure 7-9: Spectra in band 1 (VIS) from the Berlin granule.**



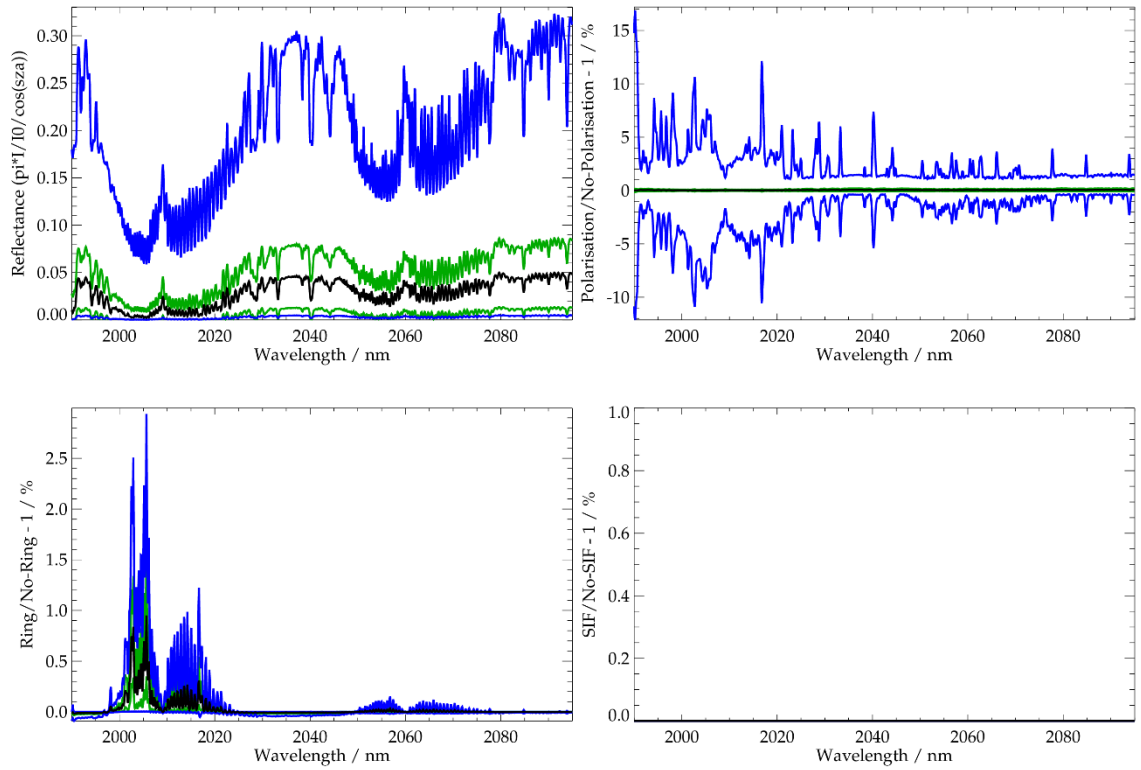

2p0/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol

**Figure 7-10: Spectra in band 2 (NIR) from the Berlin granule.**




2p0/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol

**Figure 7-11: Spectra in band 3 (SWIR-1) from the Berlin granule.**




2p0/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol

**Figure 7-12: Spectra in band 4 (SWIR-2) from the Berlin granule.**

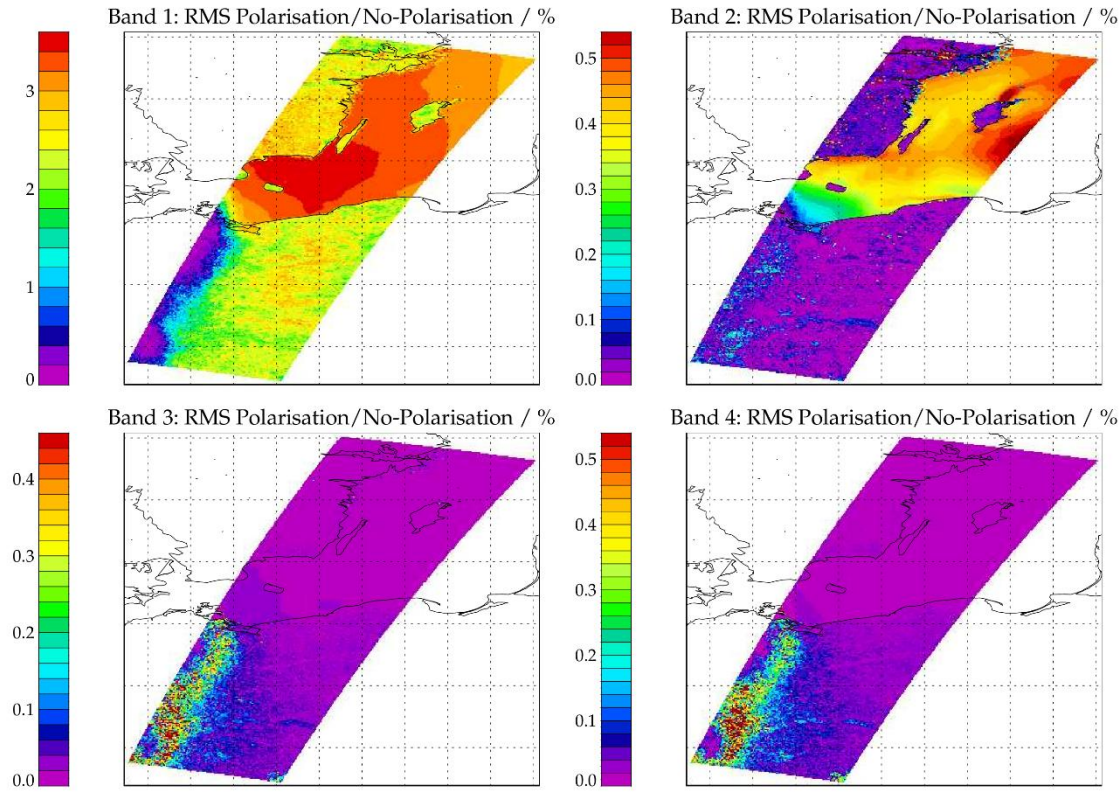



**Figure 7-13: RMS differences between spectra simulated with and without polarisation included.**

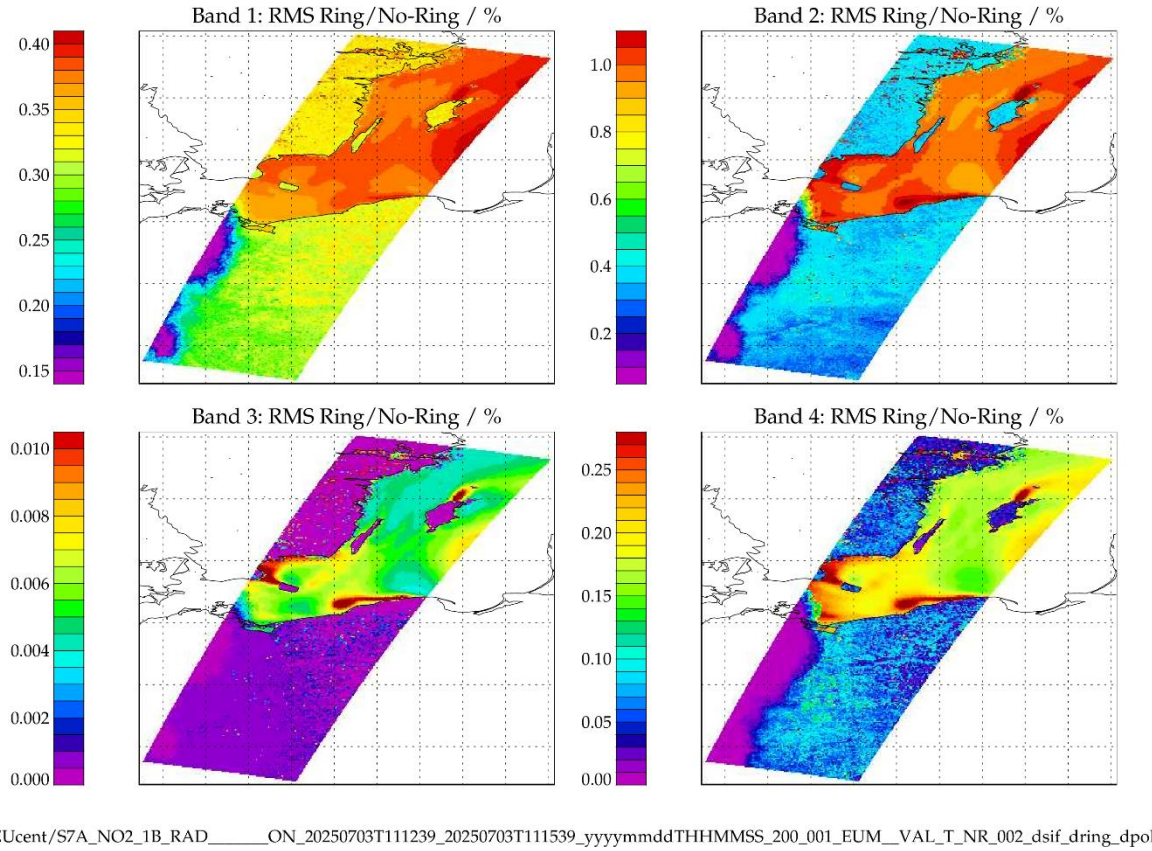



**Figure 7-14: RMS differences between spectra simulated with and without Ring effect included.**

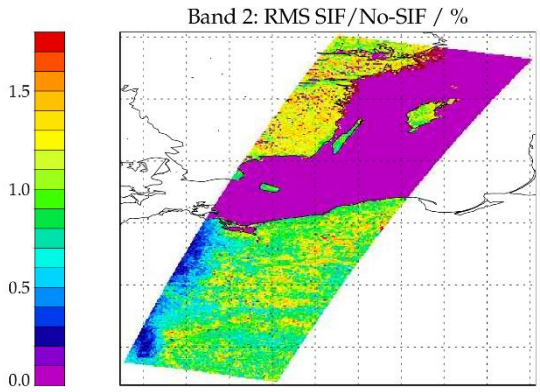



