# AEROPRO study

# Executive Summary

**P. Stammes, A. Apituley, A.F.J. Sanders (KNMI)**

**01-August-2014**

**Project: AEROPRO**
**Aerosol Profile Retrieval Concept Development and Validation for Sentinel-4**

ESA ITT AO/1-7017/11/NL/MP
Contract: 4000105882/12/NL/MP

This page is intentionally left blank.

# Executive Summary

## Overview

Sentinel-4 (S-4) will be an operational satellite mission in geostationary orbit aiming at continuous monitoring of atmospheric composition, in particular with respect to air quality, on a European basis with an hourly revisit time. The Sentinel-4 mission is part of the European initiative Copernicus/Global Monitoring for Environment and Security (GMES), which is dedicated to the implementation of a sustained operational capacity for Earth observation. The main data products of the Sentinel-4/UVN (Ultraviolet-Visible-NearIR) sounder instrument are $O_3$, $NO_2$, $SO_2$, HCHO and aerosol. The Sentinel-4 UVN instruments will be embarked on the EUMETSAT MTG-Sounder platforms, the first to be launched in 2021.

Aerosol height information is important for precise scattering corrections of the S-4 trace gas retrievals, for retrieval of elevated aerosol layers, e.g. from desert dust or volcanic eruptions, and for the determination of surface particulate matter. The S-4 aerosol profile product will be useful for operational applications, e.g. aviation safety and air pollution monitoring, and scientific applications, e.g. radiative forcing studies, long-range transport modeling and studies of cloud formation processes.

The aim of this study is the development and validation of an algorithm for the retrieval of aerosol profile information from the $O_2$ A-band around 760 nm as measured by Sentinel-4/UVN. The algorithm has to be tested on real satellite data, and validated with reference data. We use selected aerosol events with collocated satellite data and validation data, especially aerosol lidar data; this approach is called “reverse validation”.

As proxy for the S-4/UVN satellite data, we use data from the Fourier Transform Spectrometer (FTS) on the GOSAT satellite and from the GOME-2 grating spectrometer on the Metop satellite. We describe the validation data and the GOSAT and GOME-2 satellite data for selected aerosol events (cases). For the validation, data from the CARBONEXP aircraft campaign (Aug-Sept 2011) as well as from ground-based lidar stations are used. For the aerosol profile algorithm we make use of algorithm developments for the TROPOMI instrument on ESA’s Sentinel-5 Precursor mission, and extend it to be applicable for S-4. The algorithm is applied to GOSAT and GOME-2 data for the selected aerosol cases, the retrieval performance is analysed, and the retrieval results are validated. Finally, conclusions and recommendations for aerosol profile retrieval from S-4 and S-5 are given.

## 1. Existing algorithms

In TN1 a literature overview of existing algorithms for aerosol profile retrieval from the $O_2$ A-band is given. We collected and summarized the literature on aerosol profile retrieval algorithms from space-borne measurements of the $O_2$ A-band. We reviewed the capabilities to application to real data. The main features of the algorithms have been tabulated, whereas a more extensive summary and evaluation of each reference is given in the text. The retrieval of aerosol height from $O_2$ A-band is a relatively new topic, and there appears to be no existing (operational) satellite product.

In recent literature, aerosol profile is often a fitting parameter in advanced algorithms using the $O_2$ A-band for other purposes, mainly for surface pressure retrieval, $CO_2$ and $CH_4$ retrieval from the shortwave infrared, and fluorescence retrievals. Recent literature on these topics, especially regarding the effects of aerosols, has also been collected.

## 2. Case selection for GOSAT

TN2 presents the ground-based and satellite data selection procedure and the selection results. Ground-based data include AERONET data for aerosol amount and lidar data for aerosol profile. For cloud detection, GOSAT Cloud and Aerosol Imager (CAI) data were used, as well as MODIS satellite data. Selection criteria were:

1. Cloud free scenes over an area of about 10-20 $km^2$ during the GOSAT overpass
2. Aerosol profiles with total AOT around 0.5 or higher at 760 nm
3. Aerosol profiles showing lofted aerosol layers
4. Closest proximity in time and place with the GOSAT overpass (within a few hours and about 100 km).

The output is a list of cases of high and elevated aerosol loads that are matched with GOSAT overpasses (and a few GOME-2 overpasses). We describe the satellite data for the output list of cases, including quality checks. We selected 18 cases in total: 15 for GOSAT and 3 for GOME-2, in the period 2011-2012. We have ranked the list according to priority for processing, starting with three cases from the CARBONEXP campaign that took place in Greece/Turkey in August-September 2011. Thanks to support from JAXA, pointed observations were performed by GOSAT for this campaign. Cases from other locations with GOSAT data coverage, high aerosol loads and favourable cloud-free conditions were chosen. The selected cases contain different aerosol conditions, namely: Saharan dust (Santa-Cruz de Tenerife in the Atlantic Ocean), pollution type aerosol (Cabauw, the Netherlands) and long range transport aerosols (CARBONEXP).

