Global Mapping of Atmospheric Composition from Space
Retrieving Aerosol Height and Tropospheric NO2 from OMI

Chimot, Julien

DOI
10.4233/uuid:80c62d6f-0ae0-4e96-9554-841ddcd506c0

Publication date
2018

Document Version
Final published version

Citation (APA)
Chimot, J. (2018). *Global Mapping of Atmospheric Composition from Space: Retrieving Aerosol Height and Tropospheric NO2 from OMI*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:80c62d6f-0ae0-4e96-9554-841ddcd506c0

Global Mapping of Atmospheric Composition from Space

Retrieving Aerosol Height and Tropospheric NO_2 from OMI

Julien Chimot

On the cover

The front cover displays the spatial correlation between the human activities through the night lights and the NO_2 concentrations released in the troposphere over Europe and Asia. The night light picture is a composite from the NASA Suomi NPP satellite in 2016. The NASA night light images can be downloaded at `https://earthobservatory.nasa.gov/Features/NightLights/`. The tropospheric NO_2 picture is a 11-year average (2005-2015) from OMI downloaded from the NASA server `https://mirador.gsfc.nasa.gov/collections/OMNO2__003.shtml`. The back cover is a series of satellites studied by the Author during his career, from the beginning to his wishes in the future, over a night light image of America: IASI on MetOp, OMI on Aura, Sentinel-2, TROPOMI on Sentinel-5 Precursor, Sentinel-3, Sentinel-4, and Sentinel-5.

My sincere gratitude to my friend Rémi Charton for his talent and help on this cover!

Global Mapping of Atmospheric Composition from Space

Retrieving Aerosol Height and Tropospheric NO_2 from OMI

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen ergens in de verre toekomst op
maandag 10 september 2018 om 15:00 uur
door

Julien Jean-Michel CHIMOT

Master of Atmospheric Science and Oceanography (Paul Sabatier University, Toulouse, France)
Master of Environment Management and Monitoring (Paul Sabatier University, Toulouse, France)
Geboren te Agen, Lot et Garonne provincie, Frankrijk.

Dit proefschrift is goedgekeurd door:
Promotor: Prof. dr. P.F. Levelt
Copromotor: Dr. J.P. Veefkind

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. P.F. Levelt	Technische Universiteit Delft, KNMI
Dr. J.P. Veefkind,	Technische Universiteit Delft, KNMI

Onafhankelijke leden:

Prof. dr. ir. H.W.J. Russchenberg	Technische Universiteit Delft
Prof. dr. E.A.A. Aben	Vrije Universiteit (VU) Amsterdam SRON Netherlands Institute for Space Research
Prof. dr. J. Tamminen	Finnish Meteorological Institute (FMI)
Prof. dr. T. Wagner	Heidelberg Universiteit Max-Plank Institute (MPI) for Chemistry
Dr. K.F. Boersma	Wageningen Universiteit Research (WUR) KNMI

Reserve lid:

Prof. dr. ir. R.F. Hanssen	Technische Universiteit Delft

This work was funded by the Netherlands Space Office (NSO) under the OMI contract.

Keywords: trace gas, aerosol, cloud, air quality, climate, satellite remote sensing, atmospheric retrieval, spectral signature, radiation scattering, absorption

Printed by: ProefschriftMaken || `www.proefschriftmaken.nl`

ISBN 978-94-6366-010-5

An electronic version of this dissertation is available at
`http://repository.tudelft.nl`

Typeset by the author with the LaTeX Documentation System.

"Ma vie va vite, mon métier consiste à accélérer le temps, à être toujours en avance quand il s'agit de faire de mon mieux à la barre des bâteaux que je mène sur les cinq oceans. Cette fois, je ralentis un peu pour regarder le chemin parcouru [...] comme sur l'évolution d'une planète qui m'inquiète mais qui jamais ne me désespere [...] Être skipper requiert des compétences variées et permet d'explorer différents domaines de connaissances. Il faut être assez scientifique pour pouvoir défendre sa vision des choses face aux architectes et aux techniciens qui conçoivent et construisent vos bateaux. Il faut être suffisamment sportif pour résister au stress, supporter le manque de sommeil et être capable de monter au mat sans l'aide de personne. Il faut être un bon vendeur pour décrocher des budgets conséquents, un gestionnaire avisé pour ne pas envoyer par le fond la malle au tresor, et un DRH soucieux du bien-être de ceux qui l'épaulent tout en étant capable de mobiliser leurs talents. Il faut enfin et surtout être un marin qui ne craint pas de braver la tempête mais qui sait aussi apprécier le retour au calme, qui est heureux au large mais aime retrouver la terre."

Francois Gabart – Winner Vendée Globe
sailing race 2012, Rêver large

Contents

3 Aerosol layer height from OMI: development of a neural network approach based on the 477 nm O_2-O_2 spectral band 69

4 Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations 107

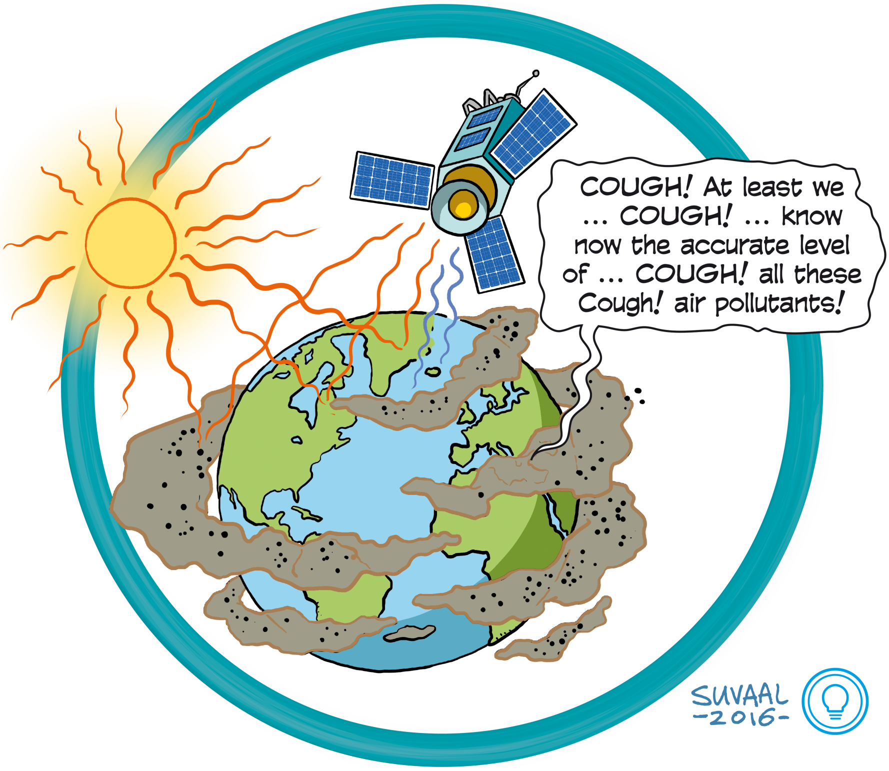


Figure 1: Summary of Julien Chimot Thesis. Cartoon made by the designer team of DIG-it!, and TU Delft Valorisation centre. Thanks to Dorien van Alphen, Robbert van Leeuwen, and Susanne Sleenhoff for the idea.

Summary

> "Notre maison brûle et on regarde ailleurs ("Our house is burning and we are looking elsewhere")."
>
> Jacques Chirac, French president 1995-2007,
> Earth Summit in Johannesburg, 2002

In the last decades, significant progress was achieved in global monitoring of the atmospheric composition from space, leading to key discoveries on the ozone layer, air quality and climate change. Recent research also shows that satellite observations can be used to monitor the effectiveness of policy measures on air quality and climate change. Therefore, it becomes more and more important to provide accurate data under any atmospheric condition. Furthermore, developing algorithms that can meet the requirements of an operational 24/7 data stream becomes challenging due to the tremendous growth of satellite measurements. Europe in general, the Netherlands specifically, have heavily invested in satellite instruments monitoring short-lived and greenhouse gases. Most notably, one can cite the Dutch-Finnish Ozone Monitoring Instrument (OMI) on-board the NASA's Earth Observation System (EOS) Aura platform, and the recently launched TROPOspheric Monitoring Instrument (TROPOMI), developed by ESA and the Netherlands, on the Copernicus Sentinel-5 Precursor.

Nitrogen dioxide NO_2 and particulate matter (referred to as aerosols) are both important constituents of the atmospheric composition. In the troposphere, NO_2 is emitted by fossil fuel combustion, biomass burning, as well as by lightning. It has not only adverse health effects, but also affects our atmosphere: 1) NO_2 plays a key role in the formation of tropospheric ozone O_3, a toxic element for humans and plants, 2), and NO_2 chemistry reactions lead to secondary aerosol formation.

Atmospheric aerosols are particles suspended in the air. Their sources are very mixed. Aerosols can be man-made or natural: e.g. smoke, desert dust, sea spray, nitrates and sulfates. Some of them are directly released in the air as particles: e.g. desert dust. Others are formed in the atmosphere resulting from chemistry reactions by precursor gases, such as SO_2 and NO_2. Because aerosols scatter and absorb sunlight, they perturb the radiative energy balance of the Earth, and thus affect climate. They also play a key role in the formation of clouds and precipitation. Aerosol effects are recognized as one of the largest uncertainties in our knowledge on climate change. Additionally, the small particles can deeply penetrate into the human respiration system leading to strong adverse health effects.

Both NO_2 and aerosols (e.g. nitrates, sulfates) are formed from combustion processes. Because their lifetime in the troposphere is comparable (NO_2 typically less than one day, aerosol 1 day to a week), the column concentrations show spatial correlation over regions where the aerosol type is dominated by large urban and industrial activities. The aerosol effects on the sunlight modify the shortwave radiation field in the atmosphere. As a consequence, they are a significant error source when exploiting satellite measurements devoted to trace gases, ocean color and vegetation. Indeed, such measurements are based on the backscattered sunlight at the top of the atmosphere in the shortwave spectrum. This mainly holds for cloud-free conditions where the aerosol signal is dominant.

The main objective of this thesis is to design a new aerosol layer height retrieval in order to improve the operational NO_2 retrieval, both in the troposphere, from space-borne instruments for highly polluted events and under cloud-free conditions. This thesis focuses on the exploitation of the OMI satellite measurements acquired in the visible wavelength range (405-490 nm). In addition, we develop numerical methods and tools (e.g. machine learning) in order to support the operational processing of big data amounts from the forthcoming new-generation satellite instruments for air quality and climate research.

In Chap. 2 of this thesis the error in the retrieved OMI tropospheric NO2 vertical column density (VCD) is quantified for current retrieval algorithms and over cloud-free scenes. Ignoring the aerosol effects in cloud-free satellite measurements, i.e. assuming aerosol-free conditions, this leads to a bias in the range of −60% to 20%, in case of high concentrations, Aerosol Optical Thickness (AOT) (550 nm) $>$ 0.6, scattering particles, and for summertime conditions. This clearly shows that a correction for aerosol effects is necessary. In the reference OMI tropospheric NO_2 dataset, named DOMINO-v2, an aerosol correction is included through the effective cloud parameters (fraction and pressure) derived by the OMI OMCLDO2 cloud algorithm. These parameters assume a Lambertian reflector model, i.e. clouds are represented as opaque layers that partly cover the observed scene. In the absence of clouds and in the presence of high aerosol loadings, the OMI cloud retrieval algorithm is sensitive to aerosol properties. This leads to an implicit aerosol correction. However, we found that in the original OMCLDO2 algorithm, as used in DOMINO v2, too coarse sampling of the look-up-tables (LUTs) leads to a too low effective cloud pressure retrieval. Consequently, retrieved tropospheric NO_2 VCDs are underestimated in the range of −40% to −20%, when aerosol particles are located at high altitude ($>$ 1.5 km), AOT(550 nm) $>$ 0.6, and in summertime conditions (chapter 2). Following this issue, the updated OMCLDO2 algorithm solves this issue.

The aerosol effect on the tropospheric NO_2 retrieval depends not only on the vertically integrated aerosol properties, such as AOT, but also how aerosols and NO_2 are vertically distributed in the troposphere. When aerosols are located above the NO_2 bulk, this leads to a reduction of the OMI measurement sensitivity to NO_2. On the contrary, an aerosol layer located below the NO_2 bulk leads to an enhancement.

In Chap. 3 of this thesis, we present a novel retrieval technique to derive the Aerosol Layer Height (ALH) from the OMI 477 nm (visible) O_2-O_2 spectral band over cloud-free scenes, as well over land as water surfaces. We make use of neural network, a specific machine learning approach that has the advantage of being able to

process the complete OMI dataset with a low time consumption. Therefore, this algorithm can deal with the big-data challenges of next generation satellite instruments. This algorithm can be used autonomously with the OMI data alone, or also benefit from the synergy with the NASA MODerate resolution Imaging Spectroradiometer (MODIS) instrument, on-board the EOS Aqua. Both OMI-Aura and MODIS-Aqua fly together in the NASA A-Train constellation. The advantage of such a synergy on the ALH retrieval accuracy is twofold: 1) filtering cloud-free OMI observation, and 2) using the reference AOT from the MODIS aerosol products as a prior information for the OMI ALH retrieval.

The performance of the OMI ALH algorithm is assessed by comparing with the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aerosol observations: first with its related climatology dataset over North-East Asia (Chap 3), and secondly on diverse selected cases (chapter 4). All the comparisons show differences between OMI and CALIOP ALH to be smaller than 800 m in case of cloud-free continental polluted cases and AOT(550 nm) > 0.5. In addition, OMI ALH demonstrates the capability of OMI visible measurements to probe the entire smoke layers, including large loading of absorbing particles, produced by intense biomass fires such as in South-America and East Russia (Chap. 4). The machine learning approach leads to the benefits of a low computing time (apart from the training task). Consequently, large ALH datasets are produced in this thesis, including 3 years of retrievals over North-East China (2005-2007) and a global yearly retrieval for 2006. The latter is obtained using a single processor during approximately 12 hours (Chap. 5). The OMI ALH neural network algorithm shows the potential to exploit the O_2-O_2 visible spectral band instead of (or in addition to) the traditional and more used O_2-A near infrared band from satellite sensors. The main limitations are due to the weak absorption of the O_2-O_2 complex dimer, the forward aerosol model uncertainty, potential cloud residuals, and the accuracy of the employed surface albedo database. The advantages of the visible O_2-O_2 band over the near-infrared O_2-A band are the higher AOT in the visible and the lower surface albedo over land surfaces.

In Chap. 6 of this thesis, we evaluate the improvements in the tropospheric NO_2 VCD retrieval thanks to different aerosol correction schemes using the new developed algorithms based on the OMI O_2-O_2 visible band: 1) the implicit aerosol correction using effective cloud parameters from the updated OMCLDO2, and 2) an explicit aerosol correction based on the OMI ALH retrieval and other selected aerosol parameters. The evaluation is done by comparing with the old OMCLDO2 algorithm. For that purpose, we reprocess 2 years of cloud-free DOMINO-v2 NO_2 data (2006-2007), in summer and winter, over north-East China and South-America. The new implicit aerosol correction shows an improved accuracy in the range of 0-20% for scenes with high aerosol loadings, scattering particles, and summertime conditions. However, such an approach still remains limited in case of more absorbing particles and does not comprehensively represent the single and multiple scattering effects inherent to aerosols. Applying an explicit correction using the OMI ALH retrieval also leads to an improved accuracy on the tropospheric NO_2 VCD. It applies a more physical modelling of the aerosol effects on the average light path thanks to the as-

sumed aerosol model in the training dataset of the NN algorithm. Therefore, more realistic vertical averaging kernel are derived, provided that correct aerosol parameters are assumed in the forward model. Higher VCD values are generally obtained by using the OMI ALH compared to OMCLDO2, between 20% and 40% depending on the seasons, regions and pollution episodes. This is likely due to differences between the considered models (i.e. Lambertian opaque reflector for the effective clouds vs. Henyey-Greenstein scattering phase function for aerosols), and/or the assumed height of aerosols and the effective cloud. Finally, the explicit aerosol correction shows higher accuracy in presence of absorbing particles, such as smoke. Overall, its quality does not only depend on the ALH accuracy, but also the set of assumed aerosol parameters (e.g. AOT, single scattering albedo, vertical profile shape, size), surface reflectance, and their resulting combination.

To develop a high-quality performance explicit aerosol correction by using the OMI ALH retrieval, several challenges need to be addressed. The main recommendations are (Chap. 7): to improve the accuracy of the OMI ALH, in particular with respect to the aerosol type knowledge, to define a consistent and accurate set of aerosol parameters that can properly be combined with the retrieved ALH, to pay attention to the OMI radiance closure budget, to improve the accuracy of surface albedo or reflectance and the NO_2 vertical profile shape.

Finally, although we mainly use OMI data for this research, all these developments and results can, in principle, be extended to other current and future satellites instruments, like TROPOMI, Sentinel-4 and Sentinel-5 sounders, and also to other trace gas retrievals, such as tropospheric SO_2 and HCHO. However, they will have to be adapted to the specificities of this new generation of instruments: e.g. the improvement in the spatial resolution. The small pixel sizes will clearly bring additional challenges such as the 3D effects of clouds, and thus the impacts on the adjacent observation pixels.

Samenvatting

In de afgelopen decennia is er significante vooruitgang geboekt op het gebied van het meten van de atmosferische samenstelling vanuit de ruimte, wat heeft geleid tot belangrijke ontdekkingen met betrekking tot de ozonlaag, luchtkwaliteit en klimaatverandering. Recent onderzoek heeft laten zien dat satellietobservaties gebruikt kunnen worden om de effectiviteit van het luchtkwaliteits- en klimaatbeleid aan te tonen. Daarom wordt het steeds belangrijker om nauwkeurige datasets te produceren onder alle atmosferische omstandigheden. Daarnaast wordt het ontwikkelen van algoritmes voor een 24/7 operationele datastroom steeds moeilijker door de enorme groei van het aantal satellietmetingen. Europa, en vooral Nederland, heeft flink geinvesteerd in de ontwikkeling van satellietinstrumenten die kortlevende gassen en broeikasgassen kunnen monitoren. De voornaamste voorbeelden hiervan zijn het Nederlands-Finse Ozone Monitoring Instrument (OMI), dat zich aan boord van de Aura Satelliet van het NASA Earth Observation System (EOS) bevindt, en het in 2017 gelanceerde TROPOspheric Monitoring Instrument (TROPOMI), dat ontwikkeld is door ESA en Nederland, en dat zich aan boord van de Copernicus Sentinel-5 Precursor bevindt. Stifstofdioxide (NO_2) en fijnstof (hierna: aerosolen) zijn beide belangrijke componenten van de atmosferische samenstelling. In de troposfeer wordt NO_2 uitgestoten door de verbranding van fossiele brandstoffen en biomassa en komt hetook vrij bij bliksem. NO_2 heeft negatieve gevolgen voor onze gezondheid, en ook effecten op de atmosfeer, zoals: 1) NO_2 heeft een belangrijke rol in de formatie van ozon (O_3) in de troposfeer, dat giftig is voor mensen en planten, en 2) chemische reacties van NO_2 dragen bij aan de vorming van secondaire aersolen.

Atmosferische aerosolen zijn deeltjes die door de lucht vliegen. Ze kunnen zowel een natuurlijke als een antropogene oorsprong hebben, bijvoorbeeld: rook, woestijnzand, zeezout, nitraten en sulfaten. Bepaalde aerosolen worden direct geemitteerd naar de atmosfeer als deeltjes, zoals woestijnzand, andere worden gevormd als gevolg van chemische reacties van gassen, zoals SO_2 en NO_2. Omdat aerosolen zonlicht verstrooien en absorberen, verstoren ze de stralingsbalans van de aarde en daardoor hebben ze een effect op het klimaat. Ze spelen ook een belangrijke rol in de formatie van wolken en neerslag. Aerosolen worden geschaard onder de grootste onzekerheden in het begrip van klimaatveranderingen. Daarnaast kunnen de kleine deeltjes diep in het ademhalingssysteem terecht komen, wat kan leiden tot negatieve gezondheidseffecten. NO_2 en aerosolen (de nitraten en sulfaten) worden gevormd bij verbrandingsprocessen. Omdat de levensduur in de troposfeer van beide vergelijkbaar is (NO_2 minder dan een dag en aerosolen maximaal een week), zijn de kolomconcentraties geografisch gecorreleerd in regio's waar de aerosolen voornamelijk voortkomen uit stedelijke of industriele uitstoot.

Naast hun effect op luchtkwaliteit en klimaat, beinvloeden aerosolen ook de nauwkeurigheid van satellietmetingen van onder andere sporegassen, oceaankleur en vegetatie. Dit komt doordat aerosolen de stralengang van zonnestraling in de atmosfeer beinvloeden. Dit heeft een effect op alle meetprincipes die gebruik maken van weerkaatst zonlicht. Het effect van aerosolen is relatief het grootst onder wolkenvrije condities.

Het belangrijkste doel van deze thesis is het ontwikkelen van een algoritme om de hoogte van aerosollagen te bepalen en daarbij een verbetering in de operationele NO_2 bepaling te bewerkstelligen. Deze thesis richt zich op de exploitatie van metingen in het visuele gedeelte van het spectrum, tussen 405 en 490 nm, gemaakt door het OMI satellietinstrument. Daarbij maken we gebruik van numerieke methoden en hulpmiddelen (bijvoorbeeld Machine Learning), die de operationele verwerking van grote datasets aan kunnen, met het oog op de aankomende nieuwe-generatie satellietinstrumenten voor luchtkwaliteits- en klimaatdoeleinden.

In hoofdstuk 2 van deze thesis wordt de fout, die gemaakt wordt in de bepaling van troposferische verticale kolomdichtheid (VCD) van NO_2, gekwantificeerd voor onbewolkte condities. Als de effecten van aerosolen buiten beschouwing worden gelaten, met andere woorden als we de effecten van aerosolen verwaarlozen, wordt bij een aerosol optische dikte (AOT)(550 nm) > 0.6 en zomerse condities een gemiddelde fout gemaakt van -60% tot +20%. Dit laat duidelijk zien dat een correctie voor aerosolen noodzakelijk is. In de OMI NO_2 troposferische referentiedataset, genaamd DOMINO-v2, is een aerosolcorrectie opgenomen door middel van effectieve wolkenparameters (fractie en druk) van het OMI OMCLDO2 wolkenalgoritme. Deze parameters nemen een zgn. Lambertiaans reflectiemodel aan, waarbij de wolken worden gerepresenteerd als ondoorzichtige lagen die voor een gedeelte de scene bedekken. Als er geen wolken zijn maar wel aerosolen, dan zijn de effectieve wolkenparameters gevoelig voor de concentratie van aerosolen en de verticale verdeling daarvan. Dit leidt tot een impliciete correctie van aerosoleffecten in de NO_2 VCD. Desalniettemin, vinden we dat een te grofmazige aanpak in de versie van het OMCLDO2 algoritme dat gebruikt wordt door DOMINO-v2, leidt tot een te lage waarde voor troposferische NO_2 VCD, tussen de -40% en -20% bij een aerosol optische dikte (AOT)(550 nm) ≥ 0.6 en zomerse condities. Het huidige verbeterde OMCLDO2 algoritme lost dit probleem op.

De aërosoleffecten op de troposferische NO_2 bepaling is niet alleen afhankelijk van de verticaal geÃŕntegreerde aerosoleigenschappen, zoals AOT, maar ook van de verticale verdeling van aërosoleffecten en van NO_2 in de atmosfeer. Als de aërosoleffecten zich boven het grootse deel van de NO_2 bevinden, leidt dit tot een vermindering van de gevoeligheid van OMI voor NO_2. Als de aerosollaag zich onder de NO_2 bevindt, leidt dit tot een verbetering van de gevoeligheid. In hoofdstuk 3 van deze thesis, presenteren we een nieuwe methode om aerosollaag hoogte (ALH) uit de OMI 477 nm O_2-O_2 spectrale band te bepalen voor wolkenvrije scenes boven land- en wateroppervlaktes. Het ALH-algoritme maakt gebruikt van neurale netwerken, een machine-learning methode, die zeer snel grote datasets kunnen verwerken. Daarmee kan het algoritme de grote datasets van de volgende generatie satellietinstrumenten gemakkelijk aan. Dit algoritme kan werken met alleen OMI data, of profiteren van de synergie met de NASA MODerate resolution Imaging Spectroradiometer (MODIS), aan boord van de EOS Aqua. OMI-Aura en MODIS-Aqua vliegen beide in de NASA A-Train satellietconstellatie. Er is een tweetal voordelen aan de synergie voor de ALH bepaling: 1) de filtering voor wolkenvrije OMI observaties verbeterd en 2) de AOT van de MODIS producten kan als a-priori informatie gebruikt worden.

De resultaten van het OMI ALH algoritme zijn getest door middel van een vergelijking met de Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aerosol observaties. Eerst met klimatologische datasets over Noordoost-Azie (hoofdstuk 3) en daarna voor verschillende geografische locaties (hoofdstuk 4). Alle vergelijkingen laten verschillen zien tussen OMI en CALIOP ALH die kleiner zijn dan 800 meter, onder wolkenvrije omstandigheden met een AOT(550 nm) > 0.5. Bovendien laat de OMI ALH zien dat OMI in staat is om in het visuele spectrum door dikke rooklagen heen te kijken, zelfs bij grote hoeveelheden absorberende deeltjes, die veroorzaakt worden door intense bosbranden zoals in Zuid-Amerika en Oost-Rusland (hoofdstuk 4). De machine-learning techniek leidt tot een snel algoritme. Vanwege de korte rekentijd, kunnen grote ALH datasets geproduceerd worden, zoals de dataset van drie jaar voor Noordoost-China (2005-2007) en een globale dataset voor heel 2006. De laatste dataset was gemaakt op een enkele processor binnen 12 uur (hoofdstuk 5). Het OMI ALH neurale netwerkalgoritme laat de potentie zien om de O_2-O_2 visuele spectrale band te gebuiken in plaats van (of naast) de traditioneel en meer gebruikte O_2-A nabij-infrarode band. De grootste beperkingen zijn de zwakke absorptie van O_2-O_2, de onzekerheid van het gebruikte aerosolmodel, niet gedetecteerde wolkenflarden, en de nauwkeurigheid van de toegepaste oppervlakte-reflectiedataset. De voordelen van de zichtbare O_2-O_2 band ten opzichte van de O_2-A band zijn een hogere AOT en een lagere oppervlaktealbedo over landoppervlaktes.

In hoofdstuk 6 van deze thesis evalueren we de verbeteringen van de troposferische NO2 VCD bepaling, met behulp van de volgende twecorrectietechnieken, gebruikmakend van de ontwikkelde algoritmes voor de OMI: 1) de impliciete aerosolcorrectie , die gebruik maakt van effectieve wolkenparameters van de verbeterde OMCLDO2 en 2) een expliciete aerosolcorrectie gebaseerd op de OMI ALH bepaling. Daarvoor zijn twee jaar aan wolkenvrije DOMINO-v2 NO_2 data (2006-2007) opnieuw geprocessed, voor de zomer en de winter over Noordoost-China en Zuid-Amerika. De nieuwe impliciete aerosolcorrectie laat een verbetering van de nauwkeurigheid zien van 0-20% voor scenario's met hoge aerosolconcentraties, verstrooiende deeltjes en zomerse omstandigheden. Daarentegen is de correctie beperkt in het geval van meer absorberende deeltjes. De expliciete correctie leidt ook tot een betere OMI ALH bepaling en leidt tot een verbeterde nauwkeurigheid van NO_2 VCD. Hiermee wordt een betere fysische beschrijving gegeven. Hierdoor wordt de schatting van de verticale gevoeligheid van de meting ook verbeterd. Het gebruik van het OMI ALH algoritme geeft 20% tot 40% hogere waardes voor de NO_2 VCD, in vergelijking met OMCLDO2, afhankelijk van de seizoenen, regio's en de aerosolconcentratie. Dit wordt waarschijnlijk veroorzaakt door het verschil tussen de modellen: een Lambertiaanse wolk in het geval van OMCLDO2 en een aerosollaag voor OMI ALH.

Uiteindelijk laat de expliciete aerosolcorrectie een betere nauwkeurigheid zien in de aanwezigheid van absorberende deeltjes, zoals rook. De gehele nauwkeurigheid hangt niet alleen af van de ALH nauwkeurigheid, maar ook van andere aangenomen aerosolparameters (te weten: AOT, verstrooiingsalbedo, verticale profielvorm, grootte), en de oppervlaktereflectie.

Om een aerosolcorrectie van hoge kwaliteit te ontwikkelen gebruik makend van de OMI ALH bepalingen, moeten een aantal problemen worden overwonnen. De belangrijkste aanbevelingen zijn (hoofdstuk 7): 1) om de OMI ALH te verbeteren, met name met betrekking tot het aerosoltype, 2) om een consistente en accurate set van aerosolparameters te definieren, die op een adequate manier gecombineerd kunnen worden met de ALH, 3) om ervoor zorg te dragen dat de aangenomen atmosfeer-en oppervlakteparameters consistent zijn met de gemeten hoeveelheid straling, en 4) om de nauwkeurigheid van de oppervlaktereflectie en de NO2 profielvorm te verbeteren. Uiteindelijk kunnen al deze ontwikkelingen en resultaten, die behaald zijn met behulp van het OMI instrument, ook worden toegepast op andere huidige en toekomstige satellietmissies, zoals TROPOMI, Sentinel-4 en Sentinel-5, en ook voor de bepaling van andere spoorgasbepalingen, zoals troposferische SO_2 en HCHO.

(Many thanks to Marcel Kleinhereinbrik for the translation, and Thomas Frederikse and Pepijn Veefkind for the reviews)

Chapter 1

Introduction

> "Indifference can be tempting – more than that, seductive [...] It is so much easier to avoid such rude interruptions to our work, our dreams, our hopes [...] Better an unjust God than an indifferent one [...] In a way, to be indifferent is what makes the human being inhuman. Indifference, after all, is more dangerous than anger and hatred. Anger can at times be creative [...] Indifference is never creative [...] Indifference elicits no response. Indifference is not a response."
>
> Elie Wiesel, the Perils of Indifference, 1999

1.1 Atmospheric composition: physical structure

The atmosphere is a 100 km layer composed of gases (i.e. air molecules with a size of about 1 nm or less), clouds (ice crystals or water droplets), and particles (larger than 1 nm) between the Earth surface and the vacuum of outer space (Burrows et al., 2011). This layer is retained by the Earth gravity. Together with the Sun and the surface, they form a complex but essential system for maintaining our environment and life. Historically, the young atmosphere, at the formation of the Earth approximately 4.54 billion years ago, was produced by out gassing from volcanic eruptions and surface. The big rise of oxygen was brought by the biosphere when life appeared about 3.8 billion years ago. And the great oxidation event took place 2.2-2.4 billion years ago.

Nowadays, more than 99 % of the dry air molecules only include 2 gases: dinitrogen N_2 (78 %) and dioxygen O_2 (21 %). Some rare gases such as Argon (0.93 %) are also present. The rest of the gases part of the dry atmosphere is called "trace gases". Their small relative amount (close to 0.04 %) is one of the main challenges of scientists working on the atmospheric composition measurements. Among these elements, one can name the green-house gases (e.g. CO_2, CH_4, N_2O, O_3), and the pollutant gaseous with a relative short lifetime (e.g. NO_2, SO_2, CO, etc...). Their very low amount does not mean they are inert in the atmosphere: most of them directly affect the atmospheric radiation and thus our climate (see Sect. 1.4.2), air quality (see Sect. 1.4.4) and the chemistry processes in our atmosphere (see Sect. 1.2).

Water vapour H_2O and ozone O_3 should be carefully distinguished from all these other trace gases:

- Water liquid and water vapour play a central part in atmospheric and hydrology processes. Water vapour abundance is very variable at the surface from almost null over desert regions to about 4 % over oceans. It is the main component of weather and climate, accounting for about 90 % of the Earth's natural greenhouse effect. When air temperature falls below the "condensation temperature", the air becomes saturated and the water vapour condenses then into cloud droplets. This usually occurs during adiabatic cooling, when moist air is lifted.

- Ozone is an important contributor of Earth's radiative balance and air quality. Depending on its location, ozone can be a life saving or a toxic gas. At high altitude (i.e. the stratosphere), its high abundance is vital for living species in the lower layers as it absorbs harmful ultraviolet (UV) radiation coming from the Sun leading then to stratospheric heating. Its concentration decrease, detected during the 1980's, has led to major concerns (see Sect. 1.4.3). But its presence close to the surface is harmful to animals and humans and damages natural systems and agricultural crops. Ozone is also an important green-house gas in the whole atmosphere due to its thermal infrared absorption and emission properties. It makes a significant contribution to the radiative balance of the upper troposphere and lower stratosphere. Its main uncertainty is related to the changes with respect to altitude that depend on the regions and other chemistry processes (see Sect. 1.4.2) (Lacis et al., 1990; Worden et al., 2008).

The atmosphere is generally assumed to be in an hydrostatic equilibrium: i.e. particles and molecules do not escape in large quantity due to the Earth gravity, and do not collapse at the surface as the gas volume is governed by the temperature and pressure conditions (i.e. the ideal gas law). As a consequence, the atmospheric pressure obeys the barometric equation and exponentially decreases with respect to altitude, about a factor of 10 every 17 km. Generally speaking, the average surface pressure at sea level is 1013 hPa but drops to about 0.02 hPa at 80 km. This means that 99.9995 % of the atmosphere density is located below this altitude, close to the surface. The dependency of temperature on altitude leads to divide the atmosphere in multiple vertical layers with either positive or negative temperature gradient (see Fig.1.1):

- The troposphere (Greek: well mixed region), or lowest part of the atmosphere, typically extends from 0 to 9 km at the poles, 17 km at the equator. The top of the troposphere is named "tropopause height". Within this layer, temperature steadily decreases at a rate of 6.5°C per km. This negative gradient leads to convective and turbulent mixing and actually provides our weather. Internally, one usually distinguish the planetary boundary layer (PBL), closest layer to the surface, and the free troposphere. The PBL height is determined by the vertical transport or buoyancy due to thermal convection: the air rises as the

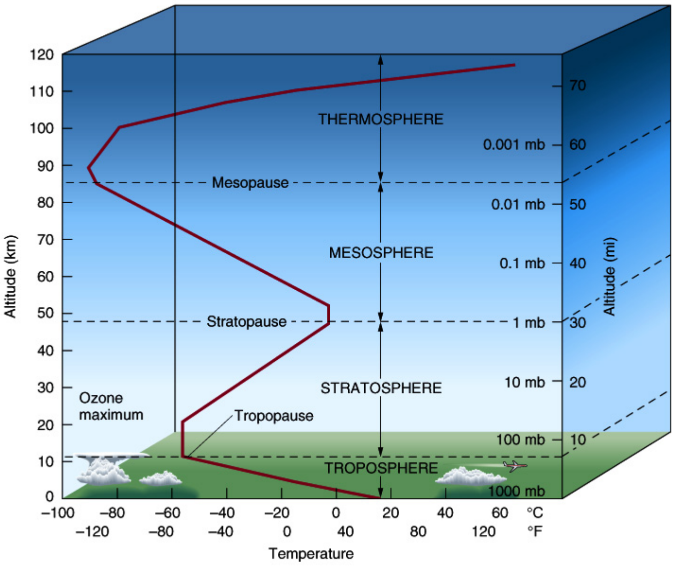


Figure 1.1: The physical vertical structure of the atmosphere as a function of altitude (University of Lagos, `http://unilaggeography2012.blogspot.nl/p/gry-101-introduction-to-physical.html`).

Sun heats the surface, expanding then its volume due to the lower pressure at higher altitude. It is quite variable, from less than 500 m in winter to 2000 m in summer, and presents a strong diurnal cycle. Air pollution is mostly produced at the surface and usually remains confined because of the temperature inversion. Transport to the free troposphere is limited although possible. Changes in the troposphere directly affect life and our environment. Therefore, observations of tropospheric composition changes is of high importance (Burrows et al., 2011).

- The stratosphere (Greek: stratified region) from above the tropopause height up to about 50 km. This atmospheric region is characterized by a positive temperature gradient: i.e. temperature rises with increasing altitude. This is primarily due to the reaction of shortwave UV radiation with O_3 and O_2. Reaction of UV with O_2, named photolysis, leads to O_3 formation and heat as a byproduct. O_3 absorbs UV radiation (because of its low stability) and emits thermal infrared and then heats the surrounding layers. Ozone concentration steadily rises from the tropopause to about 40 km, explaining then in part the constant rise of temperature. Such a temperature profile creates very stable atmospheric conditions. Furthermore, stratospheric H_2O concentration is very low. Consequently, the stratosphere is almost completely free of clouds or other forms of weather.
- And higher layers: mesosphere (Greek: middle) from 50 to 80 km, thermosphere (Greek: heated region) from 80 km to 700 km, and exosphere from

700 km.

1.2 Atmospheric chemistry and the importance of NO_x molecules

1.2.1 The atmosphere – A chemical laboratory

A chemical reaction occurs when atoms or molecules collide: i.e. they can merge, break apart or rearrange. A lot of reactions occur because of photons, and thus in presence of sunlight. The reaction rate also depends on the kinetic energy, and thus on temperature and pressure. There is a countless number of chemical reactions determining our atmospheric composition. Altogether, they form a whole complex and coupled system.

As an illustration, perhaps one of the most famous reaction is the photolysis: the Sun light energy breaks a chemical bond in a molecule which usually includes oxygen atoms. For example, the slow dissociation of dioxygen by energetic UV radiation is an important photolysis reaction in the stratosphere:

$$O_2 + h\nu \rightarrow 2O, \ \ \lambda < 240 \ nm \tag{1.1}$$

But the produced oxygen atoms quickly rearrange with dioxygen molecules to form ozone:

$$O + O_2 \rightarrow O_3 \tag{1.2}$$

This new ozone easily absorbs UV radiation, protecting life on Earth (cf. Sect. 1.1).

$$O_3 + h\nu \rightarrow O_2 + O, \ \ \lambda < 320 \ nm \tag{1.3}$$

Reactions 1.1–1.3 illustrate the rapid cycling between O and O_3 exclusively in presence of O_2 and light. But, in presence of an oxygen atom, ozone slowly transforms back to O_2:

$$O_3 + O \rightarrow 2O_2 \tag{1.4}$$

$$3O_2 + h\nu \leftrightarrow 2O_3 + h\nu \tag{1.5}$$

This is only an example of many processes occurring in the stratosphere. They are part of the Chapman mechanism explaining the formation and presence of the ozone layer as a function of the square of the O_2 concentration, around 15-30 km height (Chapman, 1930). Above this layer, not enough oxygen molecules are found to produce ozone, while below most UV radiation has been absorbed.

Many more reactions take place in our atmosphere. The later sections mostly focus on the troposphere and the nitrogen oxides NO_x molecules, the central elements of this thesis.

1.2.2 The crucial role of NO_x in tropospheric chemistry

Tropospheric constituents are regulated by natural sources such as the biosphere, exchange at the surface (land, ocean and cryosphere), lightning, natural fires and stratospheric-tropospheric exchange, and anthropogenic activities like biomass burning, the combustion of fossil fuels, and land usage (Holloway and Wayne, 2010). In general during pollution episodes, air masses contain large amounts of O_3, aerosols, acid, and other noxious chemical species compared to unperturbed air masses.

Nitrogen oxides NO_x are a family of molecules grouping both NO and nitrogen dioxide NO_2: $NO_x = NO + NO_2$. They are built upon nitrogen and oxygen atoms, the most abundant elements in the atmosphere. Their relative abundance is very low, less than one part per million. Nevertheless, scientific interest is driven by their crucial influence on the tropospheric part where humans live. The main NO_x sources are of two types, all related to combustion processes:

- **Thermal**: at temperature above 2000 K, dioxygen molecules thermolyze (e.g. in combustion chambers, biomass burning or with lightning) leading to a subsequent reaction between oxygen atoms and nitrogen N_2 molecules:

$$O_2 + (heat) \longleftrightarrow 2O \tag{1.6}$$

$$O + N_2 \longleftrightarrow NO + N \tag{1.7}$$

$$N + O_2 \longleftrightarrow NO + O \tag{1.8}$$

- **Fuel**: oxydation of nitrogen-bearing fuels such as coal and oil releases the nitrogen N bound as a free radical.

NO_x are highly reactive. Their chemical lifetime increases with altitude, from less than 1 day in polluted boundary layer to 5-10 days in the upper troposphere. At night, all NO_x are exclusively NO_2 due to absence of light. High NO_2 concentration leads to the formation of a reddish brown smog that hangs over most of the very large cities in the world (see Fig. 1.2). The negative consequences are numerous:

- Exposure to NO_2 leads to adverse health impacts (see Sect. 1.4.4) (for Europe, 2003). In particular, NO_2 enhances the effect of allergens, bronchial reactivity and increases admissions for respiratory disease (Sunyer et al., 1997).
- NO_x are the precursors of the semi-volatile ammonium nitrate NH_4NO_3, an important component of secondary aerosols and is also harmful for the health of people (see Sect. 1.4.4).

- NO_x contribute to acidification and eutrophication of soils and surface waters, and acid rain (see Sect. 1.4.1).
- The chemical budget of tropospheric O_3, which leads as well to adverse health effects for humans and stress the vegetation, is largely determined by the concentration of NO_x (Jacob et al., 1996).
- NO_x have complex and diverse effects on climate. On the one hand, they lead to a warming effect due to the greenhouse nature of O_3 (Ramanathan et al., 1985). On the other hand, they also lead to a cooling effect, notably via the secondary aerosol formation. Moreover, high concentration of NO_x leads to high abundance of OH reducing then the levels of the greenhouse gases O_3 and Methane CH_4 (see Sect. 1.4.2 and 1.5) (Shindell et al., 2009). Overall, the average resulting impact of NO_x is an indirect cooling effect on our climate, but at the detriment of a poor air quality (Shindell et al., 2009).

Figure 1.2: A high NO_2 pollution in the Los-Angeles US city, characterized by the reddish brown smog (`http:chem.libretext.org/Core/Physical_and_Theoretical_Chemistry`).

As mentioned above, NO_x play a central role in the production of tropospheric O_3. This occurs via the photolysis of NO_2 by sunlight at wavelengths ≤ 420 nm:

$$NO_2 + h\nu \rightarrow NO + O(^3P) \tag{1.9}$$

$$O(^3P) + O_2 + M \rightarrow O_3 + M \tag{1.10}$$

M is any other molecule. After this photostationnary equilibrium, NO and O_3 are usually consumed to form NO_2 again:

$$NO + O_3 + M \rightarrow NO_2 + O_2 \tag{1.11}$$

However reactions in 1.9, 1.10 and 1.11 do not lead to a net ozone production. Alternative reactions involving the hydroperoxyl radical HO_2 allow the conversion of NO into NO_2 without oxidizing O_3:

$$HO_2 + NO \rightarrow NO_2 + OH \tag{1.12}$$

The availability of HO_2 guarantees the extra-production of O_3. The production mechanisms of HO_2 are usually ensured by the presence of carbon monoxide CO and the incomplete combustion of hydrocarbons and/or the oxidation of volatile organic compounds (VOCs). Other multiple stage reactions also occur when NO molecules encounter CO and O_2, CH_4 and O_2 or non-methane hydrocarbons. For all these reactions, NO molecules evolve to NO_2 without interacting with O_3.

The regime of tropospheric O_3 can be then limited by a reduction in NO_x as well as CO or hydrocarbon, or a decrease of sunlight (e.g. in wintertime). Unfortunately, these trace gases are very abundant in city centres (Sillman et al., 1990). It also critically depends on the vertical distribution of NO_x, small amount of NO_x ending up in the free troposphere has a much longer lifetime and therefore a much stronger ozone production potential.

A significant part of the complexity of tropospheric chemistry is also related to the free radical OH which governs the oxidation of our air, and therefore subsequent transformation of chemical species either naturally released by biomass or issues from anthropogenic activities. OH has an average lifetime of 1 second and is sometimes called "detergent of the atmosphere" due to its strong oxidizer capability. OH is a particularly reactive specie and an important sink of the green-house gas CH_4 via the following slow reaction (Burrows et al., 2011):

$$OH + CH_4 \rightarrow CH_3 + H_2O, \tag{1.13}$$

and also reacts with CO and other hydrocarbons leading to the production of HO_2:

$$OH + CO + O_2 \rightarrow CO_2 + HO_2. \tag{1.14}$$

NO_x does not only trigger tropospheric O_3 production (Eq. 1.9 and 1.10) but also influences OH abundance. Indeed, O_3 is the primary source of OH in the troposphere in the presence of water vapour and sunlight with wavelengths $\leq$ 320 nm.

NO can also create OH by itself, in presence of HO_2 (cf. Eq. 1.12). Therefore, NO catalyses oxidation by OH which can be seen as a way to prevent toxic levels of CO and hydrocarbon concentrations in the atmosphere. However, at high NO_x concentration levels, NO_2 removes OH by forming HNO_3 leading then to the formation of semi-volatile ammonium nitrate aerosols in the presence of ammonia NH_3.

1.3 Atmospheric particles: Aerosols

Aerosols are small and complex chemical mixture of liquid and particles suspended in the air. In practice, referring to aerosols include then both particles and its surrounding medium. Their presence not only perturbs our climate system, human health, but also interferes with satellite observations of atmospheric trace gases (see Sect. 1.6.3). Furthermore, scattering and absorption by aerosols impact the actinic flux, and consequently modifies the photolysis rates of important processes in the atmosphere(Palancar et al., 2013).

Aerosol particles differ from gases as they are bigger than molecules. There are several classifications of atmospheric aerosols but the most widely used is according to their size. They range from the smallest superfine mode, with diameters of a few nanometers (nm), to large coarse mode particles, with diameters to more than 100 micrometers (μm) or more. Between the superfine and the coarse mode particles are the fine mode particles, with diameters ranging from 0.1 mm to a few μm (Seinfeld, 1986). In polluted conditions, they are often denoted as particulate matter (PM): e.g. PM10 referring to the dry mass of particles with a diameter less than 10 μm.

Aerosol sources combine both natural and anthropogenic processes, and are of two types:

- a direct emission resulting from dispersion of material at the Earth surface (e.g. sea spray aerosol from sea surface or waves, dust - Fig. 1.3 from desert outbreak, biomass burning aerosol, volcanic ash resulting from eruptions, primary organic aerosol, industrial debris).

Figure 1.3: Dust particles encountered when driving to the entrance of the Death valley, California, USA. Photos taken by Julien Chimot and Carole Legorgeu in April 2017.

- or indirectly from precursor trace gas emission leading to the formation of secondary aerosols (e.g. sulfates, nitrates, ammonium salts, secondary organic aerosol), or condensation or coagulation processes.

There are still some gaps in our understanding and modelling of aerosol sinks (IPCC, 2014). After their release, they undergo various physical and chemical processes modifying then their size and optical properties. They are generally removed from the air by dry or wet depositions, depending on their size and Earth's surface characteristics, at the surface (Kerkweg et al., 2006; IPCC, 2014). Wet deposition occurs by in-cloud and below-cloud scavenging. The remaining particles go through dry deposition processes such as turbulent diffusion, gravitational sedimentation, impact with obstacles and/or Brownian diffusion.

The global total aerosol mass is dominated by natural processes at the surface, in particular sea spray aerosol and desert dust. However, anthropogenic emissions (e.g. industries, vehicles, agriculture, wildfires etc...) of both primary particles and precursor gases greatly increases the total aerosol load and can locally outweigh the natural aerosols (Andreae and Rosenfeld, 2008).

1.4 Tropospheric composition changes – Life and environmental issues

Trace gas concentrations have been continuously increasing in the troposphere, since the industrial revolution in the middle of the 19^{th} century. The associated fossil fuel combustion processes perturb our air composition and lead to major impacts on our environment. Some examples are listed below.

1.4.1 Acid rain and eutrophication

Acid deposition is a phenomenon resulting from complex reactions between OH, NO_2, and SO_2. While the natural pH value of rain is lower than 7 (mainly due to dissolved CO_2), the term acid rain refers to precipitation with a pH value less than 5, which occurs in regions with large amounts of anthropogenic pollution. Aerosol and rain are expected to be slightly acidic in the presence of natural sources of SO_2 and NO_2, such as in tropical rain forests. However, atmospheric pollution enhances this phenomenon. Many soils and water surfaces can neutralise it, but their acid-neutralising capacity is often linked to the presence of ammonia NH_3, another nitrogen containing pollutant.

Acid deposition can be derived from the oxidation of sulfur material leading to the formation of sulfur dioxide SO_2 which then reacts with OH and hydrogen peroxide H_2O_2 to produce sulfuric acid H_2SO_4 (Wayne, 2000). Additionally, nitric acid HNO_3 can be produced from NO_2, either directly by reaction with OH, or indirectly by first the reaction with the nitrate radical NO_3 leading to the acid anhydride, dinitrogen pentoxide N_2O_5. HNO_3 has a high solubility property in water releasing H^+

(low pH), and therefore accumulates in aerosols and cloud droplets. It is also rained out.

Eutrophication phenomenon is driven by nitrogen fertilization of soils and wastewater from rivers. Overgrowth of surface water algae and reduction of oxygen for deeper-lying biosystems due to excess deposition of NO_x related species is stimulated by nutrient pollution.

Acid deposition and eutrophication have become an important environmental issue and research area for the last decade (Heij and Erisman, 1995). European public recognition emerged in the middle of 1980s after discovering the acidification of Scandinavian lakes. Consequently, measures have been implemented to reduce the NO_x and SO_2 emissions such as catalytic converters in cars, improved design of furnaces in power plants (in addition to sulfur scrubbers), chemical removal of acid gases. While these measures were originally driven by the acid rain problem, they also resulted in a reduction of aerosol concentrations.

1.4.2 Warming climate or climate change?

EarthâĂŹs climate has been relatively stable over the last 12-13 kyr, allowing humanity to thrive, marking the beginning of the modern climate era and of human civilization. Most of these climate variabilities have been attributed to variations in Earth’s orbit modifying the amount of solar energy received at the Earth’s surface.

Currently, anthropogenic climate change is receiving much attention, but is leading to controversial political and societal debates, as it would require major changes in our societies (which are also more vulnerable than in the past) in a short time period to mitigate it. The main evidence of the anthropogenic global change is the warming temperature: the planet’s average surface temperature has risen about 1.1 °C since the late 19th century (IPCC, 2014). Most of the warming occurred in the past 35 years. The main cause is well acknowledged by all the scientific expert of the climate system: the increase of green-house gas concentrations in the troposphere, in particular CO_2 and CH_4 that are responsible for approximately 80 % of the observed radiative forcing (IPCC, 2014) (cf. Fig. 1.4).

In addition to CO_2 and CH_4, also other greenhouse gases contribute to the warming climate such as H_2O (as a feedback of temperature rise), N_2O, CFCs and O_3. The basic processes governing the greenhouse effect, i.e. transmission of solar radiation but absorption of thermal infrared, was already identified during the 18th century. It is characterized by warming the lower atmosphere and cooling the upper atmospheric layers. Initially, the greenhouse effect is not only natural but also necessary for developing life on Earth. Indeed, without it, the average temperature would be too cold. Later, the scenario of an excessive global warming caused by the injection of gases such as CO_2 was discussed by Arrhenius (1896).

Other compelling evidences suggest additional rapid changes in many more variables that characterize the state of the whole climate system. These changes are then beyond air temperature rise:

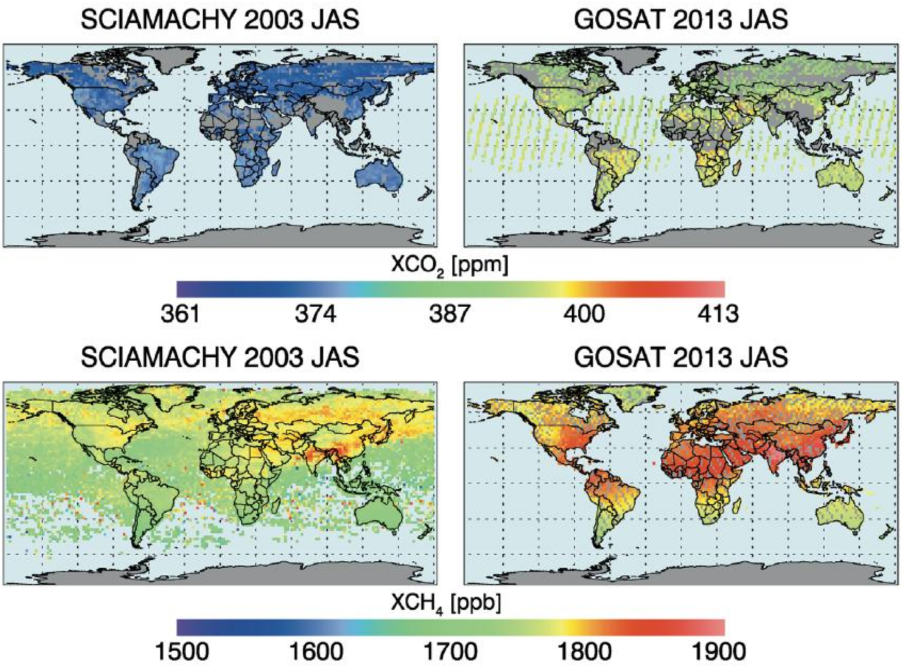


Figure 1.4: Dry column averages XCO_2 and XCH_4 from SCIAMACHY (2003) and GOSAT (2013) satellite missions (Bovensmann et al., 2015). A continued increase over the last decade is evident.

- **Warming oceans –** Oceans also absorb part of the enhanced heat balance in the atmosphere. As a consequence, increased heat has also been absorbed within the top 700 m of the ocean surface (Levitus et al., 2009).
- **Ocean acidification –** Oceans have absorbed about 25 % of emitted anthropogenic CO_2. Due to this increase, acidity of surface ocean waters has increased by about 30 % (Levitus et al., 2009).
- **Shrinking ice sheet and declining sea ice –** Greenland lost 150 to 250 km^3 of ice per year between 2002 and 2006, while Antarctica lost about 150 km^3 of ice between 2002 and 2005. Both the extent and thickness of Arctic sea ice has rapidly declined over the last decades.
- **Glacial retreat –** Everywhere in the world, including in the Alps, Himalayas, Andes, Rockies, Alaska and Africa, glaciers are slowly disappearing.
- **Decreased snow cover –** Satellite data have revealed that snow is melting earlier and the amount of spring snow cover has decreased over the past five decades in the North Hemisphere.
- **Sea level rise –** The pace of sea level rise has nearly tripled since the last two decades. Current estimations are about 1.5 mm per year globally between 1958

and 2014 (Frederikse et al., 2017a). However, such an increase is highly spatially heterogenous: e.g. about 1.6 mm per year in the North-Sea while it rises faster on the US Atlantic Coast about 3.4 mm per year since 1965 (Frederikse et al., 2016, 2017b).

- **Extreme events –** In some parts of the world, such as in the USA, there is an increase of intense rainfall events and storms due to an enhanced record high temperature episodes and a decreased number of record low temperature events (Kunkel et al., 2013). Furthermore, the year of 2017 was victime of a series of violent wildfires throughout the world, particularly in Canada British Colombia, east Russia, Europe and lately in December in California region. One of the reasons of these fires may be attributed to dry forests and warmer temperatures.

In addition to the greenhouse gases, aerosols are also an important component of climate. As they scatter and absorb the sunlight, they redistribute shortwave radiation in the atmosphere. Aerosols are an important player in the climate system by leading, on average, to surface cooling and additional atmospheric dynamical responses (Boucher et al., IPCC report, Chapter 5: Clouds and aerosols, 2015). By acting as the condensation nuclei on which clouds form, they also modify cloud formation, lifetime and precipitation: e.g. Figueras i Ventura and Russchenberg (2009); Sarna and Russchenberg (2017).

1.4.3 Ozone depletion and UV radiation

The ozone hole, discovered over Antarctica (Farman et al., 1985), is a phenomenon of severe ozone depletion at the altitude between 15-20 km (see Fig. 1.5). It occurs in late winter / early spring in presence of polar stratospheric clouds (PSCs), sunlight, strong westerly winds and when the air is isolated from other stratospheric regions.

Although located in the stratosphere, it is strongly connected to the troposphere:

- The tropospheric emissions of chlorine and bromine containing gases, most notably chlorofluorocarbons (CFCs), led to the formation of the ozone hole. They were intensively used, during the previous century, as refrigerants, propellants (aerosol applications) and solvents. Inert and volatile, the long lifetime of these molecules allow them to be transported to the stratosphere. Then, the UV sunlight initiates a photochemically conversion in halogen gases.

- Ecosystems and life in the troposphere are threatened as a lower ozone amount in the upper layers leads to a higher UV radiation at the surface. This also initiates tropospheric photo-oxidation and generation of tropospheric O_3

In September 1987, the Montreal protocol was signed by 24 countries and the European Economic Community. Its implementation has allowed to regulate and limit the production and consumption of the CFCs and related substances. In recent years, there is evidence that the ozone layer is recovering: i.e. ozone hole has started closing. However, it will take another 20–30 years to fully recover.

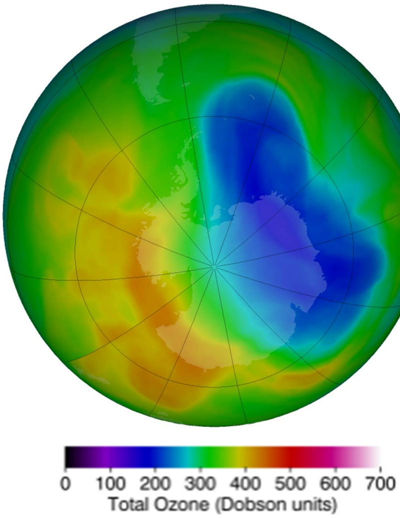


Figure 1.5: Total ozone abundance in Dobson Units (DU) over Antarctica for November 6, 2017. The blue color indicates the ozone hole (`https://ozonewatch.gsfc.nasa.gov/monthly/SH.html`).