**Figure 7-15: RMS differences between spectra simulated with and without SIF included.**

## 7.2 Results based on realistic ISRF

A set of simulated spectra has now been created using a realistic ISRF based on data in files provided by Eumetsat on 1 April 2025. Files are names "CO2I_ISRF_ESA_[BAND].nc" where "[BAND}" is "VIS", "NIR", "SWIR1" and "SWIR2" (i.e. 1 file per spectral band). Each file contains an ISRF defined for all 110 across track positions and 3 spectral positions. Here we simulate L1 radiances assuming the central ISRF in both dimensions (i.e. index 105 spatially and 1 spectrally, counting from 0). The selected ISRFs are compared to the previously assumed Gaussian functions in Figure 7-16. Although the FWHM of the functions are similar, the new ISRFs are closer to box-car shape. Figure 7-17 to Figure 7-20. Illustrate the accuracy of the PCFM which has been retrained using the new ISRFs (but identical sub-set of monochromatic points). Comparing these to Figure 6-11 to Figure 6-15, it is noted that a similar level of accuracy is retained, but differences are somewhat more spectrally structured (because of the more box-car shape of the new ISRF). Figure 7-21 summarises (for the Berlin granule) differences between the simulations with the new ISRF compared to the Gaussian shape. Note this figure shows differences after normalising radiance by solar irradiance (both simulated with the same ISRF). In the visible band relative differences in radiance are much larger as spectral structure in this band is dominated by that of the solar irradiance.