The calibration of the GOSAT spectra is an important aspect of this study. Calibration of the GOSAT FTS is a complicated procedure and has to be partly done by the data user. Details of the calibration are given in an Appendix.

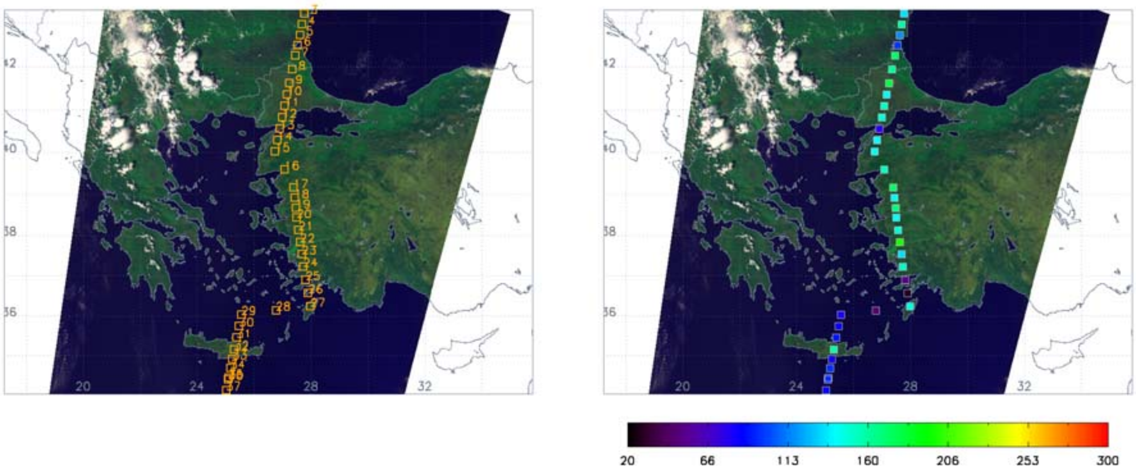
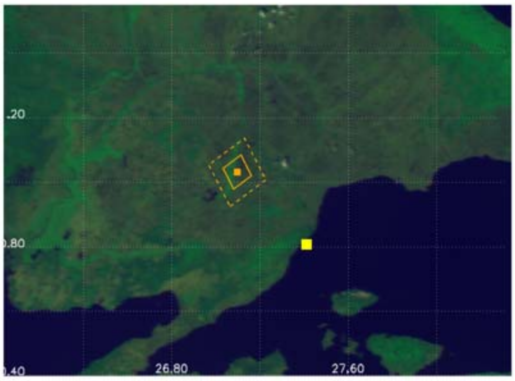



*Figure 1: GOSAT CAI imagery with FTS pixels (boxes) superimposed during the CARBONEXP campaign over Greece and Turkey on 6 September 2011. (Left) Orange open squares: GOSAT FTS scans that have passed the L1b quality filters. (Right) SNR in the continuum (758 nm) at GOSAT resolution of the FTS observations.*



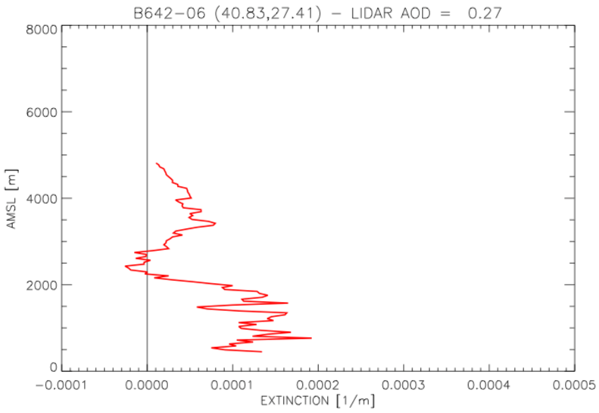




*Figure 2. (Left) Within diamond box: selected GOSAT pixel. The yellow square at the coast of the Sea of Marmara is the location of the airborne lidar profile. (Right) Airborne lidar extinction profile during CARBONEXP flight B642-06 on 6 September 2011.*

## 3. Aerosol profile retrieval algorithm

The Sentinel-4 aerosol profile retrieval algorithm concept developed within this study has the following key features:

- Spectral fit of reflectance across the $O_2$ A band (fit window 758 – 770 nm)
- Retrieval method is Optimal Estimation
- Main fit parameters are: aerosol pressure ($P$), aerosol optical thickness (AOT) and surface albedo ($A_s$)
- Error estimates are provided to improve usability of the product (e.g. for data assimilation)
- Assumed aerosol profile: single layer with a fixed pressure thickness and a constant volume extinction coefficient within the layer.