1.4.4 Air pollution and health risk

According to the World Health Organization (WHO), air pollution has become the world's single biggest environmental health risk, with around 7 million people, or nearly one in eight, deaths in 2012 (cf. `http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf?ua=1`). 3.7 million persons were affected by outdoor pollution from traffic fumes and coal-burning. The results include stroke, heart and lung diseases, and cancers. The main risks are located within the growth of cities and industrial areas, notably in south-east Asia with 2.6 million deaths with continuing deep poverty in rural areas. Rich countries are not spared by this phenomenon: about 200,000 early deaths in the United States of America (USA) and 100,000 premature deaths in Europe.

As depicted by Fig. 1.6, there is notably a strong link between the abundance of the Particulate Matter (PM) 10 and PM2.5 and the reduction of life expectancy.

Scientists of the Energy Policy Institute (EPIC) at the University of Chicago concluded that the lifetime will be reduced by 7 months for every additional 10 $\mu g/m^3$ of PM10, 1-year reduction for every additional 10 $\mu g/m^3$ of PM2.5 in the air. This is of course a detrimental conclusion for the Chinese cities covered by the Huai River Policy (i.e. free coal for indoor heating provided by the Chinese government since the 1950s) with reduction of life expectancy of about 5.5 years (Chen et al., 2013). Developing countries are also impacted where PM2.5 concentration levels are shortening lifespan of 1 month in new York and 8 months in Los Angeles (cf. Fig. 1.1) and EPIC (`https://aqli.epic.uchicago.edu/wp-content/uploads/2017/09/AQLI_1Pager_US.pdf`).

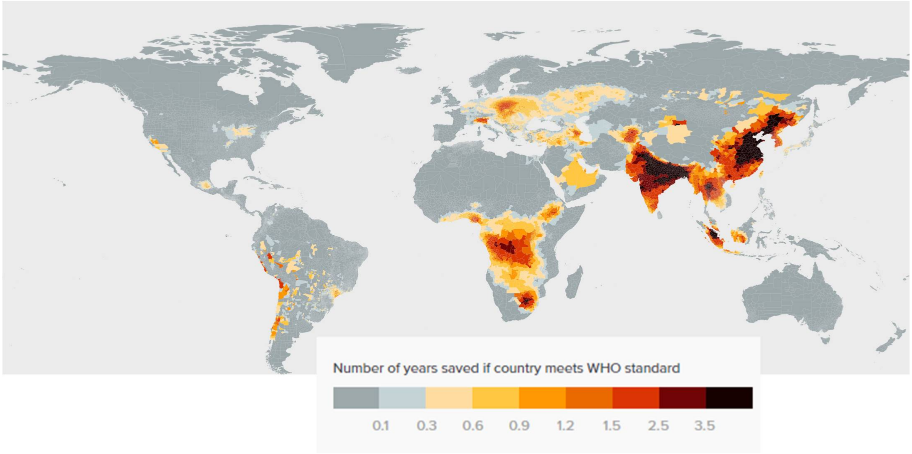


Figure 1.6: Number of years saved if country meets WHO standards: Energy Policy Institute (EPIC) (`https://aqli.epic.uchicago.edu/`) and World Economic Forum, 2017 (`https://www.weforum.org`).

The United Nations Children Fund (UNICEF) estimated that 2 billion children in 2012 lived in areas where outdoor air exceed international limits, with about 20-120 million in Europe (UNICEF, 2016). Beginning 2017, the European Environment Agency released the air and water quality review country report focused on the Netherlands. It estimated that about 11,530 premature deaths were attributable to fine particulate matter concentrations, with an additional 270 to O_3 and over 1,820 to NO_2 (`http://ec.europa.eu/environment/eir/pdf/report_nl_en.pdf`, page 16).

1.5 Needs of global tropospheric composition observations

Following the wealth and environmental issues previously listed (see Sec. 1.4), we cannot remain indifferent. To adapt to the necessary changes, the benefits of monitoring our whole global atmospheric composition, especially in the troposphere, are therefore multiple:

- Since gases and particles can be transported over long distance, air quality "nowcasting" and forecasting must be supported and their capabilities enhanced. This can be done by developing a comprehensive system describing the dynamics and chemistry evolution of air mass, such as the Copernicus Atmospheric Monitoring System (CAMS) (Eskes et al., 2015; Huijnen et al., 2016). Tropospheric composition observations can help by evaluating Earth-System model outputs or by providing them with inputs (the so-called boundary conditions) (Ciais et al., 2010; Burrows et al., 2011; IPCC, 2014). This contributes to monitor effects of health regulations and anticipate protection measures of highly populated areas.

- Exceptional events and their pollutant releases must be monitored, and their subsequent plumes should be tracked from day to day. The American Clean Air Act defines them as air quality episodes that are not reasonably controllable or preventable, caused by human activity that is unlikely to recur at a particular location or a natural event (Duncan et al., 2014). Major forest fires such as in South-America or Russia (see Sect. 4) or volcanic eruptions are considered as exceptional events.
- Long-time series (i.e. concentration trends) and knowledge of the sources and sinks of gases and particles need to be updated. This is important not only to verify the effectiveness and compliance of implemented policies and new technologies in cities (e.g. new generation of transport vehicles, electricity sources, de-nitrification systems on power plants etc.), monitor the impact of economic growth such as in South-Asian in densely urban and large industrial areas but also to anticipate the non-linear response of natural fluxes (ocean and forest) in a warming climate together with interannual sensitivity (Lu et al., 2011; Lu and Streets, 2012; Worldbank, 2015; Duncan et al., 2016). Anthropogenic emissions calculated by the bottom-up approach introduces large uncertainties in emission inventories as it uses statistics on land-use and sector specific emissions factors that can be quite outdated (Mijling and van der A, 2012; Ding et al., 2017b,a; van der A et al., 2017). Moreover, they are not independent of national declarations and their potential inaccuracies. Concentration observations, sensitive to the surface, can be used in a top-down approach through inverse modeling or data assimilation techniques to adjust emission estimates in the model, and even detect unknown sources (Martin et al., 2003; Streets et al., 2013). Trend monitoring and surface flux estimations need homogeneous long-time series of observation data.

In addition, because of the multiple interactions occurring in the troposphere, air quality and climate change characterization requires the monitoring of the ensemble components (i.e. not only one) that form the integrated part of our tropospheric composition to elude some open research questions. For example:

- Given the NO_x capability to alter equilibria of the chemically and radiatively important O_3 and OH species (cf. 1.2.2), it is crucial to estimate their global concentrations and monitor their international trends. Tropospheric O_3 observation is challenging and our knowledge is strongly limited by lack of NO and NO_2 observations in the troposphere.
- The OH abundance, the main sink of CH_4, is also driven by tropospheric NO_x and O_3 (see Sect. 1.2.2). After a period of relative stagnation in the early 2000s, atmospheric CH_4 concentration has again rapidly increased since 2007 (Saunois et al., 2016). Investigating the cause(s) of this increase is a very important question currently addressed by several scientists. The exact causes are still unclear, primarily because of uncertainties in the global CH_4 and OH

budgets, and any other processes affecting these compounds. Since 2014, CH_4 increase is now approaching the most greenhouse-gas-intensive scenarios as imagined by IPCC (2014).

- The scientific understanding of aerosol effects on the climate remains limited (IPCC, 2014). The magnitude of the aerosol radiative forcing depends on the environmental conditions, aerosol properties and horizontal and vertical distribution (Kipling et al., 2016; IPCC, 2014). While, overall, the horizontal distributions of aerosol optical thickness and size are relatively well constrained (see Sect. 1.6.5), uncertainties in vertical profile significantly contribute to the overall uncertainty of radiative effects, e.g. 25 % of the uncertainty of black carbon radiative estimations from the models is due to the uncertainty on the vertical distribution (McComiskey et al., 2008; Loeb and Su, 2010; Zarzycki and Bond, 2010; IPCC, 2014).

1.6 Tropospheric composition remote sensing from space

1.6.1 Context: past, present and future

Why satellites?

Trace gas and aerosol can be observed from a series of ground-based (in situ or remote) sensors. For instance, the global Aerosol Robotic network (AERONET) consists of several hundreds of ground-based carefully calibrated Sun photometers that measure extinction aerosol properties at seven wavelengths at several locations in the world (Holben et al., 1998). Their high accuracy is a great asset for satellite aerosol product calibrations and validations (Lee et al., 2012; Levy et al., 2013; Sayer et al., 2013; Jethva et al., 2014). MAX-DOAS technique allows to provide between 2 and 4 pieces of information on tropospheric NO_2, SO_2 and HCHO profiles, such as in China (Vlemmix et al., 2015; Wang et al., 2017). Other networks like PANDORA or ACTRIS are also of interest for trace gas observations. Ground-based instruments have several advantages:

- Related observations are usually very accurate.
- A lot of measurements contain information about concentrations at the the altitude where we breathe.
- Usually, time series are acquired with a very high frequency, thus providing insight in the temporal evolution.

However, exploitation of such instruments is hampered by a series of drawbacks:

- Measurements are not only sensitive to the local emissions but also weather conditions (i.e. the spatial extension of the gas and particle plume is driven by the wind direction). Therefore, they may not always be representative of the local implemented policies or activities. In the context of flux estimates, representativity errors are highly caused by the variability of surface concentrations (mostly governed by local conditions) and vary with respect to the station type (oceanic, coastal or continental). Furthermore, over a complex source region such as the Seoul metropolitan area (Republic of Korea), it is essential to have a high-resolution observing system with a large enough coverage to be able to differentiate changes among individual sources (such as NO_x) (Duncan et al., 2016). Ground-based sensors are usually more sensitive to their direct environment but not, or little, sensitive for a large area.

- Ground-based networks are usually limited by the spatial coverage: i.e. within a limited area close to their location. Therefore, they do not offer a full global representativity. For example, AERONET coverage remains somehow irregular: very dense over most of lands, less in Russian forests and Central Africa, and very little over ocean and narrow seas. Deployment of MAX-DOAS in China is very limited, usually restricted to specific research studies. Not all of the surface stations are properly located to determine concentration changes (such as NO_2 or CO_2) at all times (Ciais et al., 2010). Other sources, such as ship emissions, are difficult to monitor from ground. Finally, emission and sink information from politically closed countries (e.g. North-Korea) are not accessible.

- Each instrument may include a different calibration protocol leading to inhomogeneities in the time series, or between measurement sites.

- Observations require a measurement system that can sample (in daytime) up to the free troposphere, not just the affected surface layer, to avoid undue influence of local natural signals and to obtain regionally representative measurements. For example, NO_2 ground-based lidars (e.g. RIVM) are mostly well validated in the boundary layer and highly sensitive to its formation and variation, but much less to higher altitude (Volten et al., 2009; Piters et al., 2012)

The need of satellite data, in combination with ground-based networks, is crucial for global mapping of aerosol particles and trace gases in the troposphere (Burrows et al., 2011). With their global coverage and high spatial and temporal resolutions, satellite technology provides the unique potential to move from data-poor conditions toward a data-rich system. Several remote sensors have demonstrated their ability to provide relevant information on changing air quality (Richter et al., 2005; Martin, 2008; Boersma et al., 2011; Levelt et al., 2017), long-range transport of pollutants (Edwards et al., 2006), infer emission locations and strengths (Clerbaux et al., 2009; Frankenberg et al., 2008; Streets et al., 2013; Ding et al., 2015; Levelt et al., 2017; Fioletov et al., 2017), ozone hole (Levelt et al., 2017), and climate change issues (Worden et al., 2008). Furthermore, the relative long life-time of satellite instruments (typically more than 5 years) fulfill the need of trend analyses which require continuous and long-term observations to understand societal changes (e.g. economic crisis, technology implementations) (Boersma et al., 2007; Castellanos and Boersma, 2012; Duncan et al., 2014, 2016).

Figure 1.7 is an illustration of the possibility of satellite sensors (e.g. GOME-2) observing long-range transport of particles (Saharan dust and absorbing black carbon released by forest fires in Portugal in October 2017).

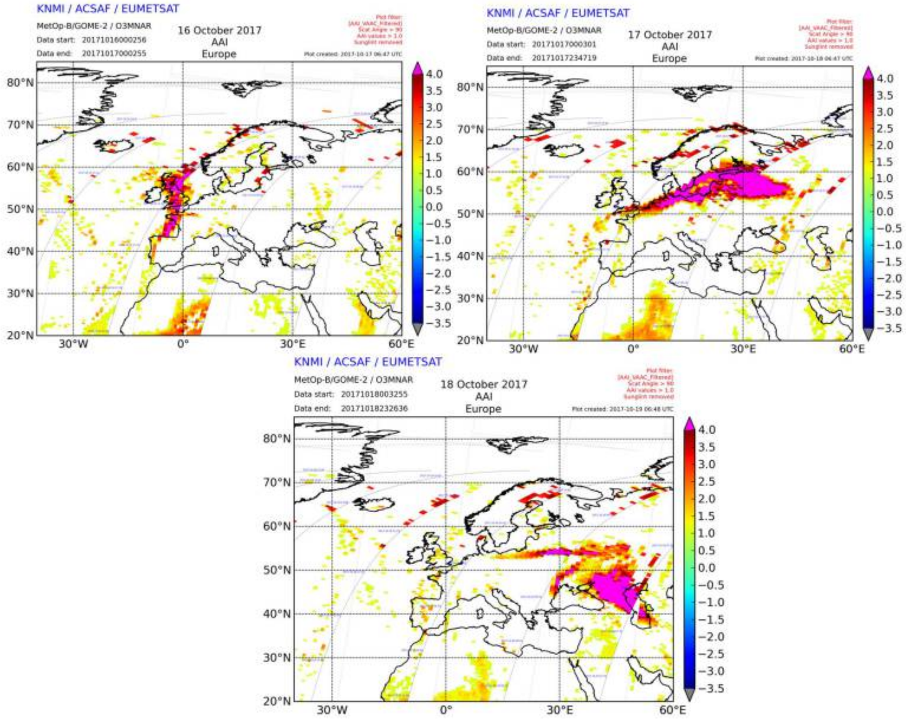


Figure 1.7: UV Aerosol Absorbing Index (UVAI) from GOME-2 sensor on-board MetOp-B platform on 16-17-18.10.2017 (see next subsection for more details). Copyright KNMI, EUMETSAT and the Atmospheric Composition (AC) Satellite Application facilities (SAF) (`http://www.temis.nl/o3msaf/vaac/vaac_gome2.php?sat=gome2b&year=2017&datatype=pics&freq=daily&filter=Filtered&region=Europe`).

From SBUV/TOMS to Sentinels

Initially, remote sensing of the chemical composition of the Earth atmosphere was focused on the stratosphere: first with the Solar Backscatter Ultraviolet Radiometer-1 and -2 (SBUV-1 and -2) in the 1970s on-board the American Nimbus 4 and then the series or Total Ozone Mapping Spectrometer (TOMS) instruments to monitor ozone. Stratospheric NO_2 was measured by a number of NASA satellite instruments since the 1980s, such as the spectrometer on-board Solar Mesosphere Explorer (SME - 1981-1989) in limb viewing (Mount et al., 1984), the series of Stratospheric Aerosol and Gas Experiment (SAGE-II/III - 1984-2005) using the Solar occultation technique (Chu and McCormick, 1979, 1986), the Halogen Occultation Experiment (HALOE - 1991-2005) on Upper Atmosphere Research Satellite (UARS) (Gordley et al., 1996), and the Polar Ozone and Aerosol Measurement (POAM - 1993-1995) on-board SPOT-

2 (Randall et al., 1998).

Over the last 22 years, Europe has led efforts to study distributions and amounts of trace constituents (especially NO_2) in the troposphere from space. This has started with the launch of the Global Ozone Monitoring Experiment (GOME) on-board ERS-2 in 1995 (Burrows et al., 1999), and then SCIAMACHY on-board ENVISAT in 2002 (Bovensmann et al., 1999). ERS-2 and ENVISAT were lead by the European Space Agency (ESA). Note that SCIAMACHY could also observe many constituents in the stratosphere thanks to its limb and occultation capabilities. GOME and SCIAMACHY relied on spectrometers with a wide spectral coverage to observe multiple atmospheric constituents simultaneously. Both instruments used the nadir technique by measuring backscattered Sun light (see Sect. 1.6.2) to determine the abundances of, notably, O_3, NO_2, and other species. GOME was fully functional until 2003 after which only limited processing was possible until 2011. SCIAMACHY stopped with the end of the ENVISAT mission in April 2012. The follow-up of GOME, GOME-2 selected by ESA and the European Organisation for Exploitation of Meteorological Satellites (EUMETSAT), was launched on-board the operational MetOp platforms (MetOp-A in 2006, MetOp-B in 2012). These platforms also include a thermal infrared emission sensor, the InfraRed Atmospheric Sounding Interferometer (IASI), which has shown high potential for tropospheric chemistry (Clerbaux et al., 2009; Clarisse et al., 2011; Bauduin et al., 2016).

Proposed by the Netherlands and lead by the Royal Netherlands Meteorological Institute (KNMI), the Dutch-Finnish Ozone Monitoring (OMI) mission was launched on the NASA Aura payload in 2004 (Levelt et al., 2006). OMI also measures the UV-visible backscattered sunlight in nadir but with a much higher spatial resolution than the previous missions: 13 x 24 km^2 (instead of 320 x 40 km^2 for GOME and 30 x 60 km^2 for SCIAMACHY). OMI has delivered a large amount of data on our tropospheric composition during the last 13 years at a global and regional scale with an almost daily-global coverage (Levelt et al., 2017). A continuous NO_2 data processing system, named DOMINO for OMI, has been operated by KNMI (Boersma et al., 2007, 2011). These data are freely distributed via the Tropospheric Emission Monitoring Internet Services (TEMIS) website (`http://www.temis.nl`). DOMINO has recently evolved through the Quality Assurance for Essential Climate Variables (QA4ECV) project (`www.qa4ecv.eu` (Lorente et al., 2017). The retrieval approach is further discussed on Sect. 1.6.4.

Overall, these satellite missions are mostly designed for trace gas observations. They are not optimized for aerosol particles, although limited information is available (see Sect. 1.6.5). However, the Aura satellite flies in the coordinated NASA A-Train constellation, where leading aerosol space-borne sensors are also present (e.g. MODIS and CALIOP - see Sect. 1.6.5). The advantage of such a constellation is the synergy between spatially and temporally co-registered measurements from different instruments for aerosol and trace gas retrievals compared to one single satellite measurement (see discussion on retrievals in Sect. 1.6.2)

Most of the dedicated satellite missions have a lifetime of at least 5 years and can provide global coverage with a high frequency and spatial resolution (up to daily-global) (Bovensmann et al., 1999; Levelt et al., 2006; Veefkind et al., 2012). In particular, operational daily maps of NO_2 columns by OMI (see `http://www.temis.nl` and Fig. 1.12) show extensive transport features that are changing from day to day within countries, but also air pollution being transported across national borders (Boersma et al., 2007, 2011).

Tropospheric composition monitoring from space will continue beyond SCIAMACHY, OMI and GOME-2. Under the leadership of the European Commission, the Copernicus programme supports the development of Earth Observation satellite and in situ data (see `http://www.copernicus.eu`). As part of this programme, the Sentinel satellites are dedicated to provide operational observations of the Earth. Launched on Friday 13th October 2017 (cf. Fig. 1.8), the Sentinel-5 Precursor (S5P) mission is the first sentinel focused on the atmospheric composition. It is a single payload including the TROPospheric Ozone Monitoring Instrument (TROPOMI), successor of OMI, and led by KNMI, as Principal Investigator (PI) and the SRON Netherlands Institute for Space Research as co-PI (Veefkind et al., 2012). TROPOMI differs from OMI in a number of important ways: 1) a higher spatial resolution (7 x 3.5 km^2 in the UV-vis-NIR, and 7 x 7 km^2 in the SWIR) which makes it possible to identify different sources of air pollution such as in a mega-cities , and 2) the observation of carbon monoxide CO and the green-house gas CH_4. After 2020, S5P will be followed by Sentinel-4 (S4) (geostationary) and Sentinel-5 (S5) (polar orbiting) (Ingmann et al., 2012).

Figure 1.8: Sentinel-5 Precursor launch on Friday 13th October 2017 in Plesetsk, Russia, with TROPOMI sensor on-board (event attended at Space Expo, Noordwijk, the Netherlands).

Meanwhile, international efforts have continued. The NASA/NOAA Suomi National Polar-orbiting Partnership (SNPP) was launched on the 28th October, 2011, with the Ozone Mapping and Profiler Suite (OMPS) on-board, which includes both a nadir and a limb sensor. It gives then the total atmospheric ozone column as well as its vertical distribution from 15 km to 60 km on a daily basis (`https://jointmission.gsfc.nasa.gov/omps.html`). In the next years, NASA will launch the Tropospheric Emissions: Monitoring of Pollution (TEMPO) instrument to monitor air quality between North America, Mexico city, Canadian soils and the Atlantic and Pacific ocean from a geostationary orbit (`https://fpd.larc.nasa.gov/tempo.html`). Similarly, the Geostationary Environment Monitoring Spectrometer (GEMS) will observe tropospheric composition over South-Korea and the Asia-Pacific region thanks to a UV-visible spectrometer (`http://www.ball.com/aerospace/Aerospace/media/Aerospace/Downloads/GEMS_0916.pdf?ext=.pdf`).

In the future, we will have an international atmospheric composition observation system from space that will, for the first time combine polar (S5P, S5) and geostationary (S4, TEMPO, GEMS) orbits.

1.6.2 Satellite measurement: breaking the Sun light

The light – Energy source – in remote sensing

Maps of trace gases or aerosols are derived from measurements of electromagnetic radiation acquired by satellite instruments at the top of the atmosphere. This radiation travelled through the atmosphere and interacted with atmospheric constituents (cf. Fig. 1.9).

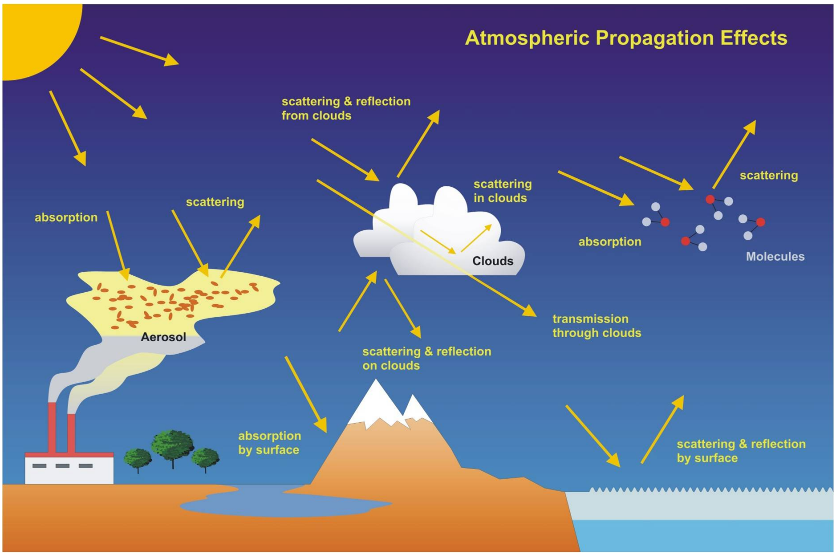


Figure 1.9: Light path issues (Bovensmann et al., 2011).

It contains all the information about their abundance and spatial distribution. In active remote sensing, the initial radiation is emitted by the satellite instrument itself (e.g. radar or lidar). In passive remote sensing, the Sun is the radiation source in the UV-visible, near and shortwave infrared spectrum, while the Earth system emits the thermal infrared radiation. Passive satellite instruments have shown to be the most successful for air quality and climate gases monitoring (Duncan et al., 2014).

Solar light interaction with the atmosphere

The sunlight can be modeled as an ensemble of electromagnetic radiation (i.e. group of photons) covering a specific spectral wavelength λ / frequency range (see Fig. 1.10). The Sun approximately radiates as a black body at a temperature of 5700 K. According to the Plank's law, solar radiation peaks in the visible spectral range from $\lambda =$ 400–700 nm, with a maximum in the green spectrum domain at $\lambda = 500$ nm. Earth itself radiates with a lower average temperature (about 255 K) having a radiation peak in the thermal infrared ($\lambda = 10\ \mu$m).

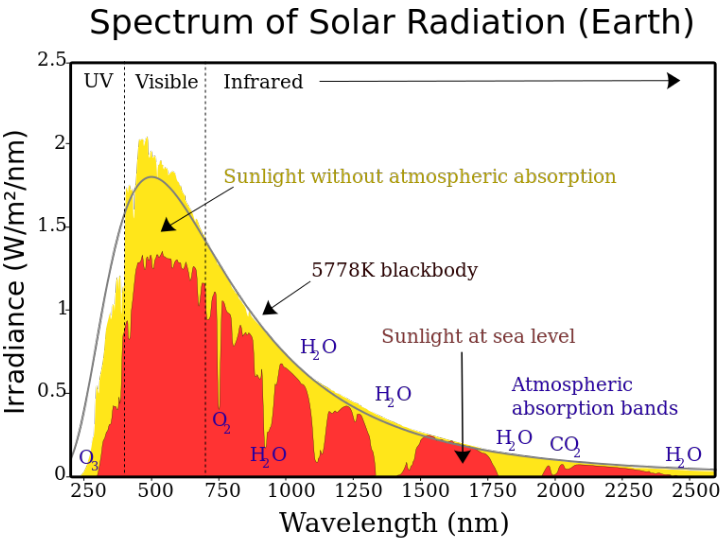


Figure 1.10: The spectrum of sunlight (light yellow) compared with a theoretical black body curve at 5778 K, and the solar radiation at sea level (Source: wikimedia).

The Earth atmosphere is not completely transparent: depending on the wavelengths, the Sun light is absorbed, remitted, reflected and/or scattered. An example of such interaction is the blue sky "visible" during daylight, which is the result of the molecular light scattering by N_2 and oxygen in all the directions (i.e. Rayleigh scattering), especially very strong for shorter wavelengths. Another example is the red sunset that can be observed sometimes. This remaining red light in the sunlight is due to the filtering of the blue part by Rayleigh scattering. In absence of atmosphere, there would be no Rayleigh scattering and our sky would be like that of the moon:

dark with only some bright stars. Aerosol and cloud particles are responsible for scattering, which is wavelength and particle type dependent. Its intensity strongly varies with the scattering angle. The specific Mie scattering (i.e. valid for spherical particles) is said to be elastic: i.e. molecules remain unchanged and the scattered photon keeps the same energy and wavelength. On the contrary, Raman scattering (weak interaction with matter) is inelastic. The outcome photon has then a different energy and wavelength.

Trace gas molecules interact with light via absorption (Einstein, 1917). Since they consist of chemical bound groups of atoms, they are driven by quantum mechanical principles. This means one can quantify their internal energy and their energy states are discrete (Burrows et al., 2011). In presence of incident solar radiation, the molecule in a lower energy state becomes excited. Incoming photon with the right energy, i.e. wavelength, is then absorbed, and the molecule undergoes to an higher energy state. Depending on the energy / frequency, the molecule stimulation is characterized by rotation of the entire molecule, vibration of the atoms or electronic transition of the energy states. The resulting spectrum contains discrete absorption lines corresponding to the discrete molecule energy states.

To be able to observe trace gas absorption, it is required to measure a light spectrum over a large enough spectral range and with a sufficient spectral resolution to identify the so-called absorption spectral lines. For that purpose, the designed UV-visible-shortwave infrared passive satellite instrument looking downward (nadir point of view) are usually spectrometers: a device that measures the upwelling radiance corresponding to the backscattered sunlight at the top of the atmosphere (TOA). The individual photons striking the instrument's detector is converted into electrons. The detected signal is then decomposed into a series of energy intensity as a function of wavelength: the radiance spectrum (Duncan et al., 2014).

Overall, a critical aspect in satellite observations of atmospheric constituents is the light path, or more precisely the average length of the ensemble of all possible light paths due to all possible scattering and absorption processes occurring in the atmosphere and at the surface, as summarized in Fig. 1.9. This is likely the key uncertainty in the quantification of atmospheric trace gases such as NO_2 (see Sect. 1.6.3 and 1.6.4).

Radiative transfer models (RTMs) simulate the propagation of the light through the atmosphere and at the surface. They can then help to investigate how the satellite spectrum measurements are affected by the state of the atmosphere and identify the most important parameters. These models are central to quantify the greenhouse magnitude, climate change effect, efficiency of photochemical reactions in the Earth atmosphere and trace gas retrievals (see Sec. 1.6.3). Several types of RTMs exist: e.g. discrete ordinate which numerically solve the radiative transfer equation (Spurr, 2008), backward Monte-Carlo based on statistical techniques applied to a large ensemble of paths for each photon (Deutschmann and Wagner, 2008), and the doubbling adding principles (de Haan et al., 1987; Stammes et al., 1989). As further described in the next chapters, this thesis research employed the doubling adding method using the DISAMAR KNMI software (de Haan, 2011). When applied to UV and visible wavelengths (in the absence of thermal emission), the central equation in the RTMs is written as:

$$\frac{\mathrm{d}I(\lambda,\theta,\phi)}{\mathrm{d}s} = -(\epsilon_a(\lambda) + \epsilon_s(\lambda)\cdot I(\lambda,\theta,\phi)) + \epsilon_s(\lambda)\cdot \int_0^{\pi}\int_0^{2\pi} I(\lambda,\theta,\phi)\cdot \frac{S(\theta,\phi)}{4\pi}\mathrm{d}\phi\cdot\sin\theta\mathrm{d}\theta, \tag{1.15}$$

with $\frac{\mathrm{d}I(\lambda,\theta,\phi)}{\mathrm{d}s}$ the spectrum change of incoming radiation moving through an infinitesimal thin layer with a thickness ds, $\epsilon_a(\lambda)$ and $\epsilon_s(\lambda)$ the absorption and scattering coefficients respectively, θ the zenith angle, ϕ the azimuth angle, and $S(\theta,\phi)$ the dimensionless scattering phase function (i.e. the angular dependency of the scattering effect).

For practical reasons, in the UV-vis, we generally focus on the reflectance spectrum R in which the extraterrestrial solar irradiance features are removed (leaving then mostly the atmospheric absorption lines of interest):

$$R(\lambda,\theta,\phi) = \frac{\pi\cdot I(\lambda,\theta,\phi)}{\mu_0\cdot I_0(\lambda,\theta,\phi)}, \tag{1.16}$$

with μ_0 the solar zenith angle.

Trace gas absorption spectroscopy

The interaction of solar radiation with matter provides a powerful way to investigate the chemical composition of the complex atmospheric mixture. In particular, absorption spectroscopy allows to focus on the spectral regions of interest for tropospheric trace gas analysis (Burrows et al., 2011) (see Fig. 1.11).

For instance, since UV radiation is absorbed at different altitudes by ozone (see Sect. 1.1 and 1.4.3), exploitation of the UV spectral band contains information of the ozone height distribution. The red-brownish color of NO_2 illustrates its transmition property in the red spectral domain while it absorbs the Sunlight in the blue band (about 405-470 nm). Green-house gases CO_2 and CH_4 interact further in the shortwave and thermal infrared range.

Trace gas research studies focusing on their absorption properties in the UV-visible spectral range (i.e. no thermal emission) focus on the Beer-Lambert law (or Bouguer-Lambert law) describing the light attenuation as a function of the travelled distance s, gas concentration and its spectral absorption intensity:

$$I(\lambda,\theta,\phi) = I_0(\lambda,\theta,\phi)\cdot \exp(\int_s -\sigma(\lambda)\cdot\rho(s)\cdot\mathrm{d}s), \tag{1.17}$$

with $I_0(\lambda,\theta,\phi)$ the initial intensity, $I(\lambda,\theta,\phi)$ the measured radiation intensity (or radiance), $\rho(s)$ the trace gas density and $\sigma(\lambda)$ the absorption cross-section of a given gas measured in a laboratory. This last parameter can be considered as the

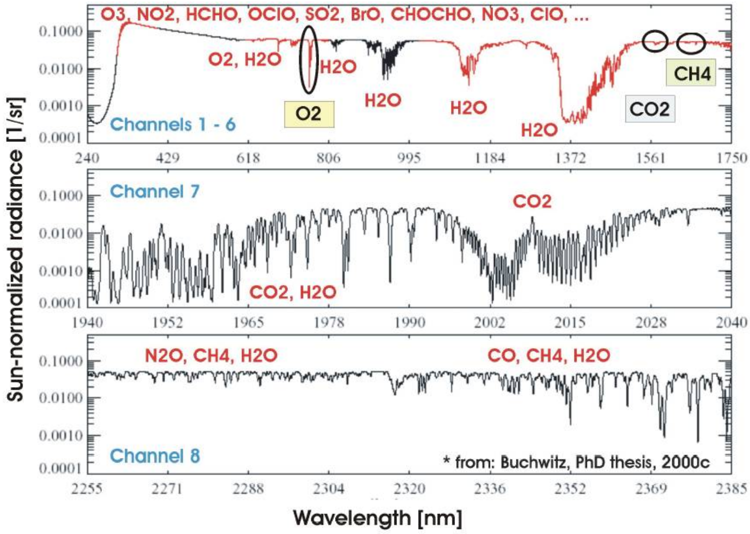


Figure 1.11: SCIAMACHY nadir spectrum expressed as Sun-normalized radiance (Buchwitz et al., 2000; Schneising et al., 2008, 2009; Burrows et al., 2011).

effective area of the molecule to remove photon's energy from incoming radiation. Note that the product $\sigma(\lambda) \cdot \rho(s) \cdot \mathrm{d}s$ gives the optical thickness.

1.6.3 Quantifying trace gases and particles: the retrieval process

Estimation principle and forward model

Since an atmospheric satellite measurement is a light spectrum containing the information on our atmospheric composition, it is necessary to convert it into a geophysical parameter: e.g. estimate the abundance of pollutant gases in the troposphere. Such a process is commonly named retrieval of an atmospheric state x (i.e. state vector) from a measurement y.

The atmospheric retrieval requires a forward model F to describe the dependency between x and y:

$$y = F(x) + \epsilon + \Delta F(x), \tag{1.18}$$

$\Delta F(x)$ is the forward model error, and ϵ is the measurement error (noise). In practice, the retrieval does not represent the true state vector x but rather a summary x' of our best understanding of the physics that explains most of the measurement. x' may include more than one geophysical parameter under consideration (e.g. simultaneous retrieval of NO_2 and aerosols). Furthermore, to close the budget of the radiance spectrum y, other parameters must be considered in the forward model: pressure, temperature, surface properties but also the instrument characteristics such as the geometry angles. They form then the set (or best of estimate) of forward model parameters b. It is important to notice that b is not fitted. x in Eq. **??** becomes then

$x = (x', b)$, such that:

$$y = F(x', b) + \Delta F(b) + \epsilon, \qquad (1.19)$$

The forward model error may combine different sources: the model inaccuracy in itself ΔF, and/or the gap between the assumed b' and the true atmospheric state parameters b that are not fitted but transported through the forward model $\frac{\mathrm{d}F}{\mathrm{d}b}(b - b')$. Equation 1.15 becomes then (Eskes and Boersma, 2003):

$$y = F(x', b') + \epsilon + \Delta F + \frac{\mathrm{d}F}{\mathrm{d}b}(b - b') + \epsilon. \qquad (1.20)$$

Inverse model approach and challenges

The atmospheric retrieval problem from a spectral satellite measurement generally faces, at least, three fundamental challenges: , 1) the choice of the forward model itself leads to systematic errors, 2) the retrieval problem is ill-posed, 3) required analytical equations of the geophysical retrieval do not exist.

A forward model can be chosen based on a full physical RTM (see Sect. 1.6.2) or relatively more simplified such as the Beer-Lambert law (cf. Sect. 1.6.2 and Eq. 1.14) as considered in the Differential Optical Absorption Spectroscopy (DOAS) (Sect. 1.6.4).

As discussed in Sect. 1.6.3, the measured backscattered solar light contains all the information on the atmospheric composition and the surface. However, this information cannot resolve all the fine structures of the atmosphere. The amount of independent (i.e. non correlated) pieces of information, also named degree of freedoms in the frame of Rodgers (2000), is generally low. The atmospheric retrieval problem is then ill-posed, an infinite number of atmospheric states (or state vectors) can lead to a same single measurement. As an example, the atmospheric vertical profile of NO_2 cannot be derived from space-borne sensors such as OMI or TROPOMI, but instead the total (stratosphere+troposphere) or tropospheric column density is estimated. This column represents the integrated number of molecules per cm^2 along the considered atmospheric layers (until the tropopause height for tropospheric NO_2, see Sect. 2.2.1). Another illustration, largely mentioned in this thesis research, is the complexity to distinguish clouds, aerosols and bright surfaces from a single passive satellite measurement (see Sect. 3 and 4). Successful retrievals from passive sensors are obtained over clear pixels, and therefore accurate cloud filtering is a major requirement. However, in practice, small cloud residuals may persist and lead to systematic biases as discussed in the next chapters.

Such ill-posed problems have two direct consequences: 1) the state vector must be carefully designed such that is consistent with the actual number of pieces with independent information, 2) accurate prior knowledge on the geophysical parameters, contributing to the measured radiance (i.e. set of forward model parameters b), is necessary to constrain the retrieval. As discussed in Sect. 1.6.4 and 2.2, retrieval of tropospheric NO_2 vertical column density from UV-visible satellite spectra must take into account the surface albedo, cloud and aerosol interferences, and the shape of NO_2 vertical profile in the vector b. We will see in the next chapters that their combined uncertainty is one of the most crucial factor affecting the tropospheric NO_2 retrieval from OMI-like sensors.

While a radiative transfer model describes the dependency of the measurement y on the state vector x, a retrieval is an inverse model where x depends on y. Such an inverse model typically does not exist by nature. To address this, there are generally speaking two approaches: variational and statistical (Blackwell, 2005). The variational approach employs a full physical forward model on-line, i.e. for each single measurement y. An $a\ priori$ state vector x_a is explicitly used, with its associated prior error, in addition to the set b. It represents our statistically best prior knowledge of the geophysical parameter to be retrieved to be propagated through the forward model, thereby producing a simulation of the at-sensor radiance (Blackwell, 2005). Both simulation and measurement are compared and the state vector is iteratively adjusted until the modeled radiance matches the observation. The minimization equation becomes (Blackwell, 2005):

$$x' = min_x(||{S_y}^{-1/2}(F(x,b')-y)||^2 + ||{S_a}^{-1/2}(x-x_a)||^2), \tag{1.21}$$

with S_y the measurement error covariance matrix, and S_a the $a\ priori$ state vector covariance matrix (i.e. the best statistical knowledge of the uncertainty related to our prior state vector). The Optimal Estimation Methodology (OEM) is likely the most famous variational approach as defined by Rodgers (2000). It seeks the statistically most likely solution by applying Bayes theorem.

The statistical approach is regression-based and does not explicitly use the forward model (that is off-line). Instead the estimation is based on an ensemble of radiance/state vectors $\mathrm{R}_{ensemble}$ and statistical characterization $[p(x'), p(y), p(y/x')]$: the probabilities p of the event x' $(p(x'))$, of the event y $(p(y))$, and of y under the condition of the event x' $(p(y/x'))$. It can be summarized as (Blackwell, 2012):

$$x' = min_x(||\mathrm{x}_{ensemble} - F(\mathrm{R}_{ensemble})||), \tag{1.22}$$

In practice, such probability density functions are difficult to obtain and alternative methods are employed like linear least-square, linear regression and non-linear least-square (Blackwell, 2005). Look-up Tables (LUT) and machine learning such as the Neural Networks (NN) or Principal Compression Analysis (PCA) are special classes of the statistical approach (Atkinson and Tatnall, 1997; Blackwell, 2005). Contrary to the variational approach, statistical techniques do not fit the state vector iteratively. Combined with an off-line forward model, this generally presents the advantage of a fast processing time.

The pros and cons of variational and statistical approaches are quite open to debate within the scientific community. Methods like the OEM are expected to give a high accuracy for an individual measurement, a retrieval error estimation together with some diagnostics variables (e.g. degree of freedom of signal, or vertical averaging kernel). Furthermore, by using prior and posterior error matrices, it can take into account potential correlations between different error sources. However, because of the need to perform some iterations and the use of a full physical model, computing time can be high, especially for a strongly non-linear problem. Furthermore, there is no guarantee that the system will converge. The so-called "divergence" problem is quite important with this approach (Sanders et al., 2015). Although, there are quite many tricks to tackle this issue (e.g. the Levenberg-Marquardt approach), no simple solution exists. Finally, the result may be quite dependent on the source of the prior knowledge and how well its uncertainty is characterized. The main advantage of a statistical approach is the fast computing time. Compared to the OEM, the dependence on prior information is lower although still significant: e.g. see the sensitivity of neural network aerosol layer height to aerosol model in Chap. 3, or the tropospheric NO_2 air mass factor (AMF) to the geophysical parameters (cf. Chap. 2). Such a system does not encounter by definition divergence problems, which means it always gives a solution for each measurement (which, on the other hand, might be more or less accurate). However, it doesn't give an optimal solution for an individual measurement but more a statistically good one for an ensemble. No error estimate is directly available for each retrieval. The error correlations in the measurement y or the assumed forward model cannot be easily taken into account. In case of a big LUT, the high memory consumption can become an important issue. Finally, there can be quite open questions on the degree of physical realism when using a machine learning approach. However, this last point actually depends on the generation of the training database and the technique used for the training process (e.g. over-training verification, identification and evaluation of the machine learning architecture etc.. see Chap. 3).

Big data challenge and operational processing needs

State-of-the-art retrieval algorithms are facing a continuous increase of the amount of observations: 1 million spectra per day from OMI (Levelt et al., 2006), 20 million spectra per day from the future TROPOMI on-board Sentinel-5 precursor mission (Veefkind et al., 2012). The next operational Sentinel-4, Sentinel-5 and the proposed greenhouse gas Sentinel-7 missions will add even more observations (Ingmann et al., 2012). Increasing amounts of satellite data lead to several challenges regarding adequate analysis and the possibility to deliver near-real time information. Since implementation of space missions are based on big investments, we have to develop novel techniques, beyond the traditional ones, to guarantee relevant air quality and climate analyses from current and future satellite missions.

The next generation of satellite retrieval algorithms needs to address the challenges of increasing data amounts, by taking advantage from developments in other research areas with high impacts on society, such as the artificial intelligence techniques and machine learning algorithms devoted to fast detection, tracking and classification (Teichman and Thrun, 2012; Held et al., 2016). These new techniques have to be applied to improve the computational speed of satellite retrievals, but with careful adaptations to keep ensuring data with high quality and relevant scientific analysis.

1.6.4 Focus on tropospheric NO_2: the DOAS method

The Differential Optical Absorption Spectroscopy (DOAS) method is a specific retrieval approach employed for UV and visible absorbing trace gases. The various DOAS techniques rely on the same key concept: a simultaneous fit of several trace gas slant column densities from the fine spectral features due to their absorption (i.e. the high frequency part) present in passive UV–visible spectral measurements of atmospheric radiation (Platt and Stutz, 2008). The Beer-Lambert law is used as forward model (Eq. 1.14). It is commonly employed for absorption spectroscopy analyses of NO_2, SO_2, HCHO and O_3 from the OMI, TROPOMI, GOME, GOME-2 and SCIAMACHY sensors (Boersma et al., 2011).

To take into account the simultaneous presence of several absorbers together with Rayleigh and Mie scattering, equation 1.17 is commonly re-written as follow:

$$I(\lambda,\theta,\phi)=I_0(\lambda,\theta,\phi)\cdot\exp(-\sum_i\sigma_i(\lambda)\cdot N_i(s)-\sigma_{Ray}(\lambda)\cdot N_{Ray}(s)-\sigma_{Mie}(\lambda)\cdot N_{Mie}(s)), \tag{1.23}$$

The spectral fit is achieved within a predefined spectral window (e.g. 405–465 nm for OMI NO_2) on the spectral reflectance R. Usually, R is divided into two parts: 1) a spectrally smooth part, modelled by a polynomial, accounting for Rayleigh and Mie scattering, surface reflection, and the low frequency of the absorption cross sections, 2) a spectrally differential part from which trace gas information is retrieved. The slant column density N^s gives then the column density along the average light path traveled by the detected photons from the Sun through the atmosphere, surface and back to the satellite sensor.

$$N^s=\int_s\rho(s)\mathrm{d}s, \tag{1.24}$$

In the case of OMI tropospheric NO_2 It is obtained through a least-squares fit approach applied to the logarithm of R over 405–465 nm, rewritten from Eq. 1.23:

$$ln(R(\lambda))=\sum_i\sigma_i(\lambda)\cdot N_i^s+polynomial, \tag{1.25}$$

Special attention is generally paid, during this fit, to the Ring effect: i.e. a so called Ring spectrum is included like a cross section at the right side of equation 1.25 (Chance and Spurr, 1997; Boersma et al., 2007, 2011). Furthermore, the temperature sensitivity of the NO_2 absorption cross section is accounted for by computing an a posteriori temperature correction term between an effective atmospheric temperature along the average photon path and the 220 K cross section spectrum used in the fitting procedure (Vandaele et al., 1998). The stratospheric NO_2 contribution must be separated from the tropospheric one in the slant column. In the case of OMI DOMINO product, this is done by assimilating slant columns into a chemistry transport model (Boersma et al., 2007, 2011)

To remove the effect of the average light path variability between different measurements, an air mass factor (AMF) must be applied to convert the slant column density into a vertical column density. The AMF is a unitless number containing information on the average light path length. It represents the sensitivity of the slant column to the actual vertical column. As further detailed in Chap. 2, for trace gases in PBL and the free troposphere, it is sensitive to the geometry angles as well as to the surface and particles.

The DOAS method is well suitable for optically thin absorbers, such as NO_2 (typical slant optical thickness of 0.005 at 440 nm (Boersma et al., 2004)). Indeed, molecules with high absorption optical thickness (e.g. O_3) highly affect by themselves the average light path, which makes then the AMF computation too much dependent on its vertical column density (initially unknown). Because it is optically thin, prior knowledge on tropospheric NO_2 vertical column density is not required.

1.6.5 Focus on aerosols

Remote sensing of particles is likely more complex than of trace gases as their description requires a multitude of parameters: size, concentration (sometimes as a function of particle size), particle shape, chemical composition, optical properties (scattering vs. absorption) and their horizontal and vertical distribution. All these parameters can generally not be simultaneously retrieved from one single satellite measurement (see Sect. 1.6.3).

Perhaps the most famous satellite sensors dedicated for aerosol observations are the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on-board the Cloud Aerosol lidar and Infrared Pathfinder (CALIPSO) since 2006, and the Moderate Resolution Imaging Spectroradiometer (MODIS) on-board Terra and Aqua platforms. Both the CALIPSO and Aqua platforms fly together with Aura (including OMI) in the A-Train satellite constellation. By measuring the upwelling Solar radiation at the top of the atmosphere in several spectral channels (covering the visible and the thermal infrared), MODIS provides daily global information on the aerosol optical thickness (AOT or τ), defined as the extinction optical thickness integrated from the surface to the top of the atmosphere (Levy et al., 2013). MODIS aerosol retrieval is based on an ensemble of algorithms, each of them specific to diverse cases (bright surfaces such as desert, dark vegetation on continents, ocean) and the detection and filtering of clouds based on the thermal infrared channels (Remer et al., 2005, 2008; Levy et al., 2013). The MODIS τ products are usually recognized as reference by many studies. However, MODIS data do not contain information on aerosol vertical profile.

CALIOP is an active lidar instrument. It provides very accurate vertical atmospheric distribution of aerosols, together with their backscattering and extinction properties. However, it only looks at the nadir point of view and thus cannot provide a global coverage with a high frequency.

Observing the aerosol vertical distribution with a high revisit frequency (ideally daily) is likely one of the main remaining challenges for aerosol satellite remote sensing. To meet this objective, passive hyperspectral sensors are generally employed as they offer adequate spatial coverage with a good temporal resolution. The traditional technique is the slant column measurement of molecular dioxygen O_2 which is well mixed and relatively stable in the atmosphere. This is a suitable proxy to determine the modified scattering height due to aerosol particles. Most of the algorithms focus on O_2-A band around 765 nm, on TROPOMI, OCO-2, GOME-2, and GOSAT spaceborne sensors (Wang et al., 2012; Sanders et al., 2015). However, exploitation of this band remains challenging due to high sensitivity to surface albedo (brightness) and saturation (Nanda et al., 2017).

Although not optimized for aerosol monitoring, OMI allows to distinguish absorbing particles (from scattering) from its UV 330-388 nm band. The so-called UV Absorbing Index (UVAI) is derived by the OMI near-UV aerosol algorithms (OMAERUV and OMAERO) in the 330-388 nm spectral band (Torres et al., 2007). This benefits from a long heritage where similar techniques were developed for past sensors (TOMS, GOME and SCIAMACHY) (Torres et al., 1998). It relies on the measured change of spectral contrast, with respect to a pure Rayleigh atmosphere. Weakly absorbing or large non-absorbing particles are associated with near-zero or negative UVAI values. Furthermore, OMI-like sensors also include the O_2-O_2 spectral features in the UV and visible. The O_2-O_2 477 nm band is the strongest one, presenting a wider (over 10 nm) although weaker spectral absorption than O_2-A. As analysed in this thesis research (see Sect. 4 and 5), this leads to high sensitivities for retrieving aerosol layer height in case of strong aerosol loading and with less challenges due to saturation.

1.7 This thesis

1.7.1 Research goal and objectives

The motivation of this thesis is to improve the quality of present and future satellite data, derived from the backscattered sunlight in the visible spectral domain and dedicated to monitor our tropospheric chemistry composition for air quality and climate research purposes. The research goal is the characterization and minimization of the aerosol scattering and absorption effects on the retrieval of tropospheric trace gas concentrations from satellite spectral measurements in the visible. Aerosol presence redistributes the atmospheric solar radiation, in particular in the short-wavelength range, and leads to complex challenges for properly characterizing the average light path of an individual trace gas satellite measurement. In addition, a secondary and more numerical goal, is the preparation for operational processing of big data amounts from the forthcoming new-generation satellite instruments in air quality and climate research communities.

Our main objective is to design and evaluate an aerosol layer height retrieval approach from a visible satellite spectrum that could contribute to the improvement of operational tropospheric NO_2 vertical column density retrieval over cloud-free scenes dominated by aerosol pollution episodes. This approach should be demonstrated with the Dutch-Finnish OMI sensor over industrial and densely populated areas (e.g.

Asia), and/or over biomass burning activities (e.g. South-America, Central Africa). As shown in Fig. 1.12, these regions display a strong spatial correlation between OMI tropospheric NO_2 and aerosol optical thickness as they share similar anthropogenic sources (see Sect. 1.2.2) (Veefkind et al., 2011). NO_2 is also precursor of secondary aerosol formation.

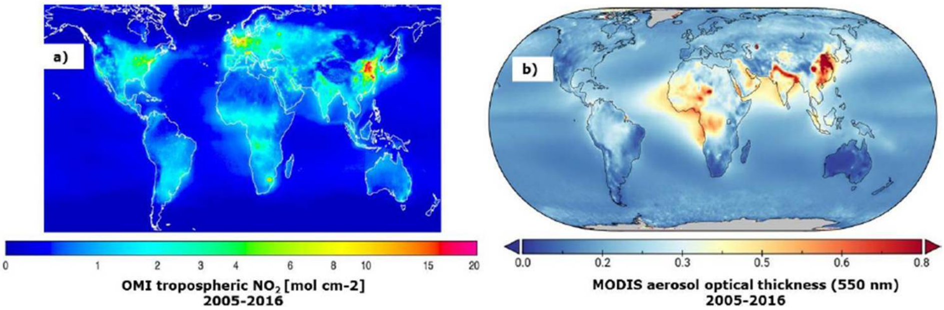


Figure 1.12: Spatial correlation between long-term averages of satellite tropospheric NO_2 and aerosol observations. Panel (a) shows tropospheric NO_2 vertical column density from OMI DOMINO product (Levelt et al., 2006; Boersma et al., 2011). Panel (b) depicts aerosol optical thickness (550 nm) from MODIS (Levy et al., 2013). Note that for several regions, for example south-east Asia, India and southern Africa, there is a strong spatial correlation due to the same underlying sources (Veefkind et al., 2011).

In particular, this thesis focuses on a simultaneous exploitation of a single satellite visible spectrum for: 1) aerosol parameters (ALH primarily, optical properties secondary) retrieval from the 477 nm O_2-O_2 absorption feature, and 2) the correction of aerosol scattering and absorption effects when retrieving tropospheric NO_2 from the 405-465 nm band. Our interest is to apply an aerosol correction based on parameters (effective or explicit) describing the particle effects on the average light path that can be determined from this visible spectrum. This thesis pays attention to the complex and simultaneous interplay between aerosol particles (presence and/or explicit retrieval), clouds (residuals, if any, and effective retrieval) and tropospheric NO_2. It also analyses in depth the accuracy of the tropospheric NO_2 AMF, one of the key parameters of the large retrieval chain for every individual OMI visible measurements, related to the description of the average light path (see Sect. 1.6.4). Overall, we will see how critical, among all the analyzed geophysical variables, the aerosol layer height knowledge is for meeting this objective.

Consequently, the following scientific questions are investigated in the next chapters:

1. **Is there a quantitative need to correct tropospheric NO_2 retrievals for the effects of aerosols?**

2. **How does the current OMI effective cloud algorithm (OMCLDO2) respond to the aerosol properties?**

3. **What is the error budget of the implemented aerosol correction (cf. the so-called implicit aerosol correction) in the OMI tropospheric NO_2, DOMINO-v2, product?**

4. **How to design an innovative concept combining global aerosol layer height retrieval from OMI visible spectra and a machine learning approach?**

5. **What is the uncertainty of the retrieved OMI aerosol layer height?**

6. **How could the developed OMI aerosol layer height algorithm be combined with the tropospheric NO_2 retrieval chain?**

1.7.2 How to read this thesis? – Overview

The following chapters present specific studies investigating the questions enumerated in Sect. 1.7.1, In particular, Chapter 2 quantifies the necessity to address and reduce the aerosol radiation effect when computing tropospheric NO_2 AMF. It shows the expected biases with assumptions of clear sky while aerosol particles are present in the observed scene. Furthermore, it analyses in depth the behaviour of the OMI effective cloud algorithm, as implemented in the OMI DOMINO NO_2 retrieval chain, and evaluates the quality of the resulting implicit error (Chimot et al., 2016).

Chapter 3 describes a novel and innovative algorithm developed for aerosol layer height (and optical thickness) retrieval from the OMI 477 nm O_2-O_2 spectral band by using neural network technique, a class of machine learning approach. It evaluates its accuracy based on a comprehensive synthetic study as well as on 3-year retrieval over East Asia compared to a reference climatology (Chimot et al., 2017).

Chapter 4 complements Chapt. 3 by assessing the spatial variability of the OMI aerosol layer height. Comparisons are made with collocated CALIOP aerosol observations and over diverse case studies: continental polluted particles over densely urban industrialized areas in east China, absorbing particles over wild forest fires in South-America and Russia, and pure desert dust from a Saharan outbreak over sea (Chimot et al., 2018b).

Chapter 5 summarizes a small task illustrating the possibility of a yearly and (almost) global aerosol layer height retrieval with a fast computing time thanks to the employed neural network technique.

Chapter 6 analyses how the upgraded OMI effective cloud algorithm and the developed neural network OMI aerosol layer height can mitigate biases due to aerosol scattering and absorption effects in the next reprocessing of OMI tropospheric O_2-O_2 product (Chimot et al., 2018a).

Finally, the concluding remarks, as well as some suggestions to extend and improve this work are the subject of Chapter 7.

Chapter 2

OMI tropospheric NO_2 retrieval and aerosol correction over industrialized regions: how accurate is a simple cloud model?

> "We need to stop relying on blind faith and start putting the science into data science. We must equip ourselves with sufficient skepticism that we can see, understand, and defend the data that goes into our algorithms and the result they produce."
>
> Cathy O'Neil, Weapons of Math Destruction, 2016

Abstract

The Ozone Monitoring Instrument (OMI) instrument has provided daily global measurements of tropospheric NO_2 for more than a decade. Numerous studies have drawn attention to the complexities related to measurements of tropospheric NO_2 in the presence of aerosols. Fine particles affect the OMI spectral measurements and the length of the average light path followed by the photons. However, they are not explicitly taken into account in the current operational OMI tropospheric NO_2 retrieval chain (DOMINO - Derivation of OMI tropospheric NO_2 - product). Instead, the operational OMI O_2-O_2 cloud retrieval algorithm is applied both to cloudy and to cloud free scenes (i.e. clear sky) dominated by the presence of aerosols. This paper describes in detail the complex interplay between the spectral effects of aerosols in the satellite observation and the associated response of the OMI O_2-O_2 cloud retrieval algorithm. Then, it evaluates the impact on the accuracy of the tropospheric NO_2 retrievals through the computed Air Mass Factor (AMF) with a focus on cloud-free scenes. For that purpose, collocated OMI NO_2 and MODIS (Moderate Resolution Imaging Spectroradiometer) Aqua aerosol products are analysed over the strongly industrialized East China area. In addition, aerosol effects on the tropospheric NO_2

The contents of this chapter are published as: **Chimot, J.**, Vlemmix, T., Veefkind, J. P., de Haan, J. F., and Levelt, P. F.: Impact of aerosols on the OMI tropospheric NO_2 retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free scenes via a simple cloud model?, Atmos. Meas. Tech., 9, 359-382, `https://doi.org/10.5194/amt-9-359-2016`, 2016.