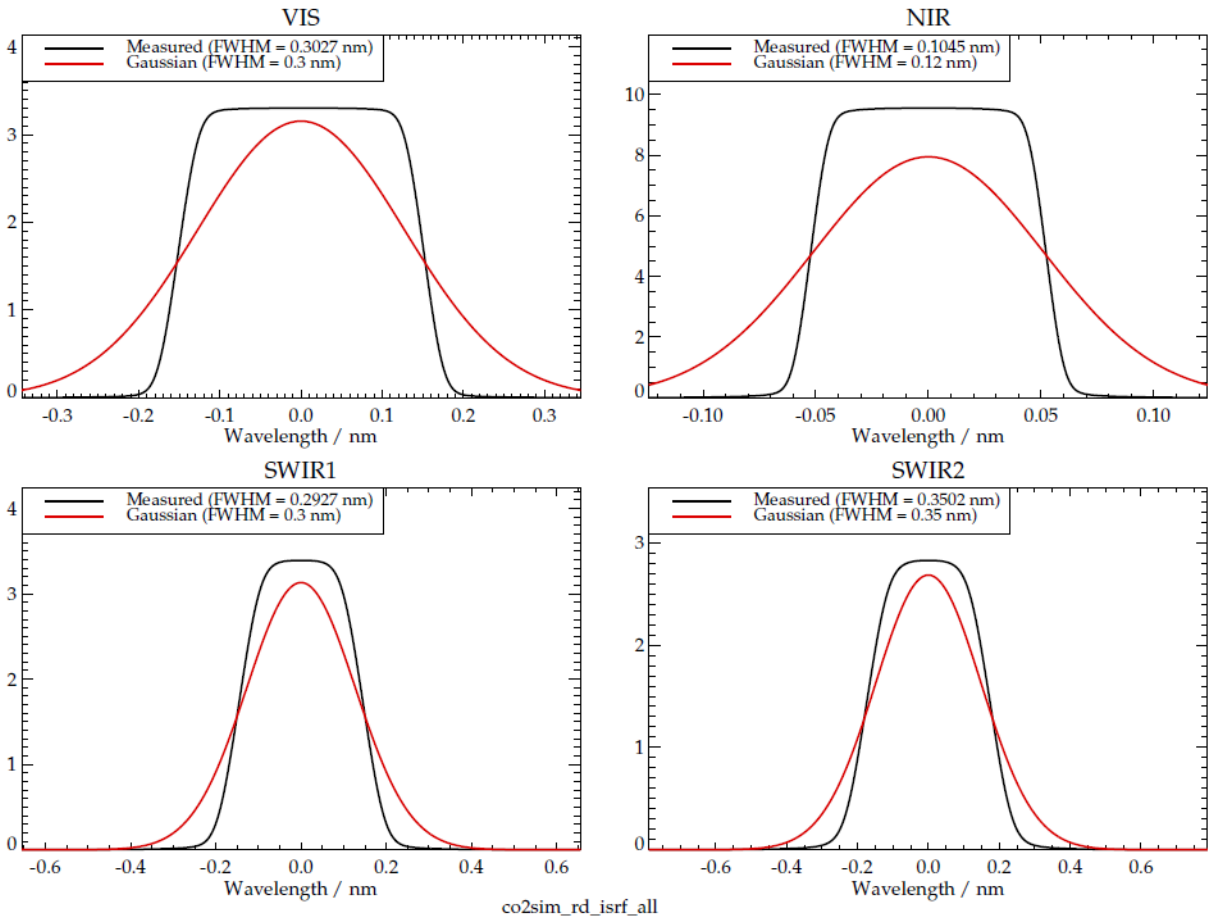



**Figure 7-16: Assumed realistic CO2M ISRFs compared to the Gaussian functions.**

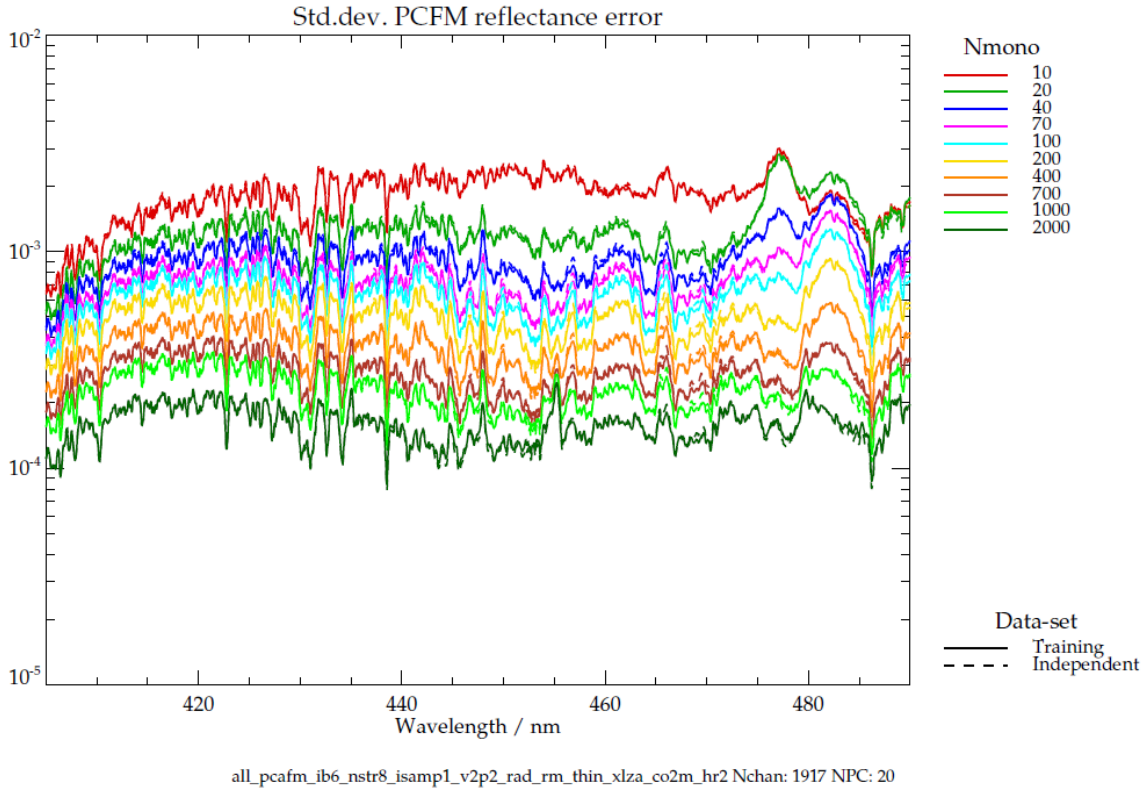



**Figure 7-17: Spectral standard deviation over all scenarios for the VIS, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