We assume that aerosols are uniformly distributed in a single layer with a fixed pressure thickness. The retrieved aerosol pressure is the mid pressure of the layer. This parameterization is most suited for aerosol profiles that are dominated by a single, elevated, optically thick aerosol layer. Examples are free tropospheric aerosols such as desert dust, biomass burning aerosols, or volcanic ash plumes. An alternative aerosol profile assumption is that of a homogeneous aerosol-laden boundary layer, where the top pressure of the aerosol layer is retrieved. The resulting aerosol profile product will be useful for operational applications, e.g. aviation safety and air pollution monitoring, and scientific applications, e.g. radiative forcing studies, long-range transport modeling and studies of cloud formation processes.

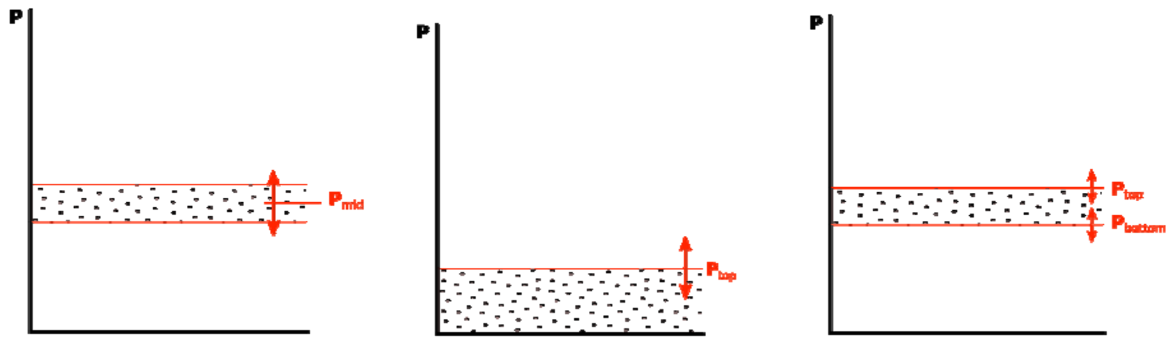



*Fig. 3. The three aerosol layer retrieval models used in this study. Model A is the baseline model (an elevated layer), in which the layer thickness is fixed (e.g. 20 or 50 hPa) and the mid-pressure is retrieved. Model B is the boundary layer model in which only the top of the layer is retrieved. In Model C both top and bottom pressure of an elevated layer are retrieved.*

Special attention in the radiative transfer model used for the retrieval is given to the derivatives. All derivatives are calculated in a semi-analytical manner. This allows for faster and more accurate calculations as compared to numerical differentiation techniques. Accurate calculation of the derivatives is particularly important for the aerosol profile algorithm, since fit parameters can be highly correlated.

Furthermore, special attention is given to the ground surface in the forward model. The surface albedo is allowed to depend linearly on wavelength, which is important for vegetated land and desert areas. In addition, the forward model accounts for fluorescence emissions from vegetation at the $O_2$ A band. Not accounting for fluorescence emissions will cause substantial biases in retrieved aerosol pressure. We show in a sensitivity study that fluorescence parameters can be fitted together with aerosol parameters from the $O_2$ A band.

The latest UVN instrument model is used in the retrieval simulations. Assuming a perfect aerosol model, the precision requirement of 50 hPa can be met.

Aerosol microphysical and optical properties show a large variation in time and space. The forward model used in the retrieval, however, can in principle assume only a single average aerosol optical model, with e.g. single scattering albedo (SSA) of 0.95 and Henyey-Greenstein phase function with $g$ of 0.7. Based on a sensitivity analysis, we show that for our aerosol profile retrieval algorithm a single aerosol optical model is sufficient for a reliable retrieval of aerosol pressure: the error in retrieved pressure remains small if the true aerosol type deviates from the assumed type.

Retrieved AOT and surface albedo, however, will show biases in response to model errors in aerosol optical properties. We show that it is essential in this respect to fit the surface albedo, next to optical thickness, as these two parameters will absorb the model biases in aerosol properties. In other words, not fitting the surface albedo will make retrieved aerosol pressure very sensitive to the aerosol model assumed in retrieval. Retrieved surface albedo and aerosol optical thickness should thus be understood as effective quantities.