AMF and the retrieval of OMI cloud parameters are simulated. Both the observation-based and the simulation-based approach demonstrate that the retrieved cloud fraction increases with increasing Aerosol Optical Thickness (AOT or τ), but the magnitude of this increase depends on the aerosol properties and surface albedo. This increase is induced by the additional scattering effects of aerosols which enhance the scene brightness. The decreasing effective cloud pressure with increasing τ primarily represents the shielding effects of the O_2-O_2 column located below the aerosol layers. The study cases show that the aerosol correction based on the implemented OMI cloud model results in biases between −20 and −40% for the DOMINO tropospheric NO_2 product in cases of high aerosol pollution ($\tau(550\text{ nm}) \geq 0.6$) at elevated altitude. These biases result from a combination of the cloud model error, used in the presence of aerosols, and the limitations of the current OMI cloud Look-Up-Table (LUT). A new LUT with a higher sampling must be designed to remove the complex behavior between these biases and τ. On the contrary, when aerosols are relatively close to the surface or mixed with NO_2, aerosol correction based on the cloud model results in an overestimation of the DOMINO tropospheric NO_2 column, between 10 and 20%. These numbers are in line with comparison studies between ground-based and OMI tropospheric NO_2 measurements in presence of high aerosol pollution and particles located at higher altitudes. This highlights the need to implement an improved aerosol correction in the computation of tropospheric NO_2 AMFs.

2.1 Introduction

Nitrogen oxides ($NO_x = NO + NO_2$) play a key role in atmospheric chemistry, regulating the level of ozone and maintaining the oxidizing capacity in the troposphere. The most important reasons to improve our knowledge of the global distributions of NO_x are (1) exposure to nitrogen dioxide leads to adverse health impacts; (2) the chemical budget of tropospheric ozone, also toxic for humans and the vegetation, is largely determined by the concentration of NO_x (Jacob et al., 1996); (3) nitrogen oxides are the precursors of (ammonium) nitrate, an important component of particulate matter, and contribute to acidification and eutrophication of soils and surface waters; and, (4) nitrogen oxides affect the global climate indirectly by affecting OH, and therefore modifying the residence time of the greenhouse gases O_3 and CH_4 (Shindell et al., 2009).

In 2004, the Dutch-Finnish Ozone Monitoring Instrument (OMI) (Levelt et al., 2006) was launched on the NASA EOS-Aura (National Aeronautics and Space Administration Earth observing System) satellite. OMI is a nadir-viewing imaging spectrometer that provides with daily global coverage of key air quality components. The retrieval technique of the OMI tropospheric NO_2 Vertical Column Density (VCD) (Boersma et al., 2004) is common to all the other similar satellite missions (Burrows et al., 1999; Bovensmann et al., 1999). The backscattered solar radiation is captured in daylight in the visible spectral domain by the instrument at the Top Of the Atmosphere (TOA) and then processed through the Differential Optical Absorption Spectroscopy (DOAS) retrieval approach. The DOAS method is based on radiative transfer modelling of tropospheric NO_2 Air Mass factor (AMF). The associated assumptions play a crucial role on the accuracy of the tropospheric NO_2 VCD.

The DOMINO (Derivation of OMI tropospheric NO_2) (Boersma et al., 2011) product contain worldwide concentrations of NO_2 in the troposphere derived from OMI. This product is used by a large number of air quality studies (e.g., Curier et al. (2014); Reuter et al. (2014); Ding et al. (2015)). The computation of tropospheric NO_2 AMF is acknowledged as the dominant source of errors in the retrieved tropospheric NO_2 column over polluted areas (Boersma et al., 2007) with important consequences for emission constraints and other applications. The overall uncertainty of the latest version of individual retrieved DOMINO tropospheric NO_2 vertical column densities is estimated to be 1.0×10^{15}molecules cm^{-2} ± 25% (Boersma et al., 2011). This evaluation takes into account the slant column precision, between 0.7 and 1.1×10^{15}molecules cm^{-2} (Boersma et al., 2007; Bucsela et al., 2006), and the uncertainty of tropospheric NO_2 AMF 10—40 % (depending on the impact of different parameters) (Boersma et al., 2007). Comparisons with in situ measurements from aircraft, ground-based observations from Multi-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) and Pandora instruments revealed that OMI tropospheric NO_2 retrievals generally agree within ±20 % (Lamsal et al., 2014). Several studies that investigated the accuracy of the DOMINO products over rural and urban areas in Russia, Asia and Indonesia, by using different long-term network observations based on MAX-DOAS, (Shaiganfar et al., 2011; Ma et al., 2013; Kanaya et al., 2014) found a negative bias between 26 and 50 % in urban and very polluted areas and when the Aerosol Optical Thickness τ is high. These underestimations have been recently confirmed over Wuxi city, area with high pollution adjoined to Shanghai (Wang et al., 2015b). These low biases could be partially attributed to the inhomogeneity of NO_2 at the spatial scale of OMI observation, and incomplete accounting of NO_2 near the surface. However, the possible additional effects caused by aerosols cannot be ignored. Over industrial regions with comparable photochemical regimes, NO_2 and aerosol concentrations are very well correlated showing similar anthropogenic sources between aerosols and short-lived trace gases (Veefkind et al., 2011).

The current version of the DOMINO algorithm does not explicitly account for the aerosol effects on the tropospheric NO_2 AMF. Similarly, these effects are not explicitly considered in most of the other UV-Vis (UltraViolet-Visible) satellite measurements (Valks et al., 2011). Martin et al. (2003) retrieved tropospheric NO_2 columns from the Global Ozone Monitoring Experiment (GOME) instrument by accounting for spatial and temporal variability of aerosols from monthly mean fields of aerosol mass concentration over 1 year simulated by the Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model. Aerosols affect the top-of-atmosphere (TOA) radiances in the visible and $O_2 - O_2$ spectral bands (Boersma et al., 2011; Leitão et al., 2010) and the light path distribution: i.e. the distribution of distances travelled by photons in the atmosphere before leaving the atmosphere. Since the impacts of aerosols on the computation of tropospheric NO_2 AMF is a function of aerosol properties (Martin et al., 2003), NO_2 and aerosols vertical distribution (Leitão et al., 2010; Bousserez, 2014), a proper characterization of aerosols' impact on the retrieval is then needed to accurately quantify trace gas amounts from satellite observations. The OMI cloud algorithm (Acarreta et al., 2004) derives the cloud fraction and cloud pressure from the O_2-O_2 absorption in the spectral window between 460–490 nm. This cloud retrieval algorithm is applied both to cloudy and cloud-free scenes with aerosols. In other words, aerosols are treated as clouds. The correction for the aerosol impact cannot be simply separated from the effects of clouds and, if a cloud retrieval scheme is applied, it will account for a large part of the aerosol effect by retrieving modified cloud fraction and pressure (Boersma et al., 2004, 2011). Wang et al. (2015a) show that OMI cloud parameters are indeed perturbed in presence of aerosols, by comparing with classified sky conditions using MAX-DOAS observations. Therefore, the cloud retrievals apply an implicit correction of aerosols presence in the computation of the tropospheric NO_2. Recent studies reprocessed the OMI tropospheric NO_2 product explicitly taking into account aerosols instead of effective cloud retrievals, either by considering model outputs or observations (Lin et al., 2014, 2015; Kuhlmann et al., 2015; Castellanos et al., 2015). These studies were performed over urban area in China or biomass burning events in South-America. Resulting changes mostly occurred in case of high aerosol pollution ($\tau(550\text{ nm}) > 0.8$) with increased or decreased tropospheric NO_2 VCDs depending on the geophysical conditions and aerosol properties. All these studies considered the magnitude of the explicit aerosol correction on the computation of tropospheric NO_2 AMFs and highlighted that aerosol properties and distribution and OMI cloud retrievals are connected. Nonetheless, there is still not a comprehensive understanding of the interplay between aerosol effects on the observations and the associated response of the operational OMI cloud retrieval algorithm. Moreover, the dependence of the retrieved cloud parameters and the associated implicit aerosol correction to the geophysical conditions and the aerosol properties and distribution are not clear. The needs to further investigate and understand the interplay between all these variables have been highlighted in different studies (Leitão et al., 2010; Lin et al., 2014). This has recently been emphasized in Castellanos et al. (2015) for the analyses of the implicit aerosol corrections in case of biomass burning aerosols.

This paper aims to analyse how aerosols affect the current operational OMI cloud products and the derived tropospheric NO_2 AMFs. For that purpose, this study uses a model vs. observation approach focused on large industrialized areas in China. This work focuses on cloud-free scenes but dominated by aerosol pollution. The paper starts with a description of the expected aerosol net effects on the tropospheric NO_2 AMFs, based on simulations in Sect. 2.2. The importance of aerosols and NO_2 vertical profiles is highlighted. In the following Sect. 2.3, the OMI DOMINO v2 and collocated MODIS (Moderate Resolution Imaging Spectroradiometer) aerosol products are confronted over large industrial area in China. This provides insights in the behaviour of OMI cloud fraction and pressure, and tropospheric NO_2 AMFs in the presence of aerosols. The following sub-section evaluates the response of the DOAS O_2-O_2 cloud algorithm on aerosol cases, as achieved by the operational OMI algorithm. The interplay between aerosol effects and the DOAS O_2-O_2 cloud retrieval can be characterized. Finally, in the Sect. 2.4, we deduce how an implicit aerosol correction is applied in the OMI retrieval chain through the Lambertian cloud model and evaluate its accuracy by comparing it to the explicit aerosol corrections simulated in Sect. 2.2.

2.2 Impact of aerosols on the tropospheric NO_2 AMF

2.2.1 Computation of the tropospheric NO_2 AMF

A two-step approach is used to determine the NO_2 VCD. First, the DOAS method (Platt and Stutz, 2008), based on the basic principle of absorption spectroscopy and the Beer-Lambert law, is applied to the measured spectra within 405–465 nm in order to derive a NO_2 Slant Column Density (SCD). This column gives the number of NO_2 molecules cm^{-2}, integrated along the average light path. Then, the NO_2 SCD is converted into NO_2 VCD. This can be done by application of the air mass factor calculated with a radiative transfer model. Note that an intermediate step is to remove the stratospheric part in the NO_2 SCD part in order to derive the tropospheric SCD component. In the DOMINO product the stratospheric SCD is derived from measurements over remote regions that are assimilated into a transport-chemistry model for the stratosphere (Boersma et al., 2011). The AMF A is defined as the ratio of the atmospheric SCD and VCD (Boersma et al., 2011):

$$A(\Psi) = N^{\mathrm{s}}(\Psi)/N^{\mathrm{v}}, \tag{2.1}$$

where N^{s} is the NO_2 SCD and N^{v} is the NO_2 VCD. The computation of $A(\Psi)$ requires accurate knowledge concerning all the parameters affecting the optical properties of the atmosphere and then impacting the length of the average light path followed by the backscattered sunlight. The list of variables describing these conditions is by Ψ and is detailed in Table 2.1.

The concept of altitude-resolved AMF was firstly introduced by Palmer et al. (2001); Eskes and Boersma (2003), although named differently (cf. "scattering weights"), and then confirmed by Wagner et al. (2007); Rozanov and Rozanov (2010); Richter and Wagner (2011). Also named block air mass factor or BAMF in other studies, it allows to generalise this definition by computing the AMF variable at discrete atmospheric layers such as:

$$a(p^*) = \left.\frac{\Delta N^s(\Psi)}{\Delta N^v}\right|_{\Delta n(p=p^*)}, \tag{2.2}$$

which describes the altitude dependence of A. The BAMF gives the change in the NO_2 slant column density for a change in the vertical column density at one atmospheric layer p (Wagner et al., 2007; Platt and Stutz, 2008; Rozanov and Rozanov, 2010; Richter and Wagner, 2011). a is the altitude-resolved AMF at atmospheric pressure p. $\Delta n(p = p^*)$ refers to a change in the trace gas profile n at pressure

Table 2.1: List of the physical variables (Ψ) required for the computation of the tropospheric NO_2 AMF, through the altitude-resolved AMF $a(p^*)$. For each of these variables, an indication about its degree of certainty is given.

Ψ		Degree of certainty
SZA (θ_0)		High
VZA (θ)		High
Relative azimuth angle ($\phi - \phi_0$)		High
Wavelength		High
Surface pressure		High
Surface albedo		Moderate
Vertical temperature profile		Moderate
Vertical pressure profile		Moderate
Clouds[1]:	Cloud (radiance) fraction	Moderate to Low
	Cloud pressure	Moderate to Low
Aerosols[2]:	τ	Low
	Altitude	Low
Vertical NO_2 profile		Low

1 Boersma et al. (2007) evaluated that an uncertainty of 0.05 on the cloud fraction and $60hPa$ on the cloud pressure respectively contribute up to 30% and 15% to the (relative) tropospheric NO_2 AMF. While inter-comparison studies show relative good agreements for the OMI cloud fractions (Stammes et al., 2008; Sneep et al., 2008) over cloudy-scenes, the retrievals can be affected by aerosols. Thus, the degree of certainty is here evaluated between low (clear and cloudy scenes) and moderate (cloudy scenes). See Sect. 2.3 for further details.

2 Sect. 2.2 shows that τ and aerosol altitude are the key parameters when computing the tropospheric NO_2 AMF. Aerosols impact several parameters (e.g., surface albedo, cloud) in addition to the tropospheric NO_2 AMF.

3 **?** and Lamsal et al. (2014) evaluated that using the high-resolution a priori NO_2 profiles impact the OMI tropospheric NO_2 columns between −43 and 22%. Aircraft and in situ measurements reveal that NO_2 profile shapes exhibit considerable day-to-day variation, such that the use of a monthly mean profile can cause significant errors in individual retrievals.

level p^*. ΔN^{v} and ΔN^{s} denote the resulting changes in VCD and SCD respectively. Currently only cloud parameters (and no aerosol parameters) are considered when computing $a(p)$ in the DOMINO product: cloud (radiance) fraction and cloud pressure. A is then computed by weighting $a(p)$ with the atmospheric vertical distribution of the trace gas profile:

$$A = \frac{\int_{p_0}^{p_{\mathrm{top}}} a(p) \cdot n(p)\,\mathrm{d}p}{\int_{p_0}^{p_{\mathrm{top}}} n(p)\,\mathrm{d}p}, \tag{2.3}$$

where $n(p)$ is the vertical distribution of trace gas density, p_{top} the pressure at the top of the atmosphere and p_0 is the surface pressure. In this paper, we define A^{tr} as tropospheric NO_2 AMF, which can be calculated from Eq. (3) by integrating a vertical profile $n(p)$ up to tropopause pressure. The ratio of the altitude-resolved air mass factor $a(p)$ to the total air mass factor A (deduced from the NO_2 shape profile) gives the averaging kernel: i.e. the sensitivity of the satellite measurement to each atmospheric layer (Eskes and Boersma, 2003; Richter and Wagner, 2011). Therefore, A^{tr} gives an indication of the sensitivity of the satellite measurement to the amount of NO_2 in the lowest atmospheric layers. Assuming identical shape of vertical NO_2 profile, a larger A^{tr} value expresses a higher sensitivity of the measurement while a reduced A^{tr} value indicates a decreased sensitivity. Indeed, in those cases, a change in A^{tr} is directly associated with a change of $a(p)$ at the atmospheric levels where the trace gas is present. $a(p)$ is in theory spectrally dependent, but the spectral dependence in case of tropospheric NO_2 retrieval is very small. The reference wavelength considered in this paper is 439 nm (Boersma et al., 2011).

Note that in the case of real OMI tropospheric NO_2 retrievals, a temperature correction should be applied on the slant column density. Since the temperature of the NO_2 absorption cross section is assumed to be fixed at 221 K, a correction term is thus implemented in the computation of A such that it represents the ratio of the NO_2 slant column density derived with a NO_2 cross section at the real temperature T to the column derived at 221 K. European Centre for Medium-Range Weather Forecasts (ECMWF) temperature fields are used for this correction (Boersma et al., 2004, 2011).

2.2.2 Qualitative description of aerosol optical effects

Similarly to cloud particles, scattering and absorbing effects induced by the presence of aerosol particles affect the total radiance observed by the satellite sensor and the light path distribution of detected photons. The presence of aerosols leads in general to higher radiance levels captured by the satellite sensor. This increase has a spectral variability depending on the aerosol properties. Its magnitude is however smaller in cases of a very bright surface and absorbing aerosols. The change of light path distribution has consequences on the sensitivity of the remote sensing measurement to the tropospheric NO_2 amount. This sensitivity can be either increased or decreased. Qualitatively, one can distinguish two effects that aerosols/clouds can have on the NO_2 absorption signal at the TOA (Leitão et al., 2010):

- Shielding effect: decreased sensitivity within and below the aerosol/cloud layers. The fraction of all detected photons coming from the top of the atmosphere that reaches the lowest part of the atmosphere (below the aerosol/cloud layer) is reduced compared to a cloud or aerosol free scene.

- Enhancement (albedo) effect: increased sensitivity within and above the aerosol/cloud layers. The fraction of all detected photons that sample the higher part of the atmosphere (above the aerosol/cloud layer) is increased, compared to a cloud or aerosol free scene, because more photons are scattered back towards the satellite.

Figure 2.1 depicts an example of vertical distribution of $a(p)$ for a cloud-free observation at 439 nm. With and without aerosols, $a(p)$ values generally decrease close to the surface. Both shielding and enhancement effects can be seen relative to the line for $\tau = 0$. The overall impact of aerosols on a measurement thus depends on the relative importance of the above mentioned effects, which mainly results from the combination of aerosol optical properties (Martin et al., 2003), amounts, surface reflectance and the vertical distribution of the particles, and the NO_2 gas. The geometry parameters like θ_0 (solar zenith angle or SZA) and θ (viewing zenith angle or VZA), and scattering angle (i.e. aerosol phase function) have an impact also as they modify the length of the light path.

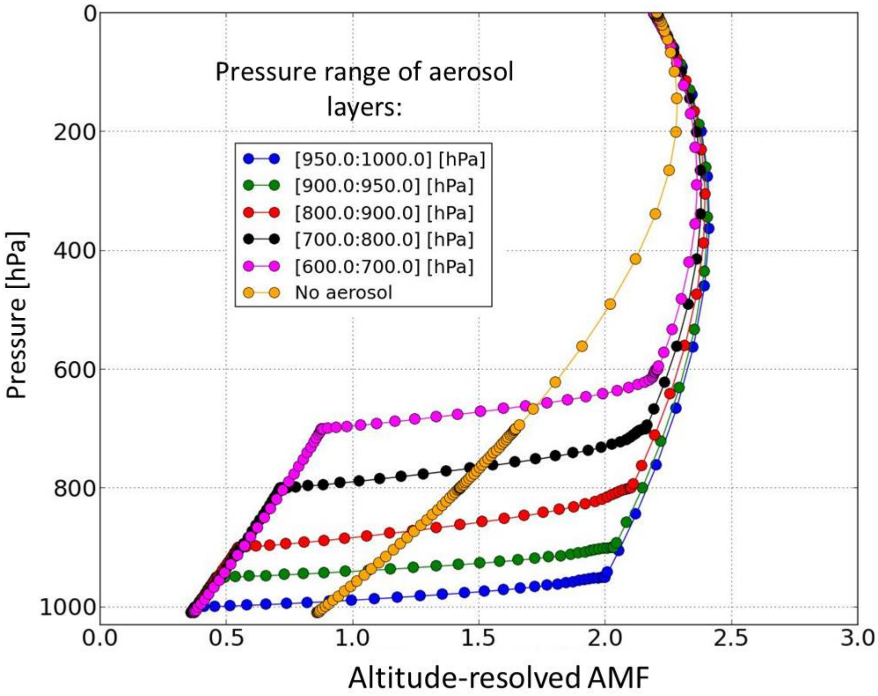


Figure 2.1: Altitude-resolved AMF (439 nm) as computed by DISAMAR for surface albedo = 0.05, SZA = 25°, VZA = 25°. Computations are done without and with aerosols at different layers. Aerosols are specified with $\tau(550nm) = 1$, SSA = 0.95, Angstrom coefficient = 1.5 and assymetry parameter 0.7.

2.2.3 Quantitative description of the impact of aerosols properties, their vertical distribution, and NO_2 profile

Methodology

This section studies the direct effects of aerosols on the computation A^{tr} without consideration of the O_2-O_2 cloud retrieval. This study uses the Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval (DISAMAR) software developed at Koninklijk Nederlands Meteorologisch Instituut (KNMI) (de Haan, 2011). This software package includes a radiative transfer model and can apply different retrieval methods, such as direct fitting (within an optimal estimation framework) and DOAS. The radiative transfer model is based on the Doubling Adding KNMI (DAK) model (de Haan et al., 1987; Stammes, 2001) and thus computes the reflectance and transmittance in the atmosphere using the adding/doubling method. This method calculates the internal radiation field in the atmosphere at levels to be specified by the user and takes into account Rayleigh scattering, trace gas absorption, and aerosol and cloud scattering. Scattering by aerosols is simulated with a Henyey–Greenstein scattering phase function $\Phi(\cos\Theta)$ (Hovenier and Hage, 1989):

$$\Phi(\cos\Theta) = \frac{1-g^2}{(1+g^2-2g\cos\Theta)^{3/2}}, \tag{2.4}$$

where Θ is the scattering angle, and $g = \langle\cos\Theta\rangle$ is the asymmetry parameter. In a standard case, an asymmetry parameter of $g = 0.7$ is used. Thus, in DISAMAR, the Angstrom exponent α gives the spectral dependence of the AOT τ. In DISAMAR, $a(p)$ is analytically determined, based on the weighting functions of the reflectances (i.e. derivatives of the reflectances to absorption cross-section and trace gas density). Indeed, $a(p)$ for an atmospheric layer can be identified as the Jacobian of the forward model $\partial F/\partial n$. This term is independent of the tracer distribution for optically thin absorbers. This methodology is conceptually equivalent to the approach discussed in Sect. 2.2.1.

The Henyey-Greenstein phase function is quite commonly used in the DOAS community for tropospheric NO_2 retrievals (Vlemmix et al., 2010; Castellanos et al., 2015) with explicit aerosol corrections. With an asymmetry parameter of $g = 0.7$, the Henyey-Greenstein functions are known to reasonably well reproduce the Mie functions. Thus, it can be used for the AMF calculation (Spada et al., 2006; Wagner et al., 2007, and Stammes, private communication). Castellanos et al. (2015) found that decreasing g from 0.7 to 0.6 in DISAMAR impacts less than 5 % the tropospheric NO_2 AMF for τ(550 nm) lower than 0.5. For larger τ values, the impacts are almost negligible. It is noted that, for τ retrievals, more realistic assumptions of the phase function should be used. However this is not the scope for an aerosol correction.

The simulations in this section are done with α 1.5 for fine particles and 0.5 for coarse particles, asymmetry parameter 0.7 and Single Scattering Albedo (SSA) $\omega_0 = 0.95$ and 0.9, assuming different altitudes, surface albedos 0.05 and 0.07 (surface reflectance is assumed Lambertian), solar zenith angle (SZA) $\theta_0 = 25°$ and viewing zenith angle (VZA) $\theta = 25°$. ω_0 and τ values are considered at the reference wavelength of 550 nm.The NO_2 profiles are taken from a model run where atmospheric chemistry and transport model Tracer Model 5 (TM5) has been integrated into the global climate model EC-Earth version 2.4 (van Noije et al., 2014). We defined the tropospheric AMF aerosol factor $f(\tau)$ as the ratio of the AMF with ($A^{\mathrm{tr}}(\tau)$) and without aerosols present ($A^{\mathrm{tr}}(\tau = 0)$).

$$f(\tau) = \frac{A^{\mathrm{tr}}(\tau)}{A^{\mathrm{tr}}(\tau = 0)}. \tag{2.5}$$

f can be interpreted as the factor by which the tropospheric NO_2 AMF of a clear scene should be adjusted to represent aerosol effects. In practice, such a factor cannot accurately be determined as not all required information, associated with instantaneous individual measurement, are available. Similarly to the DOMINO product, A^{tr} is computed at 439 nm.

Results

Figure 2.3 is an example of the computations following Eq. (2.5) based on all the individual NO_2 profiles generated by the TM5 model for the month of July 2006 at 12:00 (close to the OMI local observation time) over the region of East China (lat. 30–40°, long. 110–130°) (see Fig. 2.1). The error bars in the Figs. 2.3–2.5 represent the variability in f due to the variability of the TM5 NO_2 profiles over this region during the month of July. The curves in these figures connect the average values of f per AOT bin.

Figure 2.3 demonstrates that f lies in the range between 0.7 and 1.3. The total effect of aerosols (shielding or enhancement) strongly depends on the location of the particles in the atmospheric layers, and results from the computed $a(p)$ depicted in Fig. 2.2. Scattering of aerosols enhances the tropospheric NO_2 A up to 30 % for $\tau = 1.0$ when they are located within or below the NO_2 bulk (between the surface and 900 hPa). When a given amount is lifted to higher altitudes, aerosols thus apply a shielding effect (i.e. reduced sensitivity to the tropospheric NO_2 amount) up to 30 %. The variability of the NO_2 vertical distribution impacts the magnitude of these effects, around 10 % for $\tau = 1$.

In addition to the vertical distribution of the aerosol particles, the shape of the vertical NO_2 profile also significantly affects the magnitude of f. In winter (e.g. January in Fig. 2.1), such profile shows higher absolute values of concentrations near the surface with a higher variability. Moreover, the profile shape (obtained after normalisation to the integrated vertical profile along the atmospheric layers) depicts a small difference with a higher dynamic between the surface and the atmosphere layer at 900 hPa. Figure 2.4a shows amplified enhancement effects (up to 40 % for aerosols between the surface and 950 hPa) and amplified shielding effects (up to 45 % for aerosols at very high altitude, between 600 and 700 hPa). The transition between a net shielding or enhancement effect is also closer to the surface compared to summer (close to 950 hPa) as the aerosols are well mixed with the tropospheric NO_2 bulk only below 950 hPa. The variability of the NO_2 profile, mostly in the tropospheric layers, have a larger impact in January, where the error bars indicate a variability of around 20 % for $\tau = 1.0$. The altitudes of tropospheric NO_2 and aerosols, and so the relative altitude between both, are thus the key drivers of f.

Other parameters also contribute to the magnitude of this factor:

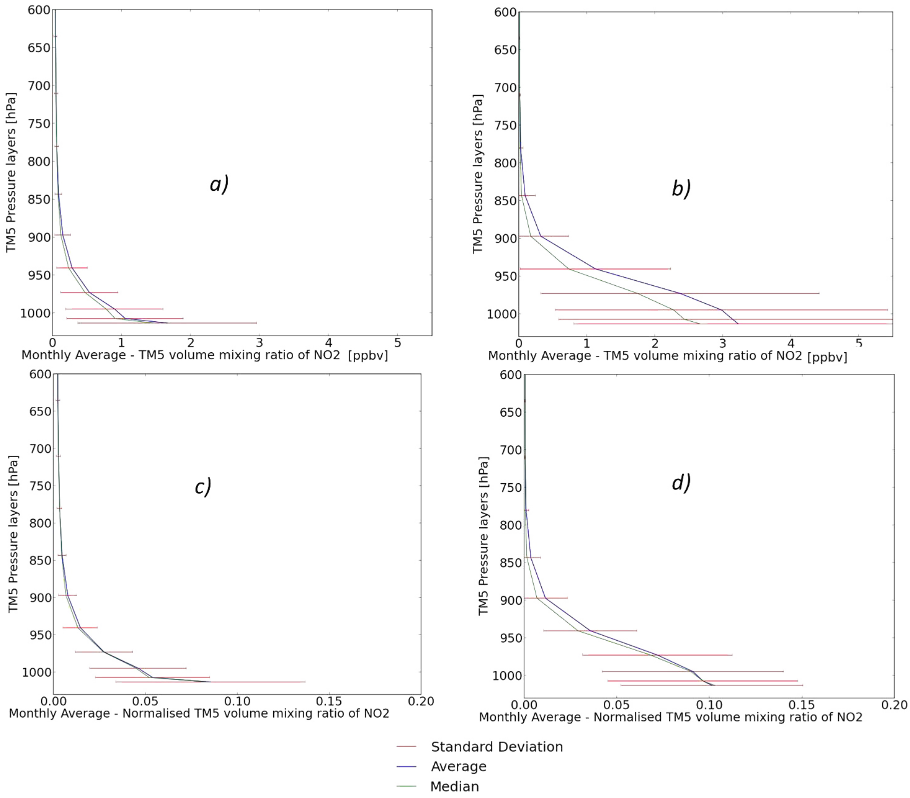


Figure 2.2: Monthly average NO_2 vertical profiles, zoomed in the troposphere, at 12:00 from TM5 simulations, 2006, East China (van Noije et al., 2014). **(a)** VMR NO_2 profile in July, **(b)** VMR NO_2 profile in January, **(c)** Normalised NO_2 profile in July, **(d)** Normalised NO_2 profile in January. Normalisations are done by dividing the VMR of each atmospheric layer to the integrated VMR profiles along the vertical atmospheric layers in the complete atmosphere, troposphere + stratosphere.

- An increase of surface albedo (cf. Fig. 2.3b), from 0.05 to 0.07, reduces the enhancement effect by 10 % and enhances the shielding effect by less than 5 % for $\tau(550nm) = 1.0$.

- The size of particles specified through α has little impact on the factor (cf. Fig. 2.5a). Decreasing α from 1.5 (fine particles) to 0.5 (coarse particles) reduces the shielding and enhancement effects between 2 % and 5 % for $\tau(550nm)$ = 1.0.

- A change of ω_0 from 0.95 to 0.9 (cf. Fig. 2.5b) leads to a reduction of the enhancement effect by 10 % (when aerosols are located below or well mixed with the tropospheric NO_2 bulk). The shielding effect is increased by 5 %.

- The increase of θ_0 from 25° to 50° (typically Winter average over China), increases the shielding effects by 10 % for $\tau = 1.0$ (assuming NO_2 profiles in Jan-

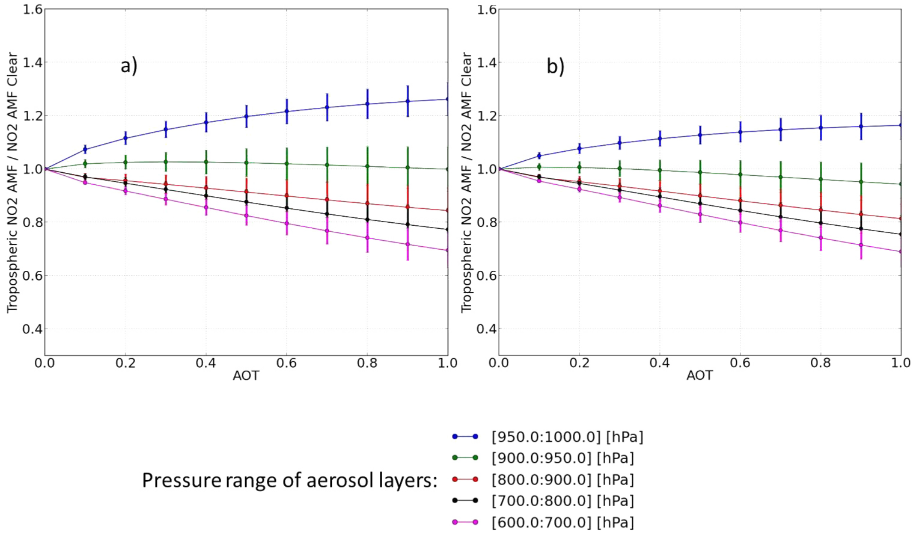


Figure 2.3: Ratio (f) of AMFs at 439 nm (cf. Eq. 2.5) as a function of AOT and aerosol layer altitude for surface albedo = 0.05 **(a)** and 0.07 **(b)**. f is determined for all the NO_2 vertical profiles from TM5 simulations over 2006, East China, July 2006 (cf. 2.1), SZA = 25°, VZA = 25°, aerosol Angstrom coefficient = 1.5, SSA = 0.95, asymmetry parameter = 0.7 for the Henyey–Greenstein phase function. The curves connect the average values per AOT bin and the error bars represent the standard deviation, of f computed for all the individual TM5 NO_2 profiles over this period and this region.

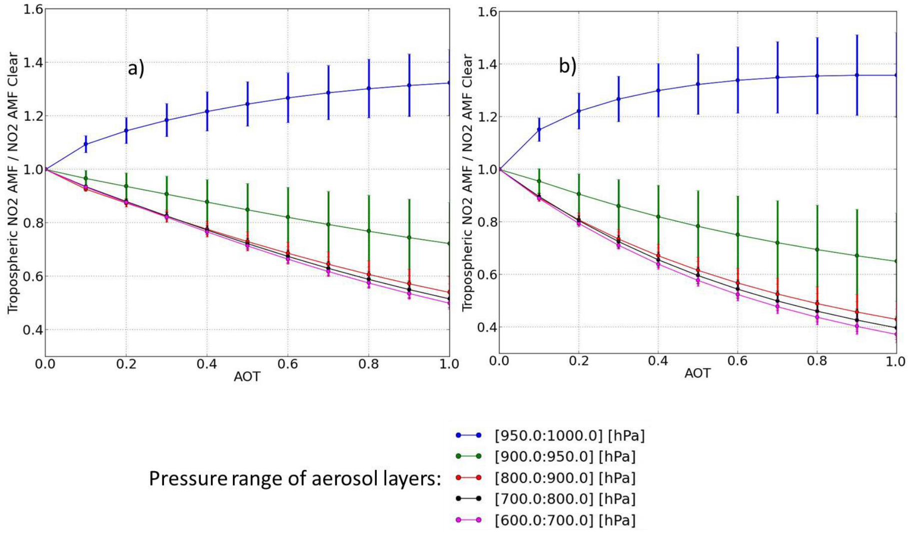


Figure 2.4: **(a)** similar to Fig. 3a, but with NO_2 profiles for January. **(b)** similar to Fig. 4a, but with SZA = 50°.

uary). Moreover, the enhancement effect increases between 5 % and 10% for $\tau(550nm)$ between 0.3 and 0.7.

- The monthly variability of the NO_2 profiles increases the variability of f, expressed by the error bars in Fig. 2.3, when aerosols are located close to the surface between 900 and 1000 hPa. This is a direct consequence of the enhancement effect induced by aerosol particles that increase the sensitivity to the NO_2 in the lower part of the atmosphere.

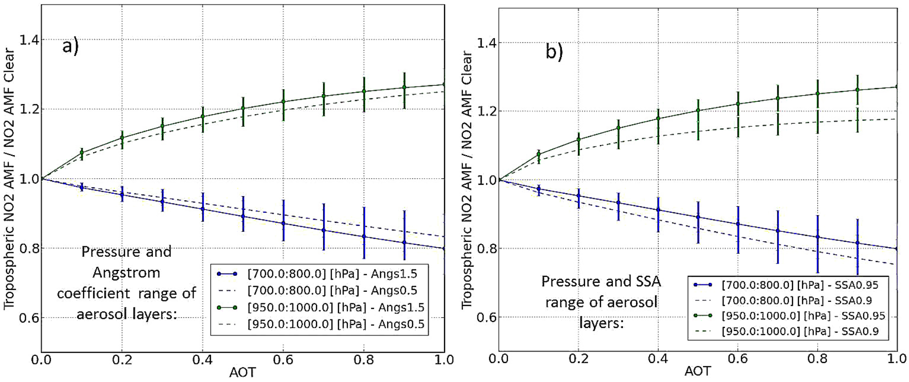


Figure 2.5: similar to Fig. 3, but with only 2 different atmospheric aerosol layers: **(a)** 2 Angstrom coefficient values: 1.5 (fine particles) and 0.5 (coarse particles), **(b)** 2 SSA values: 0.95 and 0.9.

Our results are consistent with previous findings by Leitão et al. (2010) and Bousserez (2014) who used different theoretical NO_2 and aerosol vertical distributions and optical properties. In particular, our present exercise considered various NO_2 profiles, as given by the TM5 model, representative of 2 distinctive months (July and January), and thus of two typical seasons, over urban areas in China. Moreover, the monthly variability of these profiles, and thus their impacts on the tropospheric NO_2 AMF variability, are investigated as a function of aerosol properties and vertical profile giving then complementary insights about explicit aerosol effects. These results are also in line with the work of Martin et al. (2003) where strong absorbing aerosols reduced the AMF by 40 % (over biomass burning regions) while scattering sulfate and organic aerosols increased the AMF by $5-10$ %.

2.3 Interplay between aerosols and the OMI O_2-O_2 cloud retrievals

This section explains the perturbations induced by the aerosol particles on the retrieval of cloud fraction and cloud pressure based on the OMI O_2-O_2 spectral band. This section is structured as follows: first the OMI DOMINO product is analysed in confrontation with the MODIS Aqua aerosol product. Then, the OMI DOAS cloud O_2-O_2 retrieval chain is analysed with simulated aerosol cases.

2.3.1 Comparison of OMI DOMINO-v2 with MODIS Aqua aerosol product

MODIS on-board EOS-Aqua observes the Earth's atmosphere approximately 15 min prior to OMI on-board EOS-Aura. The aerosol effects on the current OMI tropospheric NO_2 retrievals are investigated by confronting collocated OMI DOMINO with MODIS Aqua Level 2 (L2) aerosol products over large industrialized areas in China. Statistics are computed over three years (2005–2007) and two seasons: summer (June, July and August) and winter (December, January and February). MODIS L2 aerosol products have a spatial resolution of 10 km × 10 km, close then to the OMI spatial resolution (13 km × 24 km at nadir). The OMI and MODIS data are paired on a pixel-by-pixel basis if the distance between pixel centers is less than 5 km and if both observations are acquired within 15 min. Observations with a cloud fraction (OMI and MODIS) higher than 0.1 are filtered out. This threshold is applied to both OMI and MODIS, although both parameters are not identical. Applying such a threshold on the observations increases the probability to identify cloud-free scenes. Moreover, the availability of the MODIS aerosol product is a good confirmation of the identification of cloud-free scenes as MODIS Aqua τ is exclusively retrieved for cloud-free situations (Remer et al., 2008). However, it is well recognised, according to the analyses in the next section, that cloud-free observations with large presence of aerosols are filtered out as well. Tests were performed with higher cloud fraction thresholds (0.2 and 0.3) showing no statistical significant changes in the results described below.

The tropospheric NO_2 AMF (A) that is extracted from the OMI DOMINO database, shows a decreasing trend with increasing τ in Summer (cf. Fig. 2.6a). This decrease is in average 5 % for MODIS $\tau(550nm)=1$, with a variability of 20 %. A small local positive trend (around 5 %) is however noticed for τ = 0–0.2. On the contrary, in Winter, there is in average no modification of the tropospheric NO_2 AMF A with increasing τ (cf. Fig. 2.6b). By making use of the Angstrom coefficient α available in MODIS-Aqua data (cf. Fig. 2.6c), it is found that A is larger for coarse particles than for fine particles (differences of 10 %). Such statistics include as well spatial variability in τ and NO_2, such that the apparent correlations between them may be affected by other spatial factors like surface albedo or elevation.

Figures 2.7–2.9 depict the impact of aerosols on the OMI O_2-O_2 cloud fraction and pressure. Under aerosol presence and no cloud contamination in the OMI measurement, the OMI cloud fraction shows a clear linear relation with respect to τ. In average, values increase from 0.01 to 0.07 with a variability of 30 % for $\tau=1$. The magnitude of this increase depends on the surface albedo (Kleipool et al., 2008) and MODIS Aqua aerosol properties:

- The increase of cloud fraction with increasing τ is higher over dark surfaces and lower over bright surfaces (average differences of 0.03 for $\tau=1$, between OMI surface albedos of 0.04 and 0.08 (cf. Fig. 2.8)). As analysed in detail in Sect. 2.3.3., this is a direct consequence of the aerosol effects on the continuum reflectance. The attenuation of surface reflectance by particles is stronger over bright surface than over dark surface. Over a darker surface, aerosols play

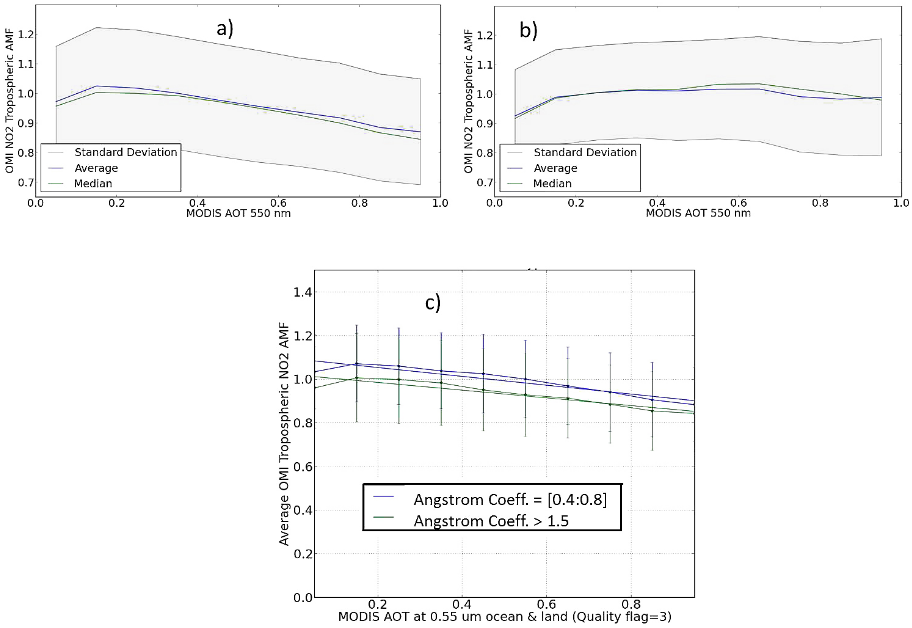


Figure 2.6: Tropospheric NO_2 AMFs from the OMI DOMINO v2 plotted against MODIS Aqua AOT. Statistics computed over 3 years (2005, 2006, 2007) and following the methodology described in Sect. 2.3: **(a)** Summer (June, July, August), **(b)** Winter (December, January, February), over East China, **(c)** Summer, distinction between 2 ranges of MODIS Aqua aerosols Angstrom coefficient.

the role of a surface layer with a higher albedo as they increase the scene brightness. As a consequence, the retrieved effective cloud fraction value is higher.

- Cloud fraction values are higher in the presence of small particles (average differences of 0.03 between MODIS Aqua α of [1.5:1.8] and [0.4:0.8]) (cf. Fig. 2.9).

The cloud pressure values show a non-linear decrease from approximately 800 to 600 hPa for $\tau = 1$, with a variability of around 100 hPa during Summer (cf. Fig. 2.8 and Fig. 2.9). However, no or little decrease is observed during Winter. The cloud pressure stays close to the surface (between 900 and 1000 hPa). The retrieved cloud pressures seem to have some sensitivity to the surface and aerosol properties. In particular, it decreases more over dark surface (difference of 100 hPa between surface albedo 0.04 and 0.07 for $\tau(550nm) = 1$) and in the presence of fine particles.

This section follows previous studies (Boersma et al., 2011; Lin et al., 2014; Castellanos et al., 2015) by analysing the OMI cloud and AMF parameters as present in the DOMINO product over scenes dominated by aerosols. It confirms that the cloud parameters respond to the presence of aerosols. The magnitude of this response is not only a function of aerosol properties but also the atmospheric and surface properties that impact the average light path (e.g. surface albedo). One may conclude that there

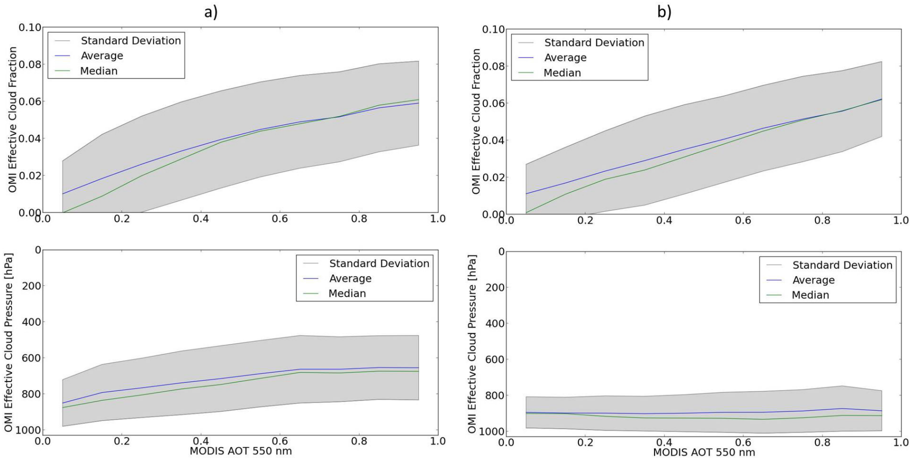


Figure 2.7: Effective cloud fraction and cloud pressure extracted from OMI DOMINO v2 compared to MODIS Aqua AOT for 2 seasons. Statistics are computed over 3 years (2005, 2006, 2007) and following the methodology described in Sect. 2.3.1: **(a)** Summer (June, July, August), **(b)** Winter (December, January, February) over East China.

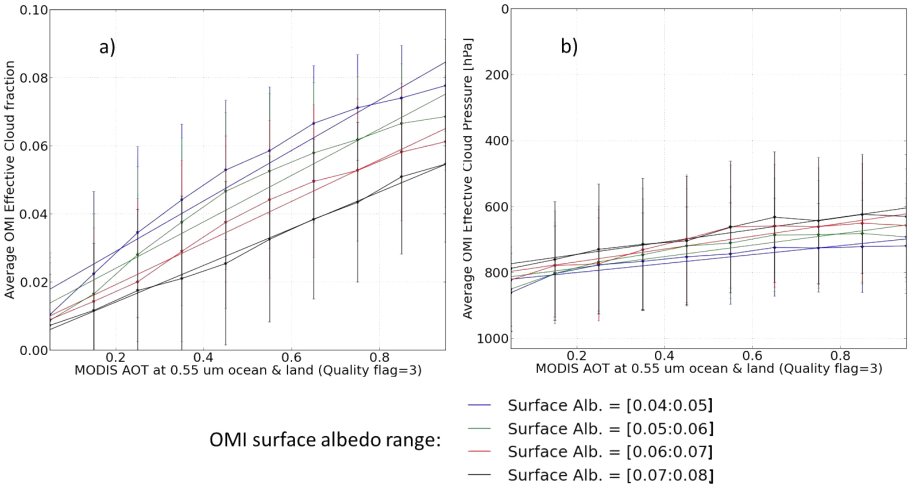


Figure 2.8: OMI effective cloud parameters extracted from OMI DOMINO v2 compared to MODIS Aqua AOT, as a function of the OMI climatology surface albedo (Kleipool et al., 2008). Statistics computed over 3 years (2005, 2006, 2007), summer (June, July, August) over East China (cf. Sect. 2.3.1): **(a)** Effective cloud fraction, **(b)** Effective cloud pressure.

is an implicit correction for the presence of aerosols in the tropospheric NO_2 retrieval chain since the OMI cloud algorithm cannot distinguish cloud and aerosol particles and retrieve effective cloud parameters. The computation of the tropospheric NO_2 AMF relies on the NO_2 vertical shape profile and aerosol properties (that drive the OMI cloud retrievals). Both parameters have different characteristics depending on the season. As a consequence, the implicit correction seems to show two different

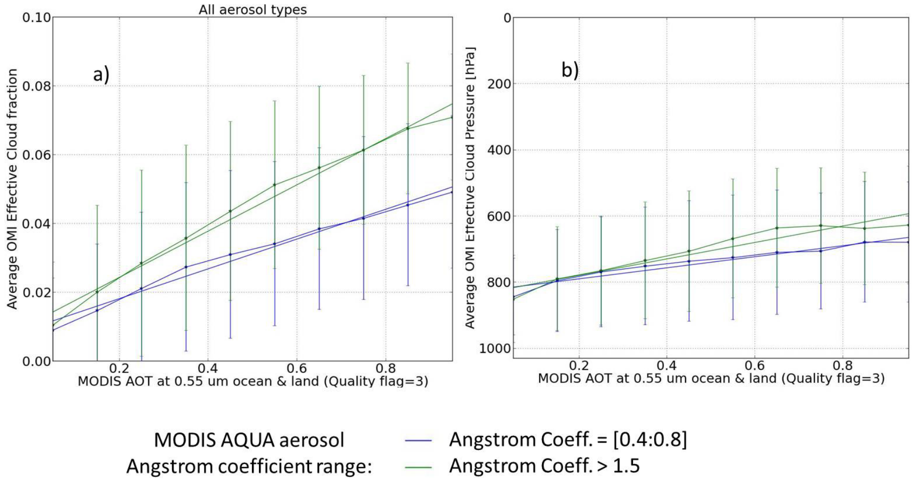


Figure 2.9: Similar to Fig. 8 but as a function of the MODIS aerosol angstrom coefficient: **(a)** Effective cloud fraction, **(b)** Effective cloud pressure.

behaviors depending on the seasons. In average, this correction applies a shielding effect in summer: i.e. the measurements are assumed to have less sensitivity to tropospheric NO_2 in the presence of aerosols. In winter, this corrections does on average not vary with increasing τ values.

2.3.2 Qualitative description of the OMI cloud algorithm

Inverse cloud model

In the context of trace gas measurements from space, the purpose of a cloud model is to describe the clouds in a way that reproduces the reflectance spectrum, and thus the distribution of photon paths, within the cloudy-scenes. For this purpose, the parameters designing such a model are cloud fraction, cloud optical thickness, cloud top altitude, and cloud vertical extent. However, instruments like OMI have limited spatial resolution (13 km × 24 km at nadir view) and do not resolve individual clouds. Therefore, cloud fraction and cloud optical thickness cannot be separated. Furthermore, OMI cannot give information on cloud microphysical properties such as cloud phase, cloud particle shape and size, and cloud vertical structure.

Because clouds are a correction step in trace gas retrievals, both the cloud retrieval algorithm and the cloud correction algorithm have to use the same cloud model. As a consequence, a simple model is used in the OMI tropospheric NO_2 retrievals, describing a cloud as a Lambertian reflector with a fixed albedo through which no light is transmitted. The associated effective cloud fraction is thus not a geometric cloud fraction but the radiometrically equivalent cloud fraction which, in combination with the assumed cloud albedo, yields a TOA reflectance that agrees with the observed reflectance. While scattering clouds have two main optical properties in the UV-Visible (namely reflection and transmission – their absorption being negligible), a Lambertian reflector has only reflection properties, determined by the cloud albedo, and no transmission properties. The OMI cloud retrieval algorithm assumes a cloud albedo of 0.8 (Stammes et al., 2008). This value has been found suitable to correct NO_2 and O_3 retrievals for partially cloudy scenes. The missing transmission of optically thin and medium thick clouds in the Lambertian cloud model is compensated for by the cloud-free part of the pixel.

Based on the properties of an opaque Lambertian cloud model, the effective cloud fraction is mainly constrained by the brightness of the cloud and how much a brighter cloud would outshine the observation scene. The effective cloud pressure is mainly constrained by the perturbation of the clouds on the O_2-O_2 collision complex absorption. A cloud located at high altitude shields the O_2-O_2 complexes that are below the cloud. As a consequence, the O_2-O_2 absorption signal, and so the associated slant column density, are attenuated (Acarreta et al., 2004; Stammes et al., 2008; Sneep et al., 2008).

Description of the O_2-O_2 DOAS retrieval algorithm

The OMI cloud retrieval chain (Acarreta et al., 2004) exploits the 460–490 nm absorption band of O_2-O_2, a collision pair of oxygen. The retrieval algorithm is based on the DOAS method and consists of two steps. In the first step, the absorption cross-section spectrum of O_2-O_2 is fitted together with a first order polynomial to the negative logarithm of the measured reflectance spectrum. The window of the spectral fit is (460–490 nm). This step can be summarised as follows:

$$-\ln(R) = \gamma_1 + \gamma_2 \cdot \lambda + N^{s}_{O_2-O_2}(\lambda) \cdot \sigma_{O_2-O_2} + N^{s}_{O_3}(\lambda) \cdot \sigma_{O_3}, \tag{2.6}$$

where $\gamma_1 + \gamma_2 \cdot \lambda$ defines the first order polynomial, $\sigma_{O_2-O_2}$ is the $O_2 - O_2$ absorption cross-section spectrum (at 253 K), σ_{O_3} is the O_3 absorption cross section spectrum, $N^{s}_{O_3}$ is the O_3 slant column density and $N^{s}_{O_2-O_2}$ is the O_2-O_2 slant column density. The O_3 cross section spectrum is included because it overlaps with the O_2-O_2 spectrum. The fitted parameters are γ_1, γ_2, $N^{s}_{O_2-O_2}$, and $N^{s}_{O_3}$. In the absence of absorbers, one may define the continuum reflectance R_c at the reference wavelength λ_0:

$$R_c = \exp(-\gamma_1 - \gamma_2 \cdot \lambda_0). \tag{2.7}$$

The reference wavelength is fixed at the middle of the DOAS fit window at $\lambda_0 = 475nm$. In the second step, a Look-Up-Table (LUT) is used to convert the retrieved $N^{s}_{O_2-O_2}$ and R_c into the cloud pressure C_p and cloud fraction C_f. This inversion step requires prior information about surface albedo, surface altitude and geometry angles (θ_0, θ and the relative azimuth angle $\phi - \phi_0$).

2.3.3 OMI cloud algorithm applied to aerosol scenes

To test the sensitivity to aerosols, the current version of the OMI DOAS O_2-O_2 algorithm was applied to simulated spectra for scenes dominated by aerosols. The implementation was performed in such a way that it is almost identical to the operational DOMINO chain at KNMI. The effective cloud fraction and cloud pressure parameters are derived following Eqs. (2.6) and (2.7) and through linear interpolation in the LUT, assuming thus an opaque Lambertian cloud model as described previously. Reflectance spectra are simulated by including only aerosol particles with the DISAMAR software. No clouds are included in the simulated reflectances. The sensitivity of the retrievals are investigated as a function of surface albedo, aerosol properties (α, ω_0, vertical distribution), θ_0 and θ. Simulated reflectances are noise-free. All the parameters (including surface albedo) are identical in the simulated spectra dominated by aerosols and the retrieval of effective cloud parameters.

Response of the cloud fraction to aerosol scenes

Figure 2.10 shows that the effective cloud fraction increases with increasing τ in cloud free scenes up to 0.09 for $\tau = 1.0$ at the wavelength of 550 nm, assuming fine particles ($\alpha = 1.5$), high single scattering albedo ($\omega_0 = 0.95$), $\theta_0 = 25^\circ$ (summer in China) and $\theta = 25^\circ$. Here, aerosols are located between 700 and 800 hPa in the atmosphere (between approximately 2 and 3 km). Similarly to what has been observed in the DOMINO product, the increase of the effective cloud fraction, in the presence of aerosols, is linear and higher with lower surface albedo (i.e. over dark surfaces). In this case, with a surface albedo of 0.07, the effective cloud fraction stays below 0.09 for $\tau = 1.0$ while, with a surface albedo of 0.03, the value is close to 0.1. Such an increase is consistent with the impact of the aerosol particles on the continuum reflectance as a function of τ and surface albedo. For these surface albedos and aerosol properties, the scattering effects of aerosols dominate over their extinction.

Figures 2.11 and 2.12 illustrate that aerosol properties (size and absorption) drive the magnitude of the increase of effective cloud fraction. Notably, low α and ω_0 values have smaller impact on the increase of the effective cloud fractions. This illustrates the reduction of scattering effects of aerosols under these conditions. Indeed, low ω_0 values increase the probability of absorption of the photons and so reduce the scattering within the layers and towards the satellite sensor. Coarse particles reduce also the scattering effects by increasing the probability of forward scattering of the photons towards the top of the atmosphere or towards the surface. With fine particles, the effective cloud fraction varies between 0.06 ($\omega_0 = 0.9$) and more than 0.1 ($\omega_0 = 0.97$) for $\tau(550nm) = 1.0$.

As a consequence, a higher cloud fraction is understood from the excess TOA reflectance caused by the additional scattering due to aerosols and the impact of the surface reflection. This represents the significant enhanced brightness of the scene (or enhanced scene albedo).

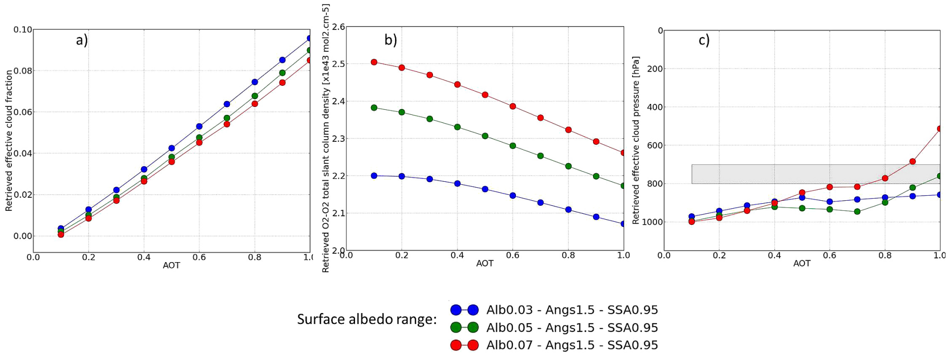


Figure 2.10: Simulated DOAS O_2-O_2 cloud retrieval results, based on noise-free spectra with aerosols, as a function of AOT and surface albedo, assuming an opaque (albedo = 0.8) Lambertian cloud forward model. The results are derived from the following geophysical conditions: average of temperature, H_2O, O_3, and NO_2 vertical profiles from TM5 month July (see Fig. 1), O_3 total column = 250 DU, SZA = 25° and VZA = 25°, surface pressure = 1010 hPa. Aerosol properties are: SSA = 0.95, Angstrom coefficient = 1.5 (fine particles), asymmetry parameter = 0.7, layers located between 700 and 800 hPa: **(a)** Effective cloud fraction, **(b)** O_2-O_2 total slant column density, **(c)** Effective cloud pressure (grey colour depicts the location of the simulated aerosol layers).

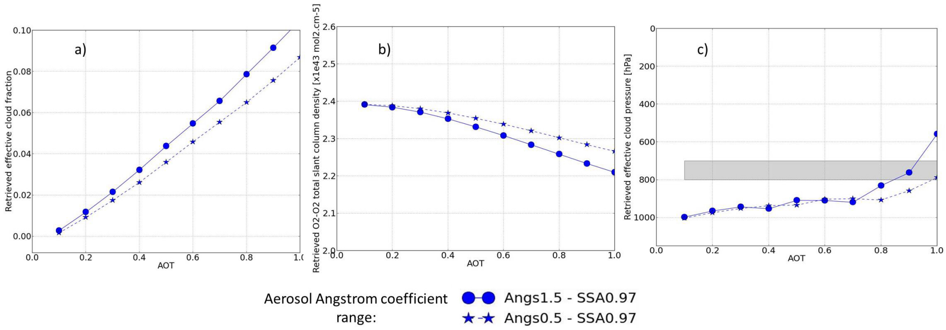


Figure 2.11: As Fig. 2.10 but the results are depicted as a function of Aerosols AOT and Angstrom coefficient. The surface albedo is here constant (0.05).