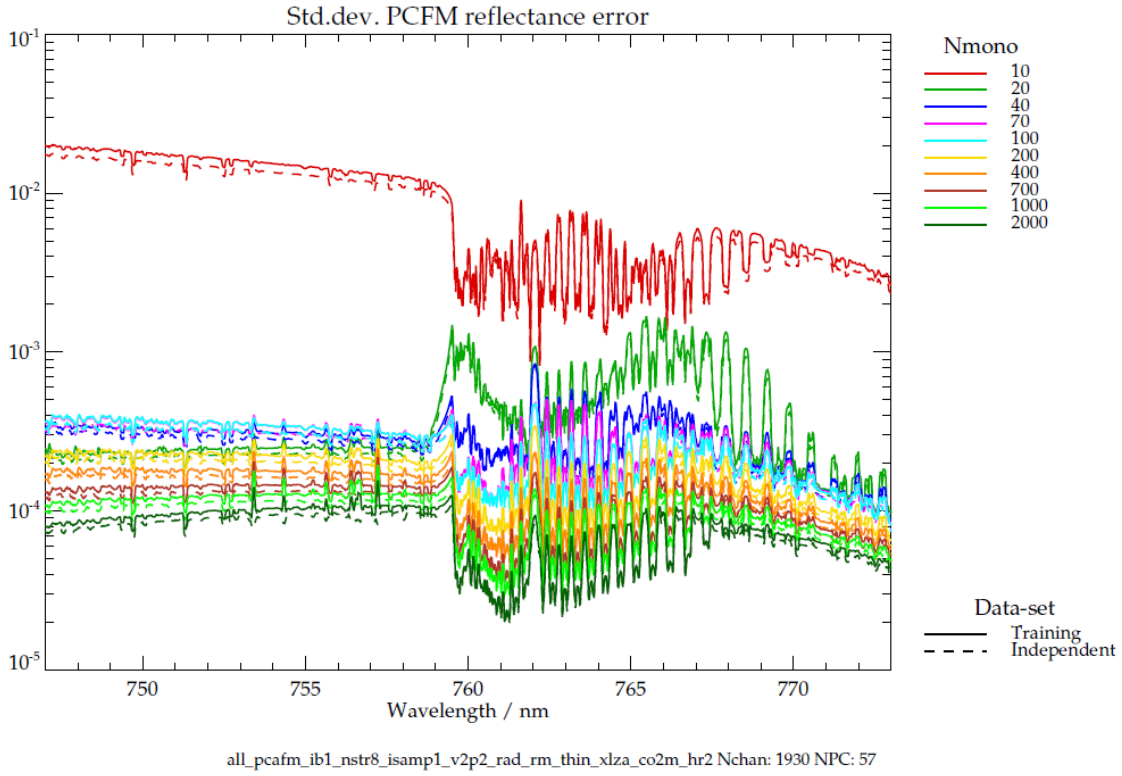



**Figure 7-18: Spectral standard deviation over all scenarios for the NIR (without SIF included), considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

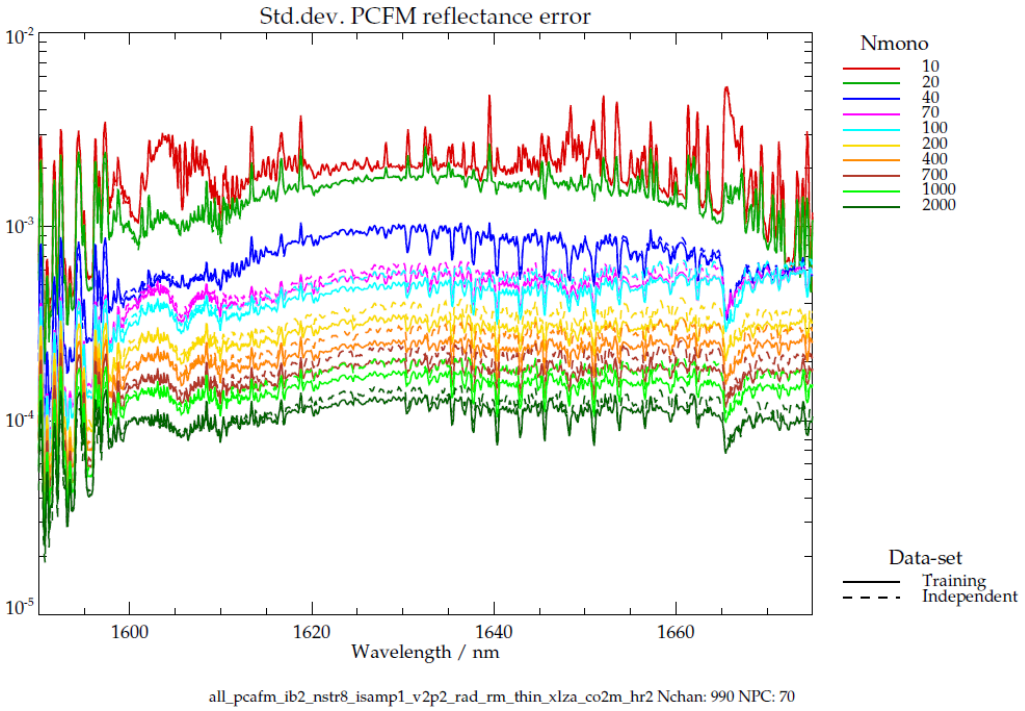



**Figure 7-19: Spectral standard deviation over all scenarios for the SWIR-1 band, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**