The baseline algorithm for the AEROPRO case studies will fit one profile parameter. However, we have provided a sensitivity analysis investigating the possibility to add another profile parameter to the state vector. In an extensive sensitivity study we have investigated whether the algorithm would benefit from external information about the reflectance of the ground surface. We have found that this benefit is very limited, as the surface albedo signal from the $O_2$ A band is strong. Precision of retrieved pressure does not improve if retrieval is constrained by an a priori error in the surface albedo that is representative of uncertainties associated with current reflectance maps or climatologies. Possibly, external surface albedo information might benefit aerosol retrieval over land by providing better starting values for the fit and help the fit reach the correct $\chi^2$-minimum.

## 4. Performance and validation of the retrieval algorithm applied to GOSAT data

In TN5 and TN6 the fifteen selected GOSAT aerosol cases from TN2 have been processed according to the given priority. First of all, the GOSAT spectra were convolved to the UVN spectral resolution (see Fig. 4).

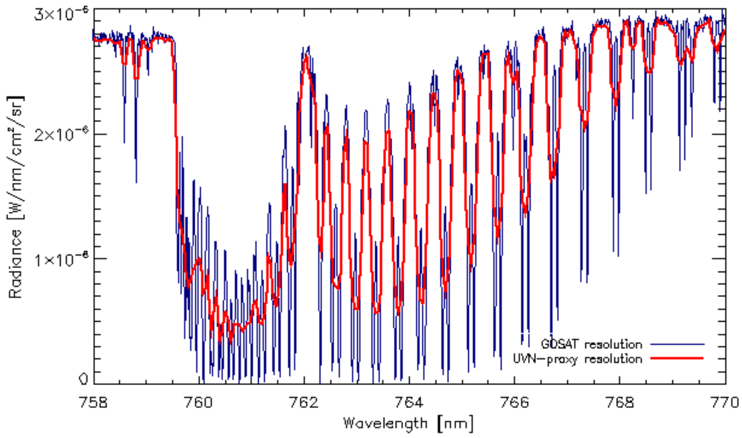



*Fig. 4: $O_2$ A-band radiance spectrum measured by GOSAT, showing the spectral convolution from the GOSAT instrument resolution (0.025 nm) to the S-4/UVN-proxy resolution (0.12 nm).*

All available information for each case was carefully taken into account, to check for e.g. absence of clouds. Next the retrievals were performed, in which many retrieval model settings were varied in order to get convergence. A priori values of the retrieved parameters were varied, and also the covariances between the retrieved parameters were calculated. The five fit parameters were: mid-pressure (model A) or top pressure (model B); AOT; surface albedo at two wavelengths; radiance bias (to correct for instrumental errors).
The following points were found:

- All cases converged but with different configurations and retrieval settings. Both aerosol layer models A and B were used; the single layer model itself was not investigated.
- If we did not fit the surface pressure or artificially raise it, or scale the $O_2$ cross section, we got almost no convergence.
- Boundary layer aerosol profiles had a difficulty to converge using the elevated layer model.

- Retrievals for cases without aerosols stopped at the lowest AOT value possible.
- Retrievals for cases relating to the same ground pixel gave similar results. Thus the algorithm gave consistent results.
- Dependence on the a-priori values of the fit parameters appeared to be case dependent.
- The spectral fit residuals have a characteristic shape (see Fig. 5). We found that line-mixing in the $O_2$ A-band has to be taken into account (see Fig. 6).

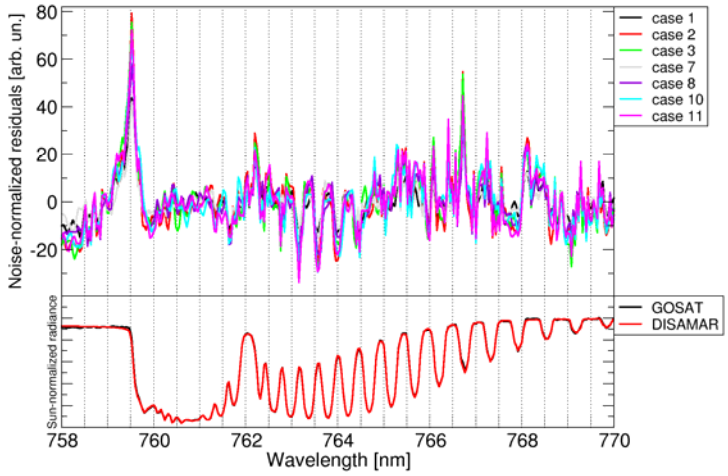



*Fig. 5: Upper plot: Spectral fit residuals (normalized to the instrumental noise) for 7 selected GOSAT cases. Lower plot: the $O_2$ A-band radiance spectrum of one of the cases, with both the measured (GOSAT) and the simulated (DISAMAR) spectrum.*