Response of the cloud pressure to aerosol scenes

Figure 2.13 shows that the retrieved effective cloud pressure decreases with increasing τ (or AOT). This decrease is linked to the O_2-O_2 shielding effect which strongly depends on τ. The O_2-O_2 absorption, below the optically thicker aerosol layer, is reduced since a high amount of particles decreases the fraction of photons reaching the lowest part of the atmosphere and increases the attenuation of the surface reflectance signal. Therefore, the length of the average light path is shortened. At high τ values, the retrieved cloud pressure correlates with the aerosol layer height. Overall, the values are close or smaller than the mean aerosol layer height which may be caused by the model error (i.e. difference between the cloud model and the aerosol

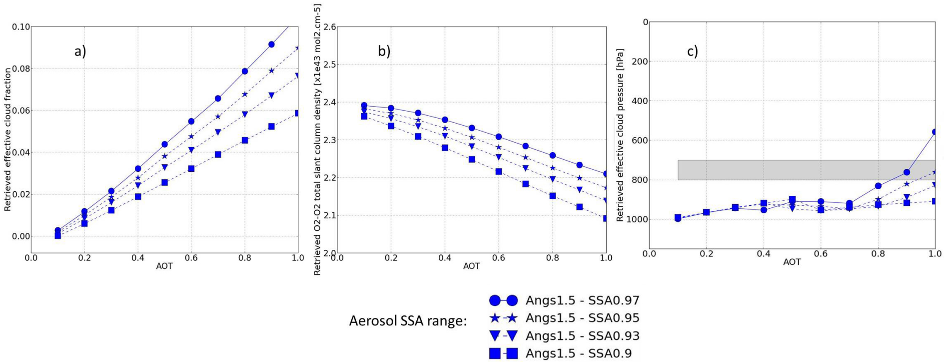


Figure 2.12: As Fig. 2.10 but the results are depicted as a function of Aerosols AOT and SSA. The surface albedo is here constant (0.05).

spectral effects). Surprisingly, at small τ values, the mean aerosol layer height has no effect on the retrieved cloud pressure. The retrieved values stay very close to the surface pressure.

The OMI cloud LUT has been intended for representing the cloud spectral effects and not those of thin aerosol layers. Thus, the sampling may be not high enough in case of low cloud fraction values (i.e. smaller than 0.1). Such values have limited effects on the average light path and the actual designed LUT is not sensitive enough to small changes on the O_2-O_2 absorption (Acarreta et al., 2004). Since low amount of aerosols have little effects on the O_2-O_2 SCD and the continuum reflectance, the overestimation of the retrieved cloud pressure, in those cases, may be caused by the coarse sampling of the employed LUT. When τ increases, the considered entry in the LUT moves from this undetermined regime to a regime where meaningful cloud pressure value can be interpreted. This can be seen in Fig. 2.10a–c where the transition between both regimes is located between $\tau(550nm) = 0.6$ and $\tau(550nm) = 0.8$ assuming $\theta_0 = 25°$ and $\theta = 25°$. This demonstrates a non-linear behavior between the cloud pressure retrieval and τ. Such behaviour is consistent with the analyses of Boersma et al. (2011), over southern and eastern US, which show that reduced OMI O_2-O_2 cloud pressure values are observed only with high τ values. Wang et al. (2015a) found that in general the effective cloud fraction of up to 15 % and cloud top pressure from the surface to 900 hPa from OMI are assigned to the condition of “clear sky with presence of aerosol particles”.

The value of $\tau(550nm)$ at which the retrieved cloud pressure starts being sensitive to the aerosol layer height mainly depends on the geometry. Figure 2.14 shows that for larger θ_0 and θ values (i.e. more than 25°), this transition triggers at smaller τ values (around $\tau = 0.4$). This can be understood as an increased average path length traveled by the photons in the atmosphere and higher retrieved effective cloud fraction values (up to 0.15). Note that change of relative azimuth angle may have similar effects.

The effects of aerosol microphysics properties on the effective cloud pressure retrieval mainly depend on the aerosol amount and the geometry. While smaller α and

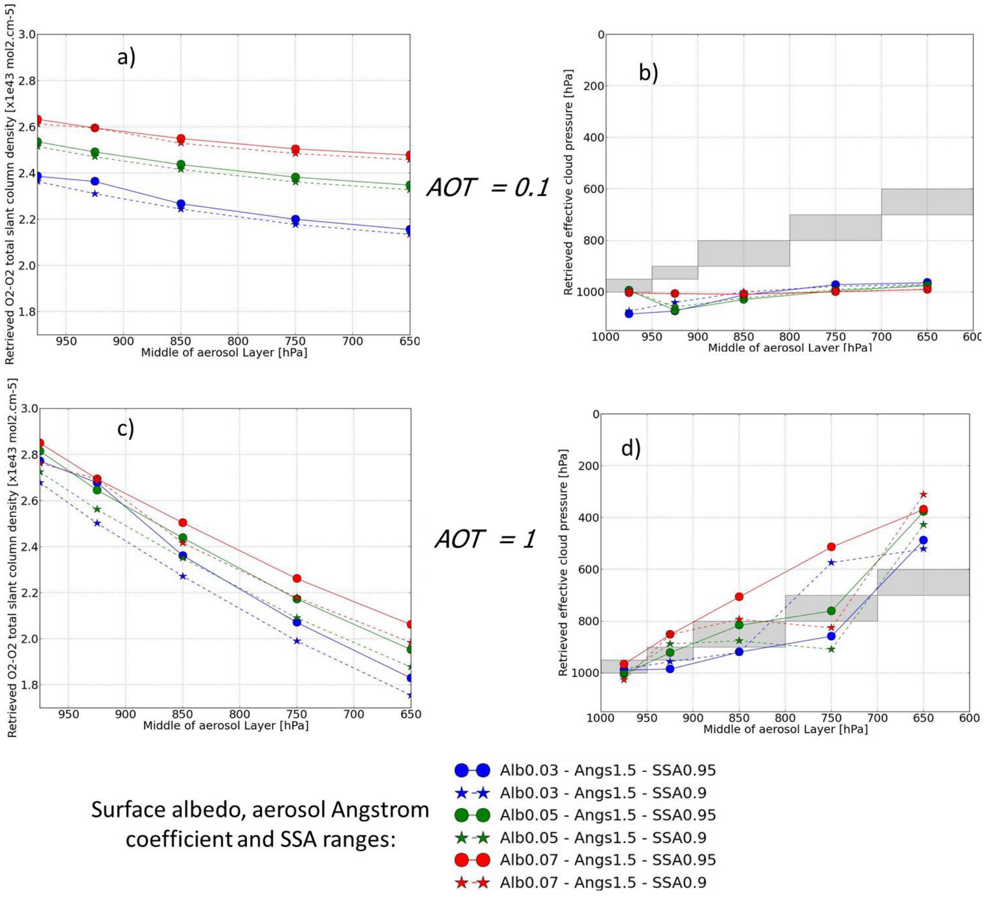


Figure 2.13: Impact of the location of atmospheric aerosols layers on the simulated DOAS O_2-O_2 cloud retrieval results as a function of AOT and surface albedo. The results are derived from conditions of Fig. 10: **(a)** O_2-O_2 total SCD for AOT $= 0.1$, **(b)** Effective cloud pressure (grey colour depicts the location of the simulated aerosol layers) for AOT $= 0.1$, **(c)** O_2-O_2 total SCD for AOT $= 1$, **(d)** Effective cloud pressure for AOT $= 1$.

ω_0 values lead to smaller O_2-O_2 SCDs (cf. Fig. 2.10b and c), the associated effective cloud pressure only decreases: for τ well above 1 in case of small angles ($\theta_0 = 25°$ and $\theta = 25°$); or between $\tau = 0.6$ and $\tau = 1$ in case of large angles ($\theta_0 = 50°$, or $\theta = 45°$). Finally, cases with high surface albedo show smaller retrieved effective cloud pressure. The brighter the surface, the more the average photon path length is reduced by a thin aerosol layer. This also highlights that a bias on the assumed surface albedo can perturb the effective cloud pressure retrieval in cases of high aerosol amount. For instance, an overestimated surface albedo (because of scattering aerosol affects) can lead to reduced effective cloud pressure.

Based on these simulations and retrievals, we can now largely understand the decrease of the effective cloud pressure in summer over China. For one part, it is a consequence of presence of fine aerosol particles (most of MODIS Angstrom coeffi-

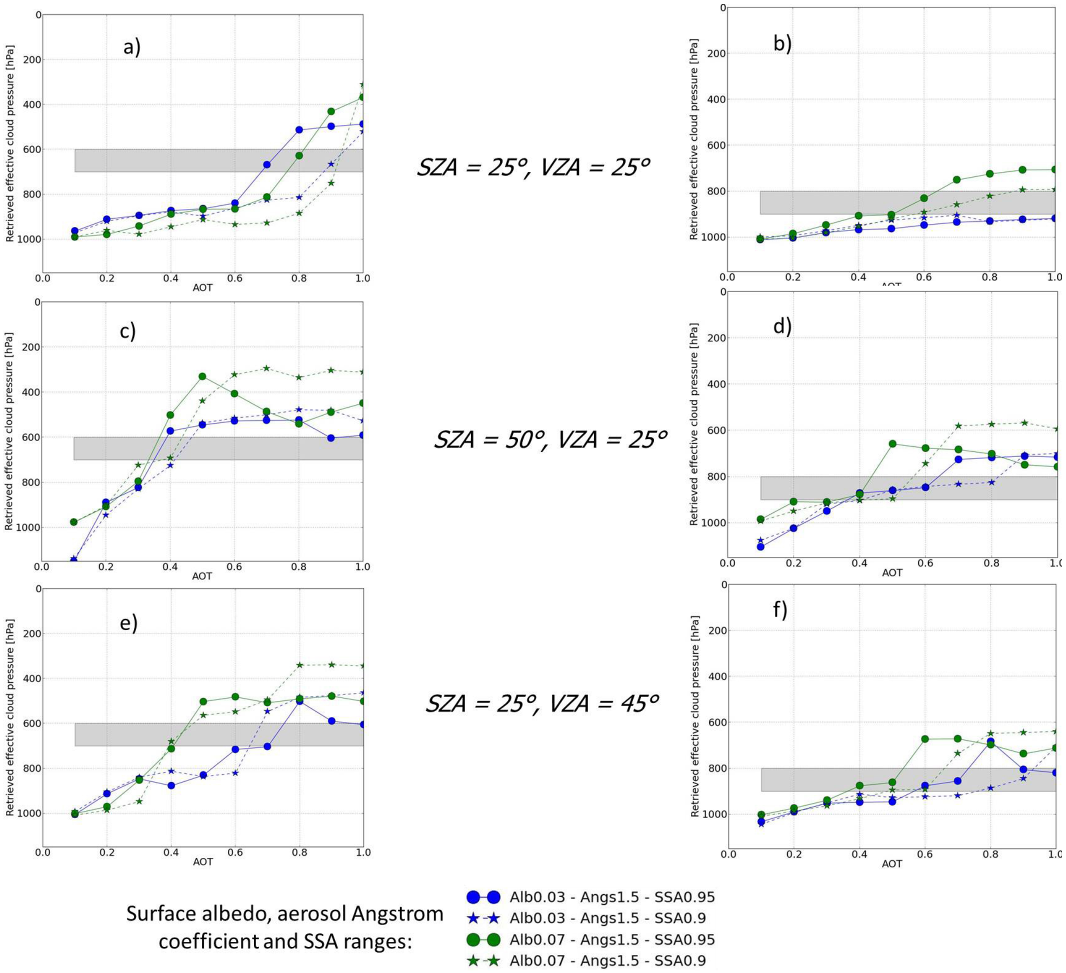


Figure 2.14: Impact of geometry angles on the effective cloud pressure retrievals as a function of surface albedo, aerosol microproperties, AOT and location of atmospheric aerosol layers (grey colour depicts the location of the simulated aerosol layers). Same conditions as Fig. 2.10 are assumed: **(a)** and **(b)** SZA = 25° and VZA = 25°, **(c)** and **(d)** SZA = 50°, VZA = 25°, **(e)** and **(f)** SZA = 25°, VZA = 45°.

cients are beyond 1.5). Moreover, the boundary layer is generally deeper in summer due to convective growth. The high cloud pressures for low τ values are largely a retrieval artifact (as discussed above); the lower cloud pressures for higher τ are probably more realistic as in the regime of high τ there is more sensitivity to the layer height (Fig. 2.10). In winter, this transition from almost no sensitivity at low τ to more sensitivity to the layer height at high τ results in an almost flat curve, probably because the boundary layer itself is quite shallow. The variability that is seen in Fig. 2.7 is related to the different effects of surface reflectance and variable viewing angles.

Table 2.2: List of values considered for the simulation nodes illustrated in Fig. 2.15: effective cloud fraction, effective cloud pressure, aerosol optical thickness and aerosol apressure.

Parameter	List of values
Effective cloud fraction	0., 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.95, 1., 1.1, 1.2
Effective cloud pressure [hPa]	1013, 963, 913, 863, 813, 763, 713, 663, 613, 563, 513, 463, 413, 363, 313, 263, 213, 163, 113, 63
Aerosol optical thickness	0., 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1., 1.05, 1.25, 1.5, 1.75, 2., 2.25, 2.5, 2.75, 3., 3.25, 3.5, 3.75, 4
Aerosol pressure [hPa]	975, 950, 925, 875, 850, 825, 800, 775, 750, 725, 700, 675, 650 550, 450, 350, 250, 150

2.3.4 General inter-comparison of cloud and aerosol impacts on the O_2-O_2 spectral band

The analyses deduced from Sect. 2.3 clearly point out the limitation of the response of the current OMI cloud algorithm, in particular the effective cloud pressure retrieval, due to the not optimised OMI cloud LUT over clear-sky scenes dominated by aerosol particles. The use of a LUT cannot be avoided since it is required to convert the continuum reflectance and the O_2-O_2 SCD into effective cloud parameters. Since different LUTs may give different results, the following exercise (summarised in Fig. 2.15) illustrates in a general way how aerosols and OMI effective cloud retrievals should be connected. O_2-O_2 spectra (from 460 nm to 490 nm) were simulated containing either an opaque Lambertian cloud (albedo = 0.8) assuming different cloud fraction and pressure values, or box fine and scattering aerosol layers with different τ and aerosol pressure values (see list of values in Tab. 2.2). A very high sampling in the simulation nodes (i.e. cloud fraction, cloud pressure, τ and aerosol pressure values) is considered. Then a DOAS fit is achieved as described in Sect. 2.3.3. Finally, a linear interpolation / extrapolation, based on the radial basis function, is performed in order to have a global overview of the variation of the simulation nodes as a function of the DOAS fit variables. The accuracy of the linear interpolation / extrapolation is here ensured thanks to the very high sampling of the simulation nodes.

As expected, Fig. 2.15 confirms that the effective cloud fraction is primarily constrained by the continuum reflectance, while the O_2-O_2 SCD maily drives the effective cloud pressure. Similarly, following the previous analyses, τ mostly impacts the continuum reflectance while the aerosol altitude (or aerosol pressure) mostly results in a change of O_2-O_2 SCD. Furthermore, in the case of low continuum reflectance (below than 0.2), which corresponds to aerosol cases and low effective cloud fraction, some correlations are observed between both DOAS fit variables.

Therefore, in the case of an ideal O_2-O_2 cloud retrieval (i.e. without the specific limitation of the current employed OMI cloud LUT), the following is expected:

- For a given aerosol altitude value, increasing τ should result in a larger continuum reflectance and thus increase the effective cloud fraction;

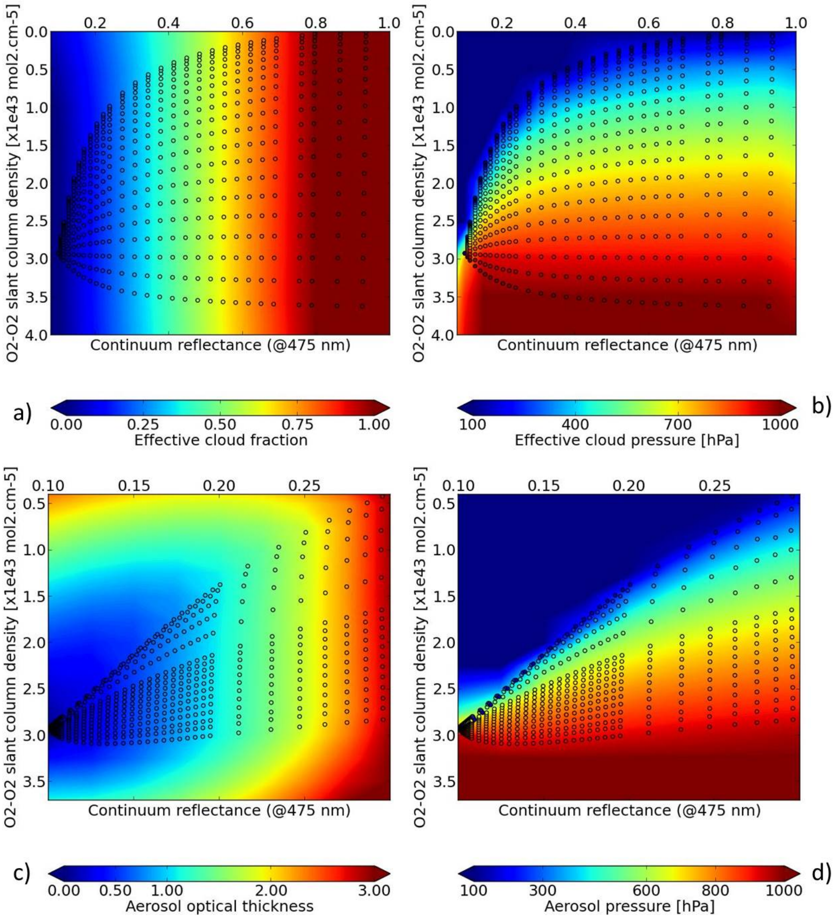


Figure 2.15: Effective cloud (Lambertian reflector, albedo = 0.8) and fine scattering aerosol ($\alpha = 1.5$, $\omega_0 = 0.95$, $g = 0.7$) parameters as a function of O_2-O_2 slant column density and continuum reflectance at 475 nm for the following conditions: climatology mid-latitude summer temperature, NO_2, O_3 and H_2O profiles, surface albedo = 0.05, SZA = 32°, VZA = 32° and surface pressure = 1013 hPa. The dots are the values specified (cf. Tab. 2.2) in the forward simulations (named simulation nodes). The background colors result from the linear interpolation / extrapolation of the DOAS fit results: **(a)** Effective cloud fraction, **(b)** Effective cloud pressure [hPa], **(c)** Aerosol optical thickness $\tau(550nm)$, **(d)** Aerosol pressure [hPa].

- For a given τ value, increasing the aerosols altitude (or decreasing pressure) should result in smaller O_2-O_2 SCD and thus decrease the effective cloud pressure;

- Since increasing τ primarily impacts the continuum reflectance but also simultaneously impacts the O_2-O_2 SCD, retrieved effective cloud pressure could theoretically either increase, decrease or stay constant depending on the aerosol altitude.This demonstrates that the magnitude of the O_2-O_2 shielding effect by aerosols is a combination of aerosol amount and altitude.

The high sampling of simulation nodes in Fig. 2.15 shows that the variation of effective cloud pressure, in case of low continuum reflectance, has very small impacts on the O_2-O_2 SCD. However, it is theoretically still possible to retrieve small values (not only values close to the surface). Low sampling of simulations would result in inaccuracy of the interpolation / extrapolation. This is why the current OMI cloud LUT only exhibits high effective cloud pressure values, i.e. close to the surface pressure.

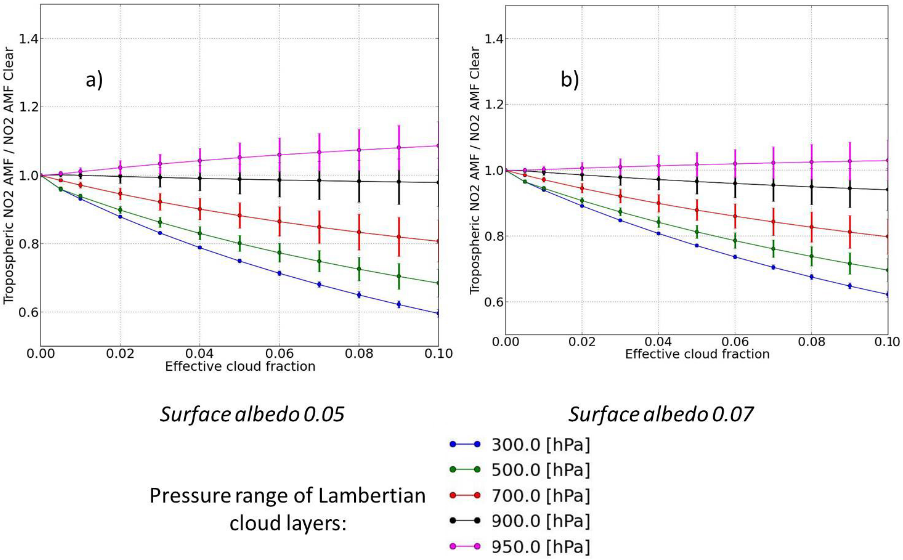
2.4 Implicit vs. explicit aerosol correction in the tropospheric NO_2 AMF

2.4.1 Tropospheric NO_2 AMF factor based on effective cloud parameters

The behaviour of the OMI cloud algorithms in response to aerosols, as analysed in Sect. 2.3.3, has consequences on the computation of the tropospheric NO_2 AMF. Indeed, as effective cloud parameters are sensitive to the presence of aerosols, their properties and their location in the atmosphere, they apply an implicit aerosol correction as observed in the DOMINO product (Sect. 2.2). This implicit aerosol correction is obtained through the altitude-resolved AMF $a(\Psi, p)$ which uses the retrieved effective cloud fraction and cloud pressure, that are impacted by the presence of aerosols, and no explicit aerosol information. This differs from an explicit aerosol correction where explicit aerosol parameters would be used.

Similarly to Fig. 2.3, Fig. 2.16 depicts the resulting tropospheric NO_2 AMF factor f following Eq. (2.5) at 439 nm, based this time on effective cloud parameters: i.e. the computation of A^{tr} is not based on τ and other aerosol properties, but on effective cloud fraction values between 0. and 0.1 and different cloud pressures. The denominator of f corresponds here to cloud-free cases (i.e. effective cloud fraction$\,=\,$0). Two surface albedo values are considered (0.05 and 0.07), $\theta_0 = 25°$, $\theta = 25°$ and NO_2 profiles from TM5 in July at 12:00 pm (cf. Fig. 2.1). In case of strong aerosol contamination (i.e. effective cloud fraction$\,=\,$0.1), the implicit aerosol factor lies in the range of 1.15–0.6: i.e. 15 % enhanced sensitivity if the cloud is retrieved close to the surface and likely well mixed (even below) the tropospheric NO_2 bulk; 40 % reduced sensitivity if the cloud is retrieved at elevated altitude. In cases of high τ values, the decrease of effective cloud pressure has more impact on the magnitude of f than the increase of cloud fraction. Indeed, an increase of effective cloud fraction from 0.08 to 0.1 has an impact of less than 10 %. At the same time, a change of cloud pressure from 900 to 700 hPa can induce a change of 20 % in the AMF factor.

Finally, the variability of the NO_2 profiles causes a higher variability of f, between 10 and 15%, for cloud pressures between the surface and 700 hPa. It is highly reduced for very high clouds (i.e. cloud pressure between 300 and 500 hPa). This reduction is caused by the absence of scattering properties in the inverse cloud model which results in an almost complete mask of the tropospheric NO_2 bulk below the



Surface albedo 0.05

Surface albedo 0.07

Figure 2.16: Tropospheric NO_2 AMF factor f at 439 nm (cf. Eq. 2.5) based on OMI effective cloud parameters (i.e. effective cloud fraction and pressure) and for 2 surface albedos, derived from all the NO_2 vertical profiles from TM5 simulations, 2006, East China, July (cf. Fig. 1). Solid lines are the average, while error bars are the standard deviation, of f computed for all the individual TM5 NO_2 profiles over this period and this region: **(a)** Surface albedo 0.05, **(b)** Surface albedo 0.07.

supposed cloud layer. This is contrary to f based on explicit aerosol properties, where even particles with strong shielding effects show a non-negligible sensitivity to the variability of tropospheric NO_2 vertical shape.

Following the sensitivity analyses of the O_2-O_2 cloud retrieval algorithm, the behavior of tropospheric NO_2 AMFs observed in the DOMINO products, over China, can be understood as following: on average, a decrease of tropospheric NO_2 AMF in summer with increasing τ is caused by the simultaneous increase of effective cloud fraction and decrease of effective cloud pressure. Qualitatively, this behaviour is in line with the expected aerosol shielding effect on tropospheric NO_2 in summer. Indeed, Vlemmix et al. (2015) has shown that in summer in China, aerosol particles are generally located above the tropospheric NO_2 layers. The probability that aerosol layers are located higher than tropospheric NO_2 bulks are also mentioned in other studies. For instance, Li et al. (2013) performed MAX-DOAS measurements during the PRIDE-PRD2006 campaign in the Pearl River Delta region,in China, for 4 weeks in July 2006. The considered site is located at 60 km north of Guangzhou in a rural area. It is clearly shown that (for this data) aerosol mixing layers are the most often deeper / higher than NO_2 mixing layers. Mendolia et al. (2013) retrieved tropospheric NO_2 vertical column densities from OMI and MAX-DOAS measurements over Canada. One key conclusion of this work is that NO_2 diurnal profiles can even be systematically lower in summer and do not follow the expected pattern of the convective boundary layer (higher in summer than in winter). Aerosols do follow this seasonal pattern since they have a longer life time.

The absence of statistic increase/decrease of tropospheric NO_2 AMF in winter with increasing τ is mainly caused by the smaller effective cloud fraction (compared to Summer) and no variation of effective cloud pressure values which stay close to 900 hPa in average. The accuracy of the implicit aerosol correction is evaluated in the next final sub-section.

2.4.2 Evaluation of the implicit aerosol correction on tropospheric NO_2 AMF

In this section, AMFs computed with the cloud model used in the OMI O_2-O_2 retrieval (Sect. 2.3.3) are compared with AMFs computed assuming aerosols instead of clouds. This comparison is applied to cloud free scenes with only aerosols in the atmosphere, i.e. effective cloud fraction = 0. Thus, the implicit aerosol correction accuracy is here evaluated and discussed through the computation of the relative bias $S_{\mathrm{A}}(\tau)$ expressed in percentage:

$$S_{\mathrm{A}}(\tau)=\frac{A^{\mathrm{tr}}(\tau)-A^{\mathrm{tr}}_{\mathrm{exp}}(\tau)}{A^{\mathrm{tr}}_{\mathrm{exp}}(\tau)}\cdot 100\,, \tag{2.8}$$

where $A^{\mathrm{tr}}_{\mathrm{exp}}(\tau)$ is the tropospheric NO_2 AMF explicitly taking into account aerosols. This relative bias is computed in two ways: (1) assuming that $A^{\mathrm{tr}}(\tau)$ includes an implicit aerosol correction (based on the retrieved effective cloud parameters), and (2) that no aerosol correction at all (i.e. $A^{\mathrm{tr}}(\tau)=A^{\mathrm{tr}}(\tau=0)$). Then, the implicit aerosol correction can be compared to the case of no aerosol correction.

Figures 2.17–2.19 show that the relative biases S_{A} induced by the implicit aerosol correction vary from −10% to 30% in most of the simulated cases. These biases are negative (i.e. underestimation of the AMFs) and minimal when aerosols are mixed with the tropospheric NO_2 bulk at the surface. They are generally positive and maximal (i.e. overestimation of the AMFs) when aerosols are elevated in the atmosphere and so not mixed with the tropospheric NO_2 peak (i.e. between 950 and 600 hPa depending on the analysed cases). In most of the simulations, the higher biases are found over scenes with elevated and high aerosol pollution, with $\tau(550nm)\geq 0.6$. This is a consequence of an insufficient shielding effect applied in the computation of A^{tr} by the implicit aerosol correction through the OMI cloud algorithm. This results from too large effective cloud pressure values. When the effective cloud pressure value significantly decreases with increasing τ, the implicit aerosol correction is then able to reproduce the aerosol shielding effect with a better accuracy. For example, in Fig. 2.15, when geometry angles are small ($\theta_0=25°$ and $\theta=25°$), with very high aerosol pollution (τ close to 1), fine particles ($\alpha=1.5$) and high SSA ($\omega_0\geq 0.95$), the biases S_{A} decrease from 30 to around 10 %. This improvement is related to the fact that only for higher τ the impact on the O_2-O_2 signal is sufficient to dominate over artifacts related to interpolation and limited sampling of the LUT. In cases of aerosols mixed with NO_2, the biases are likely related to the discrepancy between the opaque Lambertian cloud model and the aerosol properties.

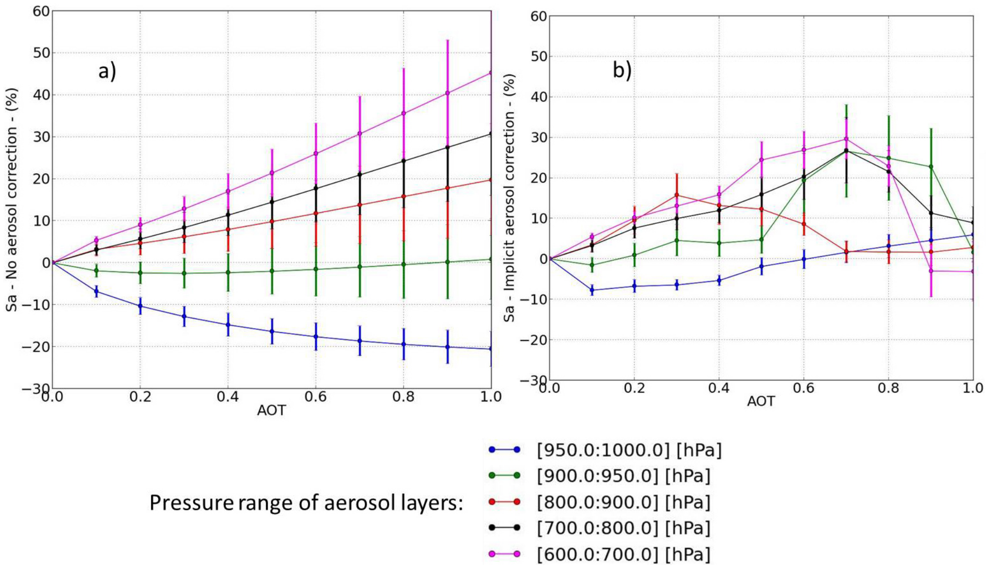


Figure 2.17: Comparison of relative tropospheric NO_2 AMF biases Sa at 439 nm (cf. Eq. 2.8) assuming different aerosol layers, surface albedo = 0.05, SZA = 25°, VZA = 25°, and TM5 NO_2 vertical profiles for the month of July at 12:00 pm over China (see Fig. 1). Aerosol properties are defined by Angstrom coefficient = 1.5, SSA = 0.95 and assymetry parameter = 0.7. Solid lines are the average, while error bars are the standard deviation, of Sa computed for all the individual TM5 NO_2 profiles over this period and this region: **(a)** Sa assuming no aerosol correction, **(b)** Sa assuming implicit aerosol correction through the OMI cloud retrieval algorithm.

Overall, the relative biases induced by the implicit aerosol correction are generally better than if no aerosol correction was applied in the computation of tropospheric NO_2 AMF. No aerosol correction would induce biases from −20% to 60% on A, assuming small geometry angles ($\theta_0 = 25°$ and $\theta = 25°$) and summer NO_2 profiles (cf. Figs. 2.17 and 2.18). Assuming winter NO_2 profiles (e.g. Fig. 2.19a) or larger angles (e.g. $\theta_0 = 50°$ in Fig. 2.19c), these relative biases can even increase up to 150 % depending on the aerosols altitude. Indeed, in those cases, aerosols apply a stronger shielding effect on the tropospheric NO_2 bulk.

Aerosols altitude and amount (i.e. τ) are the key drivers of the magnitude of the relative biases S_A. Effects of aerosols microphysics, such as associated SSA or size have a second order of magnitude. Compared to Fig. 2.15, Fig. 2.16 shows that coarser particles ($\alpha = 0.5$ instead of 1.5) and reduced SSA (0.9 instead of 0.95) mostly increase the relative biases induced by the implicit aerosol correction for very large AOT ($\tau(550nm) \geq 1$) by increasing values from around 10 to 40%. However, these values still remain lower than if no aerosol correction was applied: S_A values close to 55 % in case of high aerosols altitude. For lower τ values, no significant

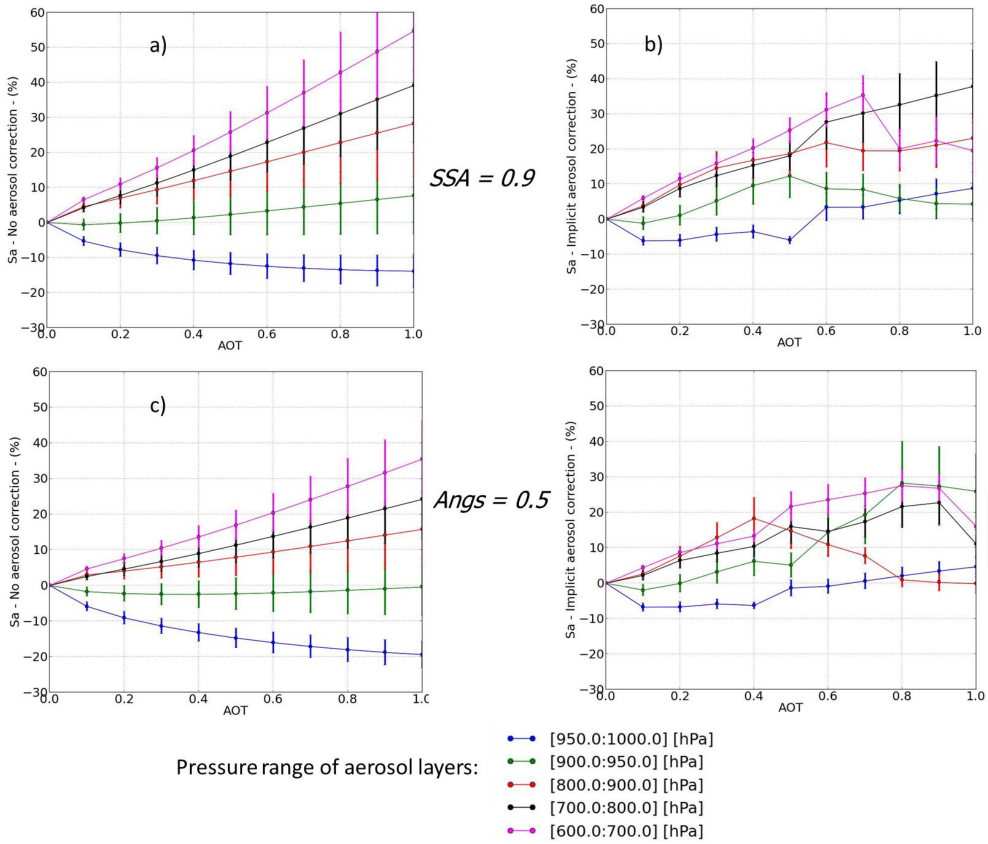


Figure 2.18: Similar to Fig. 14 but with different aerosols SSA and Angstrom coefficient values: **(a)** Sa assuming no aerosol correction, SSA = 0.9, **(b)** Sa assuming implicit aerosol correction through the OMI cloud retrieval algorithm, SSA = 0.9, **(c)** Sa assuming no aerosol correction is applied, Angstrom coefficient = 0.5, **(d)** Sa assuming implicit aerosol correction through the OMI cloud retrieval algorithm, Angstrom coefficient = 0.5.

changes are visible. Figure 2.16 depicts that the shape of NO_2 vertical profile and large angles do not significantly change the S_A values of implicit aerosol corrections for elevated aerosol layers (from 900 to 600 hPa). Only in the specific case of aerosols located between 900 and 950 hPa, the values are increased (between 50 and 70%). The cause is an enhancement effect produced by too large effective cloud pressure while aerosols actually apply a strong shielding effect.

The monthly variability of the NO_2 profile shape induces a variability on the relative biases for implicit aerosol correction between 10% and 20% (indicated by the error bars on Figs. 2.14–2.16). The magnitude of this variability depends on the distance between the aerosols layer and the peak of the tropospheric NO_2 bulk. It is generally larger when the aerosol layers are close to the maximum in the NO_2 profiles.

Relative biases associated with implicit aerosol correction shows an irregular be-

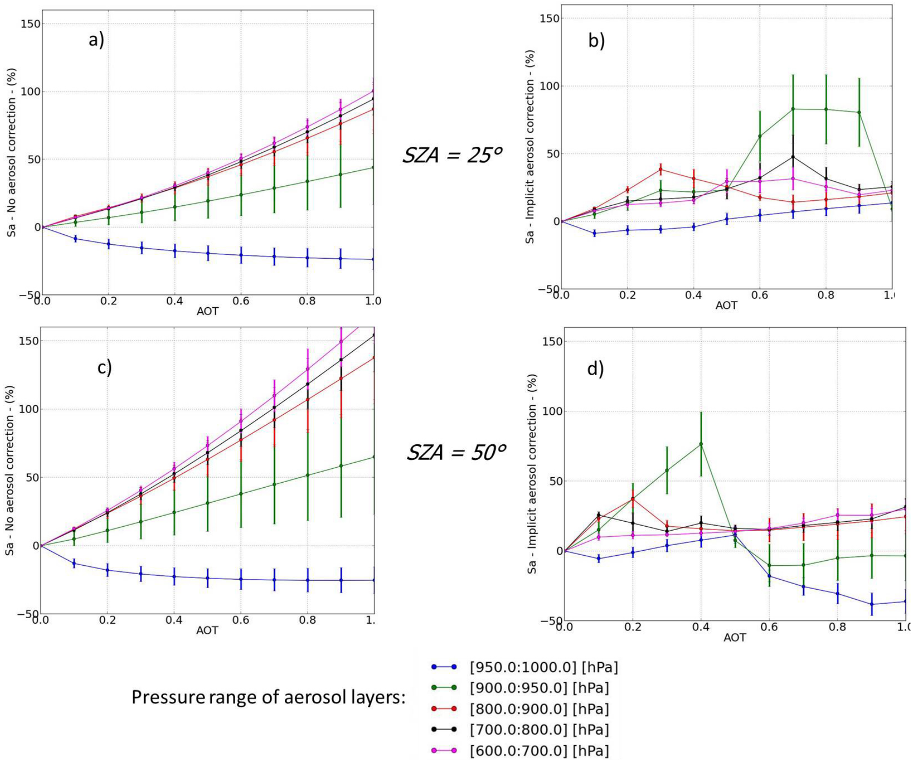


Figure 2.19: Similar to Fig. 14 but with NO_2 profiles for January and SZA = 50°: **(a)** Sa assuming no aerosol correction is applied, January NO_2 profiles, **(b)** Sa assuming implicit aerosol correction through the OMI cloud retrieval algorithm, January NO_2 profiles, **(c)** Sa assuming no aerosol correction is applied, aerosols, January NO_2 profiles and SZA = 50°, **(d)** Sa assuming implicit aerosol correction through the OMI cloud retrieval algorithm, January NO_2 profiles and SZA = 50°.

havior with respect to increasing τ values: i.e. they do not smoothly increase or decrease with increasing τ. They somewhat either increase or decrease depending on τ (and aerosol altitude) values. This differs from the smooth increase of relative biases assuming no aerosol correction with respect to τ. This is probably caused by the coarse sampling of the designed cloud LUT combined with the fact that the cloud model cannot describe aerosol dominated scenes in a perfect way. A higher sampling should be designed and tested through the OMI cloud algorithm over scenes dominated by aerosols. The behavior of these biases could lead to complex spatial and temporal patterns of the individual DOMINO tropospheric NO_2 products over highly polluted areas, not consistent with the physical NO_2 and aerosol patterns. The potential impacts on the estimation of NO_x surface fluxes should be investigated.

2.4.3 Impact of the implicit aerosol correction on observed data: comparison with recent studies

An overestimation of the tropospheric NO_2 AMF results in an underestimation of the tropospheric NO_2 VCD with the same absolute magnitude according to Eq. (2.1). The findings here identified on the biases related to the implicit aerosol correction are consistent with those identified by Shaiganfar et al. (2011); Ma et al. (2013); Kanaya et al. (2014). These studies found negative biases between −26 and −50% on the OMI tropospheric NO_2 VCDs over areas with high aerosol pollution, in particular in summer. Investigations led by Ma et al. (2013) show that these underestimations can be explained by the presence of elevated aerosol layers, which are mostly observable in summer in this region (Vlemmix et al., 2015; Li et al., 2013; Mendolia et al., 2013). Very recently, Wang et al. (2015b) analysed MAX-DOAS data over Wuxi city, area with high pollution adjoined to Shanghai. It is clearly shown that, under aerosol pollution, by using the modified cloud parameters in the collocated DOMINO products, tropospheric NO_2 AMFs are overestimated. This mostly happens when the effective cloud pressure value is larger than 900 hPa.

Kuhlmann et al. (2015) recalculated tropospheric NO_2 AMFs using high-resolution aerosol parameters over the Pearl River Delta region in southern China by the Models-3 Community Multiscale Air Quality (CMAQ) modelling system. Resulting tropospheric NO_2 VCDs increased by $+6.0 \pm 8.4\,\%$, likely because of polluted cases where the employed aerosol and NO_2 profiles show aerosol particles located higher in altitude compared to tropospheric NO_2. In addition, Lin et al. (2014) explicitly took into account aerosol optical effects from the Goddard Earth Observing System - Chemistry (GEOS-Chem) simulations where model τ is constrained by monthly MODIS Aqua τ data and validated by ground-based τ measurements. This study shows that excluding both aerosol scattering and absorption lead to changes between −40 and 90 % with AOD ≥ 0.8. Castellanos et al. (2015) have reduced the OMI NO_2 VCDs by 10 % in average by using aerosol extinction vertical profile observations from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument and AOD and SSA from the OMI/Aura level-2 near UV Aerosol (OMAERUV) database for scenes over South-America including absorbing biomass burning aerosols. According to the figures of CALIOP and collocated Tracer Model 4 (TM4) NO_2 profiles, the processed cases seem to include aerosol particles mixed (in parts) with the tropospheric NO_2 bulk. Finally, the new Peking University OMI NO_2 (POMINO) dataset which take aerosol properties from GEOS-Chem simulations, and are based on the reprocessing of all the DOMINO product, show in average a reduction of the tropospheric NO_2 VCDs by 0–40 % over most of China (Lin et al., 2015). However, it is mentioned that individual reductions or enhancements depend strongly on location and season, and thus on the occurence of the relative altitude between aerosol particles and tropospheric NO_2.

Overall, all these references which performed real retrievals show consistent numbers and conclusions with the sensitivity study performed here, and highlight the crucial role played by the actual OMI cloud algorithm and the derived implicit aerosol correction. This emphasizes that high aerosol pollution has currently large impacts on the individual OMI tropospheric NO_2 products over industrialised regions and cloud-free scenes.

2.5 Conclusions

In this paper, the behaviour of the OMI cloud model over cloud-free scenes dominated by aerosols was studied as well as the accuracy of the cloud-model based aerosol correction of tropospheric NO_2 AMFs. This study focused on the operational OMI DOMINO product for cloud-free scenes, its behavior in the presence of aerosol dominated scenes that were selected based on collocated MODIS Aqua aerosol products, and the comparison with numerical simulated study cases. The goals were to understand the behavior of the implicit aerosol correction based on the OMI cloud retrieval, and to investigate how much it improves the accuracy of the tropospheric NO_2 AMFs compared to performing no correction (and assuming clear sky conditions with no aerosols). Analyses relied on a model vs. observation approach and have specifically focused on the industrialized part of East China.

The OMI cloud algorithm cannot distinguish aerosol and cloud signals. Effective cloud parameters are retrieved over cloud-free scenes but including aerosol particles. This implies that these retrievals include considerable aerosol information (τ, optical properties, particles size, altitude) but they are treated as an opaque Lambertian reflector (albedo of 0.8). The effective cloud fraction linearly increases with increasing τ and can reach values between 0.1 and 0.15 for $\tau(550nm) = 1$. This represents the aerosol scattering effects on the 460–490 nm continuum reflectance. The slope of the linear regression of τ vs cloud is, however, dependent on the aerosol properties, the surface albedo and the SZA and VZA. The response of effective cloud pressure to aerosol scenes represents the O_2-O_2 shielding effect induced by the attenuation of photons by optically thicker aerosol layers, shortening the length of the average light path. In case of high aerosol pollution, retrieved effective cloud pressure values correlate with the mean aerosol layer height. Values smaller than the mean aerosol layer pressure may be related to the cloud model error over aerosol scenes. In cases of low τ or effective cloud fraction values, aerosols have little effect on O_2-O_2 absorption, leading to effective cloud pressure values close to the surface pressure independently of the altitude of the aerosol layer. This overestimation can be caused by the coarse sampling of the cloud LUT used by the OMI cloud algorithm to convert the O_2-O_2 continuum reflectance and slant column into effective cloud fraction and pressure. Indeed, this LUT was initially intended for retrievals over cloudy scenes, not for cloud-free scenes dominated by aerosols.

Aerosols can either decrease (shielding) or increase (enhancement) the sensitivity to tropospheric NO_2 bulk. Such effects simultaneously depend on the aerosols altitude and the shape of the NO_2 vertical profile. Shielding effects mostly occur when particles are above the NO_2 layers which should mostly happen during summer in China. Generally, if no aerosol correction was performed in the DOMINO products, relative biases of the tropospheric NO_2 VCDs would range from −60 to 20% for large τ values. These values could even decrease to −150 % in cases of large angles (e.g.

SZA ≥ 50°) or very large vertical separation between aerosols and the tropospheric NO_2 bulk.

An implicit aerosol correction is applied in the computation of the tropospheric NO_2 AMF through the use of the retrieved effective cloud fraction and pressure over scenes dominated by aerosols. After the implicit aerosol correction, relative biases in the VCDs are negative and in the range of −40 to −20% in case of elevated aerosol particles and high pollution ($\tau(550nm) \geq 0.6$). In case of aerosols located close to the surface or mixed with the tropospheric NO_2 bulk, relative biases in the VCDs are positive and in the range of 10 to 20%. These values are smaller than if no aerosol correction was applied in the OMI DOMINO products. τ and aerosols altitude are the key drivers of these biases, while aerosol microphysical properties are of secondary importance. Note that geometry angles and shape of the NO_2 profile can either increase or decrease these values. For elevated aerosols, the main cause of these biases is an underestimation of the aerosol shielding effect by the cloud algorithm. The reason of this underestimation is probably a combination of the cloud model error, used in presence of aerosols, and the employed numerical approach to convert the O_2-O_2 continuum reflectance and SCD into effective cloud fraction and pressure through a LUT. Furthermore, the coarse sampling in the employed cloud LUT leads to complex behaviors between these biases and τ. An improved LUT with a higher sampling should be implemented and evaluated. The impact on the ability to estimate the NO_x surface fluxes should be further studied. The biases in presence of aerosols located to the surface or mixed with tropospheric NO_2 are likely a consequence of difference between the opaque Lambertian cloud model and aerosol properties.

The present analyses considered box aerosol layers (i.e. discrete atmospheric layers with constant aerosol extinction value) and assumed that aerosols cover completely the OMI pixels. It is recognized that a more realistic description of the vertical distribution of aerosols and assumptions of OMI pixels partially covered by aerosols would result in different biases. Nevertheless, biases found here compare quite well to biases found in various ground-based comparison studies. Finally, all the sensitivity studies performed here assume that, in case of highly polluted regions, only non-explicit aerosol correction impact the current individual OMI tropospheric NO_2 products. It should be noted that uncertainties in the shape of vertical NO_2 profile and climatology surface albedo can also play a significant role on the estimated biases when aerosols are present in the measurement.

The results described in this paper indicate that it is worthwhile to design and evaluate an improved aerosol correction in view of retrieving tropospheric NO_2 vertical column densities. This is needed in the context of OMI measurements, but even more in the future for TROPOMI (Veefkind et al., 2012). Since the spatial resolution will be higher (7 km × 7 km), there is a significant probability that a scene will be fully covered by aerosol particles.

Chapter 3

Aerosol layer height from OMI: development of a neural network approach based on the 477 nm O_2-O_2 spectral band

> "We can only manage what we can measure."
>
> David Crisp and the Orbit Carbon observatory (OCO)-2 science team

Abstract

This paper presents an exploratory study on the aerosol layer height (ALH) retrieval from the OMI 477 nm O_2-O_2 spectral band. We have developed algorithms based on the multilayer perceptron (MLP) neural network (NN) approach and applied them to 3-year (2005–2007) OMI cloud-free scenes over north-east Asia, collocated with MODIS Aqua aerosol product. In addition to the importance of aerosol altitude for climate and air quality objectives, our long-term motivation is to evaluate the possibility of retrieving ALH for potential future improvements of trace gas retrievals (e.g. NO_2, HCHO, SO_2) from UV–visible air quality satellite measurements over scenes including high aerosol concentrations. This study presents a first step of this long-term objective and evaluates, from a statistic point of view, an ensemble of OMI ALH retrievals over a long time period of 3 years covering a large industrialized continental region. This ALH retrieval relies on the analysis of the O_2-O_2 slant column density (SCD) and requires an accurate knowledge of the aerosol optical thickness, τ. Using MODIS Aqua τ(550 nm) as a prior information, absolute seasonal differences between the LIdar climatology of vertical Aerosol Structure for space-based lidar simulation (LIVAS) and average OMI ALH, over scenes with MODIS $\tau(550\text{ nm}) \geq 1.0$, are in the range of 260–800 m (assuming single scattering albedo $\omega_0 = 0.95$) and 180–310 m (assuming $\omega_0 = 0.9$). OMI ALH retrieval inaccuracy can reach 660 m for aerosol single scattering albedo uncertainty of 0.05, and 200 m depending on the chosen surface albedo (between OMLER and MODIS black-sky albedo). Scenes with $\tau \leq 0.5$ are expected to show too large biases due to the little impact of particles on

---

The contents of this chapter are published as: **Chimot, J.**, Veefkind, J. P., Vlemmix, T., de Haan, J. F., Amiridis, V., Proestakis, E., Marinou, E., and Levelt, P. F.: An exploratory study on the aerosol height retrieval from OMI measurements of the 477 nm O_2-O_2 spectral band using a neural network approach, Atmos. Meas. Tech., 10, 783-809, https://doi.org/10.5194/amt-10-783-2017, 2017.

the O_2-O_2 SCD changes. In addition, NN algorithms also enable aerosol optical thickness retrieval by exploring the OMI reflectance in the continuum. Comparisons with collocated MODIS Aqua show agreements between -0.02 ± 0.45 and -0.18 ± 0.24, depending on the season. Improvements may be obtained from a better knowledge of the surface albedo and higher accuracy of the aerosol model. Following the previous work over ocean of Park et al. (2016), our study shows the first encouraging aerosol layer height retrieval results over land from satellite observations of the 477 nm O_2-O_2 absorption spectral band.

3.1 Introduction

The ability to monitor air quality and climate from ultraviolet–visible (UV–vis) satellite spectral measurements requires accurate trace gas (e.g. NO_2, SO_2, HCHO, O_3) and aerosol observations. Aerosols and trace gases often share similar anthropogenic sources, and their concentrations, as shown by the satellite observations, often exhibit significant correlations (Veefkind et al., 2011). The reason is that trace gases are often precursors for aerosols. The importance of measuring vertical distribution of atmospheric aerosols on a global scale is threefold. Firstly, aerosols directly impact the radiation budget of the Earth–atmosphere system through the scattering and absorption of solar and terrestrial radiation (Feingold et al., 1999). High concentrations of fine particles lead to reduced cloud droplet size, enhanced cloud reflectance (Twomey et al., 1984) and reduced precipitation (Rosenfeld, 2000; Ramanathan et al., 2001; Rosenfeld et al., 2002). Therefore, large uncertainties of aerosol optical properties limit our climate predictive capabilities (IPCC: Solomon et al., 2007). In spite of more robust climate predictions in the last years, radiative forcing (RF) induced by aerosols is still the largest uncertainty to the total RF estimate (IPCC, 2014). The vertical distribution and relative location are determining factors of aerosol radiative forcing in the long-wave spectral range (Dufresne et al., 2002; Kaufman et al., 2002).

Secondly, aerosols play a significant role in air quality, in particular near the surface. Due to the rapid growth of both population and economic activity, such as in Asia, the increase in fossil fuel emissions gives rise to concerns about fine particle formation and dispersion. Aerosols include a variety of hazardous organic and inorganic substances that reduce visibility, lead to reductions in crop productivity and strongly affect the health of inhabitants in urban regions (Chameides et al., 1999; Prospero, 1999; Eck et al., 2005).

Thirdly, slant column densities (SCDs) of trace gases, derived from UV–vis air quality space-borne sensors, have a high sensitivity to aerosol heights. For partly cloudy conditions, clouds are the main error source of trace gas measurements. But, in the absence of clouds, vertical distribution of aerosols, combined with their scattering and absorbing properties, modifies the length of the average light path of the detected photons and therefore affects trace gas air mass factors (AMFs). The application of AMFs is crucial for the conversion of SCDs from satellite line-of-sight measurements into vertical column densities. Then, aerosols strongly contribute to the uncertainties of trace gas retrievals from space-borne observations. For example, the magnitude of the error on the Ozone Monitoring Instrument (OMI) tropospheric NO_2 retrieval is, over polluted areas, mostly determined by the AMF uncertainty, not by the SCD uncertainty. It results from the combination of aerosols, clouds and the shape of the NO_2 profile (Boersma et al., 2011). Negative biases on OMI tropospheric NO_2 columns, between −26 and −50 %, are found in urban and very polluted areas in cases of high aerosol pollution and particles located at elevated altitude (Shaiganfar et al., 2011; Ma et al., 2013; Kanaya et al., 2014). HCHO AMF for GOME-2 and SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) shows about 20–50 % sensitivity to aerosols, depending whether they are located within or above the boundary layer (Barkley et al., 2012; Hewson et al., 2015). Dust aerosols (large particles, with strong absorption in UV) can reduce the AMF in the SO_2 wavelengths (310–330 nm) by half, thus doubling the retrieved SO_2 (Krotkov et al., 2008). This impacts the ability of sensors like OMI to monitor planetary boundary layer (PBL) SO_2 with a sensitivity to local anthropogenic sources. Over regions of enhanced columns, aerosols highly contribute to the total SO_2 AMF error (Lee et al., 2009). Therefore, aerosol parameters (or retrievals) are a prerequisite before retrieving trace gas vertical column densities.

State-of-the-art trace gas retrieval algorithms correct for aerosol effects either explicitly using modelled aerosol vertical profiles (e.g. Barkley et al., 2012, 2013; Kuhlmann et al., 2015; Lin et al., 2014, 2015) or implicitly via cloud algorithms. For example, the OMI O_2-O_2 absorption band at 477 nm has been widely exploited to derive cloud information (Acarreta et al., 2004; Sneep et al., 2008). However, the OMI cloud algorithm is sensitive to aerosols, and thus the retrieved effective cloud parameters are modified in their presence (Boersma et al., 2007; Castellanos et al., 2015; Chimot et al., 2016). The OMI O_2-O_2 spectral band at 477 nm contains significant information on aerosol properties and height. The retrieved effective clouds are then used to correct the computed AMF (de Smedt et al., 2008; Boersma et al., 2011). In spite of these well-considered perturbations, the use of the effective cloud parameters, assuming that the opaque Lambertian cloud model can reproduce the distribution of scattering fine particle effects, does not yet completely correct for the aerosol effects when computing the AMF, in particular for the tropospheric NO_2 columns (Castellanos et al., 2015; Chimot et al., 2016).

Characterizing the aerosol vertical distribution, in addition to the associated optical properties, using passive space-borne measurements is challenging due to the absence of spectral features in the aerosol optical properties and the combined influences of surface and cloud reflection. Contrary to effective cloud retrievals, aerosol retrieval is a more complex problem mainly because of the variability of particle microphysical properties and the lower optical thickness (typically 1–2 orders of magnitude). As a consequence, methods assuming large multiple scattering contributions, such as a simple cloud model with Lambertian properties, cannot be used. Passive radiometers like the Moderate Resolution Imaging Spectroradiometer (MODIS) can only retrieve a limited amount of independent information from their measurements, usually aerosol optical thickness, τ, and the extinction Ångström exponent, α, as a proxy for the particle size distribution (Levy et al., 2007, 2013). The near-UV technique has been widely used to map the daily global distribution of UV-absorbing aerosols such as desert dust particles as well as carbonaceous aerosols generated by anthropogenic biomass burning and wildfires. It allows us to retrieve τ, single scattering albedo ω_0 and the qualitative aerosol absorbing index (AAI) in the 330–388 nm of the Total Ozone Mapping Spectrometer (TOMS) and OMI sensors (Torres et al., 1998; Torres et al., 2002; Torres et al., 2007). However, this technique is highly affected by the dependency of the measured radiances on the height of the absorbing aerosol layer (Torres et al., 1998; de Graaf et al., 2005). OMAERUV has been upgraded by integrating a monthly climatology of CALIOP aerosol heights to retrieve aerosol parameters from OMI UV measurements (Torres et al., 2013). The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) has been providing vertical profiles of aerosols but with limited spatial coverage because of its measurements characteristics (Omar et al., 2009). Park et al. (2016) evaluated the sensitivity of the O_2-O_2 slant column density to changes in aerosol layer height (ALH) over ocean. It is demonstrated that the O_2-O_2 spectral band at 477 nm is the most sensitive to the aerosol layer effective height (compared to the O_2-O_2 absorption bands at 340, 360 and 380 nm) due to the largest O_2-O_2 absorption and reduced Rayleigh scattering. Veihelmann et al. (2007) determined that the complete OMI UV–vis reflectance measurements contain between 2 and 4 degrees of freedom of signal (DFS) about aerosol parameters. The 477 nm O_2-O_2 band adds by itself about 1 degree and therefore contains more information than any other individual band. This relative large number of DFS for UV–vis satellite solar backscatter observations is explained by the sensitivity of the reflectance to the ALH. Detailed O_2-O_2 radiative transfer simulations performed by Dirksen et al. (2009) revealed the availability of the altitude information about smoke aerosol plume, released by intense forest fires and transported over long distance, under specific conditions: high AAI and no clouds. In spite of all these efforts, no aerosol height retrieval has been done at this moment from O_2-O_2 satellite measurements at 477 nm over land.