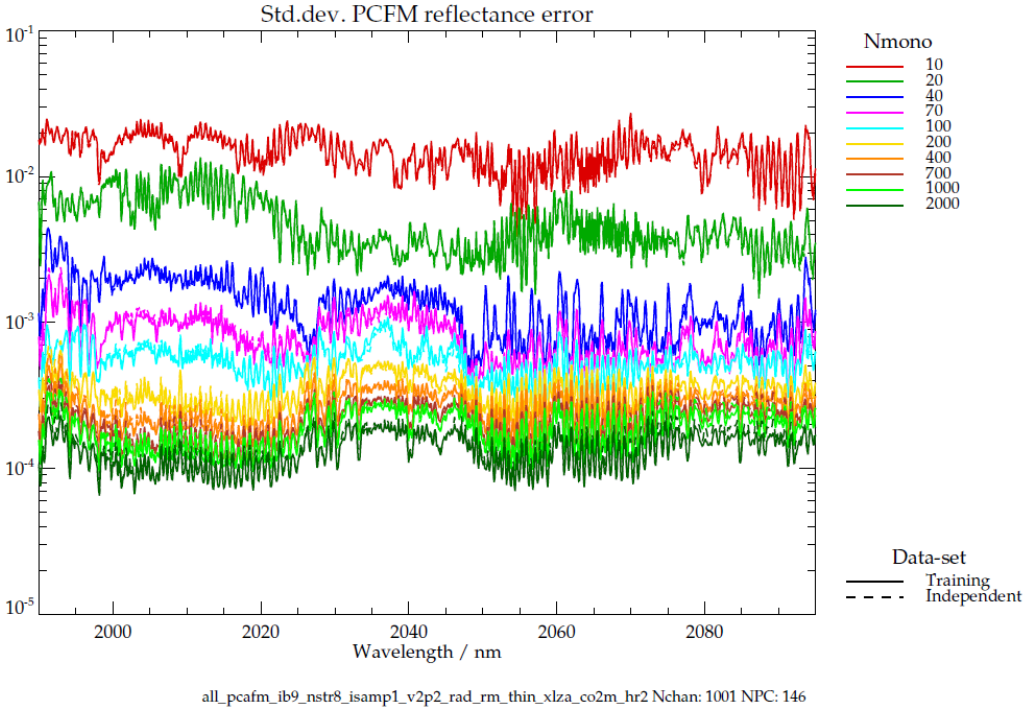
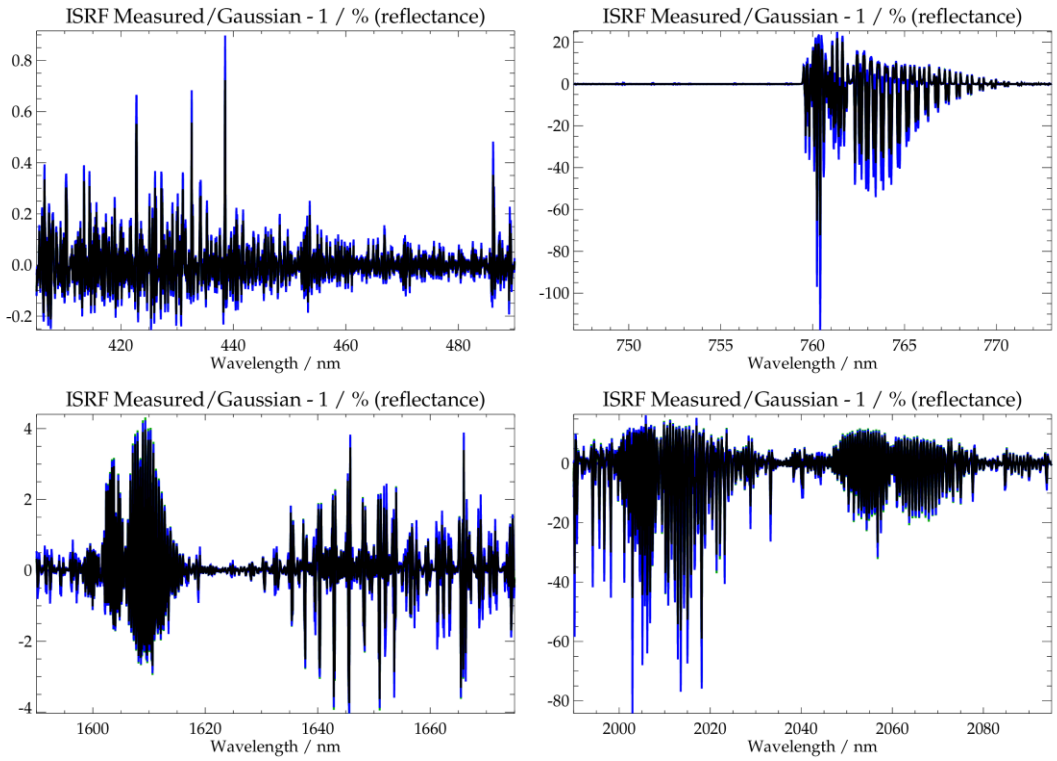
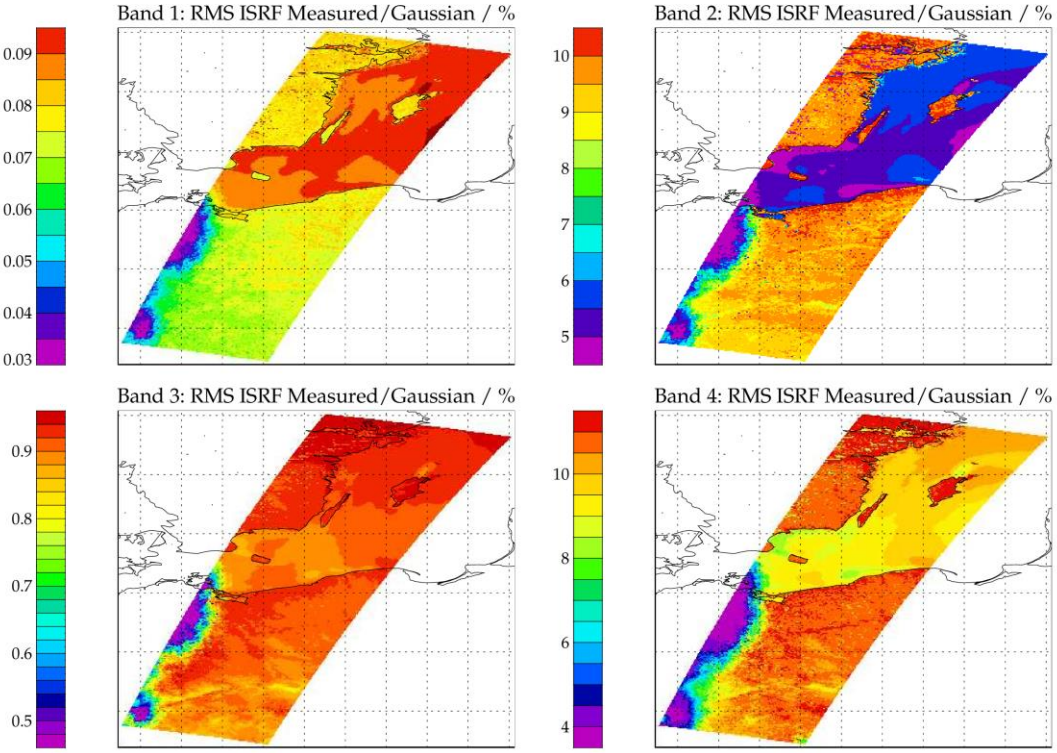



**Figure 7-20: Spectral standard deviation over all scenarios for the SWIR-2 band, considering training (solid) and independent (dashed) sets, for calculations based on varying numbers of optimised channels.**




2p0i1/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol_disrf

**Figure 7-21: Relative differences (in reflectance or sun-normalised radiance) between simulations with the new ISRF and the Gaussian function. Each panel shows a different CO2I band.**




2p0i1/EUcent/S7A_NO2_1B_RAD_____ON_20250703T111239_20250703T111539_yyyymmddTHHMMSS_200_001_EUM__VAL_T_NR_002_dsif_dring_dpol_disrf

**Figure 7-22: RMS relative differences (in reflectance or sun-normalised radiance) between simulations with the new ISRF and the Gaussian function. Each panel shows a different CO2I band.**