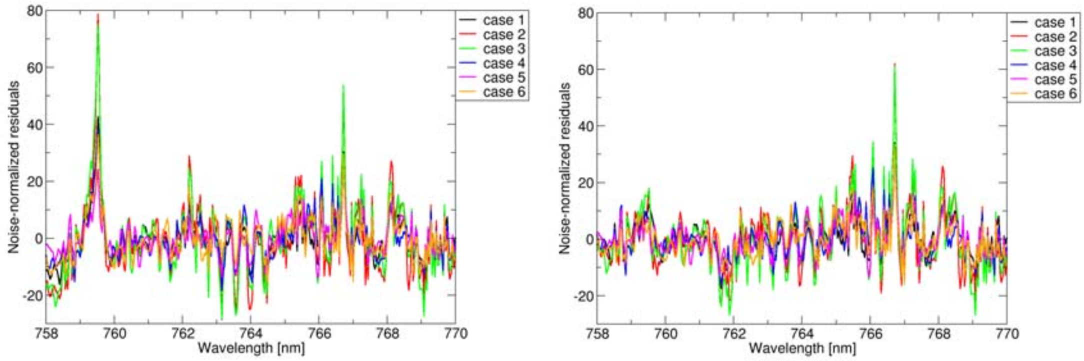



*Fig. 6. Left: Noise-normalized residuals of GOSAT retrievals for 6 cases when line-mixing is not taken into account in the simulations. Right: noise-normalized residuals when line-mixing <u>is</u> taken into account. For most wavelengths line-mixing reduces the residuals.*

As the next step, the validation was performed. In order to compare the retrieved aerosol layer heights more directly to the lidar aerosol profiles, we derived three layer height metrics from the lidar profiles: the center-of-gravity of the profile (effective height), the peak extinction height, and the profile top height. The retrievals agreed best with the lidar top height; see Fig. 7.

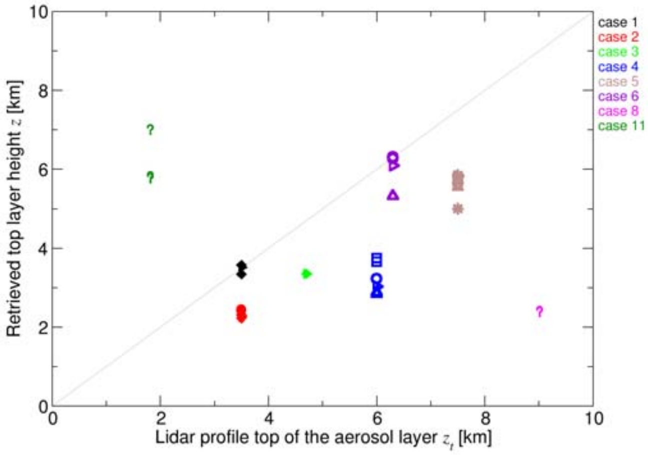



*Fig. 7: Comparison of GOSAT retrieved aerosol heights with the lidar profile top heights for 8 cases. Each case is separated by color. Multiple points per case belong to different retrieval settings. Full symbols: boundary layer model used, open symbols: elevated layer model used. Cases 8 and 11 are uncertain due to possible presence of cirrus.*

The validation results on aerosol layer height are encouraging. It seems that the algorithm is sensitive to the top height, and not to the effective height of the profile. It would be very interesting to study a volcanic plume case (truly elevated layer) to confirm this. All our cases were in some extent boundary layer cases, but the elevated layer parameterization only leads to convergence for some of them. It would be interesting to study the effect of the layer parameterization on the retrievals.

The GOSAT high resolution spectra were successfully degraded to UVN resolution. The noise levels compare quite well, except for the deepest part of the band where the GOSAT SNR is too low. But the sparse ground pixel sampling of GOSAT make it difficult to successfully find validation cases. Furthermore the GOSAT L1B product needs a lot of corrections by the user. Therefore, the project was extended to include GOME-2 data.

## 5. Performance and validation of the retrieval algorithm applied to GOME-2 data

In the extension of the AEROPRO project, described in TN8, the objective was to use GOME-2 on Metop-A data as a proxy for Sentinel-4. In the selection of cases we applied the same strategy as in TN2 albeit with a different source of satellite cloud data, namely AVHRR, also on Metop-A. The GOME-2 observations provide near global coverage in a single day, which enables us to select events that last for more than a single day, and choose an observation period that matches best with our criteria. We selected cases with different types of aerosols: boundary layer pollution, desert dust, smoke and volcanic ash. This resulted in the selection of 16 cases in total (see Fig. 8). There are 6 continental aerosol cases over CARBONEXP and Cabauw, 7 desert dust cases over Tenerife and Spain, and 3 volcanic ash cases over the North Sea area.