Since aerosol altitude, in addition to τ, is one of the key parameters affecting the computation of AMF for trace gases retrievals such as NO_2 (Leitão et al., 2010; Chimot et al., 2016), our long-term motivation is to evaluate the capability of retrieving it from the satellite O_2-O_2 absorption band at 477 nm. This exploratory study is the first step and statistically analyses an ensemble of OMI observations over a 3-year period (from 2005 to 2007) and covering a large industrialized continental region (i.e. north-east Asia). This study follows the conclusions of previous works focused on the sensitivity of this spectral band and the observed links between the O_2-O_2 effective cloud retrievals and aerosol parameters. In this paper, quite a few algorithm concepts are developed, based on the neural network (NN) approach, and then tested on a high number of OMI observations over land. Our primary focus is the retrieval performance of aerosol layer pressure (ALP) associated with scattering and fine particles over large urban, industrialized and highly polluted area and cloud-free scenes. In addition, the sensitivity of the algorithms to τ knowledge is investigated and, therefore, the capability of τ retrievals from the same OMI band is evaluated. The considered satellite observations and input data set are described in Sect. 3.2.

Section 3.3 focuses on the available OMI O_2-O_2 differential optical absorption spectroscopy (DOAS) parameters and their link with ALH and τ. The development of the different NN algorithms is described in Sect. 3.4. Their performances are evaluated in Sect. 3.5 on a synthetic and independent data set with a characterization of the main limiting factors. Finally, these algorithms are applied in Sect. 3.6 to cloud-free OMI observation over north-east Asia, where large amounts of aerosols are emitted from both natural and anthropogenic sources (Lee et al., 2012). They are then compared with other observation products, namely MODIS Aqua τ and the LIdar climatology of vertical Aerosol Structure for space-based lidar simulation (LIVAS).

3.2 Aerosol and surface albedo satellite data

In this section, we describe the three main aerosol satellite data sets that are used in this study: OMI visible measurements, MODIS aerosol product and the LIVAS climatology database. In addition, the two considered surface albedo databases: OMI Lambert equivalent spectral surface reflectance (OMLER) and MODIS black-sky albedo are also detailed.

3.2.1 OMI satellite data

The Dutch–Finnish mission OMI (Levelt et al., 2006) is a nadir-viewing push-broom imaging spectrometer launched on the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Aura satellite. It provides daily global coverage of key air quality components through observations of the backscattered solar radiation that are captured in the UV–vis spectral domain. Based on a two-dimensional detector array concept, radiance spectra are simultaneously measured on a 2600 km wide swath within a nadir pixel size of 13x24 km^2 (28x150 km^2 at extreme off-nadir). OMI has a higher spatial resolution than any other UV–vis hyperspectral spectrometers. It measures in the wavelength range of 270 to 500 nm with a spectral resolution of 0.45 in the UV-2 band (310–360 nm) and 0.63 nm in the visible band (360–500 nm).

Starting mid-2007, the so-called “row anomaly” has been perturbing OMI measurements of the Earthshine radiance at all the wavelengths. Details are given at `http://www.knmi.nl/omi/research/product/rowanomaly-background.php`. For practical reasons, this study only used the OMI data acquired during 2005–2007, i.e. before the development of this anomaly.

OMI has not been optimized for aerosol monitoring. However, the OMI near-UV aerosol algorithm (OMAERUV) independently retrieves atmospheric total columns of τ and ω_0 from two UV wavelengths, 354 and 388 nm (Torres et al., 2007, 2013). In comparison to 44 Aerosol Robotics Network (AERONET) sites, evaluated OMAERUV τ yields a root mean square error (RMSE) of 0.16 and a correlation coefficient of 0.81 over the years 2005–2008 (Ahn et al., 2014). About 65 % of these retrievals lie within the expected uncertainty. The OMAERUV ω_0 product agrees with AERONET to within 0.03 in 46,% of the collocated pairs and to within 0.05 in 69 % of the cases

(Jethva et al., 2014).

The OMI O_2-O_2 477 nm absorption band is currently operationally exploited by the OMI O_2-O_2 cloud algorithm to derive effective cloud fraction and pressure (Acarreta et al., 2004; Veefkind et al., 2016). Park et al. (2016) applied a look-up table (LUT) approach on this band to retrieve aerosol effective height over ocean, close to East Asia, within the error range of 1 km (compared to CALIOP). This approach was applied to seven case studies, each of them covering a few days. No τ was retrieved. No study has yet explicitly used this satellite band to directly retrieve ALH and τ over land. This band is available not only on OMI but also on various sensors such GOME-2, OMPS and the next space-borne TROPOspheric Monitoring Instrument (TROPOMI).

3.2.2 MODIS aerosol product

The MODIS instrument, launched on the NASA EOS Aqua platform in May 2002, is a spectrometer delivering continuous images of the Earth in the visible, solar and thermal infrared approximately 15 min prior to OMI on board EOS Aura. The considered MODIS Aqua Level 2 (L2) aerosol product is the collection 6 of MYD04_L2, based on the Dark Target (DT) land algorithm with a high enough quality flag (Xiao et al., 2016). While the MODIS measurement is acquired at the resolution of 1 km, the MODIS aerosol product is available at both 3 km x 3 km and 10 km x 10 km. Since this last one is relatively close to the OMI nadir spatial resolution, it is then used in the work below (see Sect. 3.6). The improved calibration of MODIS Aqua instrument is included in the reprocessing of the collection 6 aerosol product (Levy et al., 2013; Lyapustin et al., 2014).

The availability of the MODIS aerosol products is generally a good confirmation of cloud-free scenes as MODIS Aqua τ variable is exclusively given provided a high amount of cloud-free sub-pixels is available (i.e. the MODIS measurement resolution of 1 km).

The expected error of MODIS DT τ is about $\pm 0.05 + 15\,\%$ over land (Levy et al., 2013). The Deep Blue retrieval algorithm has been developed to complement the DT algorithm by retrieving τ over bright arid land surfaces (e.g. deserts). The typical associated uncertainties are about ±0.03 on average (Sayer et al., 2013).

3.2.3 LIVAS climatology database

LIVAS is a 3-D multi-wavelength global aerosol and cloud optical database (Amiridis et al., 2015). This database provides averaged profiles of aerosol optical properties over 9 years (1 January 2007–31 December 2015) from the Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) data on a uniform grid of 1° x 1°. LIVAS addresses the wavelength dependency of aerosol properties for many laser operating wavelengths including 532 nm. The LIVAS data set has been evaluated against AERONET in Amiridis et al. (2015), showing realistic and representative mean state aerosol optical depth values in 532 nm and making this data set ideal for synergistic use with other satellite products.

The LIVAS ALH is derived from the given averaged vertical profile of aerosol extinction (532 nm) $\sigma(l)$ over each vertical layer l defined by its altitude $h(l)$ as follows:

$$\mathrm{ALH(LIVAS)} = \frac{\sum_{\mathrm{l}} h(l)\sigma(l)}{\sum_{\mathrm{l}} \sigma(l)}. \tag{3.1}$$

Since LIVAS also provides the standard deviation associated with each averaged vertical profile of aerosol extinction (532 nm) $\partial\sigma(l)$, the equivalent standard deviation ∂ALH of each LIVAS ALH is derived as follows:

$$\partial\mathrm{ALH(LIVAS)} = \frac{\sum_{\mathrm{l}} \partial h(l)\partial\sigma(l)}{\sum_{\mathrm{l}} \sigma(l)}. \tag{3.2}$$

where $\partial h(l)$ is the geometric thickness of each vertical layer l.

3.2.4 Surface albedo data set

The standard and reference surface albedo product is the OMLER climatology derived from several years of OMI observations at the spatial resolution of 0.5° x 0.5° longitude–latitude grid for each calendar month (Kleipool et al., 2008). The OMLER algorithm is based on temporal histograms of the observed Lambert equivalent spectral surface reflectance (LER) values per grid box. Potential small residual cloud and aerosol contaminations are expected to remain in the OMLER product. As an alternative, the global and spatially complete MODIS black-sky surface albedo in the band 3 (459–479 nm) is considered. It is defined as the directional hemispherical reflectance and is a function of solar zenith angle θ_0 (Schaaf et al., 2002). It is derived by integrating the atmospheric corrected bidirectional reflectance distribution function (BRDF), derived from combined MODIS Aqua and Terra observations over every 16-day period. The downwelling flux in the MODIS black-sky albedo has no diffuse component. Collection 6 of MCD43C3 product is given on a 0.05° (5.6 km) latitude–longitude climate modelling grid (CMG). Note that Kleipool et al. (2008) demonstrated that the OMLER data set is closer to the black sky than to the white sky by evaluating the ratio between diffuse and direct illumination.

3.3 OMI O_2-O_2 DOAS analysis and aerosols

3.3.1 DOAS analysis of the OMI O_2-O_2 477 nm absorption band

In this paper, the aerosol NN retrieval algorithms allow the conversion of the continuum reflectance $R_c(475nm)$ and the O_2-O_2 SCD $N^s_{O_2-O_2}$ into $\tau(550\,\text{nm})$ and ALP (in hPa). As a consequence, the NN retrievals rely on the way that the aerosol parameters modify these two variables and thus the photons average light path.

Prior to this conversion, a spectral DOAS fit must be performed to derive $R_c(475nm)$ and $N^s_{O_2-O_2}$ from the OMI O_2-O_2 477 nm absorption band. The various DOAS techniques rely on the same key concept: a simultaneous fit of several trace gas slant column densities from the fine spectral features due to their absorption (i.e. the high frequency part) present in passive UV–vis spectral measurements of atmospheric radiation (Platt and Stutz, 2008). Here, the DOAS fit follows the same approach as in the OMI O_2-O_2 cloud algorithm (Acarreta et al., 2004; Veefkind et al., 2016): i.e. the absorption cross-section spectrum of O_2-O_2 is fitted together with a first-order polynomial:

$$\begin{aligned} -\ln(R(\lambda)) = \gamma_1 + \gamma_2 \cdot \lambda + N^s_{O_2-O_2}(\lambda) \cdot \sigma_{O_2-O_2} \\ + N^s_{O_3}(\lambda) \cdot \sigma_{O_3}, \end{aligned} \tag{3.3}$$

where $\gamma_1 + \gamma_2 x \lambda$ defines the first-order polynomial, $\sigma_{O_2-O_2}$ and σ_{O_3} are the O_2-O_2 and O_3 absorption cross-section spectra, respectively, convoluted with the OMI slit function and $N^s_{O_3}$ is the O_3 slant column density. $\sigma_{O_2-O_2}$ is based on measurements of the cross section made by C. Hermans (see `http://www.aeronomie.be/spectrolab/o2.htm` – file O4.txt). The O_3 cross-section spectrum is included because it overlaps with the O_2-O_2 spectrum. The fitted parameters are γ_1, γ_2, $N^s_{O_2-O_2}$, and $N^s_{O_3}$. In the absence of absorbers, one may define the continuum reflectance R_c at the reference wavelength λ_0:

$$R_c = \exp(-\gamma_1 - \gamma_2 \cdot \lambda_0). \tag{3.4}$$

The reference wavelength is specified as the middle of the DOAS fit window at $\lambda_0 = 475$ nm.

3.3.2 On the impact of aerosols on R_c and O_2-O_2 SCD

Figure 3.1 illustrates how aerosol particles directly drive the OMI O_2-O_2 DOAS parameters at 477 nm assuming cloud-free space-borne observations. These effects are obtained from radiative transfer simulations including aerosols and no clouds. The detailed generation of such simulations is given in Sect. 3.4.2. The DOAS fit equations following Eqs. (3.3) and (3.4) are then applied to these simulations. In this paper, the aerosol layer is assumed to be one single scattering layer (i.e. "box layer") with a constant geometric thickness (about 1 km). All the particles included in this layer are supposed to be homogeneous (i.e. same size and optical properties). ALH is then expressed by ALP, in hPa, defined as the mid-pressure of this scattering layer.

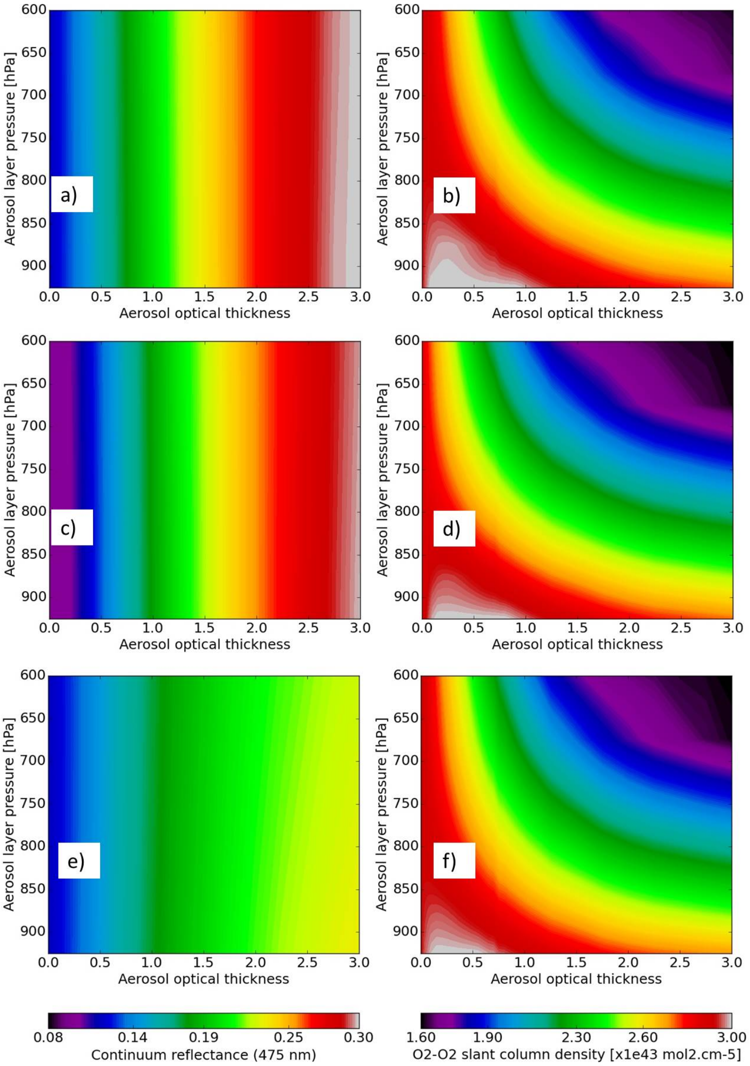


Figure 3.1: Continuum reflectance R_c(475 nm) and O_2-O_2 slant column density $N^s_{O_2-O_2}$ as a function of $\tau(550nm)$ and aerosol layer pressure for the following conditions: climatology mid-latitude summer temperature, NO_2, O_3 and H_2O profiles, $\theta_0 = 32°$, $\theta = 32°$, surface pressure of 1010 hPa and fine aerosol particles ($\alpha = 1.5$, $g = 0.7$): **(a, b)** surface albedo of 0.07 and aerosol ω_0 of 0.95; **(c, d)** surface albedo of 0.03 and aerosol ω_0 of 0.95; **(e, f)** surface albedo of 0.07 and aerosol ω_0 of 0.9.

Qualitatively, aerosols have two separate effects on the average light path and therefore on the O_2-O_2 absorption signal at the top of the atmosphere (TOA). These two effects are similar to that of aerosols and clouds on NO_2 absorption signal (Leitão et al., 2010; Chimot et al., 2016): (1) a shielding effect, i.e. a decreased sensitivity within and below the aerosol layer due to a reduced amount of photons coming from the TOA and reaching the lowest part of the atmosphere compared to an aerosol-free scene; (2) an enhancement (albedo) effect, i.e. an increased sensitivity within and above the aerosol layer as more photons are scattered back towards the sensors, and thus a larger fraction of detected photons that samples the part of the atmosphere above the aerosol layer. Shielding then leads to a reduced O_2-O_2 absorption while enhancement may increase the O_2-O_2 absorption, especially for low cloud or aerosol layers. The overall effect (enhancement vs. shielding) depends on the aerosol optical properties, the total column τ and ALP.

OMI R_c(475 nm) is directly and primarily affected by the total column τ of particles present in the observed scene. Indeed, R_c increases with increasing τ independently of the ALP (see Fig. 3.1a). This mostly results from the influence of aerosols on the number of detected photons and on the additional scattering effects observed in the scene compared to an aerosol-free scene. However, the magnitude of this increase relies on aerosol optical properties and the surface brightness. As a consequence, R_c is also affected by aerosol ω_0, phase function and the surface albedo A. Indeed, R_c decreases with decreasing ω_0 and over a darker surface (i.e. smaller A value) for all the τ values (see Fig. 3.1c and e). The importance of these parameters is further discussed in Sects. 5 and 6. Note that, in addition, the reflectance is also driven by the geometry angles: i.e. viewing zenith angles θ, θ_0 and relative azimuth angle defined as the difference between viewing and zenith azimuth angles $\phi - \phi_0$. An increase of θ or θ_0 will lead to longer average light path and thus will amplify aerosol related additional scattering effects (for a given τ).

OMI $N^s_{O_2-O_2}$ relies on the O_2-O_2 absorption magnitude along the average light path in the whole atmosphere. It is driven by the overall shielding or enhancement effect of photons by the O_2-O_2 complex in the visible spectral range due to the presence of particles. As depicted in Fig. 3.1b, $N^s_{O_2-O_2}$ decreases with decreasing ALP. This is a direct consequence of a larger shielding effect applied by aerosols located at higher altitudes (i.e. part of the O_2-O_2 complex located below the aerosol layers are shielded). Nevertheless, in the case of low τ values (i.e. ≤ 0.5), $N^s_{O_2-O_2}$ does not significantly vary with respect to ALP. This shows that a low amount of aerosols has very little impact on O_2-O_2 absorption measurements.

However, as depicted in Fig. 3.1b, d and f, not only ALP but also τ directly influences the slant O_2-O_2 absorption since both parameters simultaneously affect the average path followed by the photons, and therefore the overall shielding or enhancement effect. At a given altitude, an increase of τ leads to a decrease of $N^s_{O_2-O_2}$. The slope of this decrease depends on the aerosol altitude (i.e. higher for particles at high altitude). Note that both ω_0 and A also affect $N^s_{O_2-O_2}$, but this effect is smaller than τ. For example, a reduced ω_0 and A lead to a small decrease of $N^s_{O_2-O_2}$ (see Fig. 3.1d and f).

As a consequence, (1) the single parameter $N^s_{O_2-O_2}$ contains information on both τ and ALP. These parameters cannot be separated from this unique variable alone.

Therefore, if τ is not accurately known, there will likely be an ambiguity when analysing $N^{s}_{O_2-O_2}$ to retrieve ALP. (2) if an external or prior τ estimate is not available, then the two parameters $N^{s}_{O_2-O_2}$ and $R_c(475\,\text{nm})$ could be simultaneously and directly combined to retrieve ALP provided that one can accurately and independently retrieve τ from $R_c(475\,\text{nm})$. Then, in that condition, OMI $R_c(475\text{ nm})$ may help to distinguish both τ and ALP contributions in $N^{s}_{O_2-O_2}$. However, the simultaneous effects of aerosol ω_0 and A on R_c (as discussed above), and therefore their associated uncertainties, will impact the feasibility of retrieving τ from OMI measurements. It may then degrade the retrieved ALP performances. (3) τ, R_c and $N^{s}_{O_2-O_2}$ have a non-negligible correlation. Indeed, an increase of τ results in a simultaneous increase of R_c and $N^{s}_{O_2-O_2}$. Therefore, it has to be noted that these two last parameters are not independent and combining them does not provide two independent pieces of information.

Overall, the impact of aerosol particles on the OMI O_2-O_2 spectral band is similar to cloud particles. This explains in part the difficulty to distinguish aerosols from clouds. In cases with a mix of aerosols and clouds, there is an ambiguity between R_c, τ and the OMI effective cloud fraction on the one hand and $N^{s}_{O_2-O_2}$, ALP, τ, the OMI effective cloud pressure and fraction on the other hand (Boersma et al., 2011; Castellanos et al., 2015; Chimot et al., 2016). Therefore, this study only focuses on cloud-free reflectance to avoid this complexity.

3.4 Design of the neural network retrieval algorithms

The retrieval algorithms developed for this paper are based on the PyBrain software (Schaul et al., 2010). PyBrain is a versatile machine learning library written in Python designed to facilitate both the application of and research on first class of learning algorithms such as recurrent NNs. It includes several functions such as supervised learning algorithms, feed forward network design and error back-propagation computations. Only the main developments specific to the present study are explained and discussed in the next subsections. For more details related to the PyBrain specificities, the reader is encouraged to read Schaul et al. (2010).

The multilayer perceptron (MLP) neural networks have been widely used and acknowledged for decades in the field of remote sensing (Atkinson and Tatnall, 1997). Indeed, most retrieval problems in this field are ill posed and non-linear. Thus, the associated inverse problems can only be addressed by including a priori information and relying on statistical analysis. Since aerosol retrieval from passive spectral measurements is well known as a non-linear inverse problem, the MLP technique represents a powerful approach to design a retrieval algorithm in a fast and robust way. The basic idea is to build an optimal interpolator system to make the link between OMI 477 nm O_2-O_2 measurements and the retrieved ALP and $\tau(550\,\text{nm})$ (see Fig. 3.1). However, knowledge must be acquired by the NNs by means of a supervision database. The following sections summarize then the design of the developed algorithms (see Sect. 3.4.1), the generated supervision database (see Sect. 3.4.2) and the employed learning process (Sect. 3.4.3).

3.4.1 MLP network approach: application to the OMI O_2-O_2 aerosol retrievals

Artificial NNs are a family of models related to the machine learning and the artificial intelligence domain (Luger and Stubblefield, 1998). They are used to reduce the number of calculations of functions requiring a large number of inputs and being generally unknown (or not well defined). The idea is to approximate them by parameterized and more simple functions. Input and output signals are then interconnected by a set of activation functions and a set of weights associated with each of them (Luger and Stubblefield, 1998). In the context of this work, no invertible analytical function exists that describes the dependence of slant columns and continuum reflectances on aerosols (see Sect. 3.2.3). Thus, the idea of developing neural networks here is to identify input–output relationships directly from a well-known training ensemble. The choice of an NN approach relies on its advantages compared to more conventional methods such as linear regression, linear interpolation in a LUT or the optimal estimation method (OEM). In particular, it enables (1) very fast computations with modern computers in spite of the number of required parameters; (2) optimized interpolation technique even in the case of non-linear statistical modelling and so, potentially, lower systematic biases compared to a linear interpolation; and (3) reduced memory use compared to a LUT with a very high sampling.

As illustrated in Fig. 3.2, the designed NNs rely on a multi-layer architecture, based on the MLP technique, composed of parallel processors (i.e. neurons) organized in distinct layers (Rumelhart et al., 1986). Such an architecture allows us to separate non-linear data and generally consists of three (or more) types of layers. The first layer includes all the required input variables. The last layer includes all the desired output data (or here retrievals). The intermediate layers are usually referred as hidden layers and contain the activation functions. All these layers are connected via neural links: two nodes or neurons i and j between two consecutive layers have synaptic connections associated with a synaptic weight ω_{ij}. Each neuron j computes a weighted sum of its N x_i information sent from the neurons of the previous layer (see Eq. 3.6). Then, this weighted sum is transported through a non-linear mathematical function before being passed to the next layer (see Eqs. 3.5 and 3.6). Here use is made of the classical sigmoid function:

$$\sigma(x) = \frac{1}{1+\exp(-x)}. \tag{3.5}$$

The output z_j of the neuron j in the hidden layer is thus given by

$$z_j = \sigma\left(\sum_{i=1}^{N} \omega_{ij} \cdot x_i\right). \tag{3.6}$$

The overall set W of synaptic weights ω_{ij} contains all the information about the

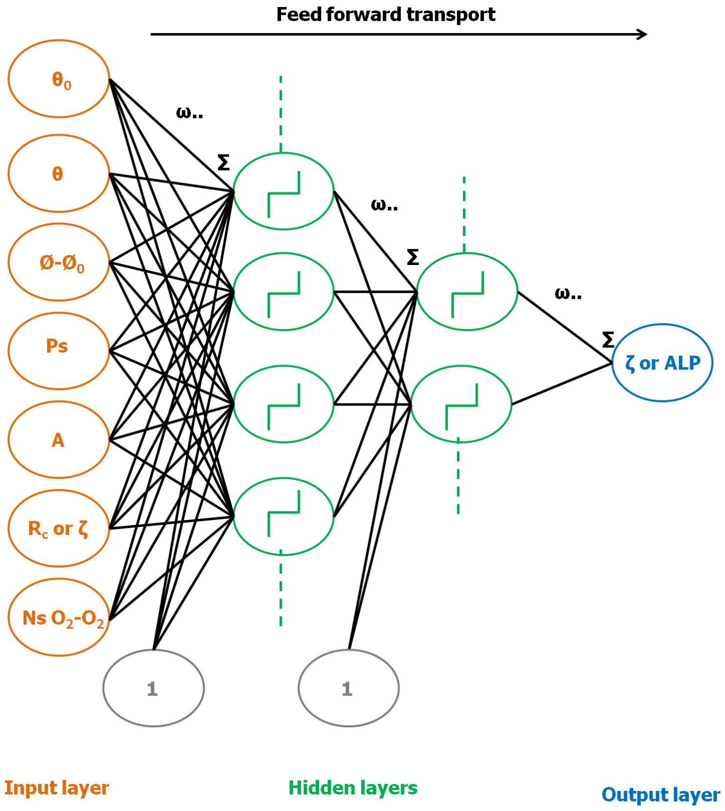


Figure 3.2: Diagram of multilayer perceptron (MLP) neural network (NN) architecture designed for aerosol layer pressure (ALP) and aerosol optical thickness τ retrieval algorithms from the OMI O_2-O_2 spectral band at 477 nm. The input parameters are based on the list given in Table 3.1. The different considered approaches for the MLP design and their applications are more detailed in Sect. 3.3. Each circle represents a specific processor (named neuron) including either an input–output variable (in the input–output layer) or the activation function (i.e. sigmoid function in the hidden layer). The synaptic weights ω ensure the connections of neurons between two consecutive layers. A weighted sum $\sum$ is performed before the transport through the activation function. Note the presence of the bias neurons, prior to the activation functions in the hidden layers. For simplicity, bias neurons are commonly visualized as values added to each neuron in the input and hidden layers of a network, but in practice are treated in exactly the same manner as other weights: all biases are simply weights associated with vectors that lead from a single node whose location is outside of the main network and whose activation is always 1. While the synaptic weights essentially change the steepness of the activation functions, the bias neurons allow to modify the origin of these functions from 0 to positive or negative values.

network (i.e. its neural architecture defined by a specified number of layers, neurons and connections). When the transport reaches the output layer, it forms the network output.

The chosen neural architecture is the following (see Fig. 3.2). The input layer is composed of seven parameters that include (1) θ, θ_0, $\phi - \phi_0$, surface pressure P_s, surface albedo A, and (2) either R_c(475 nm) and $N^s_{O_2-O_2}$ for τ(550 nm) and ALP retrieval (named $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$) or $\tau(550nm)$ and $N^s_{O_2-O_2}$ for ALP retrieval (named $\mathrm{NN}_{\tau, N^s_{O_2-O_2}}$).

The output layer is, for each NN retrieval algorithm, composed of only one output variable: either τ(550 nm) or ALP. In total, three NN retrieval algorithm configurations are then selected and used at the end: $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ for τ(550 nm) retrieval and $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ and $\mathrm{NN}_{\tau, N^s_{O_2-O_2}}$ for ALP retrieval.

The choice to use either $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ or $\mathrm{NN}_{\tau, N^s_{O_2-O_2}}$ will impact the accuracy of the ALP retrieval results (see Sects. 5.2 and 5.3).

3.4.2 Generation of the supervision database: aerosol properties and simulations

The MLP neural networks must be accurately trained from a well-known data set. They are then able to generalize the inverse problem by predicting the aerosol retrievals from input observations that have never been seen before. For that purpose, a learning database must be carefully designed and generated. It must be representative of the entire distribution of (input–output) values that can likely be encountered in the OMI observations. As a consequence for the MLP algorithms, a large quantity of data are often required for the learning process. However, very large learning data set can be extremely time-consuming in terms of generation and then NN training.

Training a neural network based on a large ensemble of synthetic data set has been widely employed in atmospheric retrieval science such as for CO_2 and CH_4 (Crevoisier et al., 2009a,b), aerosol (Di Noia et al., 2015) and cloud properties (Loyola, 2004; Loyola et al., 2007, 2010). This study created our own training data set based on simulations from the Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval (DISAMAR) software of KNMI (de Haan, 2011). DISAMAR includes a radiative transfer model and different retrieval methods. The radiative transfer model is based on the Doubling Adding KNMI (DAK) model (de Haan et al., 1987; Stammes, 2001) and thus computes the reflectance and transmittance in the atmosphere using the adding/doubling method. This method calculates the internal radiation field in the atmosphere at levels to be specified by the user and takes into account Rayleigh, aerosol and cloud scattering and trace gas and aerosol absorption. Scattering by aerosols is simulated with a Henyey–Greenstein scattering phase function $\Phi(\Theta)$ (Hovenier and Hage, 1989):

$$\Phi(\Theta) = \frac{1 - g^2}{(1 + g^2 - 2g\cos\Theta)^{3/2}}, \tag{3.7}$$

where Θ is the scattering angle. The phase function is then parameterized by the asymmetry parameter g, which is the average of the cosine of the scattering angle. It can vary between -1 and 1, from back-scattering through isotropic scattering to forward scattering. Following the DISAMAR configuration, τ values in the simulations are specified at the reference wavelength of 550 nm. The Ångström exponent α describes the spectral dependence of τ.

ALP is the main target parameter since this is one of the main parameters describing the average light path distribution in the tropospheric NO_2 AMF computation. The second target is $\tau(550\ \mathrm{nm})$ since this information may be requested for a good ALP retrieval quality. We thus assume that we do not need at this level to define more realistic aerosol models for every aerosol scene. With a reference asymmetry parameter of $g = 0.7$, the intermediate value typically observed (Dubovik et al., 2002), the Henyey–Greenstein function is known to be smooth and reproduce the Mie scattering functions reasonably well for most of aerosol types. This approach is also used for the preparation of the operational aerosol layer height retrieval algorithm from Sentinel-5 Precursor (Sanders et al., 2015) and for explicit aerosol corrections in the AMF calculation when retrieving trace gases such as tropospheric NO_2 (Spada et al., 2006; Wagner et al., 2007; Castellanos et al., 2015).

The ensemble of parameters and associated values used for generating the learning database is detailed in Table 3.1. About 460 000 spectral simulations, over the O_2-O_2 spectral band (460–490 nm), were generated, assuming different satellite viewing and solar geometries, A, P_s and aerosol pollution levels. Scenes with too large angles (i.e. $\theta_0 \geq 65°$) and too-bright surfaces (i.e. $A > 0.1$) are excluded. For each of these simulations, $R_c(475\,\mathrm{nm})$ and $N^s_{O_2-O_2}$ were deduced from the DOAS fit equations Eqs. (3.3) and (3.4). Aerosols are specified for a standard case, assuming fine particles with a unique value of $\alpha = 1.5$ and $g = 0.7$. Aerosol profiles are parameterized by scattering layers with constant aerosol volume extinction coefficient and ω_0 and with a fixed pressure thickness. In order to investigate the assumptions related to the single scattering albedo properties ω_0, two typical values are considered: $\omega_0 = 0.95$ and 0.9. Contrary to the other variables, ω_0 is not known for each OMI pixel and thus cannot be used as an explicit input parameter by the designed NNs. Moreover, it cannot be retrieved from this band since it is supposed to affect R_c(475 nm) and $N^s_{O_2-O_2}$, similarly to τ. Therefore, two sets of NN models are developed for different purposes:

- one set of three MLP NN algorithms ($NN_{R_c,N^s_{O_2-O_2}}$ for $\tau(550\,\mathrm{nm})$ retrieval, $NN_{R_c,N^s_{O_2-O_2}}$ and $NN_{\tau,N^s_{O_2-O_2}}$ for ALP retrieval) is trained with a learning database including aerosol spectral simulations assuming $\omega_0 = 0.95$;
- one set of three MLP NN algorithms ($NN_{R_c,N^s_{O_2-O_2}}$ for $\tau(550\,\mathrm{nm})$ retrieval, $NN_{R_c,N^s_{O_2-O_2}}$ and $NN_{\tau,N^s_{O_2-O_2}}$ for ALP retrieval) is trained with a learning database including aerosol spectral simulations assuming $\omega_0 = 0.9$.

The choice to use one of these sets will impact the accuracy of the retrieval results.

3.4.3 Optimization of the learning process and selection of the best NN architecture

Prediction of the optimal NN architecture (i.e. number of neurons and hidden layers) is generally not possible as these values are strongly specific to the given problem

Table 3.1: Ensemble of parameters and values associated with the synthetic learning data set (see Sect. 3.3.2). Aerosols are simulated with a Henyey–Greenstein scattering phase function (Hovenier and Hage, 1989).

Parameter	List of values
Solar zenith angle (θ_0) (°)	9.267, 21.167, 32.892, 44.217, 54.940, 64.814
Viewing zenith angle (θ) (°)	0.0, 9.267, 21.167, 32.892, 44.217
Relative azimuth angle ($\phi - \phi_0$) (°)	0., 30., 60., 90., 120., 150., 180.
Surface pressure (P_s) (hPa)	1013., 963.
Surface albedo (A)	0.025, 0.05, 0.075, 0.1
τ(550 nm)	0.0, 0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.9, 1.25, 2.0, 3.0
Aerosol layer pressure (ALP) (hPa)	975., 925., 850., 750., 700., 650., 550., 350., 150.
Aerosol single scattering albedo (ω_0)	0.9, 0.95
Ångström coefficient (α)	1.5
Asymmetry parameter (g)	0.7

(Atkinson and Tatnall, 1997). While it has been theoretically proven that a single-hidden-layer MLP network with non-linear activation functions may represent any non-linear continuous function (Haykin, 1999), a two-hidden-layer MLP may approximate any function to any degree of non-linearity taking also into account discontinuities (Sontag, 1992). To identify the best NN architecture for each aerosol retrieval parameter (τ and ALP) and for each configuration ($\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ or $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$; see Sect. 3.3.1), several architectures are trained and then evaluated: one single hidden layer with a variable number of neurons (between 9 and 70) and two hidden layers (between 15 and 70 neurons on the first layer and between 10 and 40 on the second layer). Then, the optimal NN architecture is selected based on the best computed evaluation score. In total, about 96 different MLP architectures, for each configuration, were evaluated.

For one given NN architecture, the training process is the optimization technique that estimates the optimal network parameters W of synaptic weights σ_{ij} (see Sect. 3.3.1). For that purpose, an error function E must be minimized. This error function measures, for a set of p representative situations for which inputs and outputs (i.e. τ(550 nm) and ALP) are known, the mismatch between the neural network outputs o_i and the true outputs t_i as follows:

$$E = \frac{1}{2}\sum_{i=1}^{\mathrm{p}}(\mathrm{o_i} - \mathrm{t_i})^2. \tag{3.8}$$

This error function minimization follows here the error back-propagation algorithm as specified by Rumelhart et al. (1986). It is a stochastic steepest descent algorithm well adapted to the MLP hierarchical architecture. The learning step is made sample by sample, iteratively and stochastically selected in the training data set. The network is initialized with random synaptic weights. At each iteration, the error is computed and then propagated backwards from the output layer. The processes of error back-propagating and feeding forward signals are repeated iteratively until the error function is minimized or the maximum number of iterations is reached (i.e. 500).

During the training phase, the considered network architecture must obtain an optimal generalization performance: i.e. the network performance should not degrade significantly when data set other than the training one is analysed. Standard NN architectures, like the fully connected MLP, generally have a too large parameter space and are prone to overfitting. Although the network performance seems to constantly improve on the training sets at each iteration, it can actually begin to worsen (in terms of errors) on unseen data sets, i.e. data not present in the training set and thus not used for for the synaptic weight computation. Therefore, a verification step is performed, over the last 15 iterations, to detect this overfitting moment (i.e. no significant variation of E) and stop the training phase. This process is called early stopping. Finally, to ensure that the system is not trapped in local minima during the error function minimization, the learning phase (training plus verification) is repeated three times, the synaptic weights being randomly initialized at the beginning of each training phase. The network system presenting the best evaluation score (see Eq. 8) is then selected. All these precautions are carried out by randomly splitting the learning data (see Table 3.1) into three independent sets: training, verification and evaluation. They consisted of 70, 15 and 15 %, respectively.

Figure 3.3 depicts the box–whisker plots of the square of errors obtained over the ensemble of training–verification–evaluation data set for $\tau(550\,\text{nm})$ and ALP retrieval and for $\text{NN}_{R_c,N^s_{O_2-O_2}}$ configuration, assuming $\omega_0 = 0.95$. For $\tau(550\,\text{nm})$ retrievals, although the NNs with 40 and 70 neurons on one hidden layer do reasonably well, the scores show improved values when two hidden layers are used. The ALP retrieval scores are significantly larger than for τ (a factor of 100). This is a direct consequence that ALP is less well constrained by the spectral measurements: lower pieces information are available compared to τ, in particular for scenes dominated by low τ values. While the NNs with one hidden layer do not show any significant improvements when increasing number of neurons, better scores are only obtained with two hidden layers. Overall, the similar behaviour of training–verification–evaluation scores validate that the trained NNs are generalized enough to be able to reproduce similar variation of the scores on other independent data set. The identified best NN algorithms are thus found with two hidden layers, including between 25 and 70 neurons on the first layer and between 10 and 20 neurons on the second layer, depending on the considered configuration and retrieved parameter (see Sect. 3.2.3).

3.5 Sensitivity analyses on synthetic data set

The robustness of the trained and selected NN retrieval algorithms is assessed by applying them to independent simulations not present in the learning (training–verification–evaluation) database. Simulated spectra are noise free and only include aerosol particles (no clouds). The sensitivity of $\tau(550\,\text{nm})$ and ALP retrievals is verified for different A and aerosol properties (ω_0, τ, ALP). $R_c(475\,\text{nm})$ and $N^s_{O_2-O_2}$ are derived from the spectra and provided as inputs to the NNs. The impact of uncertainties on surface albedo, aerosol model and $N^s_{O_2-O_2}$ are analysed. Consistent geophysical conditions (temperature, NO_2 and O_3 profiles) are considered between these

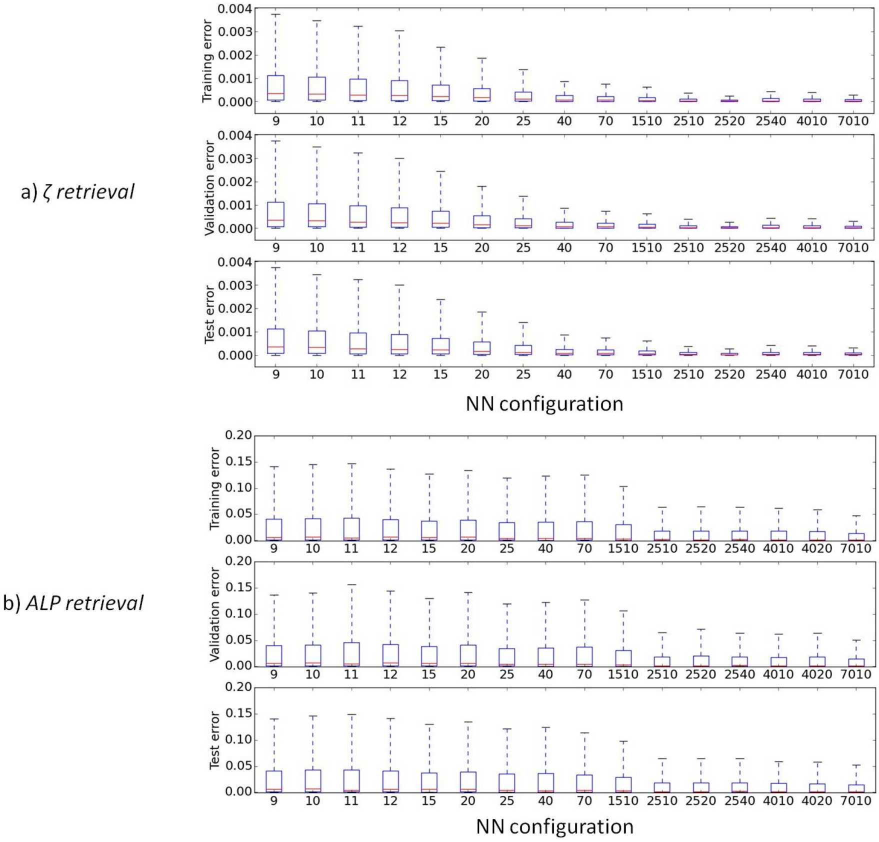


Figure 3.3: Box–whisker plots of the square of errors (see Eq. 7) obtained for different neural network (NN) configurations, at the end of their training, over the supervised data set (training validation test). The NNs XX have one hidden layer, where XX indicate the number of neurons. The NNs YYXX have two hidden layers, where YY and XX are the number of neurons in the first and second hidden layer, respectively: **(a)** NNs for τ retrieval and **(b)** NNs for ALP retrieval. Note that errors are computed over normalized output and true τ(550 nm) and ALP values (between -1 and 1) due to the definition of the sigmoid functions (see Sect. 3.3.1).

simulations and those included in the learning database. All the analyses performed here are summarized in Table 3.2.

3.5.1 Aerosol optical thickness retrievals

Figure 3.4 compares the retrieved to the true $\tau(550nm)$ values of the simulated spectra, and how uncertainties on ω_0, g and A degrade the retrieval quality. Overall, retrieved and true $\tau(550nm)$ values are very well correlated for all the types of surface, assuming no error in the assumed surface albedo and aerosol properties. This confirms the success of the learning process implemented in Sect. 3.4.2 and 3.4.3 and the use of the NN approach.

Table 3.2: Summary of OMI $\tau(550\text{ nm})$ and ALP retrieval error sources (either from $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ or from $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ combined with MODIS $\tau(550\text{ nm})$) and budget as evaluated by the sensitivity analyses on synthetic data set (see Sect. 3.5) or on 3-year (2005–2007) observation measurements over north-east Asia (see Sect. 3.6).

Error source	$\tau(550\text{ nm})$	From $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$: ALP in hPa ALH in m	From $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ ALP in hPa ALH in m
$\tau(550\text{ nm})$:			
True $\tau(550 nm) \leq 0.5$ (see Sect. 3.5)	–	$[250:400]$ hPa	$[150:400]$ hPa
True $\tau(550\text{ nm}) = [0.5{:}1.0]$ (see Sect. 3.5)	–	100 hPa	100 hPa
True $\tau(550\text{ nm}) \geq 1.0$ (see Sect. 3.5)	–	≤ 50 hPa	≤ 50 hPa
MODIS $\tau(550\text{ nm}) \geq 1.0$ (see Sect. 3.6)	–	650–1140 m	260–800 m
$\partial\tau(550 nm) = 0.25$ (see Sect. 3.5)	–	–	50 hPa(true $\tau = 0.6$–1.0)
	–	–	Almost zero (true $\tau \geq 1.0$)
Surface albedo:			
$\partial A = 0.05$ (see Sect. 3.5)	0.25–0.5	> 100 hPa (true $\tau = 0.5$–1.0)	50–100 hPa (true $\tau = 0.5$–1.0)
		0–50 hPa (true $\tau \geq 1.0$)	
OMLER vs. MODIS black sky (see Sect. 3.6)	0.05–0.1	≤ 730 m (MODIS ($\tau \geq 1.0$)	≤ 180 m (MODIS ($\tau \geq 1.0$)
Aerosol single scattering albedo ($\partial\omega_0 = 0.05$)			
(see Sect. 3.5)	0.8 (true $\tau = 2$)	> 100 hPa (true $\tau = 0.5$–1.5)	0–50 hPa (true $\tau \geq 0.5$)
	0.1 (true $\tau = 0.5$)	0–100 hPa (true $\tau \geq 1.0$)	
(see Sect. 3.6)	0.5 (MODIS $\tau = 1.5$)	540–1200 m (MODIS ($\tau \geq 1.0$)	560–660 m (MODIS $\tau \geq 1.0$)
Asymmetry parameter ($\partial g = 0.1$) (see Sect. 3.5)	0.5 (true $\tau = 1.5$)	200–400 hPa (true $\tau = 0.5$–1.0)	0–50 hPa (true $\tau \geq 0.5$)
	0.25 (true $\tau = 0.5$)	50 hPa (true $\tau \geq 1.0$)	
O_2-O_2 SCD ($\partial N^s_{O_2-O_2}$) (see Sect. 3.5):			
$\partial N^s_{O_2-O_2} = 0.05\ \mathrm{mol^2cm^{-5}}$	–	19 ± 29 hPa	19 ± 29 hPa
$\partial N^s_{O_2-O_2} = 0.25\ \mathrm{mol^2cm^{-5}}$	–	57 ± 31 hPa	57 ± 31 hPa
O_2-O_2 SCD temperature correction (see Sect. 3.6)	–	50–300 m (MODIS ($\tau \geq 1.0$)	50–300 m($MODIS \tau \geq 1.0$)

The assumed aerosol properties (ω_0 and phase function through g), and so the choice of the trained NN algorithm, are of high importance. They change the slope between retrieved and true τ values and drastically affect the retrieved $\tau(550\,\text{nm})$ accuracy. If the assumed ω_0 (i.e. 0.95) through the simulations in the learning database is too high (true $\omega_0 = 0.9$), retrieved τ is then underestimated: i.e. a bias of -0.8 for $\tau(550\,\text{nm}) = 2$ and -0.1 for $\tau(550\,\text{nm}) \leq 0.5$ (see Fig. 3.4a). This is a direct consequence of reduced scattering efficiency as more photons are absorbed instead of being scattered back towards the satellite sensor. The measured $R_c(475\,\text{nm})$ is then lower (see Fig. 3.1). Reciprocally, an assumed too low aerosol ω_0 (i.e. true $\omega_0 = 1.0$) leads to an overestimation of retrieved $\tau(550\,\text{nm})$: i.e. a bias of 1.0 for $\tau(550\,\text{nm}) = 2$ (see Fig. 3.4a).

Figure 3.4b illustrates retrieved $\tau(550\,\text{nm})$ bias due to the difference between the assumed g in the learning database and in the synthetic spectra. While $g = 0.7$ is the reference value for most of aerosols, scenes with $g = 0.6$ are related to finer and weakly absorbing particles with a somewhat reduced forward scattering direction such as carbonaceous aerosols, desert dust and volcanic dash models as given by the ESA aerosol CCI project (de Leeuw et al., 2015). Values of $g = 0.8$ are associated with larger particles and an increased forward scattering direction such as cirrus (Sanders et al., 2015). An overestimation of g (i.e. assumed $g = 0.7$ while true $g = 0.6$) leads to an increased retrieved τ value (i.e. positive bias) because of less photons scattered towards the surface, and therefore more photons scattered back towards the satellite sensor, compared to what is theoretically assumed. Reciprocally, an underestimation

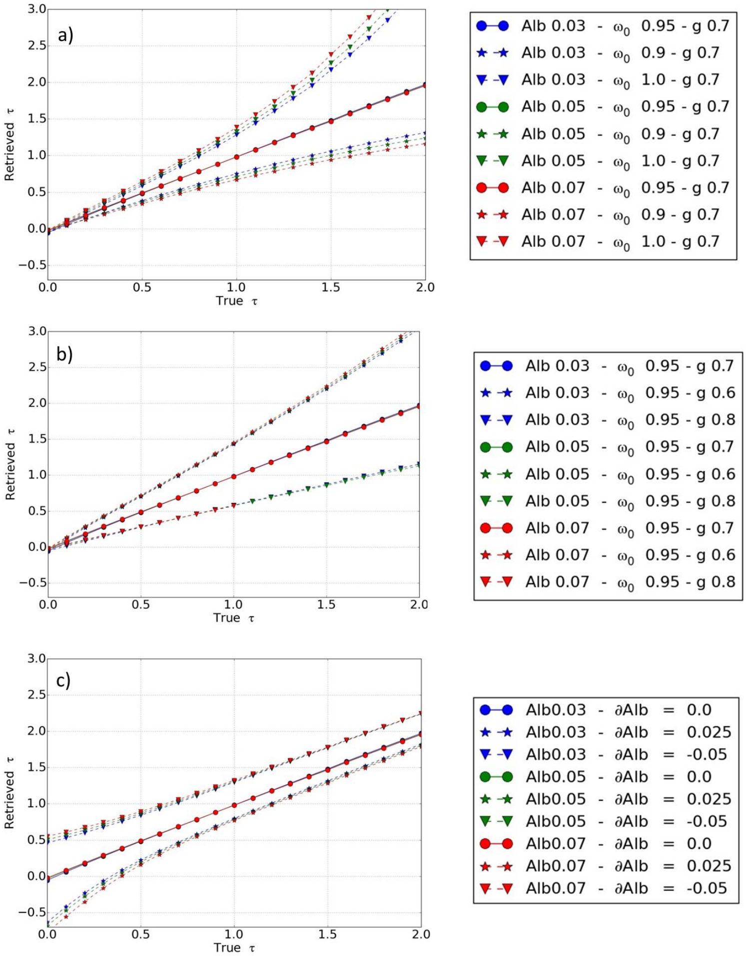


Figure 3.4: Simulated $\tau(550\,\mathrm{nm})$ retrievals, based on noise-free synthetic spectra with aerosols, as a function of true $\tau(550\,\mathrm{nm})$. The assumed geophysical conditions are temperature, H_2O, O_3 and NO_2 from climatology mid-latitude summer; $\theta_0 = 25°$, $\theta = 25°$ and $P_s = 1010$ hPa. All particles are located between 800 and 900 hPa and $\alpha = 1.5$. Note that the scenarios with lines and similar symbols general tend to fall on top of each other. The reference aerosol scenario is plotted with continuous lines and circle symbols and includes consistent aerosol optical properties with the supervision data set used to train the neural network algorithm: i.e. $\omega_0 = 0.95$, $g = 0.7$. All the retrievals are achieved with the NN algorithm trained with $\omega_0 = 0.95$: **(a)** sensitivity of $\tau(550\,\mathrm{nm})$ retrievals to the aerosol single scattering albedo (true $\omega_0 = 0.95$, 0.9 or 1.0); **(b)** sensitivity of $\tau(550\,\mathrm{nm})$ retrievals to the aerosol asymmetry parameter (true $g = 0.6$, 0.7 or 0.8); **(c)** sensitivity of $\tau(550\,\mathrm{nm})$ retrievals to a surface albedo bias ($\partial\mathrm{Alb} = 0.0$, 0.025, -0.05) with $\omega_0 = 0.95$, $g = 0.7$.

of g (i.e. assumed $g = 0.7$ while true $g = 0.8$) leads to a decreased retrieved τ value (i.e. negative bias) due to less photons scattered back towards the satellite sensor and more towards the surface compared to the assumption. Absolute bias values can exceed 0.5 for $\tau(550nm) = 1.5$ while they stay close to 0.25 for $\tau(550nm) = 0.5$.

Errors in surface albedo also lead to biases in retrieved $\tau(550nm)$ (see Fig. 3.4c). Overall, biases are larger over scenes with small $\tau(550nm)$ values. The reason is the dominance of surface reflection in this regime. Only in cases of high amount of aerosols do aerosol scattering signals become dominant and surface reflection uncertainties have less impact. An underestimated (overestimated) surface albedo results in a negative (positive) retrieved $\tau(550\,\mathrm{nm})$ bias. This is directly related to the change in the measured OMI Rc (see Fig. 3.1). Surface albedo uncertainties in the range of 0.025–0.05 lead to absolute biases close to 0.5 for $\tau(550\,\mathrm{nm})$ in the range of 0.0–0.5, smaller than 0.25 for $\tau(550nm) = 2.0$. Typical differences in climatological surface albedo from the TOMS and the global ozone monitoring experiment (GOME; Koelemeijer et al., 2003), or between OMLER and the MODIS black-sky albedo (Kleipool et al., 2008), are known to be up to 0.02.

3.5.2 Aerosol layer pressure retrievals

Figures 3.5 and 3.6 depict the expected performances of the developed NN algorithms for ALP retrievals. Over scenes with $\tau(550\text{ nm})$ in the range of 0.5–1.0, and assuming no error in prior A and the employed aerosol model, ALP retrievals are relatively stable presenting biases close to 100 hPa. Only for $\tau(550\text{ nm}) \geq 1.0$, biases are smaller than 50 hPa. The accuracy of the retrieved ALP generally increases with increasing τ. Indeed, assuming true ALP of 850 hPa (see Fig. 3.6), positive biases larger than 400 hPa are found for $\tau(550nm) \leq 0.5$. Note that this behaviour is observed for all the NN configurations ($\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ and $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$). A box–whisker plot, in Fig. 3.7, illustrates the variability of the ALP NN biases as a function of $\tau(550\,\mathrm{nm})$ over all the simulations contained in the entire learning database (as defined and used in Sect. 3.3). This confirms that, in spite of the strict training–verification–evaluation process achieved in Sect. 3.4.3, the NN ALP retrievals are not expected to be accurate for small $\tau(550\,\mathrm{nm})$ values, especially below 0.5. The reason is directly linked to the magnitude of the O_2-O_2 shielding effect and its combined dependence on aerosol amount (or τ) and aerosol altitude (see Fig. 3.1 and Sect. 3.3.2). Because low amount of aerosols have very limited effects on the O_2-O_2 absorption (see Sect. 3.3.2), even advanced interpolation techniques like NNs have difficulties to interpret the associated signal. When τ increases, the O_2-O_2 shielding effect amplifies and the algorithms are more able to link the O_2-O_2 absorption signal to ALP. Overall, even for small $\tau(550nm)$ values (like 0.5), the retrieved aerosol pressures correlate with the true values in spite of very poor accuracy (see Fig. 3.6a and c).

A very accurate prior τ information as input is required to generally improve the ALP retrieval performances. As explained in Sect. 3.3.2, using $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ implicitly relies on the $\tau(550\,\mathrm{nm})$ retrieval capability from OMI $R_c(475\text{ nm})$. Using $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ with the true $\tau(550\,\mathrm{nm})$ value as input (thus no error) is expected to deliver higher performances. Firstly, it allows us to improve the accuracy of the ALP retrieval over scenes with low τ and particles located at high altitude (above 800 hPa or ~ 2 km). Indeed, in Fig. 3.5b, for $\tau(550nm) = 0.5$ and ALP between 750 and 850 hPa, ALP biases are reduced from 250–350 hPa with $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ to 150–250 hPa with $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$

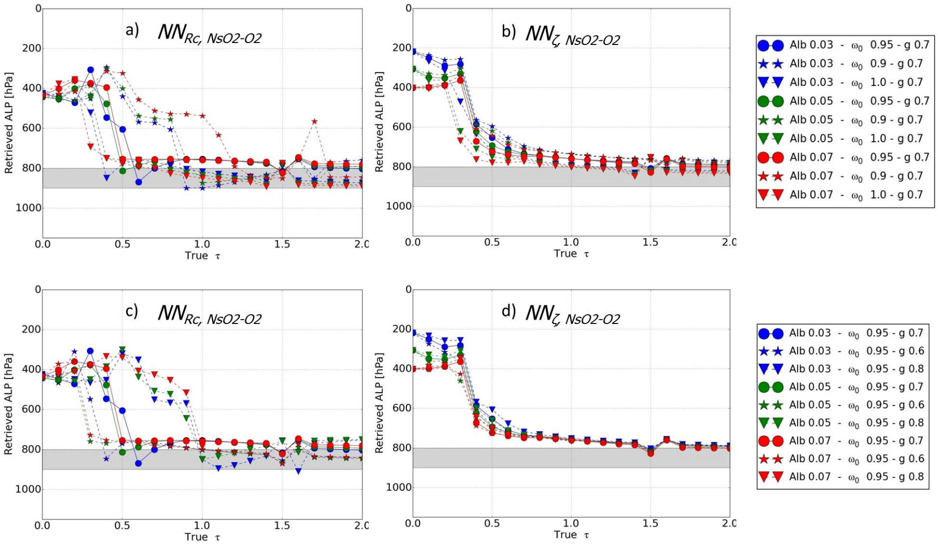


Figure 3.5: Simulated ALP retrievals, based on noise-free synthetic spectra with aerosols, as a function of true τ. All the retrievals are achieved with the two NN configurations ($\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ and $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$; see Sect. 3.3.1) trained with $\omega_0 = 0.95$. The assumed geophysical conditions are temperature, H_2O, O_3 and NO_2 from climatology mid-latitude summer, $\theta_0 = 25°$, $\theta = 25°$ and $P_s = 1010$ hPa. The reference aerosol scenario assumes fine scattering particles ($\alpha = 1.5$, $\omega_0 = 0.95$, $g = 0.7$) located between 800 and 900 hPa: **(a, b)** sensitivity of ALP retrievals to the aerosol single scattering albedo (true $\omega_0 = 0.95$ or 0.9) in the synthetic spectra; **(c, d)** sensitivity of ALP retrievals to the aerosol asymmetry parameter (true $g = 0.6$, 0.7 or 0.8) in the synthetic spectra.

and true $\tau(550\,\mathrm{nm})$ value. For particles higher than 650 hPa (or ~ 3.5 km), however, no improvements are observed. The low sensitivity to retrieve ALP when particles are located at a very high altitude is directly due to the O_2-O_2 complex and its vertical distribution. This was demonstrated by Park et al. (2016): O_2-O_2 concentration exponentially decreases with increasing atmospheric altitude.