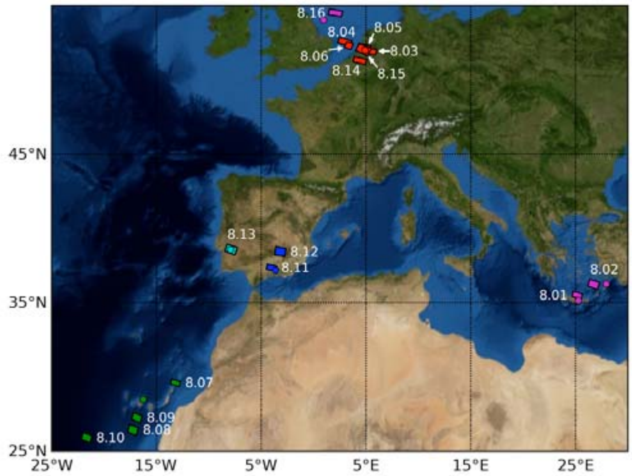



*Fig. 8: Location of the 16 selected cases for GOME-2.*

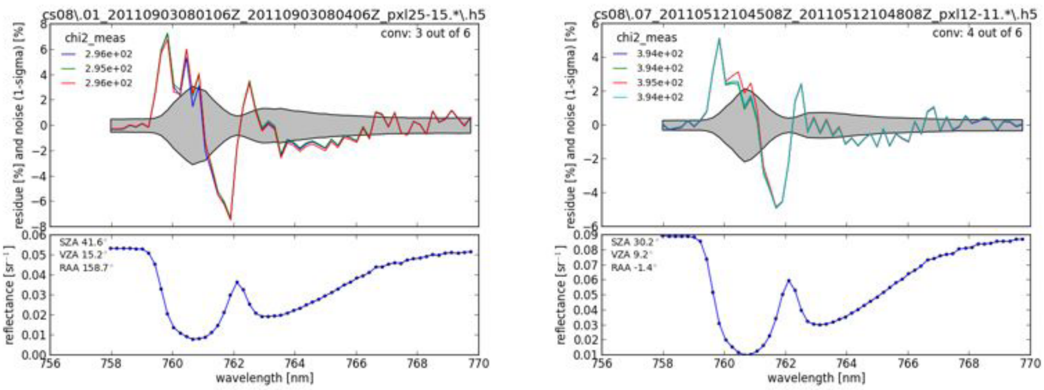



*Fig. 9: Example: fit residuals and fitted reflectance spectra of GOME-2 cases 1 and 7.*

From the selected cases, the GOME-2 pixels of 40x80 $km^2$ had to be chosen, in order to meet the criteria, especially absence of clouds. This was carefully done using different ground-based and satellite data sources: AERONET, MODIS, GOME-2 PMD, GOME-2 AAI, and AVHRR data. One case is shown in Fig. 8.

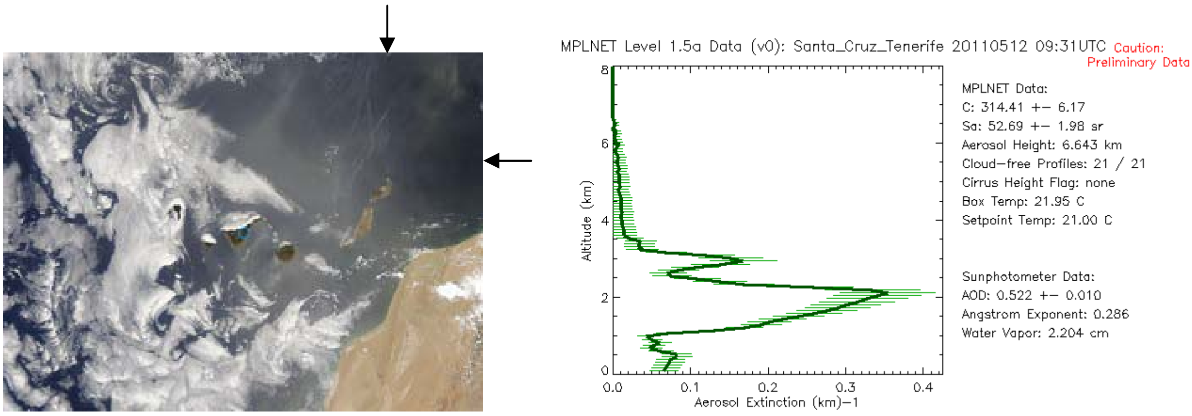



*Fig. 10: (Left) MODIS rgb image of GOME-2 case 7 (Tenerife, 12 May 2011), showing a dust layer. The location of the selected GOME-2 pixel is indicated by the arrows.*

*(Right) Lidar profile of the dust layer from the Santa Cruz de Tenerife station, indicated on the left plot with a blue circle.*

The performance of the retrieval algorithm regarding convergence and stability was systematically investigated. All 16 cases were treated in the same way. The fit parameters are: surface albedo (two wavelength nodes), AOT, and aerosol layer mid pressure. Convergence rate was highest when the oxygen absorption cross section was increased by 3% and a temperature offset was included in the fit. A systematic fit residual was found of the order of +/- 7 %; see Fig. 11.