Secondly, impacts due to uncertainties on the chosen surface albedo and aerosol model are reduced. Assumptions on aerosol ω_0 drive the interpretation of the shielding of the O_2-O_2 dimers by aerosols. ω_0 can perturb ALP retrievals obtained with $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ more than 100 hPa (see Fig. 3.5a). These perturbations are reduced to the range of 0–100 hPa over scenes with high $\tau(550\,\mathrm{nm})$ values (larger than 1) only for particles close to the surface, i.e. true ALP ≥ 850 hPa (see Fig. 3.6a and c). Using $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ with true $\tau(550$ nm) value helps to mitigate these biases. All the ALP retrievals present the same behaviours with respect to the particles altitude and τ and biases lie in the range of 0–50 hPa (see Fig. 3.5b). Similar conclusions are observed regarding uncertainties on g (see Fig. 3.5c and d). Too high g values impact the ALP retrievals from $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ over scenes with $\tau(550nm) \leq 1.0$. Such a bias is largely reduced with the $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ configuration.

Surface albedo contributes to the length of the average light path and thus affects

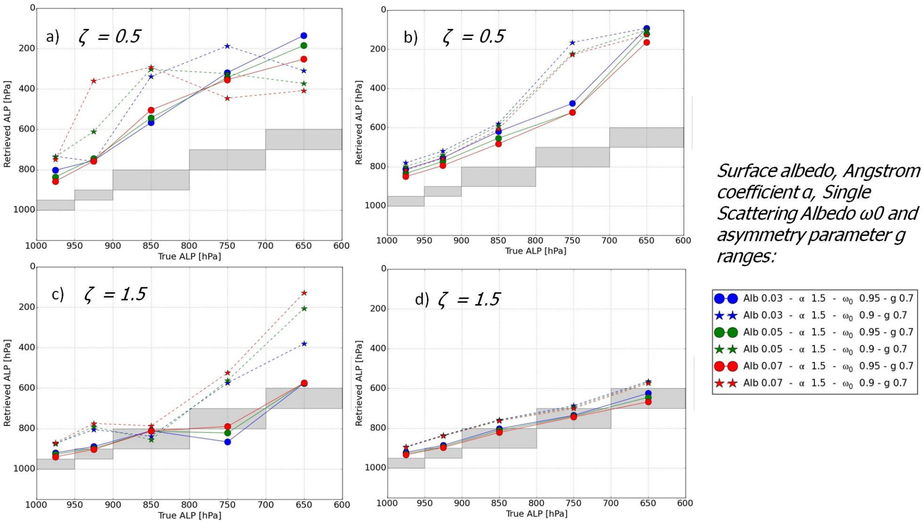


Figure 3.6: Simulated ALP retrievals vs. true ALP for 2 τ values (0.5 and 1.5) and the two NN configurations (see Sect. 3.3.1) and for the following conditions: temperature, H_2O, O_3 and NO_2 from climatology mid-latitude summer, $\theta_0 = 25°$, $\theta = 25°$, surface pressure of 1010 hPa and fine scattering aerosol particles ($\alpha = 1.5$, $\omega_0 = 0.95$, $g = 0.7$): **(a)** $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ and $\tau(550nm) = 0.5$; **(b)** $\mathrm{NN}_{\tau, N^s_{O_2-O_2}}$ with true $\tau(550nm)$ value as input and $\tau(550nm) = 0.5$; **(c)** $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ and $\tau(550\text{ nm}) = 1.5$; **(d)** $\mathrm{NN}_{\tau, N^s_{O_2-O_2}}$ with true $\tau(550\text{ nm})$ value as input and $\tau(550nm) = 1.5$.

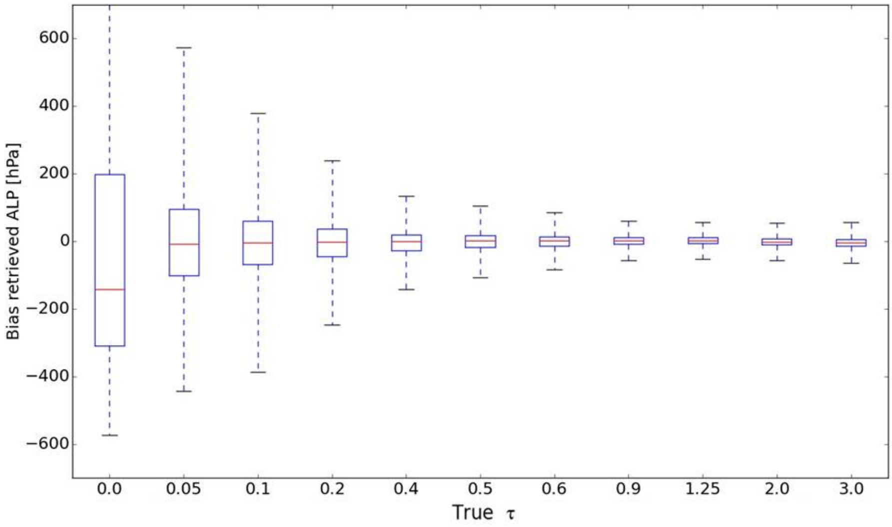


Figure 3.7: Box–whisker plots of retrieved aerosol layer pressure (ALP) biases as a function of true $\tau(550\,\text{nm})$ from $\mathrm{NN}_{R_c, N^s_{O_2-O_2}}$ configuration over the whole learning data set.

$N^{s}_{O_2-O_2}$. Retrieved ALP biases are maximum (several hundreds hPa) for $\tau(550nm) \leq 0.5$ (see Fig. 3.8a and b). For $\tau(550nm)$ in the range of 0.5–1.0, retrieved ALP are impacted by lower absolute values (between 50 and 100 hPa on average) with $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$, while they remain too high with $\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$. Over scenes with $\tau(550$ nm$) \geq 1.0$, biases are reduced to 0–50 hPa since aerosol scattering signals dominate over surface reflection. The main cause of all these improvements is that using an accurate prior τ information (or at least more than retrieved OMI $\tau(550\,\mathrm{nm})$ from $R_c(475$ nm$)$) allows a better distinction of τ and ALP effects on the O_2-O_2 slant column density and reduce impacts of A and ω_0 uncertainties.

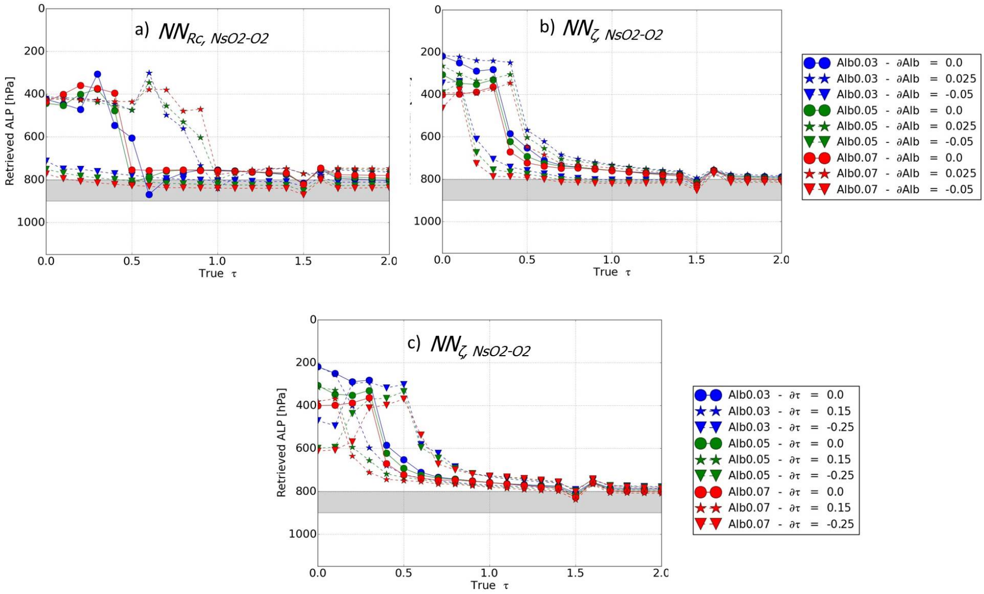


Figure 3.8: Simulated ALP retrievals, based on noise-free synthetic spectra with aerosols, as a function of true τ. The retrievals are achieved with the NN configurations ($\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$ and/or $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$; see Sect. 3.1) trained with $\omega_0 = 0.95$. The assumed geophysical conditions are temperature, H_2O, O_3 and NO_2 from climatology mid-latitude summer, $\theta_0 = 25°$, $\theta = 25°$ and $P_s = 1010$ hPa. The reference aerosol scenario assumes fine scattering particles ($\alpha = 1.5$, $\omega_0 = 0.95$, $g = 0.7$) located between 800 and 900 hPa: **(a)** and **(b)** sensitivity of ALP retrievals to a surface albedo bias ($\partial Alb = 0.0$, 0.025, -0.05) with $\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$ and $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$; **(c)** sensitivity of ALP retrievals to a $\tau(550\,\mathrm{nm})$ bias ($\partial\tau(550nm) = 0.0$, 0.15, -0.25) with $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$.

An accuracy better than 0.2 must be required on prior τ information (see Fig. 3.8c). Indeed, a $\tau(550$ nm$)$ bias of 0.25 can impact, in absolute, the retrieved ALP up to 50 hPa for $\tau(550$ nm$)$ in the range of 0.6–1.0. For $\tau(550$ nm$) \geq 1.0$, impact on ALP becomes almost null. Therefore, using MODIS τ as prior to $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$ is likely expected to show retrieved ALP with a higher quality than with $\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$. Indeed, the current retrieved OMI $\tau(550\,\mathrm{nm})$ from $R_c(475$ nm$)$ does not present a better accuracy than MODIS $\tau(550nm)$.

Figure 3.9 depicts the box–whisker distribution of ALP precision $\epsilon(\partial N^{s}_{O_2-O_2})$ due to $N^{s}_{O_2-O_2}$ precision.

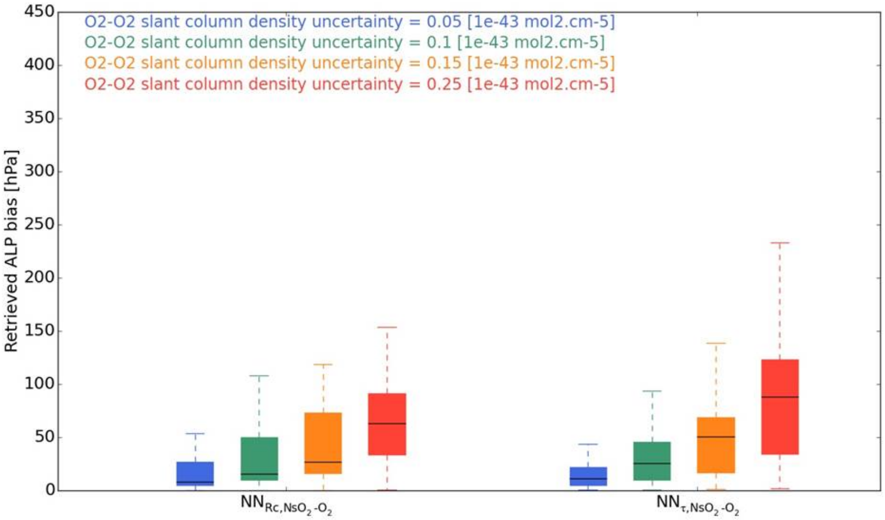


Figure 3.9: Box–whisker plots of ALP retrieval biases induced by O_2-O_2 SCD $N^{s}_{O_2-O_2}$ uncertainties. The assumed conditions are fine scattering aerosols ($\alpha = 1.5$, $\omega_0 = 0.95$, g O_3 and H_2O profiles, surface pressure of 1010 hPa, surface albedo of [0.03–0.05–0.07] and combination of $\theta_0 - \theta$ of [25–25, 50–25, 25–45°].

Estimations are obtained for fine and scattering particles ($\alpha = 1.5$, $\omega_0 = 0.95$, $g = 0.7$). $\epsilon(\partial N^{s}_{O_2-O_2})$ is obtained from the half of ALP differences between adding and deducting uncertainties of the variables as follows:

$$\epsilon(\partial N^{s}_{O_2-O_2}) = \frac{1}{2} \mid \mathrm{ALP}(N^{s}_{O_2-O_2} + \partial N^{s}_{O_2-O_2}) - \mathrm{ALP}(N^{s}_{O_2-O_2} - \partial N^{s}_{O_2-O_2}) \mid, \tag{3.9}$$

where $\partial N^{s}_{O_2-O_2}$ is the uncertainty applied to $N^{s}_{O_2-O_2}$. $\epsilon(\partial N^{s}_{O_2-O_2})$ values are computed for all combinations of surface albedo 0.03–0.05–0.07 and θ_0–θ = [25–25, 50–25, 25–45°]. The reason to use this approach here is that, since $N^{s}_{O_2-O_2}$ precision is a random error (opposite to systematic), it will directly impact the retrieved ALP precision instead of leading to a systematic bias. A precision of $N^{s}_{O_2-O_2}$ lying in the range of 0.05–0.25 10^{-43} $\mathrm{mol}^2\mathrm{cm}^{-5}$ (i.e. at a first order, 2–7 % of $N^{s}_{O_2-O_2}$) results in ALP uncertainties between 19 ± 29 and 57 ± 31 hPa on average for both NN configurations (see Fig. 3.5c).

Overall all the estimated NN retrieval uncertainties are in line with the theoretical sensitivity analyses of Park et al. (2016), who found that the O_2-O_2 at 477 nm is significantly influenced by aerosol optical properties (including ω_0), τ, particle size and A. In particular, a ω_0 uncertainty of 10 % was demonstrated to lead to the aerosol effective height (AEH) retrieval error ranging from 270 to 1440 m, depending on the aerosol types. Errors were found larger for high particle altitude and low τ cases. A surface albedo uncertainty of 0.02 was expected to impact AEH retrievals between 154 and 434 m on average. AEH error was frequently larger only for low $\tau(550nm)$ (≤ 0.4) and high AEH (≥ 1 km).

3.6 Application to OMI observation measurements

3.6.1 Methodology

Aerosol retrievals, as described in the previous sections, are performed on the OMI O_2-O_2 477 nm observations over large industrialized continental areas in north-east Asia over 3 years, 2005–2007, and cloud-free scenes. All the associated results are summarized in Table 3.2. The considered north-east Asia area is defined by the range of latitude 25–40° N and longitude 110–130° E, excluding the part over the Gobi desert which presents a too-bright surface (as further explained below).

Only OMI observations collocated with MODIS Aqua L2 aerosol product collection 6 are considered (see Sect. 3.2). The reason is threefold: (1) to maximize the probability of the selection of cloud-free OMI observation pixels dominated by aerosol pollution; (2) to evaluate the retrieved OMI $\tau(550nm)$ products by comparing with collocated MODIS $\tau(550nm)$; and (3) to use the MODIS $\tau(550nm)$ as input of the $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$ algorithm for retrieving the OMI ALP product, assuming then this is the most accurate τ information available for each collocated OMI observation pixel–MODIS aerosol grid cell.

MODIS data are paired on a OMI pixel-by-pixel basis when the distance between OMI pixel centre and MODIS aerosol grid cell is smaller than 5 km and when both observation and product are acquired within 15 min. A threshold of 0.1 is applied to both OMI and MODIS cloud fraction: i.e. if the OMI effective cloud fraction and/or the MODIS geometric cloud fraction (given in the same MODIS aerosol product at 10 km resolution) has a cloud fraction value higher than 0.1, the OMI pixels are filtered out. However, since the OMI effective cloud fraction is sensitive to the scattering aerosols, it is well recognized that cloud-free observations with large presence of scattering aerosols are frequently excluded as well. In addition, a threshold of 0.1 is applied to the OMLER database in order to filter out too-bright surfaces (either desert or snow-covered pixels).

The NN retrieval algorithms developed and selected in Sect. 3.3 are used and evaluated here: $\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$ for $\tau(550\text{ nm})$ and ALP retrievals and $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$ with MODIS $\tau(550\text{ nm})$, from DT algorithm over land, as input for ALP retrieval. As a reminder, retrieving ALP from $\mathrm{NN}_{R_c,N^{s}_{O_2-O_2}}$ is implicitly similar than from $\mathrm{NN}_{\tau,N^{s}_{O_2-O_2}}$ but with the retrieved OMI $\tau(550\text{ nm})$ as input (see Sects. 3.2 and 5.2). All the NNs designed and trained with the two different ω_0 (i.e. 0.9 and 0.95) are considered in order to investigate the impact of ω_0 assumptions.

These algorithms are applied on the OMI DOAS O_2-O_2 observations, available in the OMCLDO2 product (Acarreta et al., 2004) which can be downloaded at `http://`

`disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omcldo2_v003.shtml`. As explained in Veefkind et al. (2016), $N^{s}_{O_2-O_2}$ depends on the temperature profile due to the nature of dimers of which the absorption scales with the pressure-squared instead of being linear with pressure. Therefore, a simple temperature correction is here applied by using seasonal mean temperature profiles given by the National Centers for Environmental Prediction (NCEP) analysis data. This correction is performed through the computation of the γ factor (Veefkind et al., 2016):

$$\gamma = \frac{N^{\mathrm{s\,Ref}}_{O_2-O_2}(\lambda)}{N^{\mathrm{s\,Meas}}_{O_2-O_2}(\lambda)}, \tag{3.10}$$

with $N^{\mathrm{sRef}}_{O_2-O_2}$, the O_2-O_2 SCD associated with the reference temperature profile employed in the learning database, and $N^{\mathrm{sMeas}}_{O_2-O_2}$, the measured O_2-O_2 SCD related to the actual temperature conditions. As a first and simple approximation, no prior knowledge on aerosols is considered here. The main reason is the low sensitivity to aerosol loading and altitude of this γ factor, for $\tau(550nm) \leq 2.0$, compared to the change of temperature profiles for the considered OMI observations.

Finally, retrievals are performed based on different assumed surface albedo databases: either OMLER or MODIS black-sky albedo (see Sect. 3.4). The MODIS black-sky albedo is resampled to match the OMI pixel resolution by calculating the average of all MODIS pixels falling within the processed OMI pixel.

3.6.2 Aerosol optical thickness accuracy: on the importance of the surface albedo and the assumed aerosol properties

Figure 3.10 compares collocated retrieved OMI and MODIS τ(550 nm). Similarly to the analyses on synthetic cases (see Sect. 3.5.1), the change of assumed ω_0 mostly perturbs retrievals of high τ values and thus the slope between OMI and MODIS τ(550 nm). Increasing ω_0 from 0.9 to 0.95 reduces the retrieved τ values of about 0.5 for MODIS $\tau = 1.5$. Overall a very good agreement is obtained assuming $\omega_0 = 0.9$ for the seasons spring, autumn and winter (see Figs. 3.10 and 3.13): differences (OMI-MODIS) of $\tau(550nm)$ lie between -0.18 ± 0.24 in winter and -0.02 ± 0.45 in spring. In summertime, the best agreements are found by assuming $\omega_0 = 0.95$ with differences in the range of -0.06 ± 0.31 (see Figs. 3.8 and 3.12).

From the end of autumn to spring, westerly winds transport mineral dust from the Taklimakan and Gobi deserts in northern China and Mongolia. These dust particles are then frequently mixed with the local anthropogenic aerosols released from the industrial activities, vehicle emissions and coal burning (Eck et al., 2005). Southeast Asia is affected in spring by biomass-burning activity (mostly over the peninsular), which is a major source of carbonaceous aerosols in the world. Jethva et al. (2014) show that AERONET and OMAERUV retrieve aerosol ω_0 values on average between 0.9 and 0.95 in these regions: most of the sulfate particles have ω_0 close to 0.95, while smoke and dust present lower values (closer to 0.9, even below in

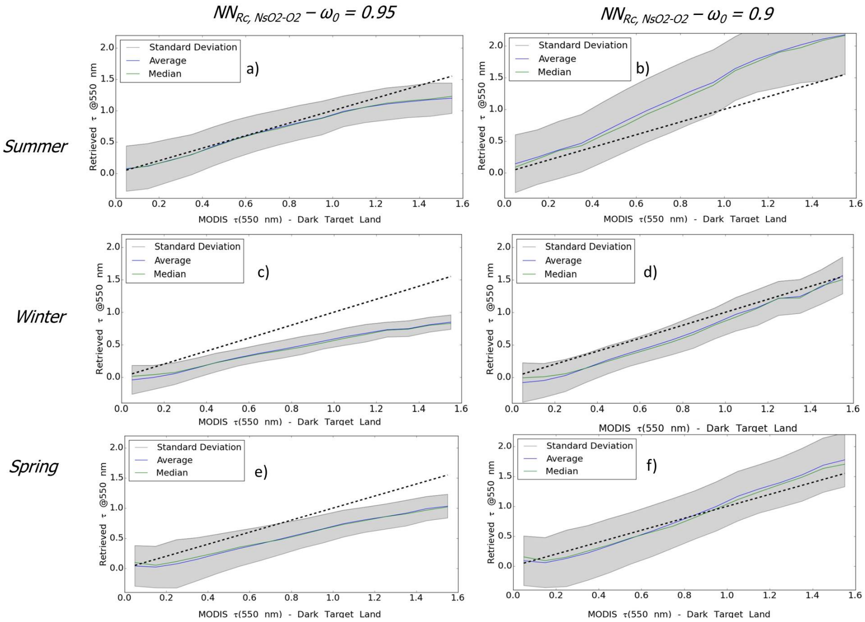
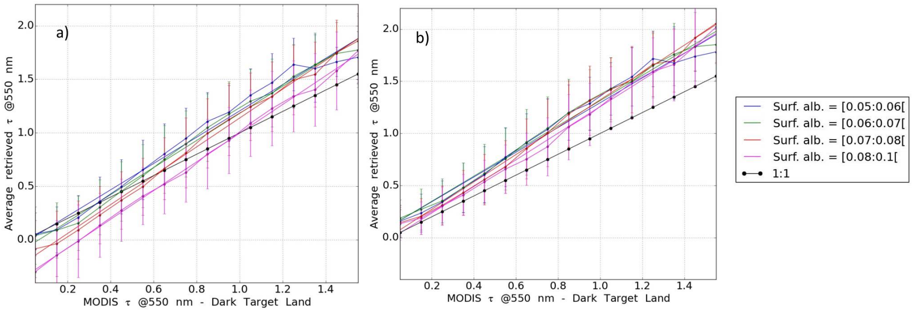


Figure 3.10: Collocated MODIS Aqua and retrieved OMI $\tau(550nm)$ based on the OMLER surface albedo over East China for cloud-free scenes and summer, winter and spring seasons. Statistics are computed over 3 years, 2005–2007: **(a, c, e)** assumed aerosol model with $\omega_0 = 0.95$; **(b, d, f)** assumed aerosol model with $\omega_0 = 0.9$.

some cases). These analyses confirm that the assumption of $\omega_0 = 0.95$ should be considered as an upper limit for the OMI τ retrievals in autumn, winter and spring, while a lower ω_0 (i.e. 0.9) is likely more appropriate and thus allows, on average, more reliable τ retrievals. In summertime, because of reduced amounts of dust particles, τ values are more representative of local anthropogenic urban pollution, with a higher daily variability in the optical and scattering properties. Lee et al. (2007) and Lin et al. (2015) also found higher ω_0 values over north-east Asia in summer (0.95–0.96) and lower for the other seasons (0.88–0.92). Overall, assuming the same constant value (i.e. average) for all the acquired OMI pixels probably leads to some errors since aerosol scattering and absorption properties likely vary day to day and even month to month.

Figure 3.11 depicts the dependence of the retrieved OMI $\tau(550\,\text{nm})$ on the OMLER surface albedo values. As discussed in Sect. 3.5.1, error in surface albedo directly creates a bias on the retrieved τ. Most of the retrievals depict higher values over darker surfaces or lower surface albedo. In the range of OMI surface albedo values 0.05–0.1, such behaviour should not be observed assuming no systematic bias on the surface properties. Moreover, OMI $\tau(550\,\text{nm})$ shows values that are too small for scenes with MODIS $\tau(550nm) \leq 0.4$ in autumn and winter using OMLER (see Figs. 3.11 and 3.14). MODIS black-sky surface albedo allows us to reduce this dependence in summer and spring and reasonably increases OMI τ retrievals over scenes with low MODIS τ values in autumn and winter (see Fig. 3.14). Furthermore, the standard deviation of differences (OMI-MODIS) τ(550 nm) shows a net improvement of the retrievals precision, with a mean reduction of 0.05 from OMLER to MODIS black-sky (see Fig. 3.12). An exception is, however, noticed in winter, which may be due to remaining snow-covered pixels in spite of the applied filtering. Overall, spatial patterns better match between collocated MODIS and OMI products when employing MODIS black-sky albedo with higher values over the high density population areas (i.e. north-east and south-west of selected Chinese regions, south-western Korea) and lower values over south-eastern China (see Figs. 3.13 and 3.14). These improvements may be due to a more accurate atmospheric correction in the MODIS black-sky surface albedo and potential remaining aerosol residuals present in the OMLER database.



OMLER input *MODIS Black Sky input* *OMLER surface albedo range*

Figure 3.11: Collocated retrieved OMI retrieval and MODIS Aqua τ(550 nm) from Dark Target algorithm over land, over north-east Asia for cloud-free scenes in autumn. Retrievals are depicted as a function of OMLER surface albedo ranges (Kleipool et al., 2008; see Sect. 3.5.1). The dotted black line is the reference 1 : 1. Statistics are computed over 3 years, 2005–2007: **(a)** OMI retrievals based on OMLER; **(b)** OMI retrievals based on MODIS black-sky surface albedo.

In spite of these improved precisions, using the MODIS black-sky albedo does not always improve the accuracy of the OMI τ(550 nm) retrievals. In particular, summer and spring seasons present τ values that are too high compared to the use of OMLER. This emphasizes that applying the MODIS black-sky albedo to OMI measurements may be not fully optimal because (1) MODIS albedo is the integral value over the full hemisphere, which is not in line with the range of angles (θ_0 and θ) encountered by OMI; and (2) the MODIS black-sky albedo is valid for local solar noon zenith angle of each location, which does not match the 1345 ascending node equator crossing time of OMI. An ideal surface albedo database should be aerosol and cloud free, as well as representative of the viewing and solar angles encountered by the space-borne sensor. Problems related to uncertainties in surface albedo climatology for the aerosol retrieval problem are well known and were recently highlighted by Sanders et al. (2015), although a different spectral band is used (O_2-A at 758–770 nm).

As a conclusion, mostly because uncertainties on assumed aerosol ω_0 parameter and prior surface albedo are dominant, our retrieved OMI τ(550 nm) shows lower performances than MODIS τ(550 nm) from the DT algorithm over land in collection

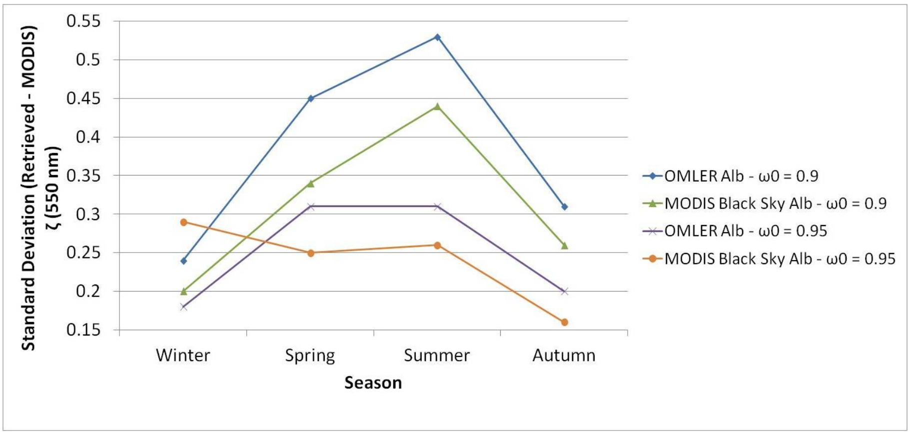
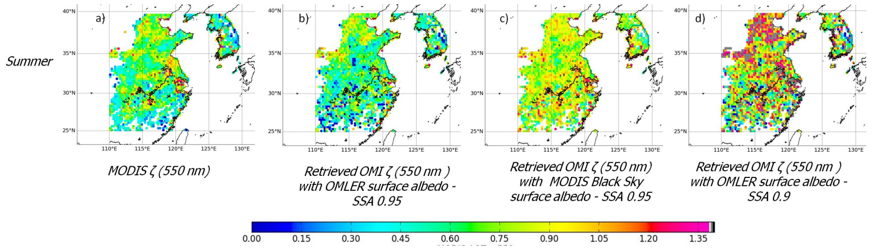


Figure 3.12: Standard deviation of the differences between OMI τ(550 nm) retrievals and MODIS τ(550 nm) from Dark Target land algorithm over land for all the individual cloud-free observations over north-east Asia. The retrievals are obtained over 3 years, 2005–2007, and for the four seasons. Aerosol single scattering albedo $\omega_0 = 0.95$ and $\omega_0 = 0.9$ are assumed. OMLER and MODIS black-sky surface albedo is alternatively considered.



MODIS ζ (550 nm)

Retrieved OMI ζ (550 nm) with OMLER surface albedo - SSA 0.95

Retrieved OMI ζ (550 nm) with MODIS Black Sky surface albedo - SSA 0.95

Retrieved OMI ζ (550 nm) with OMLER surface albedo - SSA 0.9

Figure 3.13: Spatial averages of τ(550 nm) values over north-east Asia for cloud-free scene. Statistics are computed over 3 years, 2005–2007, for summer.

6 (see Sect. 3.2.2). Furthermore, errors in the phase function or not taking into account the effect of polarization can play a role. These aspects should be further investigated.

3.6.3 Long-term analyses of the aerosol layer pressure retrievals

Figure 3.15 shows the retrieved OMI ALP as a function of collocated MODIS τ(550 nm). MODIS τ(550 nm) is considered for OMI ALP retrievals since, at this stage, they are considered as the best prior information available with higher accuracy than OMI τ(550 nm) (see Sect. 3.6.1). While ALP retrievals over scenes with MODIS $\tau(550nm) \leq 0.5$ exhibit large variability (more than 400 hPa) and are systematically very high, they start converging to more realistic values with increasing MODIS τ. At MODIS τ(550 nm) ≥ 1.0, retrieved ALP lies in the range of 800–1000 hPa depending on the season, with lower variability (between 50 hPa and 200 hPa maximum). As discussed in Sect. 3.5.2, scenes with $\tau \leq 0.5$ are expected to present substantial large

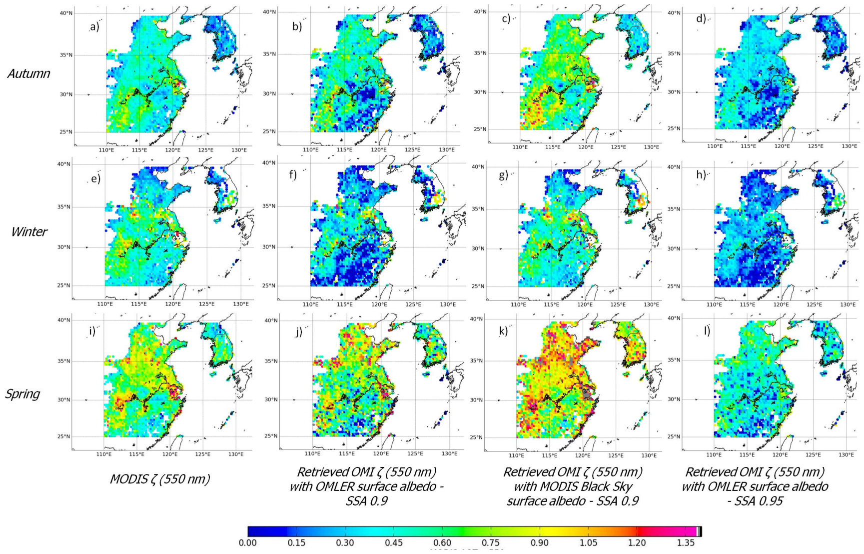


Figure 3.14: Spatial averages of $\tau(550\text{ nm})$ values over north-east Asia for cloud-free scenes. Statistics are computed over 3 years, 2005–2007, for the three seasons, autumn, winter and spring.

biases because of the minor impact on the O_2-O_2 changes. Part of the variability can be related to uncertainties of surface albedo and non-constant and inhomogeneous aerosol properties from OMI pixel-to-pixel (e.g. aerosol ω_0 in the OMI observations).

When considering $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ with MODIS $\tau(550\text{ nm})$, from DT algorithm over land, as input, the retrievals globally show a reduced variability, especially for $\tau(550\text{ nm})$ values in the range of 0.6–2.0 compared to the $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ configuration (see Fig. 3.16). Over scenes with MODIS $\tau(550\text{ nm}) \geq 1.0$, the variability of the OMI ALH, derived from Eq. (11) as explained in the next subsection, greatly decreases from the range of 1.1–2.7 km ($\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$) to 0.7–1.9 km ($\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$) depending on the season. When the OMLER is replaced by the MODIS black-sky albedo database, the ALH variability continues to decrease of about 0.1 km (see Fig. 3.16).

3.6.4 Comparison of OMI aerosol layer height with LIVAS climatology

The results of 3 years of OMI ALP retrievals over north-east Asia can be statistically compared to a climatology. Although the years of the OMI "climatology" and LIVAS do not strictly overlap, the average ALH is assumed not to change significantly between both periods. The comparison is done per season. Spatial average of LIVAS ALH is done over the same area where retrievals are performed. Since large biases are expected at low τ, only OMI retrievals acquired for MODIS $\tau(550\text{ nm}) \geq 1.0$ are taken into account and then spatially and temporally averaged per season. About 17 % in summer and spring, and between 5 and 6 % in winter and autumn, of the

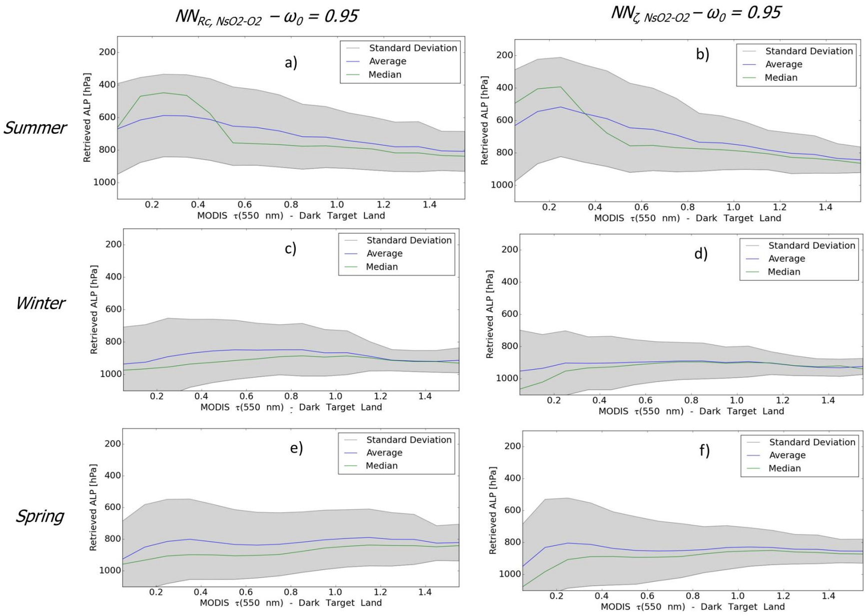


Figure 3.15: Collocated retrieved OMI ALP (assumed aerosol model with $\omega_0 = 0.95$ and based on the OMLER surface albedo) and MODIS Aqua τ(550 nm) over north-east Asia for cloud-free scenes and summer, winter and spring seasons. Statistics are computed over 3 years, 2005–2007: **(a, c, e)** OMI retrievals are from the $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ configuration; **(b, d, f)** OMI retrievals are from the $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ configuration with MODIS Aqua τ(550 nm), from Dark Target land algorithm, as input (see Sect. 3.5.1).

OMI retrievals over the 3 years were then selected. As a first and simple approximation, OMI ALP retrievals are converted into ALH in km above sea level, assuming the atmosphere is in hydrostatic balance, scale height of 8 km and a surface pressure at the sea level of 1013 hPa:

$$\mathrm{ALH\ (OMI)} = -8 \times \ln(\mathrm{ALP}/1013). \tag{3.11}$$

Assumptions on the forward aerosol model (used then in the creation of the supervision database) lead to the highest impact on the spatial–seasonal averaged ALH retrievals. This is mostly related to the ability of the corresponding NN algorithm to interpret the scattering vs. absorption aerosol effects on the $N^s_{O_2-O_2}$. Assuming OMLER surface albedo, differences between average ALH retrievals with $\omega_0 = 0.95$ and $\omega_0 = 0.9$ are in the range of 540–1200 m with $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ and 560–660 m with $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ using MODIS τ(550 nm). Assumptions on prior surface albedo (OMLER vs. MODIS black-sky albedo) also affect the spatial–seasonal averaged ALH retrievals. Related differences are in the range of 0–730 m with $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ and 0–180 m with $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$. Note that a remarkable change is noticed using MODIS black-sky albedo

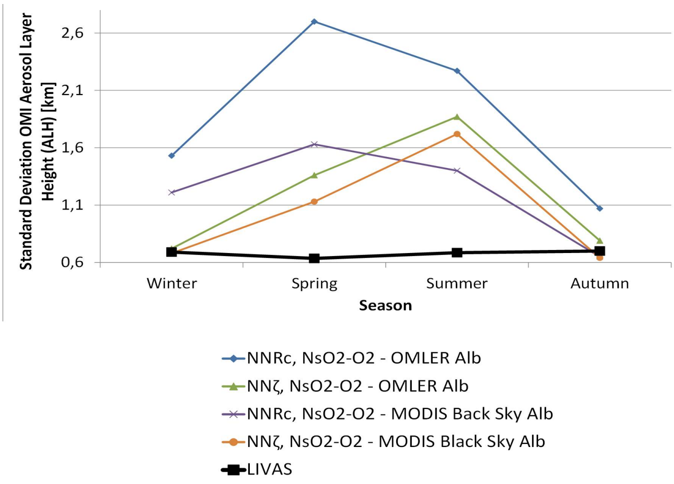


Figure 3.16: Standard deviation of the OMI ALH retrievals obtained for the cloud-free scenes with MODIS $\tau(550\ \text{nm}) \geq 1.0$ over north-east Asia. The retrievals are done for cloud-free scene over 3 years, 2005–2007. Aerosol single scattering albedo $\omega_0 = 0.95$ is assumed. OMLER and MODIS black-sky surface albedo, and the NN configurations ($\text{NN}_{R_c, N^s_{O_2-O_2}}$ and $\text{NN}_{\tau, N^s_{O_2-O_2}}$ with MODIS $\tau(550\ \text{nm})$ as input) are alternatively considered (see Sect. 3.6.1).

in winter with $\text{NN}_{R_c, N^s_{O_2-O_2}}$, which is likely due to non-filtered snow-covered pixels. Associated impacts are then lower than those related to assumed ω_0. Since only scenes with MODIS $\tau(550\ \text{nm}) \geq 1.0$ are selected, aerosol scattering signal is expected to dominate over the surface brightness. Although not shown here, we remarked that applying the temperature correction on $N^s_{O_2-O_2}$ is crucial as it corrected the retrievals between 50 and 300 m.

Because MODIS $\tau(550\ \text{nm})$ has a better accuracy than OMI $\tau(550\ \text{nm})$, seasonal-spatial average ALH retrievals from $\text{NN}_{\tau, N^s_{O_2-O_2}}$ combined with this first product shows a clear reduced impact and higher stability with respect to uncertainties on surface albedo and aerosol model. This is because using the most accurate prior information on aerosol amount provides then with the best ALH retrieval performance.

By comparing these best ALH results with the seasonal spatial averaged LIVAS ALH values, maximum differences in the range of 260–800 m are obtained, depending on the seasons, assuming $\omega_0 = 0.95$ and OMLER surface albedo. These differences are reduced to 180–310 m with $\omega_0 = 0.9$ regardless of the prior surface albedo. Furthermore, when comparing with the LIVAS ALH variability (see Eq. 2) spatially averaged over the area, the results from $\text{NN}_{\tau, N^s_{O_2-O_2}}$ with MODIS $\tau(550\ \text{nm})$ and prior MODIS black-sky albedo have the closest variability with a spread in the range of 0.0–1.0 km (see Fig. 3.16).

Consistent seasonal patterns can be observed between averaged OMI and LIVAS ALH with higher values in spring and summer, probably due to long-range transport during the dust activity from the desert (see Sect. 3.6.2) and smaller values in autumn and winter (see Fig. 3.17).

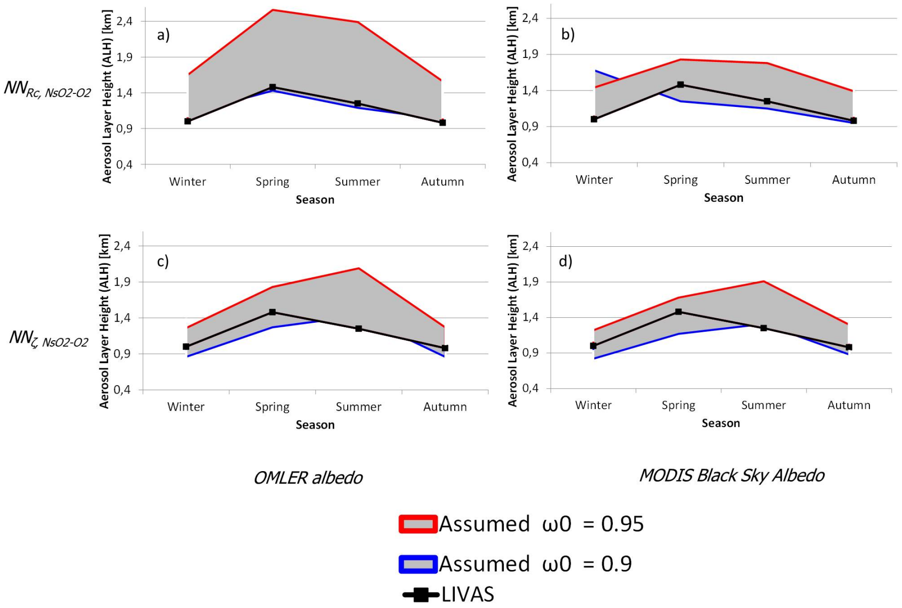


Figure 3.17: Comparison of the average of the OMI ALH retrievals obtained over scenes with collocated MODIS $\tau \geq 1.0$, with the LIVAS ALH climatology database. The retrievals are achieved over north-east Asia for cloud-free scenes, over 3 years, 2005–2007, and for the four seasons. OMI and MODIS black-sky surface albedo is alternatively considered. The two NN OMI ALH algorithms are used (see Sect. 3.5.1): $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ based on OMI R_c(475 nm) and $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ based on MODIS τ(550 nm). A temperature correction is applied to the OMI $N^s_{O_2-O_2}$ prior to the retrievals (see Sect. 3.6.1): **(a)** $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ algorithm and OMLER surface albedo, **(b)** $\mathrm{NN}_{R_c,N^s_{O_2-O_2}}$ algorithm and MODIS black-sky surface albedo, **(c)** $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ algorithm, and OMLER surface albedo and **(d)** $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ algorithm and MODIS black-sky surface albedo.

Nevertheless, while the LIVAS ALH depicts higher aerosol heights in spring than in summer, the OMI ALH shows the opposite. Several explanations are possible: (1) exclusion of OMI scenes with strong aerosol pollution episodes because of a too strict threshold applied on the OMI effective cloud fraction (see Sect. 3.6.1), (2) a more rigorous temperature correction should be applied on measured $N^s_{O_2-O_2}$ (assuming daily instead of seasonal temperature profiles), (3) inaccuracies of MODIS τ, or (4) remaining small cloud residuals in OMI observations in spite of the strict applied filtering. All these elements should be further investigated.

Assuming that LIVAS ALH are here the reference, these results seem to present higher accuracy than the exercise of Park et al. (2016), showing a bias of 1 km of between retrieved OMI and the CALIPSO AEH values over ocean, during the Asian dust event on 31 March 2007. The reasons can be multiple: use of NNs instead of linear interpolation within a LUT, aerosol retrieved over land instead of ocean surfaces, consideration of variable surface albedo as inputs instead of a single value, application of a temperature correction on $N^s_{O_2-O_2}$, use of longer data records.

3.7 Conclusions

In this study, different MLP NN algorithms were developed and evaluated in order to retrieve ALH over land from the OMI 477 nm O_2-O_2 spectral band. The aerosol height was here retrieved as aerosol layer pressure (ALP) and defined as the mid-pressure of an homogeneous scattering layer with a constant geometric thickness. The focus was on north-east Asia and cloud-free scenes dominated by scattering aerosol fine particles with Ångström coefficient $\alpha = 1.5$, single scattering albedo ω_0 in the range of 0.9–0.95 and asymmetry parameter $g = 0.7$. The algorithms were trained with a large ensemble of synthetic data set and several precautions were taken into account to avoid overtraining or local minima problems. The key concept of OMI ALP retrievals is the link between the measured O_2-O_2 SCD $N^s_{O_2-O_2}$ and the aerosol altitude as a consequence of shielding effect applied by the particles on the O_2-O_2 dimer complexes that are at lower altitudes. ALP was retrieved from 3 years, 2005–2007, of OMI cloud-free observations collocated with MODIS Aqua aerosol product in north-east Asia. The main objective of this work is first to evaluate the feasibility of a direct retrieval of this key aerosol parameter from a statistical point of view: i.e. over a long time period and large industrialized continental area, and therefore a high number of observations. All the evaluated performances are summarized in Table 3.2.

Analysis show that a good ALP retrieval requires an accurate prior knowledge of aerosol optical thickness τ as input information. Indeed, both τ and ALP parameters simultaneously contribute to the shielding of O_2-O_2 dimers. The analyses of $N^s_{O_2-O_2}$ alone lead to an ambiguity since aerosol extinction and aerosol altitude cannot be distinguished.

Because low amounts of aerosols have very little impact on $N^s_{O_2-O_2}$ changes, large biases are expected over scenes including aerosol particles with $\tau(550nm) \leq 0.5$. This τ(550 nm) value should be considered as a threshold for a good ALP retrieval quality. Moreover, the algorithms are expected to present a very low sensitivity to particles located at an altitude higher than 4 km. This is because of the nature of the O_2-O_2 complex of which the absorption scales with the pressure-squared instead of being linear with pressure.

Different NN configurations were tested. Sensitivity analyses on synthetic cases show that ALP accuracy with the best NN configuration (i.e. $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ algorithm with true τ(550 nm) value as input) lies in the range of 50–200 hPa (i.e. about 500 m and 1 km) over aerosol scenes with τ(550 nm) in the range of 0.5–1.0. The accuracy is improved to 0–50 hPa over scenes with $\tau(550\ \mathrm{nm}) \geq 1.0$. Using the best available τ information is theoretically expected to limit the impact of uncertainties due to (1) aerosol model, i.e. bias in the range of 0–100 hPa if $\tau(550nm) \leq 1.0$, for a difference of 0.05 in ω_0 or for a difference of 0.1 in g, and (2) surface albedo uncertainty in the range of 0.025–0.05, which leads to absolute ALP biases in the range of 50–100 hPa for τ between 0.5 and 1.0 and 0–50 hPa for $\tau(550\ \mathrm{nm}) \geq 1.0$. Real ALH retrievals were performed over 3 years of OMI O_2-O_2 visible observations over cloud-free scenes with MODIS $\tau(550\ \mathrm{nm}) \geq 1.0$ and using $\mathrm{NN}_{\tau,N^s_{O_2-O_2}}$ combined with MODIS Aqua

τ(550 nm). Comparison of seasonal and spatial averages with the LIVAS climatology database shows maximum (minimum) differences in the range of 260–800 (180–310) m, depending on the season and assuming $\omega_0 = 0.95$ ($\omega_0 = 0.9$). Aerosol model assumptions (in particular ω_0) are the most critical as, in that configuration, they impact OMI seasonal-spatial averaged ALH in the range of 560–660 m. Changes due to the prior surface albedo database (OMLER or MODIS black sky) have a second-order impact, up to 200 m.

In addition, algorithms should take into account that the O_2-O_2 SCD precision, resulting from the DOAS spectral fitting, affects the ALP retrieval. O_2-O_2 SCD precision lying in the range of 0.05–0.25 × 10^{-43} mol^2 cm^{-5} leads to ALP precision between 19 ± 29 and 57 ± 31 hPa. Due to the nature of the O_2-O_2 collision complex, a temperature correction must be applied to the SCD prior to retrievals. Other parameters should be further investigated such as polarization effects and assumptions about the vertical distribution of particles.

An accuracy better than 0.2 must be required on prior τ(550 nm) information. Indeed, a τ(550 nm) bias of 0.25 is expected to bias the retrieved ALP up to 50 hPa for τ(550 nm) in the range of 0.6–1.0. For $\tau(550 \text{ nm}) \geq 1.0$, related ALP impacts become almost null. If no prior accurate τ information, such as from the MODIS aerosol DT algorithm, is available, then this input parameter can be replaced by the OMI continuum reflectance R_c(475 nm) (see the $\text{NN}_{R_c,N^s_{O_2-O_2}}$ algorithm). Indeed, this parameter is primarily affected by the aerosol amount and therefore contains information on τ(550 nm). It may then help to analyse $N^s_{O_2-O_2}$ for retrieving ALP provided that we can retrieve τ with a good quality. Similarly to ALP retrieval, a NN algorithm was also developed to retrieve τ(550 nm) information from the OMI 477 nm O_2-O_2 spectral band. However, its capability is strongly affected by uncertainties on the assumed aerosol model. An overestimation of aerosol single scattering albedo, from $\omega_0 = 0.9$ to 0.95, induces a negative bias of 0.8 for $\tau(550 \text{ nm}) = 2$. The impact is much lower for smaller τ (lower than 0.1 for $\tau(550 \text{ nm}) \leq 0.5$). Similar conclusions were found regarding uncertainty of the asymmetry parameter and thus the phase function characterization. Another major challenge when retrieving aerosol properties from passive satellite sensors is to separate the atmospheric and surface contributions in the total observed reflectance.

Similarly to aerosol ω_0, an overestimation of surface reflection leads to an underestimation of retrieved τ(550 nm). Surface albedo uncertainty below 0.025 should limit OMI retrieved τ(550 nm) bias smaller than 0.5 for τ(550 nm) in the range of 0.0–0.5, 0.25 for $\tau(550 \text{ nm}) = 2.0$. Comparisons of OMI retrievals with collocated MODIS τ(550 nm) show agreements between -0.02 ± 0.45 and -0.18 ± 0.24 depending on the seasons. Further improvements should be made before being able to use these OMI τ(550 nm) products as prior information to ALP retrievals.

Using the $\text{NN}_{R_c,N^s_{O_2-O_2}}$ algorithm for ALP retrieval is, in practice, similar to $\text{NN}_{\tau,N^s_{O_2-O_2}}$ combined with retrieved OMI τ(550 nm). Since the retrieved OMI τ accuracy is lower than MODIS τ accuracy from the DT land algorithm, $\text{NN}_{R_c,N^s_{O_2-O_2}}$ shows reduced performances and higher sensitivity to aerosol model and surface albedo uncertainties. Associated 3-year OMI ALH retrievals over north-east Asia are impacted by ω_0

uncertainties (0.9–0.95) in the range of 540–1200 m and by surface albedo (OMLER vs. MODIS black-sky albedo) up to 730 m.

The NN approach presents, at this stage, quite promising results for a future operational processing of the OMI O_2-O_2 spectral band and the next UV–vis satellite missions such as the TROPOMI (Veefkind et al., 2012). In spite of the high computing time due to the learning database creation and the training of these algorithms, very fast operational processing is allowed. Such processing is much faster than approaches relying on the OEM and employs more optimized interpolation techniques than a classical linear interpolation within a LUT. For future processing of the OMI data, the OMLER climatology database should be optimized by filtering out small aerosol residuals.

Our study indicates that it is worthwhile to design and evaluate aerosol height retrieval algorithm exploiting the satellites 477 nm O_2-O_2 absorption band. Our long-term motivation is to evaluate the feasibility of replacing the effective clouds by more explicit aerosol parameters in the computation of trace gas AMF. This is relevant not only for OMI but for most of the UV–vis satellite missions devoted to air quality monitoring. For that purpose, further analyses must be performed by focusing on significant geophysical variability cases: e.g. pixel-by-pixel variability over smaller regions. Furthermore, single OMI ALH retrievals should be compared with reference aerosol vertical profile measurements (ground-based and/or satellites) over some remarkable case studies.

Chapter 4

Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations

> "Measure what can be measured, and make measurable what cannot be measured."
>
> Galileo Galilei

Abstract

A global picture of atmospheric aerosol vertical distribution with a high temporal resolution is of key importance not only for climate, cloud formation and air quality research studies, but also for correcting aerosol radiation effect in absorbing trace gas retrievals from passive satellite sensors. Aerosol layer height (ALH) was retrieved from the OMI 477 nm O_2-O_2 band and its spatial pattern evaluated over selected cloud-free scenes. Such retrievals benefit from a synergy with MODIS data to provide complementary information on aerosols and cloudy pixels. We used a neural network approach previously trained and developed. Comparison with CALIOP aerosol level 2 products over urban and industrial pollution in eastern China shows consistent spatial patterns with an uncertainty in the range of 462-648 m. In addition, we show the possibility to determine the height of thick aerosol layers released by intensive biomass burning events in South America and Russia, and of a Saharan dust outbreak over sea from OMI visible measurements. Complementary detailed analyses show that the assumed aerosol properties in the modeling are the key factors affecting the accuracy of the results, together with potential cloud residuals in the observation pixels. Furthermore, we demonstrate that the physical meaning of the retrieved ALH scalar corresponds to the weighted average of the vertical aerosol extinction profile. These encouraging findings strongly suggest the potential of the OMI ALH product, and in more general the use of the 477 nm O_2-O_2 band from present and future similar satellite sensors, for climate studies as well as for future aerosol correction in air quality trace gas retrievals.

The contents of this chapter are published as: **Chimot, J.**, Veefkind, J. P., Vlemmix, T., and Levelt, P. F.: Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations, Atmos. Meas. Tech., 11, 2257-2277, https://doi.org/10.5194/amt-11-2257-2018, 2018.

4.1 Introduction

Aerosols are small particles suspended in the air (e.g. desert dust, sea salt, volcanic ashes, sulfate, nitrate and smoke from biomass and fossil-fuel burning). Aerosol sources and sinks are heterogeneously distributed. Due to their scattering and absorption effects on solar and thermal radiation, they redistribute shortwave radiation in the atmosphere. Their presence not only perturbs the air thermal state and stability, our climate system, air quality and meteorological conditions but also interferes with satellite observations of atmospheric trace gases. Aerosols are an important player in the climate system by leading to atmospheric warming, surface cooling and additional atmospheric dynamical responses (IPCC, 2014). By acting as the condensation nuclei on which clouds form, they also modify cloud formation, lifetime and precipitation (Figueras i Ventura and Russchenberg, 2009; Sarna and Russchenberg, 2017). Overall, the climate effects of aerosols are large, but the scientific understanding of their effects remains challenging as their radiative properties is one of the main uncertain components in global climate models (Yu et al., 2006; IPCC, 2014). Finally, the scattering and absorption by aerosols impact the actinic flux and consequently modify the photolysis rates of important processes in the atmosphere (Palancar et al., 2013).

In addition, scattering and absorption of shortwave radiation by aerosols modify the average light path in the atmosphere and therefore interfere with satellite observations of gases, such as NO_2, SO_2, O_3, CO_2, and CH_4, which are important for air quality and climate science objectives. Europe is heavily investing in the development of polar-orbiting and geostationary satellite systems in the Copernicus program (Ingmann et al., 2012), which will form an important component of air quality and climate observing systems on urban, regional, and global scales (Martin, 2008; Duncan et al., 2014). However, inaccurate aerosol correction on these satellite measurements leads to misinterpretations and incorrect evaluations of the implemented emission regulation controls.

The magnitude of the radiative forcing by aerosols depends on the environmental conditions, aerosol properties and horizontal and vertical distribution (IPCC, 2014; Kipling et al., 2016). Its determination requires satellite data in addition to models (IPCC, 2014). While, overall, the horizontal distributions of aerosol optical depth (AOD or τ) and size are relatively well constrained, uncertainties in vertical profile significantly contribute to the overall uncertainty of radiative effects: e.g. 25 % of the uncertainty of black carbon radiative estimations from the models is related to an inaccurate knowledge on the vertical distribution (McComiskey et al., 2008; Loeb and Su, 2010; Zarzycki and Bond, 2010; IPCC, 2014). Knowledge of aerosol vertical profiles allows the computation of related heating rates: e.g. particles located above clouds can increase the liquid water path and geometric thickness of clouds and the subsequent atmospheric heating, and advection of light-absorbing aerosols over the ocean and clouds from rice straw burning in China can strongly reduce clouds and Earth radiant energy (Hsu et al., 2003; de Graaf et al., 2012; Wilcox, 2012). Therefore, aerosol layer height (ALH) drives not only the magnitude but also the sign of aerosol direct and indirect radiative effects (Kipling et al., 2016). Current ALH simulated by climate models can differ in the range of 1.5-3 km (Koffi et al., 2012; Kipling et al., 2016).

Furthermore, in the absence of clouds, vertical distribution of aerosols is one of the most significant error sources in trace gas retrievals from satellites (Leitão et al., 2010; Chimot et al., 2016). Major biases on the pollutant tropospheric NO_2 measured by satellites, depending on AOD and ALH, can be expected if no aerosol correction is applied. Because such information is not available for every observation, aerosols are approximated via a simple cloud model (Acarreta et al., 2004; Boersma et al., 2011; Veefkind et al., 2016). This only leads to a first-order correction for short-lived gases NO_2, SO_2, and HCHO), that does not comprehensively assume the full scattering and absorbing effects of aerosol particles on the average light path followed by the detected photons (Boersma et al., 2011; Chimot et al., 2016). In particular, current uncertainties on ALH lead to substantial biases in areas with high AOD (> 0.5), and absorbing and elevated particles: between -26 % and -40 % on the retrieved tropospheric NO_2 columns from the Dutch-Finnish Ozone Monitoring instrument (OMI) (Castellanos et al., 2015; Chimot et al., 2016), 20-50 % on Global Ozone Monitoring Experiment-2 (GOME-2) and SCIAMACHY HCHO (Barkley et al., 2012; Hewson et al., 2015), and about 50 % on OMI SO_2 (Krotkov et al., 2008). ALH also remains one of the largest error sources for greenhouse gas retrievals: e.g. CO_2 from the American carbon OCO-2 mission (Crisp, 2015; Connor et al., 2016; Wunch et al., 2017) and CH_4 from the future TROPOMI on board Sentinel-5 Precursor (Hu et al., 2016).

Consequently, determining ALH with a large coverage (ideally daily and global) and an uncertainty better than 1 km (as a first approximation), for every single absorbing trace gas atmospheric satellite pixel, is ideally needed. Active satellite sensors, such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), allow us to probe detailed vertical aerosol profile, but with a limited coverage as they only look towards the nadir. This can lead to a gap up to 2200 km(in the tropics and subtropics) between adjacent orbital tracks. As an alternative, passive satellite sensors, with a high spectral resolution such as OMI, offer adequate spatial coverage with a good temporal resolution (up to daily global before the OMI row anomaly development) thanks to a wide swath. Thus, passive hyperspectral instruments can provide great contribution even if they do not achieve the same level of accuracy as active instruments (i.e. limited vertical resolution, only cloud-free scenes). Because molecular oxygen (O_2) is well mixed, its slant column measurement provides a suitable proxy for the determination of the modified scattering height due to aerosols, in the absence of clouds. Most of the developed ALH retrieval algorithms from backscattered sunlight satellite measurements focus on the absorption spectroscopy of the O_2 A band around 765 nm, relatively close to the CO_2 and CH_4 absorption bands (Wang et al., 2012; Sanders et al., 2015). Some studies also focus on the use of the O_2 B band (Ding et al., 2016; Xu et al., 2017). So far, only a few studies have yet worked on using the O_2-O_2 satellite absorption bands, within the ultraviolet (UV) and visible (vis) spectral ranges, to retrieve ALH and τ (Park et al., 2016; Chimot et al., 2017). These bands are spectrally closer to the NO_2, SO_2 and HCHO absorption lines. Contrary to the O_2 A band, the O_2-O_2 477 nm band presents a wider (over 10 nm) but weaker spectral absorption. This leads to high sensitivities in the case of strong aerosol loading and less challenges due to saturation. Moreover, in the visible spectral range, AOD values are generally higher while surface albedo or reflectance is lower, leading to a higher contrast between aerosol and surface scattering signals. The 477 nm O_2-O_2 is not only present in the present GOME-2, OMI and TROPOMI satellite sensors, but will be also included in the future Sentinel-4 and Sentinel-5 instruments (Ingmann et al., 2012; Veefkind et al., 2012). They should be then more studied in detail in terms of information content about particles.

This paper follows the exploratory study of Chimot et al. (2017), in which a neural network (NN) algorithm has been developed to investigate the feasibility of deriving ALH from the OMI 477 nm O_2-O_2 spectral band over cloud-free scenes. The main objective was the study of anthropogenic particles emission and their precursors from vehicles, coal burning, and industries. It has allowed to retrieve ALH over land for the first time from this specific spectral band. A statistic evaluation of 3-year cloud-free OMI observations over eastern Asia, focusing on urban and large industrialized areas, has shown maximum differences below 800 m with a reference climatology database. In order to complete this first and statistically focused evaluation, the present study evaluates the spatial distribution of the OMI 477 nm O_2-O_2 ALH product on a pixel-by-pixel basis. It therefore focuses on its variability for single days. For that purpose, specific cloud-free case studies are selected, including 3 winter days with strong anthropogenic pollution over eastern China. In addition, to extend the performance assessment of such an approach beyond the initial objective of Chimot et al. (2017), new types of aerosol pollution episodes are investigated: 4 summer days with large biomass burning events in South America and east of Russia and 1 day of wide desert dust transport over sea. The OMI ALH retrieval is compared with CALIOP level 1 (L1) measurements and level 2 (L2) aerosol retrievals.

4.2 OMI, MODIS and CALIOP aerosol observations

4.2.1 The OMI sensor and O_2-O_2 477 nm spectral band

The Dutch-Finnish OMI mission (Levelt et al., 2006) is a nadir-viewing push-broom imaging spectrometer launched on the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Aura satellite. It delivers global coverage with a high temporal resolution of key air quality components derived from measurements of the backscattered solar radiation acquired in the UV-vis spectral domain (270-550 nm) with approximately 0.5 nm resolution. Based on a two-dimensional detector array concept, radiance spectra are simultaneously measured on a 2600 km wide swath within a nadir pixel size of 13x24 km^2 (28x150 km^2 at extreme off-nadir).

The O_2-O_2 477 nm absorption band is currently operationally exploited by the OMI O_2-O_2 cloud algorithm (OMCLDO2) to derive effective cloud parameters (Acarreta et al., 2004; Veefkind et al., 2016). This spectral band directly measures the absorption of the visible part of the sunlight induced by the O_2-O_2 collision complex along the whole light path. A spectral fit, prior to the OMI effective cloud retrieval algorithm, is performed over the 460-490 nm spectral range to derive the continuum reflectance $R_c(475\,\mathrm{nm})$ and the O_2-O_2 slant column density $N^s_{O_2-O_2}$. This spectral fit relies on the differential optical absorption spectroscopy (DOAS) approach (Platt and Stutz, 2008). $N^s_{O_2-O_2}$ represents the O_2-O_2 absorption magnitude along the average light path through the atmosphere. This is the key input variable for the OMI ALH retrieval by the NN algorithm.

4.2.2 The OMI aerosol layer height neural network algorithm

The OMI ALH retrieval algorithm (Chimot et al., 2017) is based on the exploitation of $N^s_{O_2-O_2}$ derived from the DOAS fit, and relies on a NN approach. Here, the main elements of this algorithm are summarized but for more details about their technical development and implementation, see Chimot et al. (2017).

This algorithm relies on how aerosols affect the length of the average light path along which the O_2-O_2 absorbs. $N^s_{O_2-O_2}$ is then driven by the overall shielding or enhancement effect of photons by the O_2-O_2 complex in the visible spectral range, due to the presence of particles. An aerosol layer located at high altitudes will apply a large shielding effect on the O_2-O_2 located in the atmospheric layers below: i.e. the amount of photons coming from the top of the atmosphere (TOA) and reaching the lowest part of the atmosphere is reduced compared to an aerosol-free scene. This shielding effect is then larger when the aerosol layer is located at an elevated altitude than close to the surface.

The designed NNs belong to the family of machine learning and the artificial intelligence domain and rely on a multi-layer architecture, also named multilayer perceptron. The input and output variables are interconnected through a set of sigmoid functions present in the hidden layers and the synaptic weights W. For each single sigmoid function, two simple operations are performed: (1) a weighted sum of all the inputs given by the previous layer and (2) a transport of this sum through the sigmoid functions. The ALH retrieval problem then becomes a simple series of analytical functions.

Aerosol profiles are assumed as one single scattering layer (also called "box layer") with a constant geometric thickness (100 hPa, or about 1 km). The particles included in this layer are homogeneous (i.e. same size and optical properties). ALH is then defined as the mid-altitude (above sea level) of this scattering layer. Furthermore, aerosol particles are assumed to cover the entire satellite observation pixel. The input layer contains seven parameters: viewing zenith angle θ, solar zenith angle θ_0, relative azimuth angle ϕ_r, surface pressure Ps, surface albedo A, aerosol optical thickness $\tau(550\ \mathrm{nm})$ and the OMI $N^s_{O_2-O_2}$. As explained in Chimot et al. (2017), a prior $\tau(550\ \mathrm{nm})$ is required as input as both ALH and $\tau(550\ \mathrm{nm})$ simultaneously affect $N^s_{O_2-O_2}$ and need to be distinguished.

The optimal weights were estimated through a rigorous training task following the error back propagation technique and a training dataset that includes a set of representative situations for which inputs and outputs are well known. The quality of the training dataset was ensured by full physical spectral simulations, dominated by aerosol particles without clouds, generated by the Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval (DISAMAR) software of KNMI (de Haan, 2011). Aerosol scattering is simulated by a Henyey–Greenstein (HG) scattering phase function $\Phi(\Theta)$ parameterized by the asymmetry parameter g, which is the average of the cosine of the scattering angle (Hovenier and Hage, 1989). Aerosols were specified for a standard case, assuming fine particles with a unique value of the extinction Ångström exponent $\alpha = 1.5$ and $g = 0.7$. In order to investigate the assumptions related to the aerosol single scattering albedo ω_0 properties, two training datasets were generated with a different typical value: one with $\omega_0 = 0.95$ and one with $\omega_0 = 0.9$ in the visible spectral domain. Therefore, two OMI ALH NN algorithms have been created, one for each aerosol ω_0 values.

The aerosol models in the training database were based on a HG scattering phase function for two reasons. First, both Chimot et al. (2017) and this paper are exploratory studies focusing on the potential of exploiting the O_2-O_2 spectral band for aerosol retrievals from a satellite sensor. Second, our first long-term objective is the potential use of the ALH parameter for future tropospheric NO_2 and similar trace gas retrievals over cloud-free scenes. Several studies emphasized that ALH is the key variable affecting the length of the average light path in the computation of the related air mass factor (AMF) computation through the DOAS approach (Boersma et al., 2004; Castellanos et al., 2015; Chimot et al., 2016). This is because the only quantity that is relevant for absorption by trace gases in the visible is the light path distribution, i.e. the distribution of distances travelled by photons in the atmosphere before leaving the atmosphere. The absolute radiance at the TOA is less important. The second variable of interest is τ. This light path distribution is mostly governed by ω_0 and g, and of course ALH and τ, and much less by details in the phase function. Studies by Leitão et al. (2010); Castellanos et al. (2015); Chimot et al. (2016) showed the lower sensitivity of the AMF to α, ω_0, and g. These scattering parameters ω_0 and g are included in HG scattering, parameterized the phase function, and therefore can be used for AMF calculations. At this level and with respect to these mentioned objectives, it is then assumed one do not need to define more realistic aerosol models for every single OMI pixel. With $g = 0.7$, the HG function is known to be smooth and reproduce the Mie scattering functions reasonably well for most of aerosol types (Dubovik et al., 2002). Such an approach is used for the preparation of the operational ALH retrieval algorithms for Sentinel-4 and Sentinel-5 Precursor (Leitão et al., 2010; Sanders et al., 2015; Colosimo et al., 2016; Nanda et al., 2017), and for various explicit aerosol corrections in the AMF calculation when retrieving trace gases, such as tropospheric NO_2, over urban and industrial areas dominated by anthropogenic pollution, for example in eastern China (Spada et al., 2006; Wagner et al., 2007; Castellanos et al., 2015; Vlemmix et al., 2010). The potential impact of the modelled scattering phase discussion is kept in mind and further discussed in Sect. 4.4.4. However, reperforming the whole NN training process with more complex particle shape models is computationally very demanding and beyond the scope of this paper. Instead, more elements on specific error analysis are further discussed in Sect. 4.4.

Maximum seasonal differences between the LIdar climatology of Vertical Aerosol Structure for space-based lidar simulation studies (LIVAS) and 3 year OMI ALH, over cloud-free scenes in north-eastern Asia with MODIS $\tau(550m) \geq 1.0$, are in the range of 180–800 m (Amiridis et al., 2015; Chimot et al., 2017). The previous extended sensitive study has shown the following: (a) Due to the nature of the O_2-O_2 spectral band, a minimum particle load (i.e. $\tau(550nm) = 0.5$) is required to be able to exploit the aerosol signal as, below this threshold, low amounts of aerosols have negligible impacts on $N^{s}_{O_2-O_2}$ shielding and lead to high ALH bias, (b) The aerosol model assumptions are the most critical, in particular ω_0, as they may affect ALH retrieval uncertainty up to 660 m, (c) In addition, potential aerosol residuals in the prior surface albedo may impact up to 200 m, (d) An accuracy of 0.2 is required on prior τ(550 nm) to limit ALH bias close to zero when $\tau(550nm) \geq 1.0$ and below 500 m for τ(550 nm) values close to 0.6.

4.2.3 The CALIOP and MODIS aerosol products

CALIOP sensor is a standard dual-wavelength elastically backscattered lidar on board the CALIPSO satellite platform, flying since 2006. Equipped with a depolarization channel at 532 nm, it probes the aerosol and cloud vertical layers, from the surface to 40 km above sea level, with a high vertical resolution (Winker et al., 2009). Level 1 scientific data products, distributed by the Atmospheric Science Data Center (ASDC) of NASA, include the lidar calibrated and geolocated measurements of high-resolution vertical profiles (between 30 and 60 m in the troposphere) of the aerosol and cloud attenuated backscatter coefficients at 532 and 1064 nm with horizontal resolutions of 1/3 km, 1 km and 5 km (Winker et al., 2009).

The CALIOP aerosol L2 product contains the retrieved aerosol backscatter and extinction coefficient profiles at 532 and 1064 nm, for each identified and well-located aerosol layer, at 5 km horizontal resolution. These retrievals are performed after calibration, range correction, feature detection and classification, and assumptions on lidar extinction-to-backscattering ratio (Winker et al., 2009; Young and Vaughan, 2009).

The MODIS spectrometer was launched on the NASA EOS Aqua platform in May 2002 and has been delivering continuous images of the Earth in the visible, solar and thermal infrared approximately 15 min prior to OMI. The considered MODIS Aqua L2 aerosol product is the Collection 6 of MYD04_L2, based on the Dark Target (DT) and Deep Blue (DB) algorithms with a high enough quality assurance flag and an improved calibration of the instrument (Levy et al., 2013). While the MODIS measurement is acquired at the resolution of 1 km, the used MODIS aerosol τ(550 nm) is at 10 km x 10 km, relatively close to the OMI nadir spatial resolution. The expected uncertainties of MODIS τ(550 nm) are about $\pm 0.05 + 15\,\%$ over land for DT (Levy et al., 2013), and about ±0.03 on average for DB (Sayer et al., 2013).

4.3 Case studies: results and discussion

4.3.1 Methodology

OMI ALH retrievals are here obtained using MODIS L2 aerosol τ(550 nm) from the combined DT and DB product as prior input, collocated within a distance of 15 km and where τ(550 nm) ≥ 0.55 (see Sect. 4.2.2). Mitigating the probability of cloud contamination within the OMI pixel is one of the first criteria for a successful ALH retrieval. For that purpose, we rely on the availability of the MODIS aerosol product with the highest quality assurance flag ensuring that MODIS Aqua τ(550 nm) is exclusively estimated when a sufficient high amount of cloud-free sub-pixels is available (i.e. at the MODIS measurement resolution of 1 km) (Levy et al., 2013). However, since this may be not completely representative of the atmospheric situation of the OMI pixel, two thresholds are added for each collocated OMI-MODIS pixel: the geometric MODIS cloud fraction to be smaller than 0.1, and the effective OMI cloud fraction lower than 0.2. For this last parameter, it was shown that values higher than 0.3 are generally likely contaminated by clouds while values between 0.1 and 0.2 may be cloud-free but contain a substantial amount of very scattering particles enhancing then the scene brightness (Boersma et al., 2011; Chimot et al., 2016).

The ALH retrievals are applied to the OMI DOAS O_2-O_2 observations, available in the last reprocessed OMCLDO2 product version (Acarreta et al., 2004; Veefkind et al., 2016). A temperature correction is taken into account on the $N^{s}_{O_2-O_2}$ variable, using the information available in the OMCLDO2 product which is itself based on the temperature profiles of the National Centers for Environmental Prediction (NCEP) analysis data (Veefkind et al., 2016).

The selected case studies include 1) urban and industrial aerosol pollution over eastern China during 3 days between October and November 2006, 2) large wildfire episodes in South America in August 2006 and September 2007 and in eastern Russia in August 2010 and June 2012, and 3) a Saharan dust transport over sea in June 2012. OMI ALH retrievals are compared with collocated CALIOP products within a distance of 50 km for the cases over east China and South America, 300 km for west Sahara and eastern Russia. The larger OMI-CALIOP distance over these two last regions is due to the so-called "row anomaly" which has been significantly perturbing OMI measurements of the earthshine radiance at all the wavelengths since 2009. This leads to a reduced number of valid OMI ground pixels close to the CALIOP track. Details are given at `http://www.knmi.nl/omi/research/product/rowanomaly-background.php`.

For each case study, the most likely suitable NN algorithm (see Sect. 4.2.2) is selected by hand. We decided to rely on 1) the OMI UV aerosol absorbing index (UVAI) and 2) their well-known absorbing properties (according to the literature) in the visible spectral range in order to approximate the assumption on aerosol ω_0 at the visible (460-490 nm) spectral wavelengths. OMI UVAI is derived by the OMI near-UV aerosol algorithm (OMAERUV) in the 330-388 nm spectral band (Torres et al., 2007). It allows us to detect and distinguish UV absorbing from scattering aerosols through the measured change of spectral contrast, with respect to a pure Rayleigh atmosphere. Weakly absorbing or large non-absorbing particles are associated with near-zero or negative UVAI values. A threshold of 1 on UVAI is then specified to detect absorbing particles in the UV and then potentially in the visible.

4.3.2 Urban aerosol pollution

Fossil-fuel combustion is the main source of air pollution in the large urban and industrialized area of eastern China. With decreasing temperatures in autumn, coal-burning power plant activity is increased due to a higher energy consumption of heating systems. Consequently, excessive amounts of aerosol particles and their precursors are emitted (Chameides et al., 1999). Moreover, crop residue burning in the agricultural areas of eastern Asia may enhance aerosol concentrations (Xue et al., 2014). Mineral dust particles, from the Taklamakan and Gobi deserts between middle of spring and end of autumn, are transported through westerly winds (Eck et al., 2005; Proestakis et al., 2017). Collectively, the mix of all these pollutants contributes to the formation of regional brown hazes greatly threatening public health, over the North China Plain during the dry season (from October to March). They have been frequently detected by satellite and ground-based observations (Ma et al., 2010).

Three typical days between October and November 2006 in eastern China were selected to illustrate the performance of the NN algorithm over scenes with strong urban aerosol pollution: day 1 of 2006.10.02, day 2 of 2006.10.06 and day 3 of 2006.11.01. As illustrated by the maps in Fig. 4.1, these days are characterized by high τ values over land as shown by MODIS Aqua: τ(550nm) in the range of 0.5-1.6 in October 2006, and 0.5-1.3 in November 2006.

Lin et al. (2015) estimated ω_0 values in summer (and likely beginning of autumn) in the range of 0.94-0.96 in the visible. This is likely a consequence of lower black carbon particle amounts at that time (compared to winter and spring) and a high dominance of anthropogenic particles such as nitrate and sulfate. These particles may also be mixed, in part, with desert dust. Consistently, OMI UVAI depicts for the selected days values lower than or close to 1 (see Fig. 4.1). Therefore, we use the NN algorithm trained with $\omega_0 = 0.95$ assuming low abundance of UV and visible particles.

Figure 4.2 depicts the spatial distribution of retrieved OMI ALH for all the selected collocated OMI-MODIS pixels, with a variability between 0.5 and 3 km. The CALIPSO suborbital tracks were mostly located inland in days 1 and 3, and between inland and over sea in day 2 (see Fig 4.1 and Fig. 4.2).

The aerosol layers in the CALIOP L2 product, based on the total backscatter coefficients (532 nm), are generally located between the surface and 1.5 km height (see Fig. 4.3). Maximum top heights do not exceed 2 km on 2006.10.06 or 3 km on the 2 other days. Collocated OMI ALH are mostly located in the middle aerosol layers, and rarely exceed the top and bottom layer limits (see Fig. 4.2). Overall, for the 3 selected days, the OMI NN retrievals reproduce the spatial CALIOP L2 patterns. On

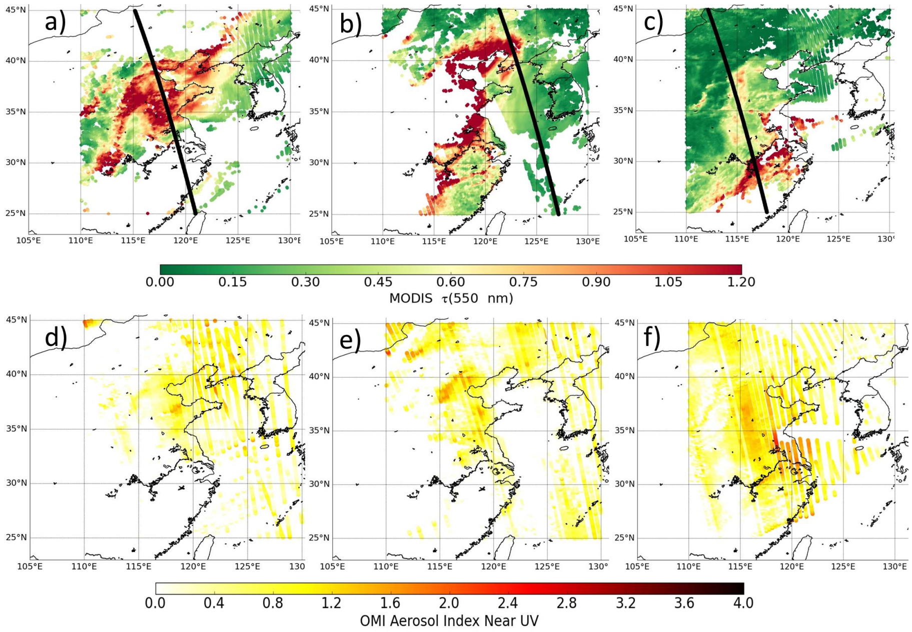


Figure 4.1: Maps of MODIS Aqua τ(550 nm) from the combined DT and DB Collection 6 (see Sect. 4.2.3), and collocated OMI aerosol index from near-UV (UVAI) values (see Sect. 4.3) over cloud-free scenes for the urban and industrialized cases in eastern China. The dark thick lines represent the track of CALIPSO space-borne sensor over the selected case studies: **(a, d)** 2006.10.02, **(b, e)** 2006.10.06, and **(c, f)** 2006.11.01.

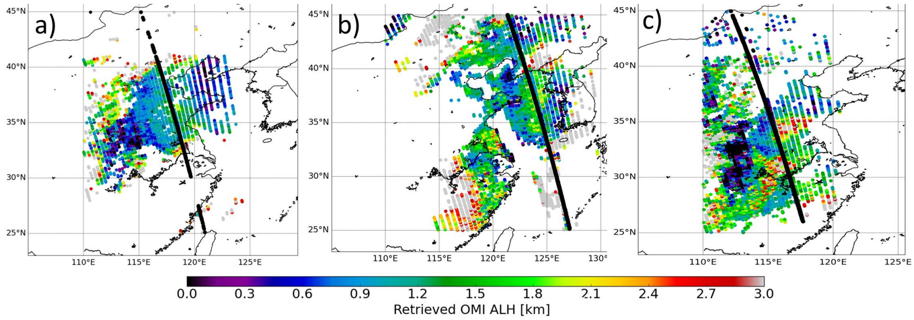


Figure 4.2: Maps of retrieved OMI aerosol layer height (ALH) from all the cloud-free pixels collocated with MODIS Aqua τ(550 nm) (see Fig. 1). The dark thick lines represent the track of CALIPSO space-borne sensor over the selected case studies: **(a)** 2006.10.02, **(b)** 2006.10.06, and **(c)** 2006.11.01.

2006.10.02 in particular, OMI ALH remains relatively stable at the average altitude of 1 km, within the CALIOP L2 aerosol layers (see Fig. 4.3a). Only at the latitude 36.5° N do both products simultaneously show an increased altitude close to 3 km. On the 2 other selected days, OMI ALH and CALIOP L2 show simultaneously descending slopes from south to north: a slope of about 2 km over 2.5° latitude on 2006.10.06, and around 1.5 km over 8° latitude on 2006.11.01 (see Fig. 4.3b and Fig. 4.3c).

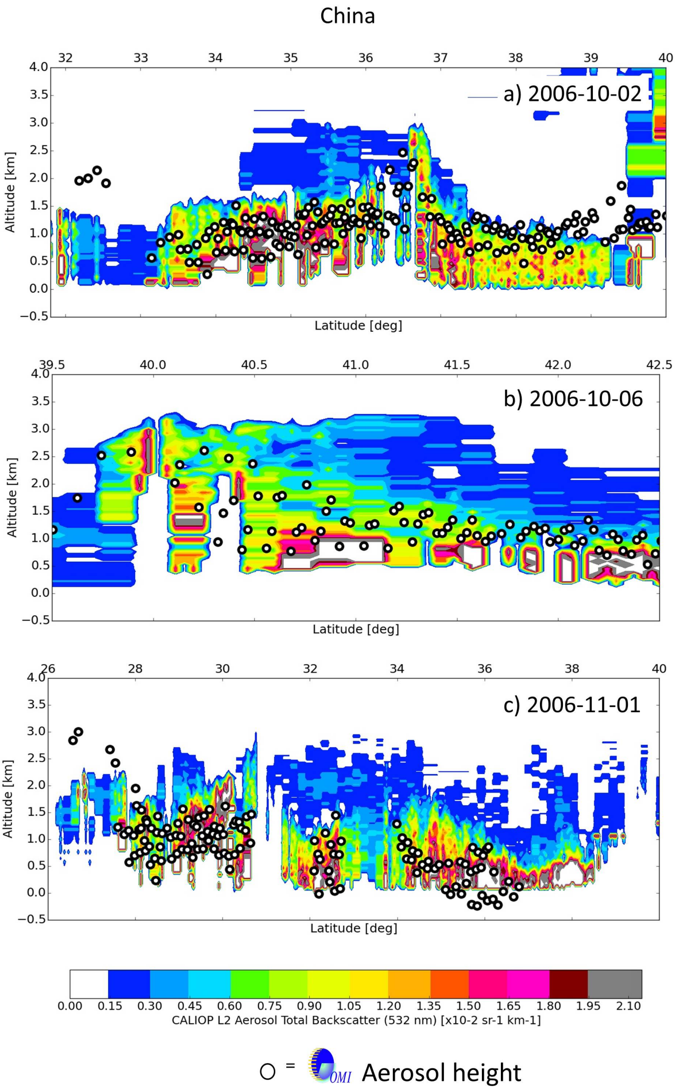


Figure 4.3: Retrieved OMI ALH compared with vertical profile of aerosol total backscatter coefficient (532 nm) from the CALIOP L2 product. Maximal distance between OMI pixels and CALIOP ground track is 50 km. Only cloud-free OMI pixels, collocated with MODIS-Aqua Collection 6 aerosol cells, $\tau(550nm) \geq 0.55$ (from the MODIS DT and DB algorithms), are selected: **(a)** 2006.10.02, **(b)** 2006.10.06, and **(c)** 2006.11.01.

An equivalent CALIOP L2 ALH can be derived by calculating an aerosol extinction weighted average altitude as follows:

$$\mathrm{ALH(CALIOPL2)} = \frac{\sum\limits_{\mathrm{l}} h(l)\sigma(l)}{\sum\limits_{\mathrm{l}} \sigma(l)}. \tag{4.1}$$

where $\sigma(l)$ is the CALIOP aerosol extinction (532 nm) of the vertical layer l defined by its mid-altitude $h(l)$.

In Figure 4.4, root-mean-square deviation (RMSD) between OMI and CALIOP L2 ALH lies in the range of 462-648 m when the maximum distance between the selected OMI and CALIOP ground pixels is lower than 50 km and with collocated MODIS $\tau(550nm) \geq 0.55$ (see Fig. 4.3).

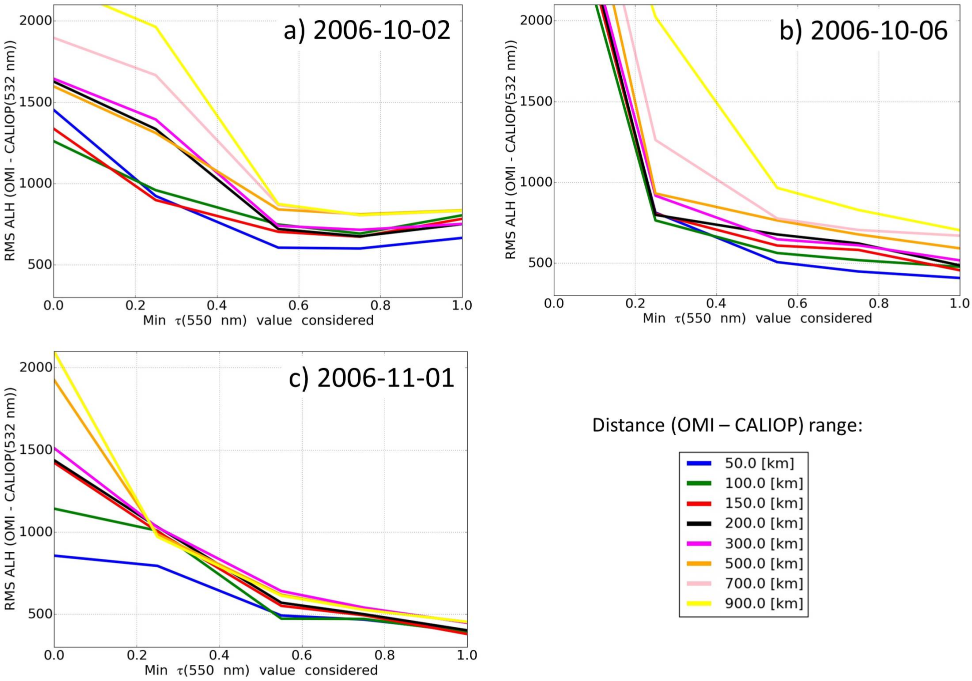


Figure 4.4: Root-mean-square (RMS) deviation given in m between collocated retrieved OMI ALH and derived CALIOP ALH (532 nm) (see Sect. 4.4.1) for urban and industrialized cases over eastern China as a function of minimum MODIS τ(550 nm) and distance between OMI and CALIOP ground pixels: **(a)** 2006.10.02, **(b)** 2006.10.06, and **(c)** 2006.11.01.

Associated bias values (i.e. average difference between OMI and CALIOP ALH per day) are between -86 and -128 m. These results significantly deteriorate, firstly when specifying a lower threshold on collocated MODIS τ(550 nm) (e.g. RMSD $\geq$ 1000 m with all MODIS τ(550 nm) values included), and secondly with a more flexible distance criterion (e.g. RMSD in the range of 594-888 m with a maximum distance of 500 km between the selected OMI and CALIOP ground pixels). The relatively low impact, noticed here, on the distance between OMI and CALIOP pixels is probably related to the large spatial extent of aerosol plumes and their relative spatial homogeneity. The impact of distance between collocated OMI-CALIOP pixels would be more detrimental over scenes with smaller and/or more heterogeneous plumes. Figure. 4.5 shows the one-to-one comparison between OMI and CALIOP L2 ALH within a distance of 50 km per case study and as a function of associated MODIS τ(550 nm). The correlation coefficient (R) between OMI and CALIOP ALH varies per day, between 0.4 and 0.6 for all scenes with MODIS $\tau(550nm) \geq 0.55$.

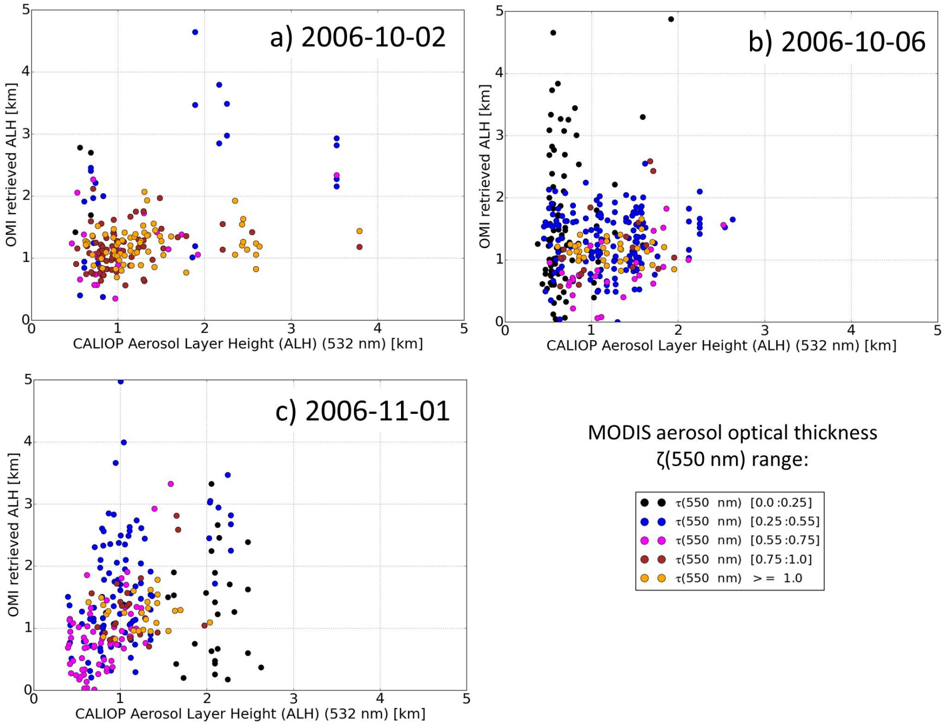


Figure 4.5: Scatter-plot of collocated retrieved OMI ALH and derived CALIOP ALH (532 nm) (see Sect. 4.4.1) for urban and industrialized cases over eastern China as a function of MODIS τ(550 nm). Distance between OMI and CALIOP pixels is 50 km: **(a)** 2006.10.02, **(b)** 2006.10.06, and **(c)** 2006.11.01.

4.3.3 Smoke and absorbing aerosol pollution from biomass-burning

Intensive biomass burning releases large amounts of carbonaceous and black carbon aerosols. The resulting dense smoke layers have a predominance of fine and strongly light-absorbing particles, especially in both the UV and visible visible spectral ranges. Combined with large τ values, this yields large light extinction and Angstrom exponents (≥ 1.5) (Torres et al., 2013; Wu et al., 2014). Figure 4.6 shows the location and associated MODIS τ(550 nm) and OMI UVAI values for the selected biomass burning episodes: the two first events are over South America on 2006.08.24 and 2007.09.30; the two last events are over eastern Russia on 2010.08.08 and 2012.06.23. Due to the very high load of absorbing particles with MODIS $\tau(550nm \geq 1.1)$, OMI UVAI values are generally higher than 2 and can locally reach 4, suggesting then the use of the NN algorithm trained with $\omega_0 = 0.9$ (see Fig. 4.6).

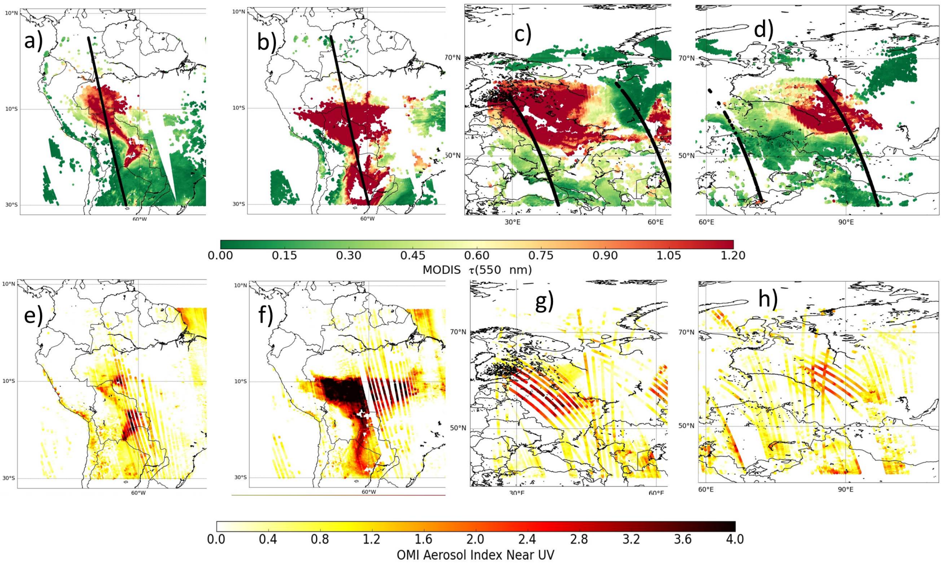


Figure 4.6: Maps of MODIS Aqua $\tau(550 \text{ nm})$ from the combined DT and DB Collection 6 (see Sect. 4.2.3), and collocated OMI aerosol index from near-UV (UVAI) values (see Sect. 4.3) over cloud-free scenes and intensive biomass burning episodes. The dark thick lines represent the track of CALIPSO space-borne sensor over the selected case studies: **(a,e)** South America on 2006.08.24, **(b,f)** South America on 2007.09.30, **(c,g)** eastern Russia on 2010.10.08, and **(d,h)** eastern Russia on 2012.06.23.

Several studies have identified loss of sensitivity of CALIOP attenuated backscatter profile measurements at 532 nm over scenes with dense smoke layers, such as over Canadian boreal and Amazonian fire events (Kacenelenbogen et al., 2011; Torres et al., 2013; Wu et al., 2014). Light extinction due to these layers is much larger at 532 nm than at 1064 nm (Pueschel and Livingston, 1990). Since CALIOP does not directly measure the aerosol backscattering but rather the attenuated backscattering, the range-dependent reduction in CALIOP lidar signals due to attenuation occurs more rapidly in the short wavelengths. Therefore, over scenes with heavy smoke particle loads, the attenuated backscatter coefficients (532 nm) in the lower part of the aerosol layer fall below the CALIOP's detection threshold, preventing the identification of the full vertical extent of the aerosol layers (from the top to the bottom). Being a down-looking observation lidar system, CALIOP tends then to mostly detect the top height compared to the base height of the aerosol layer as the laser's energy undergoes substantial attenuation when the beam travels through an optically thick layer (Vaughan et al., 2005; Kim et al., 2013). Therefore, the identified layers that are fully attenuated at 532 nm, in the L1 product, are filtered out in the L2 product. As a consequence of this filtering, CALIOP's τ of smoke layers is generally underestimated due to an overestimation of the layer base altitude (Wu et al., 2014; Kim et al., 2013).

Figure 4.7 depict an example of the loss of sensitivity for a biomass burning case in South America. The CALIOP aerosol total backscatter (532 nm) and backscatter (1064 nm) coefficients in the L2 product mostly show the top layer of carbonaceous aerosols in the range of 3-4 km altitude with maximum thickness of 1 km, between 14° S and 11° S (see Fig. 4.7a and 4.7b). We found that the layers located below are flagged as totally attenuated at the wavelength of 532 nm, according to the CALIOP vertical feature mask. On the northernmost end of the detected plume, the aerosol load is around 2 km height. In contrast, the CALIOP L1 attenuated backscatter (1064 nm) profile detects an aerosol layer between the surface and 1.5 km, at the latitudes 11-14° S, not visible the CALIOP L1 total attenuated backscatter (532 nm) profile, (see Fig. 4.7c and 4.7d) likely due to a better sensitivity of this channel to the particles located close to the surface.

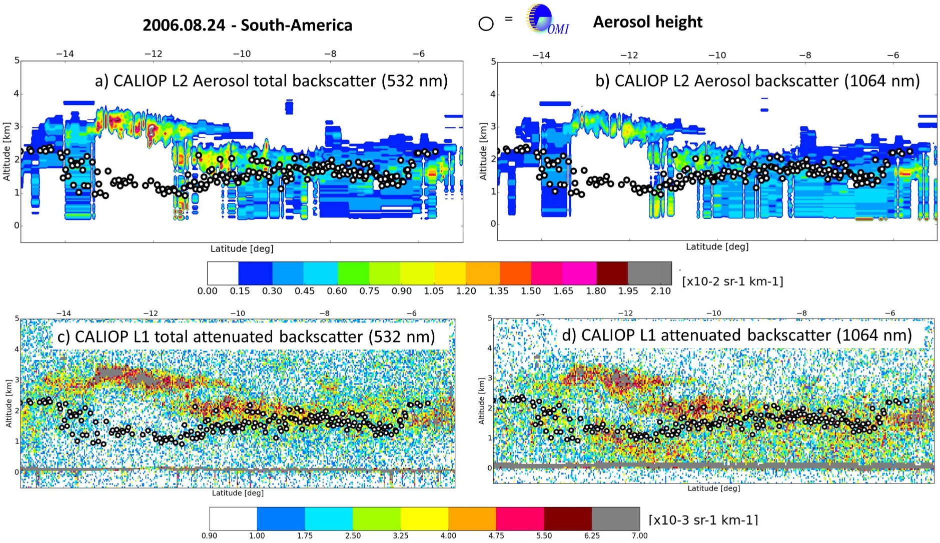


Figure 4.7: Retrieved OMI ALH compared with CALIOP along-track vertical profile observations for biomass burning case over South America: **(a)** CALIOP L2 aerosol total backscattering (532 nm), **(b)** CALIOP L2 aerosol backscattering (1064 nm), **(c)** CALIOP L1 attenuated backscattering (532 nm), and **(d)** CALIOP L1 attenuated backscattering (1064 nm).

The case of 2006.08.24 over South America shows the retrieved OMI ALH being well located, i.e. below the first elevated aerosol layer (at about 3 km) and at the top of the second aerosol layer, close to the surface (see Fig. 4.7d). Contrary to the CALIPSO L1 measurement (532 nm), our retrievals based on OMI visible measurements are not restricted to the top of the smoke or absorbing layer but correctly match with the middle of the layers detected by CALIPSO L1 (1064 nm). The reason that our OMI ALH seems closer to the top of the second layer may be due to a higher aerosol load and/or different layer properties (see. Sect. 4.4.2).

Figure 4.8 shows that similar full CALIOP attenuation processes occur with the other selected biomass burning cases. The CALIOP L1 total attenuated backscatter (532 nm) vertical profiles mostly correlate with the top of the detected aerosol layers, while the CALIOP L1 attenuated backscatter (1064 nm) profiles reveal lower layers. On days of 2007.09.30 and 2010.08, the top layers are at elevated altitudes (higher than 3 km), while the lower ones extend from the surface to 1-2 km. On the last day, 2012.06.23, the top layer is lower (between 1 and 2 km). Similarly, all the OMI ALH retrievals are not restricted to the top layers but match, most of the time, with the middle of the layers, sometimes a bit closer to the base of the top layer or the top of the bottom layer. This may depend on the differences in terms of AOD and/or optical properties of each layer (see. Sect. 4.4.2). In addition, it is worth to noticing: the similar vertical variability (around 500 m) on 2007.09.30 in South America at the latitudes 8-15° S, and the remarkable descending slope, on 2012.06.23 in eastern Russia, from north to south at the latitudes 56-58° S present in both OMI ALH and CALIOP L1 products.

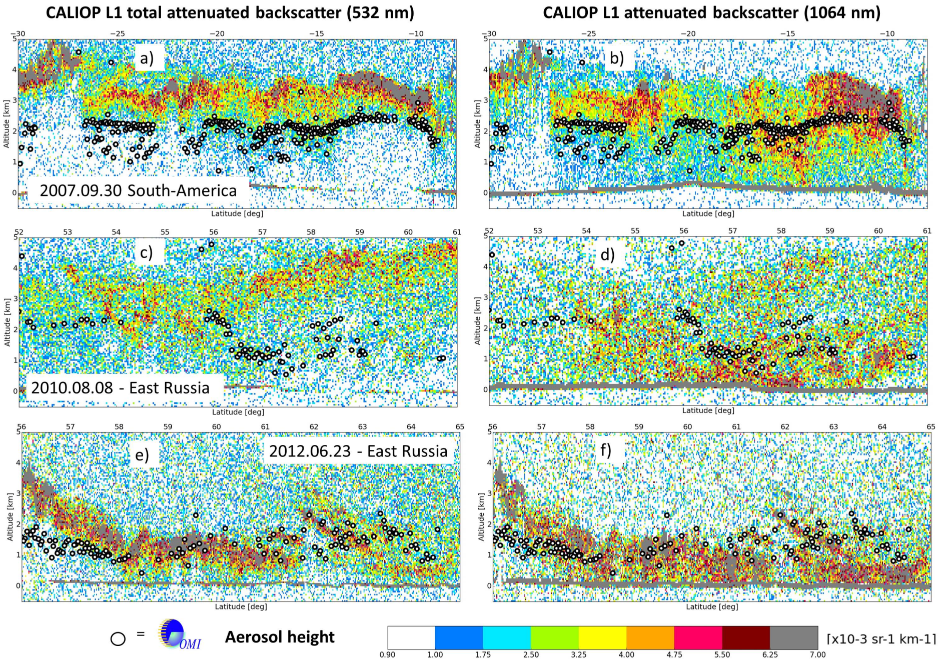


Figure 4.8: Retrieved OMI ALH compared with CALIOP L1 along-track vertical profile observations (532 and 1064 nm) for biomass burning cases: **(a,b)** 2007.09.30 in South America, **(c,d)** 2010.08.08 in eastern Russia, and **(e,f)** 2012.06.23 in eastern Russia.

Three reasons may explain why OMI visible spectra allow us to probe an entire absorbing aerosol layer, contrary to the active satellite visible measurement of CALIOP: (1) OMI measurements rely on the sun irradiance which is much more intense than the laser pulse of CALIOP; (2) OMI measurements are largely issued from multiple scattering effects occurring at different altitudes, allowing a higher number of photons to reach the lower atmospheric layers; (3) the relatively higher signal-to-noise ratio (SNR) of OMI likely allows us to better detect and exploit the upcoming signal from smoke layers. In contrast, contributions of multiple scattering to the CALIOP backscattered signals are lower than single scattering effects within moderately dense dust layer and insignificant within smoke aerosol extinction (Winker, 2003; Liu et al., 2011). Furthermore, retrieving vertical profile of smoke layers from CALIOP requires a high aerosol extinction threshold due to the large associated lidar ratio and thus low SNR (Winker et al., 2013).

4.3.4 Desert dust transport

The case illustrated in Fig. 4.9a is a large desert dust plume over ocean surface, with MODIS $\tau(550\text{ nm})$ values up to 1.1, released from the Sahara and transported through westerly winds along the African coast. It occurred in summer on 2012.06.28. Since the Sahara desert is the most important source of mineral particles, associated dust aerosols include hematite and other iron oxides. Spectrally, desert dust is a UV-absorbing particle but quite highly scattering in the visible (contrary to smoke) and longer wavelengths, leading to the appearance of relatively bright plumes (light brown) over the dark marine surface from a satellite point of view. The NN algorithm trained with aerosol $\omega_0 = 0.95$ is therefore used here.

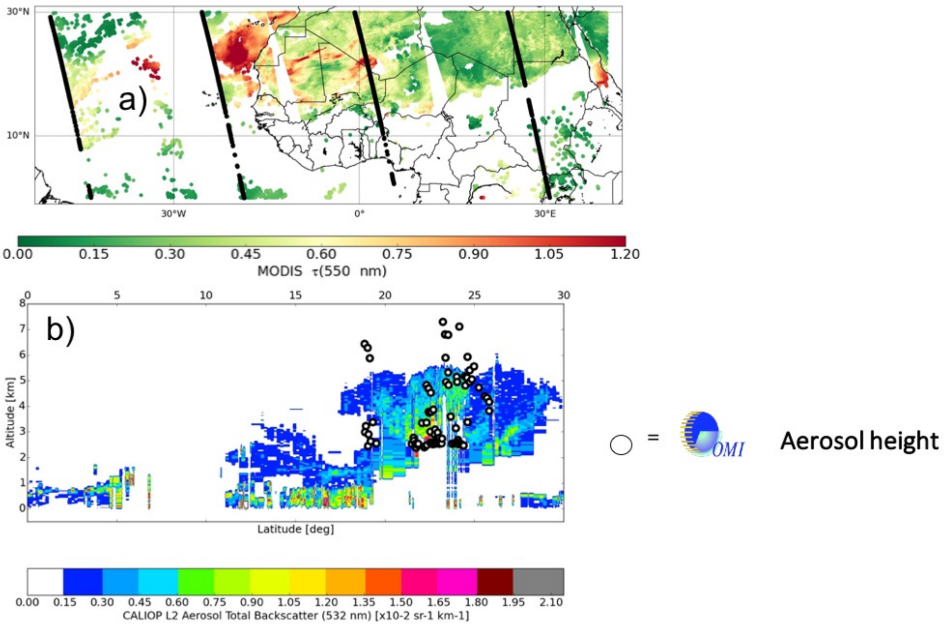


Figure 4.9: Cloud-free case of an elevated layer due to a Saharan dust outbreak transported to western Mediterranean region over sea on 2012.06.28: **(a)** Map of MODIS Aqua $\tau(550\text{ nm})$ from the combined DT and DB Collection 6 (see Sect. 4.2.3), **(b)** Retrieved OMI ALH compared with vertical profile of aerosol total backscatter coefficient (532 nm) from the CALIOP L2 aerosol total backscatter (532 nm) associated with the econd left CALIPSO track over sea in Fig. 4.9a.

The vertical profile of CALIOP L2 aerosol total backscatter (532 nm) shows elevated layers, ranging from 1-2 km at 15° S to 3-4 kmat 23° S (see Fig. 4.9b). Such a slope likely results from large-scale circulation governed by subtropical subsidence of the Intertropical Convergence Zone's northern branch, dry air from the desert, and the Saharan intense sensible heating effect perturbing the temperature inversion layers and thus creating convection uplifting dust from the surface (Prospero and Carlson, 1980). Generally, the OMI ALH results are consistent with CALIOP observations, with elevated values lying between 2 km and 4 km. Average OMI ALH is about 2.5 km which is then the middle of the large uplifted dust plume from south to north. The median difference between OMI and CALIOP L2 ALH, collocated within a distance of 200 km, is -557.8 m. The standard deviations is however quite large, about 1.5 km.

Several elements likely contribute to the difficulties encountered in this case study. Due to the well developed row anomaly (see. Sect. 4.3.1), selected OMI pixels are quite distant from the CALIPSO track. As analyzed in Sect. 4.3.2, large distance can significantly deteriorate the score of the ALH comparisons depending on the aerosol properties, the layer horizontal extension and its homogeneity. Furthermore, desert dust particles can be relatively coarse (thus low α value) and are irregularly shaped. Their optical modeling in the NN training dataset regarding their assumed size and the employed phase function model (see Sect. 4.2.2) may contribute to the higher ALH uncertainties than in urban cases of Sect. 4.3.2 (see further discussions in Sect. 4.4.3 and 4.4.4).

4.4 Specific error analysis

The analysed OMI ALH in Sect. 4.3 may include uncertainties due to assumptions made on the aerosol models used in the NN training dataset. The following subsections focus on some specific uncertainty sources that are relevant for these specific case studies. They provide further detailed error analysis and are complementary to the evaluations performed in (Chimot et al., 2017). Most of these analyses are based on synthetic scenarios. Simulations are performed in a similar way as in the NN training dataset (see Sect. 4.2.2). No bias is introduced in the geophysical parameters such as surface, temperature profile and atmospheric trace gases. The true prior aerosol $\tau(550)$ nm value is given for all the retrievals. Aerosols are assumed to cover the full ground pixel. The key analysed variable is the aerosol layer pressure (ALP) which corresponds to ALH expressed in hPa in order to be consistent with all the input parameter specifications (e.g. vertical grid) in the radiative transfer simulations.

4.4.1 Aerosol single scattering albedo

Aerosol ω_0 represents the scattering vs. absorption efficiency of the particles and therefore directly drives the magnitude of the applied shielding effect on the O_2-O_2 dimers (Chimot et al., 2016). An overestimated ω_0, in the training database, directly leads to an overestimation of ALH (or underestimation of ALP) as the measured $N^{s}_{O_2-O_2}$ is lower (i.e. stronger shielding) than expected if one knows the true extinction profile and assumes a biased ω_0 (Chimot et al., 2017).

Dense smoke layers from wildfires, such as those analysed in Sect. 4.3.2, may contain particles that are more absorbing that the assumed aerosol model. Figure 4.10 illustrates the impact of particles with $\omega_0 = 0.8$ while the NN algorithm trained with $\omega_0 = 0.9$ is used (same as in Sect. 4.3.2). No errors are introduced in all the other geophysical parameters.

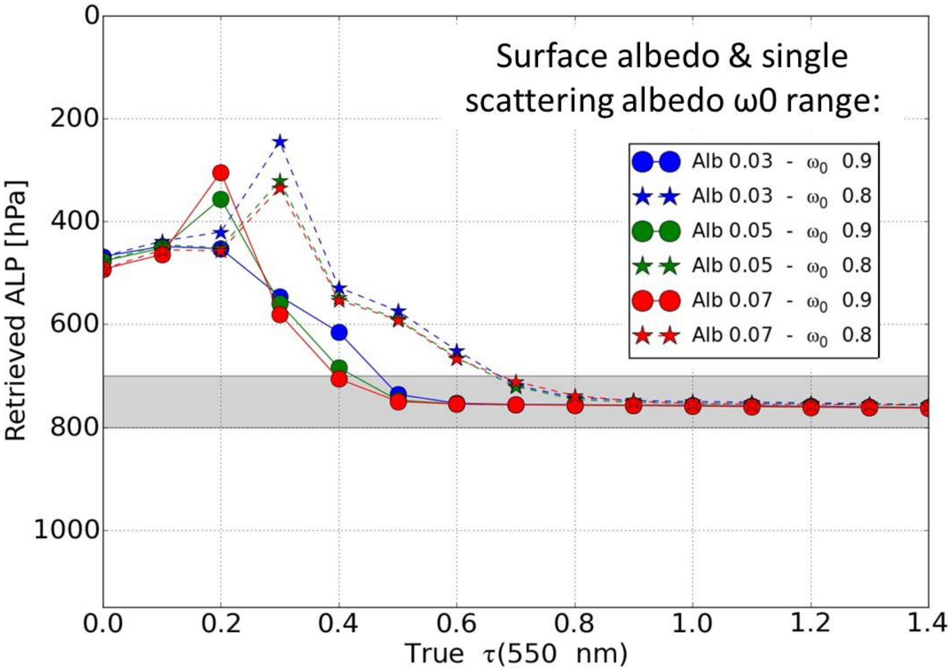


Figure 4.10: Simulated ALP retrievals, based on noise-free synthetic spectra with aerosols, as a function of true τ(550 nm). All the retrievals are achieved with the NN algorithm trained with aerosol $\omega_0 = 0.9$ and true prior τ(550 nm) value. The assumed geophysical conditions are temperature, H_2O, O_3 and NO_2 from climatology mid-latitude summer, $\theta_0 = 25°$, $\theta = 45°$ and $P_s = 1010$ hPa. The reference aerosol scenario assumes fine scattering particles ($\alpha = 1.5$, $g = 0.7$), and two aerosol ω_0 values: 0.9 and 0.8. Its location is depicted by the grey box, between 700 and 800 hPa.

The resulting ALP values are overestimated with a bias up to 100 hPa (around 900 m) for scenes with τ(550 nm) in the range of 0.5-0.9. ALP biases are almost null over scenes with higher aerosol load as the shielding effect due to the already high amount of particles clearly dominates over their optical properties, and thus regardless of ω_0 assumptions. These results are in line with those estimated from the use of the NN algorithm trained with $\omega_0 = 0.95$ in Chimot et al. (2017).

4.4.2 Aerosol vertical distribution

Due to the specific limitations of passive satellite sensor, the OMI ALH retrievals summarizes the description of the aerosol extinction profile in a single scalar value assuming a specific profile shape. However, aerosol profiles in the observed scene may considerably deviate from this simplified profile description. It is then legitimate to ask the meaning of the retrieved ALH. As explained in Sect. 4.2, the NNs were trained based on a single "box layer" with a constant geometric thickness of about 1 km (100 hPa exactly), ALP / ALH being then the mid-pressure / mid-altitude of this layer. Several of the analysed cases in Sect. 4.3 depict more extended aerosol layers (e.g. up to 3.5 km in Fig. 4.8b), or two separate layers (e.g. Fig. 4.7b).

Figure 4.11 illustrates the retrievals in a case of an extended aerosol laye with thickness of 300 hPa, located between 700 hPa and 1000 hPa.

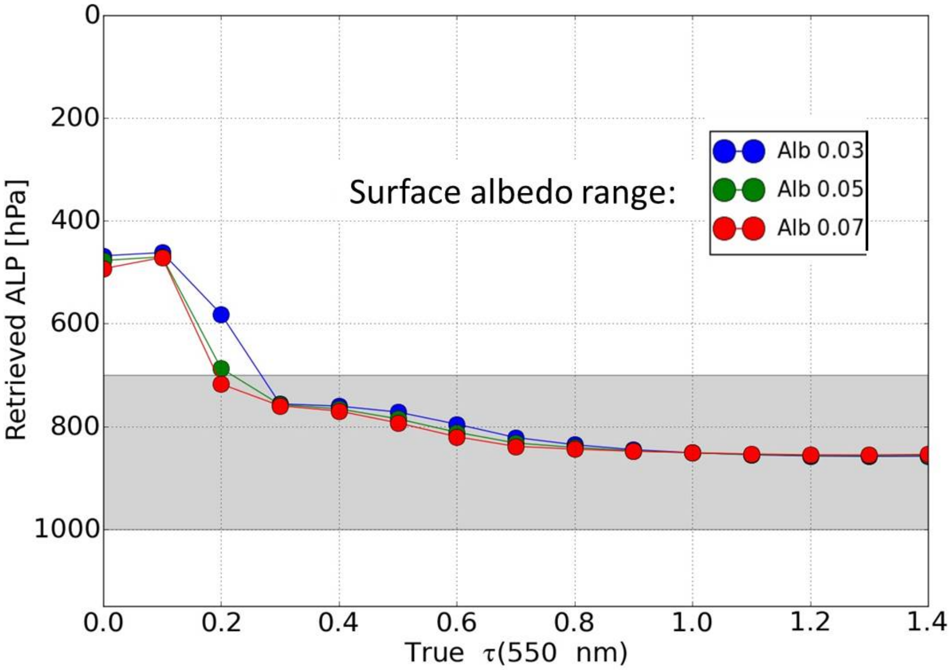


Figure 4.11: Same as Fig. 4.10 but with one unique aerosol ω_0 value (= 0.9) and a larger geometric extension of the aerosol layer included in the simulated spectra: i.e. between 700 and 1000 hPa.