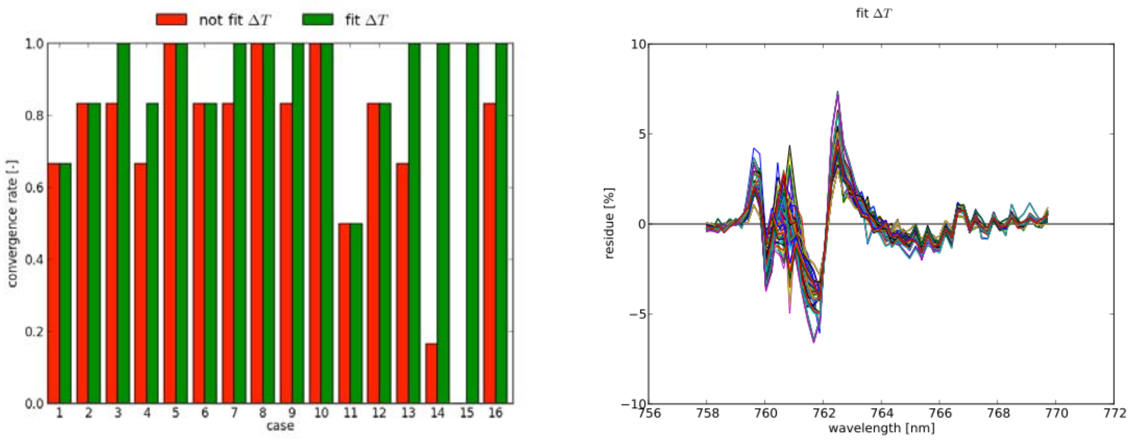



*Fig. 11. (Left) Convergence rate for the 16 GOME-2 cases in a retrieval experiment in which the temperature profile offset is fitted. In each case, 6 different a-priori values were used. (Right) GOME-2 fitting residuals for this experiment.*

For converging retrievals, the solution was stable but the retrieved mid pressures were typically unrealistically low; see Fig. 12. We find retrieved mid pressures between 200 hPa and 600 hPa, corresponding to altitudes between about 12 km and 4 km. However, we also see that once retrieval converges solutions are very stable, which suggests that these solutions are global chi-square minima.

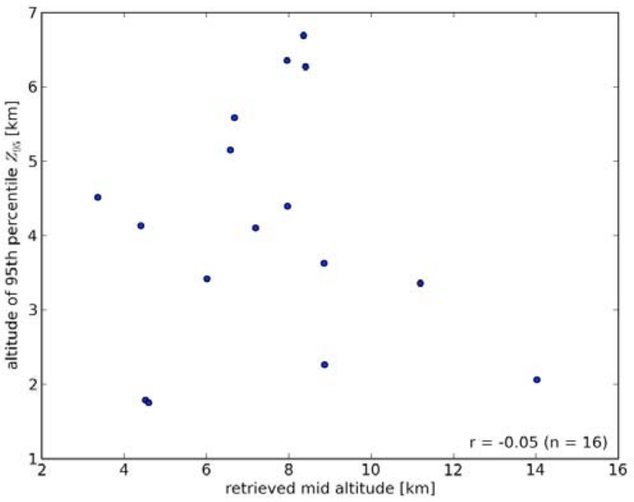



*Fig. 12: Scatter plot of GOME-2 retrieved layer mid altitudes and lidar-derived 95-th percentile altitudes (as proxy of the layer top altitudes). The correlation coefficient is -0.05.*

Subsequent retrieval simulations suggest that this bias of the retrieval towards high altitudes is due to boundary layer aerosols interacting with the surface. If aerosol is present in the boundary layer, the retrieved pressure of an elevated aerosol layer is not some average aerosol pressure, Instead, it is surprisingly biased strongly towards low pressures, especially for an optically thin elevated aerosol layer. The sensitivity of the retrieved aerosol layer pressure to the boundary layer AOT is shown in Fig. 13. The aerosol model in the retrieval algorithm should be extended to account for this boundary layer effect.

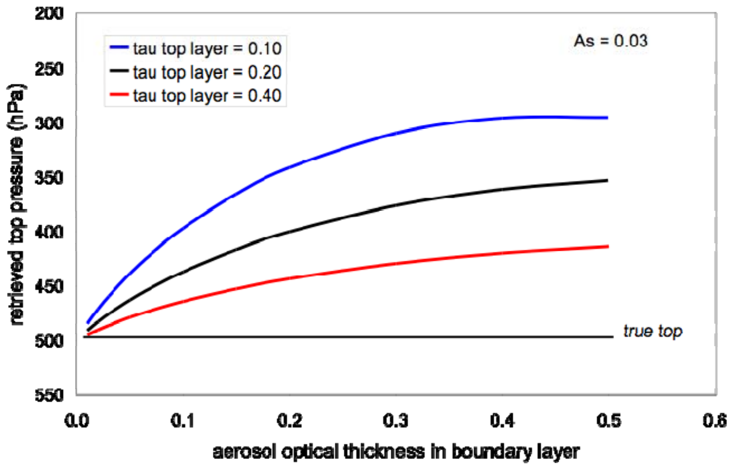



*Fig. 13. Simulation of retrieved top pressure of a single aerosol layer for a system consisting of two aerosol layers, as a function of the boundary layer AOT. It appears that the retrieved aerosol layer height rises (!) when the boundary layer AOT increases.*

## 6. Conclusions and recommendations

Much has been learned for algorithm development of aerosol height retrieval from the $O_2$ A-band by applying the AEROPRO algorithm to two sources of real data: the GOSAT Sentinel-4 proxy data and the GOME-2 data. Generally speaking, we think that the outcome of the AEROPRO retrieval performance and validation study using both sources of satellite data are relevant for both S-4 and S-5 processor development, regarding the retrieval model set-up and settings, the inversion approach, and the validation approach.

Sensitivity studies like those presented in TN3 provide a good view of what we can achieve in aerosol profile retrieval when using the $O_2$ A band. They can also provide a good theoretical understanding of the physics behind the algorithm. However, the forward model can never capture all the physics of the measurement: real data will be subject to unanticipated or unknown atmospheric model biases as well as instrument errors. Therefore, application of the algorithm to real data is needed, as performed here for GOSAT in TN5 and TN6 and GOME-2 data in TN8.

Both for GOSAT and GOME-2 data we find too high retrieved aerosol layers as compared to validation measurements. Furthermore, the residuals have a characteristic shape, which is different for GOSAT and GOME-2.  We conclude that boundary layer aerosol should be included in the a priori aerosol profile model. It appears that boundary layer aerosol plays an important role in the retrieval of an elevated aerosol layer, since fitting the surface albedo is not sufficient to represent boundary layer aerosol scattering. An option could be to extend the retrieval algorithm so that properties of two atmospheric intervals can be fitted. One layer represents the elevated aerosol layer and the other layer represents boundary layer aerosol. This option may work for most cases in practice. Fitting an extinction coefficient profile (possibly with a-priori smoothness constraint) and an output profile with correlations and averaging kernel might be another option. This is the cleanest approach from a theoretical point of view. Whether this option is possible in practice needs to be investigated. We furthermore found that we had to scale the $O_2$ cross-sections to get good convergence, both for GOSAT and GOME-2 data. More accurate $O_2$ A-band cross-section data are clearly needed.

The operational application of the retrieved aerosol height is data assimilation into the MACC/Copernicus Atmosphere Service model for the aerosol model forecasts. This product would also be useful for climate research. Two important issues that should be given attention when developing an operational algorithm are cloud contamination and computational speed. Screening pixels for optically thin cirrus clouds is particularly important, since these clouds can cause large pressure biases, as aerosols are optically thin too. Computational speed has to be increased, mainly by increasing the speed of the radiative transfer calculations.

For Sentinel-4 a three-step retrieval approach would be useful in practice:

1. Fast cloud/aerosol height retrieval from the Sentinel-4 $O_2$ A-band using a FRESCO-type look-up-table approach, to have a first guess of cloud fraction and cloud/aerosol height for all pixels.
2. Cloud filtering, using a high resolution cloud detection method using sub-satellite pixel imagery data and a dedicated high resolution cloud mask. This would prepare for filtering the data for the AEROPRO algorithm.
3. Perform the AEROPRO aerosol profile retrieval for the cloud-free scenes.

**Acknowledgements**

We would like to thank our colleagues M.O. Vieitez, J.F. de Haan, O. Tuinder, and C. Koning for their contributions to this project.

We gladly acknowledge the following agencies and persons for their help and assistance:

- ESA for funding; B. Veihelmann, Y. Meijer, and R. Koopman for many fruitful discussions.
- NIES: for providing GOSAT data and for pointed observations; H. Watanabe and T. Yokota for kind support.
- JAXA: for providing information on the GOSAT calibration; A. Kuze and H. Suto for much advice and early access to L1 data.
- EUMETSAT: for providing GOME-2 data; R. Lang for advice.
- DWD: for providing AVHRR L2 cloud data.
- NASA: for providing MODIS data.
- AERONET: for providing sunphotometer data.
- Earlinet/ACTRIS: for providing lidar profiles.