The derived ALP values are close to 850 hPa for scenes with $\tau(550nm) \geq 0.5$, which thus corresponds to the mid-level of the simulated layer. This result may be understood as the true aerosol vertical extinction profile being lower than the assumption. The retrieval then reaches the average altitude where most of the O_2-O_2 is actually shielded.

In Figure 4.12, ALP is retrieved when two separate aerosol layers with same thickness (i.e. 100 hPa) are simulated: an elevated one between 600 hPa and 700 hPa and a lower one between 900 and 1000 hPa.

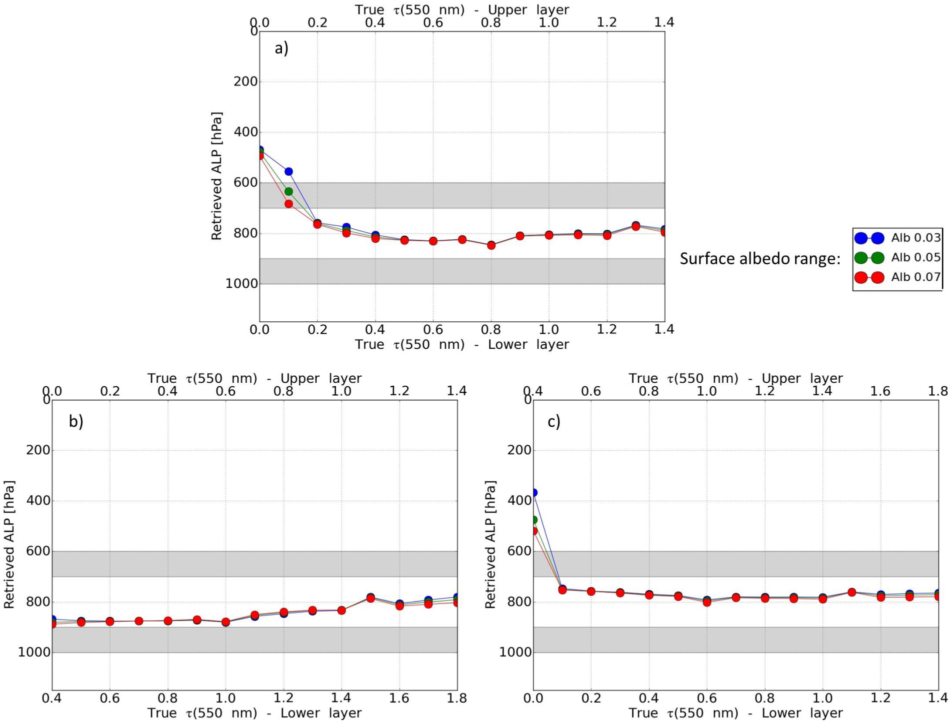


Figure 4.12: Same as Fig. 10 but with one unique aerosol ω_0 value (= 0.9) and two separate aerosol layers included in the simulated spectra. The bottom x-axis corresponds to the τ(550 nm) value of the lower layer, while the top x-axis is the τ(550 nm) value of the upper layer. Both layers have same geometric thickness (i.e. 100 hPa). The first is located between 600 and 700 hPa, and the second is between 900 and 1000 hPa. **(a)** Both aerosol layers have same optical properties and τ(550 nm) values, **(b)** Both aerosol layers have same optical properties but different τ(550 nm) values: the lower layer has systematically a higher τ(550 nm) (i.e. +0.4 for each scenario), **(c)** Both aerosol layers have same optical properties but different τ(550 nm) values: the upper layer has systematically a higher τ(550 nm) (i.e. +0.4 for each scenario)

Assuming that both layers have same optical properties, ALP is retrieved close to 800 hPa for $\tau(550nm) \geq 0.5$ (see Fig. 4.12a). Here, the retrieval corresponds to the average height of both layers. However, when one of these layers has a higher aerosol load (i.e. a higher value for τ(550 nm)), the retrieval is: close to the optically thicker aerosol layer for total τ(550 nm) in the range of 0.-1.6 and reaches the average height (i.e. 850 hPa) for total $\tau(550nm) \geq 1.6$. This demonstrates the sensitivity of the retrieval to the extinction properties of the particles and its vertical distribution driving the location where most of the O_2-O_2 dimers are shielded. As a consequence, the retrieved ALP or ALH actually represent a weighted average of the actual aerosol vertical distribution, the weights being the vertical extinction values.

4.4.3 Aerosol size

Within the HG phase function model, particle size is primarily governed by α which describes the spectral variation of the aerosol load τ. While the NNs were trained for fine particles emitted from anthropogenic activities such as power plants and vehicles (i.e. $\alpha = 1.5$), other particles such as dust can be coarser.

Figure 4.13 depicts the ALP retrievals assuming scattering particles with $\omega_0 = 0.95$ (same as in Sect. 4.3.4) but with different α values: 1.5 (consistent with the training dataset) and 0.5.

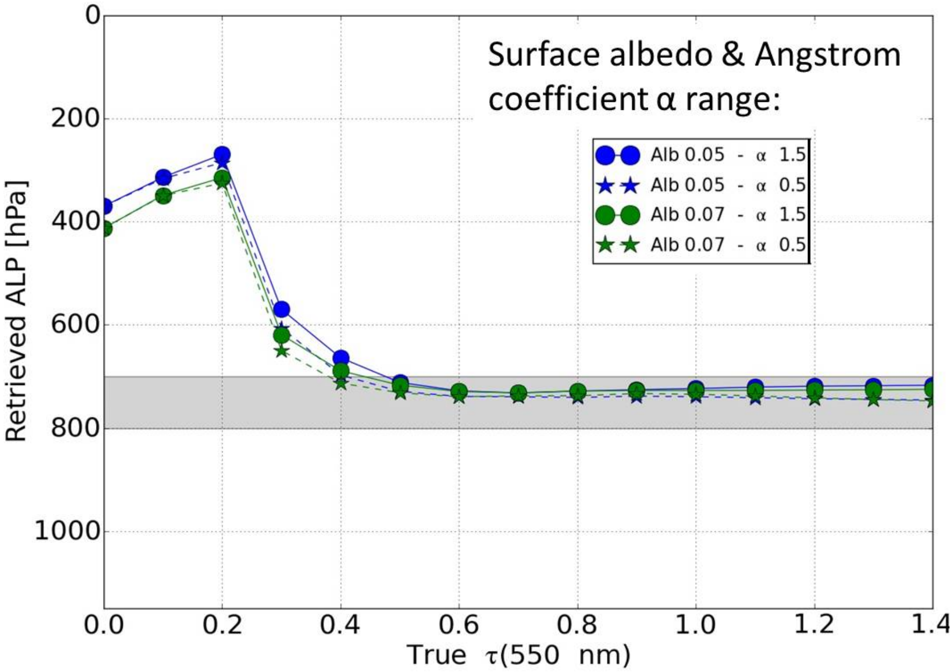


Figure 4.13: Same as Fig. 10 but with one unique aerosol ω_0 value (= 0.95) and two aerosol α (1.5 and 0.5) in the reference aerosol scenarios. ALP retrievals are estimated from the NN algorithm trained with aerosol $\omega_0 = 0.95$ similarly to the desert dust case in Sect. 4.3.4.

Overestimating α (i.e. underestimating particle size) leads to a increase (decrease) of retrieved ALP (ALH). This is because coarser particles generally extend the length of the average light path, due to reduced multiple scattering, lower the O_2-O_2 shielding, and thus increase the measured $N^{S}_{O_2-O_2}$ as shown in Chimot et al. (2016). The ALP change is nevertheless about 25 hPa.

4.4.4 Scattering phase function

Modelling the aerosol scattering phase function requires precise information not only on their size and optical properties but also on their shape and the phase function modelling theory itself. As an example, optical modelling of desert dust can, for some applications, be done using Mie theory which is mostly valid for homogeneous and spherical particles whereas for other applications one could better consider alternative spheroids or T-matrix/geometric optics traditionally used for non-spherical particles (de Graaf et al., 2007; Xu et al., 2017).

For reasons explained in Sect. 4.2.2, the HG was employed in the NN training database. This may lead to some errors in ALH / ALP retrievals due to inaccurate scattering angular dependence depending on the particle type and the assumed g parameter. The shape of the phase function is parameterized by g in HG modelling. In Chimot et al. (2016), we demonstrated that bias on ALP does not exceed 50 hPafor a typical uncertainty of 0.1 on g over scenes with $\tau(550\text{nm}) \geq 0.5$, assuming no additional bias on α or ω_0. Comparison between Mie and HG modelling would mix errors caused by these three parameters altogether, which would then make complex to identify the actual error source. Colosimo et al. (2016); Sanders et al. (2015) show with simulation studies comparing phase function models, although in different spectral bands, that using a scattering layer with constant particle extinction coefficient does reasonably well without additional biases than those analysed in the previous sections.

Pure desert dust particles are known to be irregularly shaped, and thus the use of the HG forward model may be inappropriate in Sect. 4.3.4 and Fig. 4.9. Furthermore, by using a prior τ parameter from MODIS, that may also be derived from an inaccurate and different model, can add some inconsistencies in the OMI ALH retrieval. This may explain in part the larger uncertainties found in Sect. 4.3.4. In further steps, to confirm the real performances of ALH retrievals over a long time series of OMI measurements and/or a potential implementation in the OMI processing chain, new NN algorithms should be designed and trained with a larger dataset that includes accurate aerosol parameters (size and ω_0) combined with different detailed models of the phase function. Each of these algorithms should be evaluated on a high number of specific observations to conclude on the exact aerosol model type to be assumed for the OMI visible spectral measurements.

4.4.5 Cloud contamination

When backscattered solar light measurements from UV-vis passive satellite sensors are exploited, detecting cloud-free pixels is one of the most crucial prerequisite for aerosol retrievals. In spite of a strict cloud filtering applied in Sect. 4.3.1, some small cloud residuals may remain in the analysed scenes, especially over biomass burning episodes where the distinction of dense smoke particles and small cloud layers can be difficult.

Presence of cloud layers have similar effects as aerosols on the OMI visible measurements and the O_2-O_2 molecules although associated optical thickness are an order of magnitude higher. In Figure 4.14, cloud layers were added to an aerosol layer located between 700 hPaand 800 hPa. Clouds were simulated as an opaque Lambertian bright layer with an albedo of 0.8 and different effective cloud pressure and fraction values. Such a model is similar to what is employed in the OMCLDO2 algorithm to detect and characterize the presence of clouds within the OMI pixel or to implicitly correct aerosol effect in trace gas retrievals (Acarreta et al., 2004; Veefkind et al., 2016; Boersma et al., 2011; Chimot et al., 2016). It should be noted that aerosols are assumed to cover the whole scene in the simulations.

Figure 4.14 shows that the impact on the ALP retrieval strongly depends on the cloud altitude. If the aerosol layer is located below a cloud with an effective fraction of 0.3, the ALP is strongly biased low (i.e. ALH high) for $\tau(550\text{nm}) \leq 0.8$, while it tends towards the effective cloud pressure for $\tau(550nm) \geq 0.8$ (see Fig. 4.14c). Such a behavior may be explained by the high O_2-O_2 shielding caused by the clouds, much higher than what is anticipated by the retrieval algorithm through the given aerosol $\tau(550$ nm$)$. The assumed optical thickness of the scene is too low to match with the strongly reduced $N^s_{O_2-O_2}$ measurement, especially over scenes with a small aerosol load. Moreover, the opaque and bright cloud shields part of the scattering layer located below and dominates over the aerosol signal. The retrieval compensates then with a strongly reduced ALP value. When a high aerosol load (both in the scene and in the prior information) roughly corresponds to the optical thickness of the scene, the ALP represents the altitude where most of the O_2-O_2 shielding occurs: at the cloud level. This last effect is also visualized in Sanders et al. (2015) with an optically thick aerosol layer and a cirrus above it, although a different spectral band, in the near infrared, is employed.

In contrast, if aerosols are located above the cloud with an effective fraction of 0.3, retrieved ALP is located between both layers (see Fig. 4.14a). Part of the cloud signal is attenuated this time. Similarly to Sect. 4.4.2, ALP likely represents a weighted average of the extinction vertical profile. This average is weighted not only by the aerosol properties and the cloud altitude but also by the effective cloud fraction. In the presence of a reduced effective cloud fraction (0.1 instead of 0.3), the estimated ALP is lower (decrease of 40 hPa), close to the base height of the aerosol layer.

4.5 Conclusions

Following the study of Chimot et al. (2017), aerosol layer heights (ALHs) were retrieved from OMI cloud-free pixels using the O_2-O_2 visible absorption band at 477 nm, based on a neural network approach. The physical principle relies on the dependency of the shielding of the O_2-O_2 dimers on the aerosol height. Three days with urban and industrial pollution episodes in eastern China, 4 days with widespread biomass burning events in South America and Russia and 1 day of a Saharan dust plume transport event over ocean were studied in detail. The goal was to evaluate the OMI ALH spatial patterns over case studies. Prior aerosol optical thickness $\tau(550$ nm$)$ information was used from collocated MODIS L2 product (Dark target and Deep Blue algorithms). The retrievals were compared with CALIOP along-track product. The selection of events largely depends on the availability of coinciding OMI and CALIOP data over relevant cases.

Good agreement was found between OMI and CALIOP ALH, where the latter

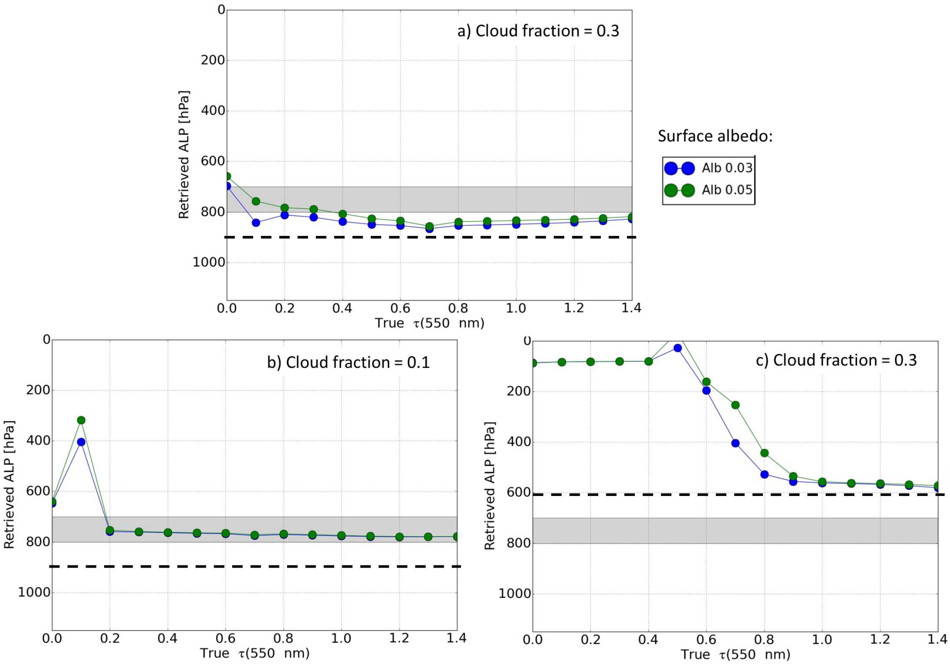


Figure 4.14: Same as Fig. 4.10 but with one unique aerosol ω_0 value (0.9) and the inclusion of a cloud (dashed thick black line) in addition to the aerosol layer. The cloud reflectance is simulated via a simple opaque (cloud albedo of 0.8) and Lambertian layer: **(a)** Effective cloud fraction of 0.3 and cloud pressure of 900 hPa, **(b)** Effective cloud fraction of 0.1 and cloud pressure of 900 hPa, **(c)** Effective cloud fraction of 0.3 and cloud pressure of 600 hPa.

was derived from the level 2 (L2) aerosol extinction profile, over urban and industrial pollution episodes: we find root-mean-square deviation in the range of 462-648 m for distances between OMI and CALIOP ground pixels smaller than 50 km and with collocated MODIS $\tau(550\text{ nm}) \geq 0.55$. Similar spatial patterns are also observed between both sensors. Carbonaceous and black carbon particles within dense smoke layers over biomass burning events strongly attenuate the CALIOP backscatter signal (532 nm). As attenuated backscatter profiles decrease more rapidly in the short than in the long wavelengths, only CALIOP L1 measurements (1064 nm) allow us to probe the entire vertical extent of smoke aerosol layers. OMI ALH retrievals match well with these last CALIOP measurements. The higher sensitivity of visible spectral measurements acquired by passive satellite sensors, such as OMI, to capture information from lower altitudes of an optically thick absorbing layer is probably due to the photon measurements resulting from multiple scattering effects from different atmospheric altitudes combined with a high signal-to-noise ratio. While scattering leads to a strong reduction of the active signal (i.e. lidar) in penetration depth, this reduction is much lower for a remote sensing sensor using the solar light source as the fraction of detected photons that reached the lower part of the aerosol layer is considerably higher. Finally, although OMI ALH shows in general consistent results with respect to CALIOP over the transport of the Saharan dust plume over ocean (difference median of -557.8 km), it remains locally limited likely due to 1) the large distance between OMI and CALIOP ground-based pixels, and 2) the potential artifacts due to inaccurate modeling of this particle type in the NN training database.

Detailed analyses and discussions on specific error sources confirm that prior assumptions on aerosol optical properties are the key crucial factor affecting the OMI ALH retrieval accuracy over cloud-free scenes. In particular, the combination of aerosol single scattering albedo, particle size and shape, and the angular dependency of the scattering phase function assumptions may impact up to 500 m for each individual parameter. The reason is the direct impact of these variables on the O_2-O_2 dimers shielding applied by aerosols. Furthermore, a strict cloud filtering is required to distinguish aerosol from cloud effects. The impact of cloud residuals is a function of the cloud coverage and location with respect to the aerosol layer. Finally, the true meaning of the retrieved ALH parameter depends on the actual aerosol vertical distribution. It can be summarized as the weighted average of the optical (or extinction) particle properties along the vertical atmospheric layers, an optically thick (or strongly absorbing) layer having more weights than an optically thin (or highly scattering) particle layer.

Future works should include further comparisons with multiple sensors (satellite, ground-based and airborne), generation of yearly series, trend analysis, and then evaluation of aerosol effect correction in support of satellite UV-vis trace gas retrievals (e.g. tropospheric NO_2). The use of satellite O_2-O_2 visible absorption band should be further studied for aerosol retrievals in addition to the consideration of the more traditional O_2 band in the near infrared as it may bring additional relevant information. Moreover, expectation for air quality and climate research from a future global OMI ALH product, with a high temporal resolution, should be further investigated, and required improvements should be implemented for an optimal exploitation.

Chapter 5

Yearly global OMI aerosol retrievals dataset

> "Traveling through hyperspace ain't like dusting crops, farm boy."
>
> Han Solo, Star Wars IV - A New Hope, 1977

5.1 Introduction and motivation

Following, the developments and analysis in Chap.3 and 4, we decided during this thesis research to generate a global 1-year OMI ALH dataset over cloud-free scenes. This task was achieved in parallel of the research work written in the other chapters. The motivation of this task is double: 1) to demonstrate the feasibility, from a technical point of view, to generate a global yearly satellite ALH product, based on the developed OMI aerosol neural network (NN) algorithm, and 2) to illustrate the temporal and spatial behaviour of this research product. Such a dataset, if well validated, could then be further used for correcting of aerosol effects in trace gas retrievals (cf. Chap. 5) or for other specific scientific analyses.

5.2 Methodology and description of the OMI aerosol datasets

The developed OMI aerosol dataset contains a series of ALH (as key parameter), and τ(550 nm) (as a secondary parameter). All aerosol parameters are derived from the OMI 477 nm O_2-O_2 spectral band, over cloud-free scenes for the whole globe and year 2006. The retrieval principle follows the descriptions in Chap.3.

The following items deserve to be emphasized:

- **Cloud-free pixels**: As explained in Chap.3, the OMI aerosol NN is only applicable to cloud-free scenes. Therefore, it is essential to identify such scenes, and to minimise as much as possible cloud residuals in the selected OMI pixels. For that purpose, the merged OMI and MODIS cloud product, named here after

---

This chapter shortly describes a yearly (2006) and global OMI aerosol retrieval dataset, generated at the end of this thesis work. Part of it was presented in: **Chimot J.**, Veefkind P., van Ouwerkerk G., Vlemmix T., Levelt P., Aerosol layer height retrieval from OMI and neural network- Possibility for a 13-year time series?, 20th OMI Science Team Meeting, NASA Goddard, Greenbelt MD, Virginia, USA, 2017.09.12.

OMMYDCLD was used to maximise the probability of cloud-free scene representative of the OMI pixel. OMMYDCLD was developed by Science Systems Applications, Inc. (SSAI) and NASA Goddard Space Flight Center. To identify an OMI pixel as cloud-free, a threshold of 0.02 was specified on the associated MODIS geometric cloud fraction.

- **Prior τ**: As explained in Chap.3, the quality of the retrieved OMI ALH depends on the prior τ knowledge. Similarly to Chimot et al. (2017, 2018b), the Collection 6 of MODIS aerosol τ(550 nm) product at the resolution of 10 km, combining the Dark Target (DT) and Deep Blue (DB) algorithms was considered (Levy et al., 2013). Collocation between OMI pixels and MODIS aerosol grid was ensured within a distance of 15 km and a difference in time observation of 15 minutes.
- **Aerosol models**: Both aerosol models employed for the NN training were considered for each OMI ALH and τ(550 nm) retrieval. Consequently, two series of OMI aerosol datasets were created: one based on aerosol ω_0=0.9, and the other one based on aerosol ω_0=0.95.
- **Aerosol scenes**: the retrievals were performed for all selected OMI scenes (with respect to the criteria mentioned above) regardless of the aerosol abundance (i.e. MODIS τ). However, since we demonstrated that the OMI ALH accuracy critically depends on true τ, the next maps only depict the ALH retrievals for MODIS τ(550 nm) $\geq$ 0.5 (cf. Chap.3). Figure 5.1 depicts the global yearly average of MODIS τ for 2006. We can then see the location of such high values, and therefore where high ALH accuracy may be expected. Note that aerosol τ values also have a seasonal dependency (cf. Fig. 5.3-5.4).

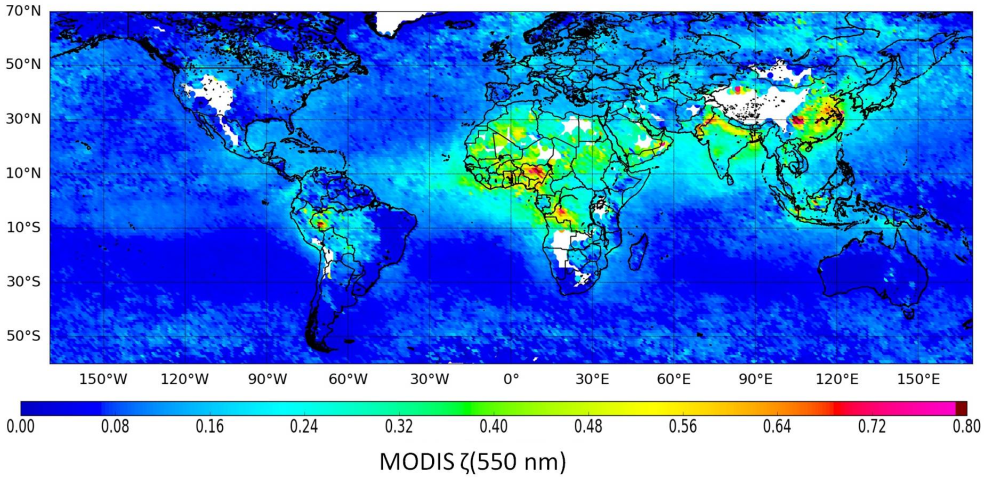


Figure 5.1: Global MODIS τ(550 nm) for the whole year 2006 average for 2006 over cloud free scenes (see Sect.), and OMI surface albedo $\leq$ 0.15.

- **OMI O_2-O_2**: The OMI O_2-O_2 datasets as well as the associated OMI climatology surface albedo from the improved OMI cloud algorithm (OMCLDO2) were considered for this task (Veefkind et al., 2016).

5.3 Discussions

Overall, the whole NN retrievals took about 12 hours on one single CPU for each aerosol model. This means that the whole 13-year cloud-free OMI time series could be processed in almost 1 week on 1 CPU. Such a computing time definitely emphasizes the high advantage of using a NN approach for aerosol (both ALH and τ) retrievals.

Figure 5.2 shows the global OMI ALH for the year 2006 over cloud-free scenes and MODIS τ(550 nm) $\geq$ 0.5. With respect to these criteria, OMI ALH is mostly available over land. Average values are in the range of 2-2.5 km but with a large spatial variability: the largest values are mostly present over Central Africa where they could be associated with biomass burning activities. The smallest values are found over the bright Sahara desert, but are likely not realistic. Cases with very high surface albedo values (i.e. $\geq$ 0.15) were not included in the NN training dataset. In such a case, the NN extrapolation may be then lead to large errors. Furthermore, such bright surfaces remain generally very challenging for measurements from optical sensors (Nanda et al., 2017). Finally, as explained in Chap.3, OMI NN ALH values derived from aerosol ω_0=0.95 are about 500 higher higher than from ω_0=0.9.

Figures 5.3-5.4 illustrate the OMI cloud-free ALH seasonal variability over 2006 for collocated MODIS τ(550 nm) $\geq$ 0.5. ALH appear over South-America only during June-July-August (JJA) and September-October-November (SON), corresponding to the biomass burning season in this region. At that time of the year, heaving amounts of carbonaceous particles are then released from widespread fires. Associated plumes can reach elevated altitudes, about 2-3 km (cf. Chap. 4). OMI ALH pixels are also found over Europe and North-America mostly during JJA, when large urban and industrialized areas encountered aerosol peak episodes. Finally, the seasonal ALH variability, already mentioned in Chap.3, is also found here with larger values during JJA, and lower values during DJF and SON.

Finally, Fig. 5.5 shows the retrieved OMI τ(550 nm) for the whole 2006 over the selected cloud-free scenes. The interest of this parameter is double: 1) it illustrates the possibility to derive the τ parameter from OMI visible measurements, and 2) in absence of reference ALH validations, it shows the potential problems remaining in both OMI τ and ALH parameters since reference collocated MODIS τ product is available. Overall, OMI τ shows larger values over land and lower over sea, similarly to MODIS. However, as discussed in Chap.3, OMI τ remains limited: i) it looks obvious that, despite the use of the OMMYDCLD, cloud residuals remain in the selected OMI pixels, especially over sea; ii) additionally, problems due to foam and wind may be significant over sea (such parameters are not considered in the OMI aerosol retrievals); iii) inacurate aerosol model (i.e. ω_0 parameter) may highly bias the OMI ALH and τ retrievals.

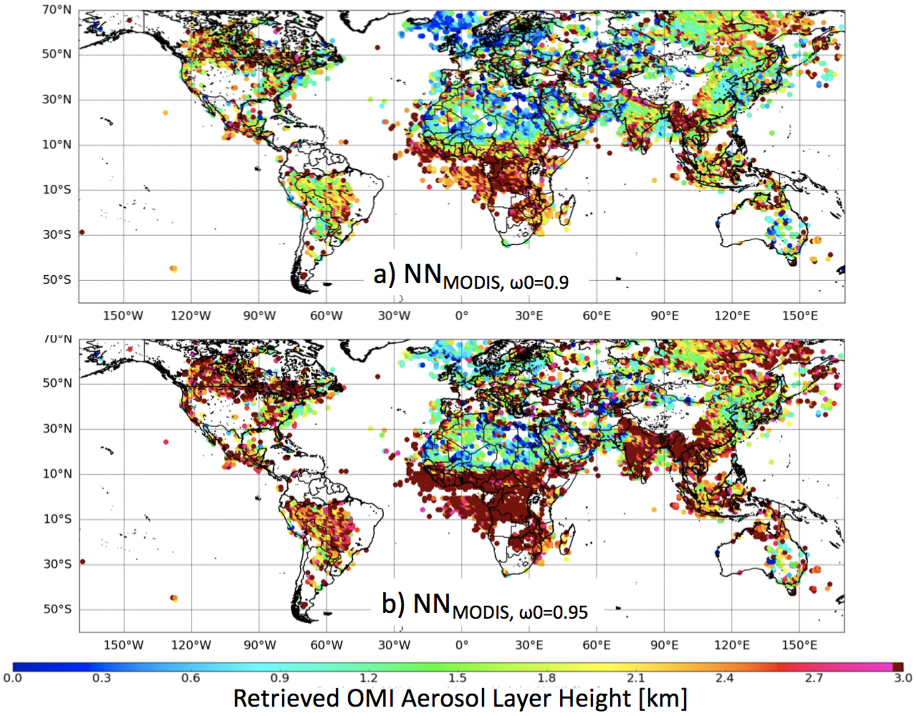


Figure 5.2: Global retrieved OMI ALH for the whole year 2006, over cloud-free scenes and for MODIS τ(550 nm) $\geq$ 0.5, and OMI surface albedo $\leq$ 0.15: **(a)** Retrieval with $NN_{MODIS,\omega_0=0.9}$, **(b)** Retrieval with $NN_{MODIS,\omega_0=0.95}$.

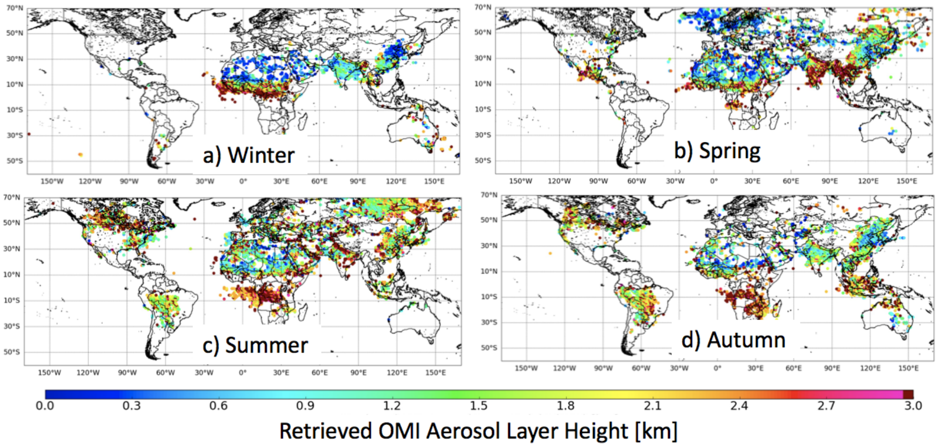


Figure 5.3: Global retrieved OMI ALH for the year 2006 and per season with $NN_{MODIS,\omega_0=0.9}$, over cloud-free scenes and for MODIS τ(550 nm) $\geq$ 0.5, and OMI surface albedo $\leq$ 0.15: **(a)** December-January-February, **(b)** March-April-May, **(c)** June-July-August, **(d)** September-October-November.

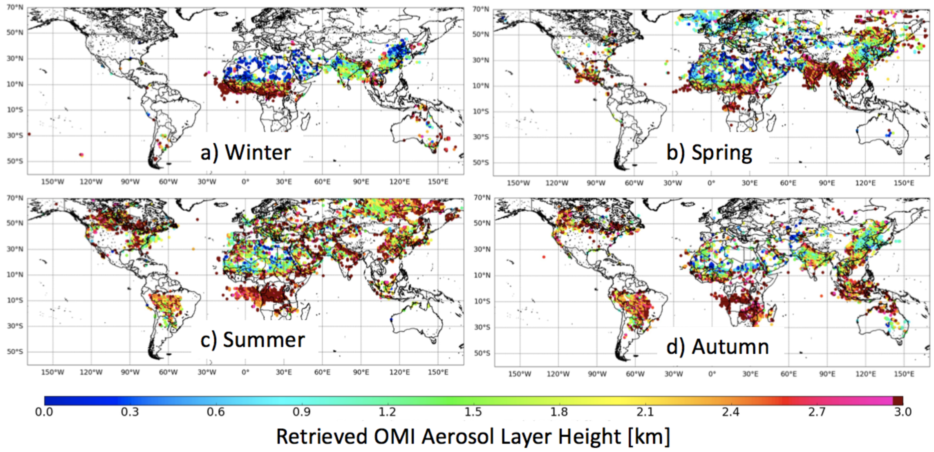


Figure 5.4: Same as Fig. 5.3 but with $NN_{MODIS,\omega_0=0.95}$.

Overall, an in-depth validation is necessary to appreciate more quantitatively these maps, but was not the priority as explained in Sect. 5.1. Statistical evaluation based on the CALIOP climatology dataset in Chap. 3, or specific case studies like in Chap. 4 can already give some ideas: on average, OMI ALH accuracy is expected to lie in the range of 200-800 m depending on the areas and seasons.

5.4 Conclusions

This chapter shows a yearly (2006) global ALH dataset derived from the OMI 477 nm O_2-O_2 spectral band over identified cloud-free scenes. Additionally, OMI τ(550 nm) was retrieved. Both parameters display remarkable spatial and seasonal patterns depending on the related events (large biomass burnings, or urban aerosol pollution). Retrievals were performed assuming two aerosol models, based on two different ω_0 values: 0.9 and 0.95. Overall, after filtering cloudy pixels and identifying collocated OMI pixels and MODIS aerosol product cell, the retrievals were obtained within 12 hours using one single CPU. This confirms the possibility to process OMI time series for aerosol purposes within a low computing time.

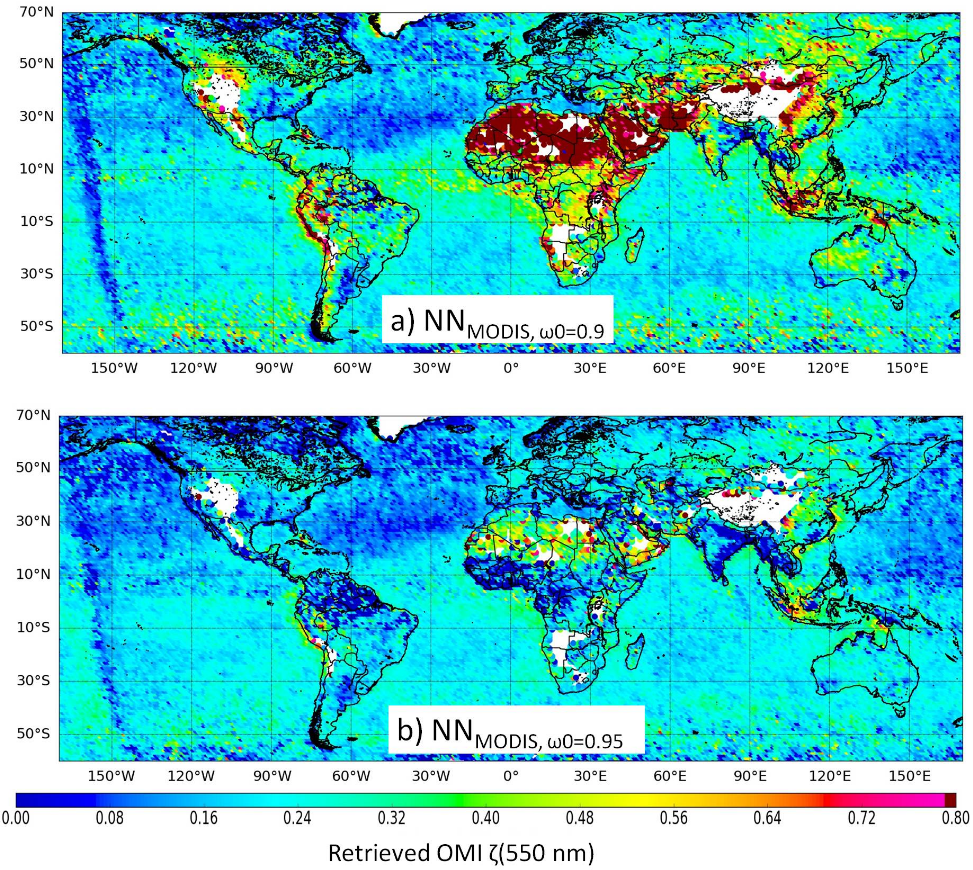


Figure 5.5: Global retrieved OMI τ(550 nm) for the whole year 2006 average for 2006 over cloud free scenes (see Sect. 5.2), and OMI surface albedo ≤ 0.15: **(a)** Retrieval with $NN_{MODIS,\omega_0=0.9}$, **(b)** Retrieval with $NN_{MODIS,\omega_0=0.95}$.

Chapter 6

Minimizing aerosol effects on the OMI tropospheric NO_2 retrieval – An improved use of the 477 nm O_2-O_2 band

> "The idea of neutral objectivity is at best misleading and often fraudulent. We cannot help but approach complex and controversial questions – especially those of human significance – with a definite point of view, with an ax to grind [...], and that ax should be apparent right up front so that those we address can see where we are coming from in our choice and interpretation."
>
> Noam Chomsky, Optimism over Despair – On Capitalism, Empire and Social Change, 2017

Abstract

Global mapping of satellite tropospheric NO_2 vertical column density (VCD), a key gas in air quality monitoring, requires accurate retrievals over complex urban and industrialized areas. However, the high abundance of aerosol particles in regions dominated by anthropogenic fossil fuel combustion, mega-cities and biomass burning affects the space-borne spectral measurement. This leads to one of the main challenges in tropospheric NO_2 VCD retrieval which is known to have a high bias in presence of optically thick aerosol layers. For the first time, reference Ozone Monitoring Instrument (OMI) DOMINO-v2 product was reprocessed over cloud-free scenes, by applying new aerosol correction parameters issued from the 477 nm O_2-O_2 band, over east China and South-America and 2 years (2006-2007). These new parameters are based on two different and separate algorithms developed during the last two years in view of an improved use of the 477 nm O_2-O_2 band: 1) the updated OMCLDO2 algorithm which derives improved effective cloud parameters, 2) the aerosol neural network (NN) giving explicit aerosol parameters by assuming a more physical aerosol model. The OMI aerosol NN is a step ahead to OMCLDO2 by retrieving primarily an explicit aerosol layer height (ALH), and a secondly the optical thickness τ.

The contents of this chapter are intended to be submitted as: **Chimot, J.**, Veefkind, J. P., de Haan, J. F., Stammes, P., and Levelt, P. F.: Minimizing aerosol effects on the OMI tropospheric NO_2 retrieval – An improved use of the 477 nm O_2-O_2 band, To be submitted in Atmos. Meas. Tech. Discuss., July 2018.

Overall, it was found that the biases identified in DOMINO-v2 over scenes in China with high aerosol abundance in summertimes will be reduced from −20-−40% to 0-20% regardless of the considered aerosol model type. The use of the aerosol parameters retrieved from the NN algorithm leads in general to a more explicit aerosol correction and higher tropospheric NO_2 VCD values than from OMCLDO2: in the range of 20-40%. Such a correction scheme also includes more realistic consideration of aerosol multiple scattering effects, especially over scenes dominated by strongly absorbing particles, where the correction based on OMCLDO2 seems to remain insufficient. However, the use of ALH and τ from the OMI NN aerosol algorithm is not a straightforward operation and future studies need to further investigate the optimal methodology. Overall, we demonstrate the possibility to apply a more explicit aerosol correction by considering aerosol parameters directly derived from the 477 nm O_2-O_2 spectral band, measured by a same satellite instrument: i.e. OMI. Such an approach can easily transposed to future instruments (e.g. TROPOMI), applied to a fast reprocessing of tropospheric satellite NO_2 data over cloud-free scenes (cloudy pixels need to be filtered out), and for other trace gas retrievals (e.g. SO_2, HCHO).

6.1 Introduction

Long-time series of UV-visible (UV-vis) satellite measurements are a great asset for monitoring the distribution and evolution of pollutants such as NO_2, HCHO, or SO_2 and aerosol particles in the troposphere. With the forthcoming new generation of sensors like TROpospheric Ozone Monitoring Instrument (TROPOMI) on-board Sentinel-5-Precursor (Veefkind et al., 2012), Sentinel-4-UVN, and Sentinel-5-UVNS within the Copernicus programe (Ingmann et al., 2012), they will become an important tool for verifying the effectiveness of implemented technology to protect environment and health population (Duncan et al., 2016). While current generation of space instruments have a pixel size between 13 x 24 km^2 for the Ozone Monitoring Instrument (OMI) or 80 x 40 km^2 for the Global ozone Monitoring Experiment (GOME-2), the new and future generations will have smaller pixel sizes (about 7 x 3.5 km^2 for TROPOMI), allowing air quality mapping of complex urban and city areas. This will also reduce the probability of cloud contamination. However, the significant probability of aerosol contamination in areas such as India, China or regions dominated by biomass burning episodes will likely remain, or may even increase.

OMI is the Dutch-Finnish push-broom spectrometer flying on the National Aeronautics and Space Administration (NASA)'s Earth Observation Satellite (EOS)-Aura satellite since July, 15th 2004. Its Sun-synchroneous orbit has a local equator crossing-time of approximately 13:40h. Operational tropospheric NO_2 products derived from the visible backscattered spectral light (405-465 nm) such as the OMI DOMINO-v2 (Boersma et al., 2011) or the very recent Quality Assurance for Essential Climate Variables (QA4ECV) are nowadays a reference. Their related global mapping of tropospheric NO_2 concentrations have been used by a countless air quality research studies focusing on NO_x emissions, secondary pollutant formation, as well as tropospheric NO_x chemistry and transport: e.g. (Curier et al., 2014; Reuter et al., 2014;

Ding et al., 2015).

A critical element for an accurate tropospheric NO_2 vertical column density (VCD) retrieval is our capability to reproduce the average light path along which the photons travelled before being detected by the satellite sensor in the visible spectral window. In particular, scattering induced by atmospheric aerosol particles over cloud-free scenes are known to lead to very complex light path. Because they are emitted by the same sources, high NO_2 and aerosol concentrations are often spatially correlated (Veefkind et al., 2011). Therefore, the presence of aerosol needs to be properly addressed in the retrieval algorithms. In the frame of tropospheric NO_2 retrievals from visible spectral measurements which use the differential optical absorption spectroscopy (DOAS) approach, the aerosol correction has to be done on the air mass factor (AMF): a unitless number representative of the length of the average light path.

Over cloud-free scenes, a full explicit aerosol correction ideally requires a comprehensive set of parameters describing aerosols: i.e. optical properties such as the single scattering albedo ω_0, size, scattering phase function, load, and vertical distribution (Martin et al., 2003; Leitão et al., 2010; Bousserez, 2014). Among all these variables, many studies emphasized the importance of the aerosol layer height (ALH) knowledge (Leitão et al., 2010; Castellanos et al., 2015; Chimot et al., 2016). Assuming no aerosol correction (i.e. aerosol-free scene (Richter and Burrows, 2002)) would clearly create high biases on the OMI tropospheric NO_2 retrievals (Chimot et al., 2016).

There are basically two strategies for applying the aerosol correction in the AMF: 1) either by considering external data, or 2) by using the available particle parameters that can be simultaneously derived from the same UV-visible spectral space-borne measurement. Studies that reprocessed DOMINO-v2 dataset using external data usually combined atmospheric transport model outputs such as GEOS-Chem in POMINO (Lin et al., 2014, 2015), or observations issued from different satellite platforms such as the Cloud-Aerosol Lidar with orthogonal Polarization (CALIOP) (Castellanos et al., 2015). Resulting changes mostly occurred in case of high aerosol pollution (τ(550 nm) > 0.8) with increased or decreased tropospheric NO_2 VCDs depending on the geophysical conditions and aerosol properties and distributions. However, the resulting AMF computation becomes dependent on these data sources, their quality and the possibility (or not) to combine them altogether. In general, spatial and temporal co-registration between the different instruments or due to different resolution between the observation pixel and the model grid cell may become an issue. In the frame of an operational processing, it is generally preferred to maximize the exploitation of the spectral measurement acquired by a same instrument representative of the considered observation pixel. Exploitation of the 477 nm O_2-O_2 dimer absorption for aerosol retrieval is very promising. It is not only measured by OMI, but also by GOME-2, TROPOMI, Sentinel-4-UVN and Sentinel-5-UVNS. Several studies based on ground-based and satellite instrument have demonstrated its relative high sensitivity to aerosols, in particular to ALH (Wagner et al., 2004; Castellanos et al., 2015; Park et al., 2016; Chimot et al., 2016, 2017).

Because of the difficulty to easily distinguish clouds from aerosols, and to identify the right aerosol model to use, it has always been preferred to retrieve effective clouds assuming a Lambertian and opaque reflector model (Acarreta et al., 2004; Stammes et al., 2008) and consequently to compute the resulting troposphere NO_2 AMF for all the OMI scenes, regardless of the presence of clouds, aerosols or both together. Such correction is historically named an "implicit aerosol correction" (Boersma et al., 2004, 2011). Chimot et al. (2016) clearly demonstrated that, in spite of its implicit nature, such a correction allows to mitigate biases on the OMI DOMINO NO_2 product over cloud-free scenes compared to an aerosol-free pixel assumption. However, limitations were identified: 1) a numerical artifact is present due to a too coarse sampling employed in the OMI cloud look-up-table (LUT) leading then to a strong underestimation on the OMI tropospheric NO_2 VCD over scenes with strong aerosol load (τ(550 nm) $\geq$ 0.6) and particles located at high altitude, 2) the Lambertian cloud model, in spite of its benefits, remains somehow too simple and likely does not fully reproduce all the multiple scattering effects inherent to aerosol particles. The OMI effective cloud algorithm was then updated in order to remove these numerical artifacts (Veefkind et al., 2016). It also includes many additional relevant improvements of the OMI 477 nm O_2-O_2 band. But its impact on the correction for aerosols has not yet been evaluated.

To move one step further, Chimot et al. (2017) developed a novel machine learning algorithm, based on the neural network (NN) technique, that allows to retrieve ALH together with aerosol optical thickness (AOT) τ from the same OMI 477 nm O_2-O_2 spectral band over cloud-free scenes. These retrievals were performed on various cases both over land and sea and compared with reference CALIOP observations and related climatology (Chimot et al., 2017, 2018b). They benefit from a strong synergy with MODIS on-board the NASA Aqua platform flying together with Aura in the same NASA A-Train constellation, in order to identify the cloud-free scenes and to better constrain the ALH retrieval quality. For such a purpose, the 477 nm O_2-O_2 band represents some advantages compared to the more traditional O_2-A band: 1) it is spatially closer to the NO_2, HCHO, or SO_2 absorption lines, 2) it has a wider spectral range (over 10 nm) although weaker signal leading to high sensitivities in case of high aerosol loading, and 3) less challenges due to saturation. Moreover, the NN technique development allows a very fast OMI data processing which is an important requirement within an operational environment. The aerosol retrievals performed with this algorithm are expected to lead to an explicit aerosol correction over cloud-free scenes by using the OMI 477 nm O_2-O_2 measurement simultaneously acquired with the 405-465 nm NO_2 band.

This paper aims to evaluate the benefits of our improved use of the OMI 477 nm O_2-O_2 band for correcting aerosol effects in tropospheric NO_2 VCD retrieval from the same visible observation. We evaluate the potential of directly using the OMI NN aerosol ALH (and τ) in view of an explicit correction. We also assess the expected changes in the implicit aerosol correction based on the improved OMI effective cloud algorithm. To compare the different aerosol correction strategies, we reprocessed 2 years (2006-2007) of the DOMINO-v2 data: over the large urban and industrialized east China region in summer (June-July-August) and winter (December-January-February) times dominated by continental polluted particles ; and over South-America during the biomass burning season associated with heavy absorbing aerosol emission (August-September). Sections 6.2 and 6.3 describe the algorithms and the reprocessing methodology. Section 6.4 evaluates the results of the applied aerosol corrections in the reprocessed tropospheric NO_2 retrievals. To complete the analyses, Sect. 6.5 includes specific discussions based on academic simulations to better understand the behaviour of the new OMI tropospheric NO_2 VCD. Similarly to Chimot et al. (2016), the advantage of such simulations is to determine, on well controlled cases, the expected new biases on the OMI tropospheric NO_2 VCD and to identify the key factors driving them. At the end, in Sect. 6.6, we conclude about the benefits and challenges of each aerosol correction.

6.2 The OMI $O_2 - O_2$ algorithms

6.2.1 $O_2 - O_2$ DOAS spectral fit

In this paper, both effective cloud and aerosol algorithms are based on the same OMI 477 nm $O_2 - O_2$ spectral band. More specifically, they use the continuum reflectance R_c(475 nm) and the O_2-O_2 SCD $N^s_{O_2-O_2}$. These variables are derived from the DOAS spectral fit approach, which is then a prerequisite before applying either the OMI cloud look-up-table (LUT) (see Sect. 6.2.2) or the aerosol neural networks (see Sect. 6.2.3).

The DOAS method is a specific retrieval approach following basic principle of absorption spectroscopy employed for UV and visible absorbing trace gases. The various DOAS techniques rely on the same key concept: a simultaneous fit of several trace gas slant column densities from the fine spectral features due to their absorption (i.e. the high frequency part) present in passive UV–visible spectral measurements of atmospheric radiation (Platt and Stutz, 2008). The assumed Beer-Lambert (or Bouguer-Lambert) law describes the light attenuation as a function of the traveled distance in the atmosphere, gas concentration and its spectral absorption intensity. It is commonly employed for absorption spectroscopy analyses of NO_2, SO_2, HCHO and O_3 from the OMI, TROPOMI, GOME, GOME-2 and SCIAMACHY sensors: e.g. (Boersma et al., 2011; De Smedt et al., 2017).

To take into account the simultaneous presence of several absorbers together with Rayleigh and Mie scattering, the measured reflectance R is usually written as follow:

$$\frac{\mathrm{d}R(\lambda)}{\mathrm{d}s} = -\sum_i \sigma_i(\lambda) \cdot N_i(s) + \sigma_{Ray}(\lambda) \cdot N_{Ray}(s) + \sigma_{Mie}(\lambda) \cdot N_{Mie}(s). \tag{6.1}$$

with λ the wavelength, σ the absorption cross-section of gases i, or Rayleigh and Mie scattering, and ρ the distribution of trace gas density. The spectral fit is achieved within a predefined spectral window and the slant column density N^s is defined as the column density along the average light path travelled by the detected photons from the Sun through the atmosphere, surface and back to the satellite sensor.

Here, the OMI 477 nm O_2-O_2 DOAS fits together the absorption cross-section spectrum of O_2-O_2 with a first-order polynomial over the (460–490 nm) spectral band (Acarreta et al., 2004; Veefkind et al., 2016). The continuum reflectance R_c at the reference wavelength $\lambda_0 = 475$ nm is the reference which would be measured in the absence of absorbers.

Recent improvements include outlier removals and a necessary temperature correction on $N^s_{O_2-O_2}$ prior to the effective cloud or aerosol retrievals (Veefkind et al., 2016).

In absence of clouds, both OMCLDO2 and OMI aerosol algorithms rely on how aerosols affect the length of the average light path along which the O_2-O_2 absorbs. R_c(475 nm) is known to represent the enhanced scene brightness due to the additional scattering effects induced by the particles. In particular, R_c(475 nm) directly increases with increasing τ. This enhancement depends on aerosol properties as well as the surface albedo (Boersma et al., 2011; Chimot et al., 2016; Castellanos et al., 2015) $N^s_{O_2-O_2}$ is governed by the overall shielding or enhancement effect of the absorption of the photons by the O_2-O_2 complex in the visible spectral range along the average light path. A reduction of the length of the average light path, i.e. shielding effect, reduces the absorption possibility by the O_2-O_2. The aerosol layer height is the primary driver (Castellanos et al., 2015; Chimot et al., 2016, 2017). An aerosol layer located at high altitudes applies a large shielding effect on the O_2-O_2 located in the atmospheric layers below, by reducing the amount of photons coming from the top-of-atmosphere and reaching the lowest part of the atmosphere compared to an aerosol-free scene. As a second order, aerosol properties, load and surface conditions additionally contribute to $N^s_{O_2-O_2}$.

6.2.2 OMI cloud algorithm OMCLDO2

The OMI cloud algorithm also named OMCLDO2 (Acarreta et al., 2004) derives the effective cloud fraction cf and cloud pressure cp assuming the cloud model as an opaque Lambertian reflector with a constant albedo of 0.8 (Stammes et al., 2008) and the independent pixel approximation (IPA) (Martin et al., 2002; Acarreta et al., 2004). The measured reflectance R is formulated as a linear combination of a clear-sky R_{Clear} and a cloudy reflectance R_{Cloud} (Martin et al., 2002; Acarreta et al., 2004):

$$R(\lambda) = \mathrm{c}_f \cdot \mathrm{R}_{Cloud} + (1 - \mathrm{c}_f) \cdot \mathrm{R}_{Clear}. \tag{6.2}$$

A look-up-table (LUT) enables the conversion of R_c(475 nm) and $N^s_{O_2-O_2}$ into effective cloud parameters. It requires knowledge on the surface reflectance and surface pressure in addition to satellite and sun geometry configurations (Acarreta et al., 2004; Veefkind et al., 2016). Because of the low impact of small clouds on the O_2-O_2 band, effective cp has large uncertainties in case of low effective cf (Acarreta et al., 2004). The term "effective" here means these cloud parameters do not represent actual clouds, but our best explanation of the measured radiance by combining these variables with the assumed approximate model (Sneep et al., 2008). Therefore, the retrieved cf and cp values adapt to each observed scene to match with the measurement summarized by (R_c(475 nm)- $N^s_{O_2-O_2}$), such that the (460–490 nm) radiance budget is comprehensively closed (apart from instrument noise). For example, true optically thin clouds will be retrieved as a an opaque and bright Lambertian reflector covering only a small part of the OMI pixel, mostly because of the large assumed cloud albedo value (Veefkind et al., 2016).

The main motivation of this cloud retrieval scheme has been the correction of cloud effects in trace gas retrievals (Stammes et al., 2008). However, this algorithm is actually applied both to cloudy and cloud-free scenes with aerosols, without any prior distinction. Many studies demonstrated that OMCLDO2 accounts for a large part of aerosol effects by retrieving the effective cf and cp (Boersma et al., 2004, 2011; Castellanos et al., 2015; Chimot et al., 2016; Wang et al., 2015a). Under these conditions, the OMI cloud parameters become then "more effective" as they do not represent anymore cloud but aerosol effects on the (460–490 nm) radiance. One could claim that OMCLDO2 becomes then an approximate aerosol model, independent of those considered in Sect. 6.2.3. Chimot et al. (2016) demonstrated how OMCLDO2 responses to aerosols: 1) The effective cf is mostly driven by R_c(475 nm) and increases with increasing aerosol load, regardless of its altitude. Its magnitude is weighted by aerosol properties and surface conditions. 2) The effective cp represents beforehand the degree of shielding effect applied by aerosols, which results from a complex combination of ALH as a first order, aerosol load τ and type, surface properties and geometry angles as a second order. A stronger shielding effect leads to a reduced cp. In general, over scenes with high τ values, the effective cp correlates well with ALH although the values are not identical. Furthermore, regardless of true aerosol layer altitude, absorbing particles lead to a decrease of cp, while the presence of more scattering particles increases cp values (Castellanos et al., 2015; Chimot et al., 2016).

Initially, the coarse sampling of the OMI cloud LUT associated with the OMCLDO2-Old version created a numerical artefact: the effective cloud pressure was increasing with decreasing effective cloud fraction (or aerosol τ) without any physical explanation (Chimot et al., 2016; Veefkind et al., 2016). To remove this, OMCLDO2 has recently been updated (cf. OMCLDO2-New version) by generating a LUT with a higher sampling, especially in the continuum reflectance dimension (Veefkind et al., 2016). It has to be noted that the OMCLDO2 version employed in DOMINO included not only the coarse OMI cloud LUT (like OMCLDO2-Old) but also a an older version of the OMI surface albedo climatology (see further details in Sect. 6.3.3). A summary of all the OMI cloud algorithms is given in Tab. 6.1.

6.2.3 OMI aerosol neural network

The OMI O_2-O_2 aerosol algorithm relies on a NN multilayer perceptron approach to retrieve primarily the key parameter ALH over cloud-free scenes, but also aerosol

Table 6.1: Summary of the different aerosol correction algorithms considered in this study with a list of the associated parameters: input OMI Lambertian Equivalent Reflectivity (LER) climatology, OMI cloud look-up table (LUT) for the effective cloud retrievals, Aerosol parameters (see for more details Sect. 6.2.2-6.2.3):

Aerosol Correction algorithm	LER dataset	Effective clouds	Aerosols
DOMINO	3-year climatology	Coarse LUT	
OMCLDO2-Old	5-year climatology	Coarse LUT	
OMCLDO2-New	5-year climatology	High-sampling LUT	
$NN_{MODIS,XX}$	5-year climatology	–	MODIS τ, OMI ALH
$NN_{OMI,XX}$	5-year climatology	–	OMI τ and ALH

τ(550 nm) as a secondary parameter (Chimot et al., 2017, 2018b). Since aerosol vertical profiles cannot be retrieved from OMI UV-visible measurements, they are assumed as one box layer with a constant geometric thickness (100 hPa). ALH is the mid-altitude of this layer in km over sea level but can also be expressed in pressure. Here, the strategy differs from the OMI effective clouds of Sect. 6.2.2. The main motivation is to try to reproduce aerosol scattering effects in the visible spectrum via a more explicit model. Therefore, the idea of retrieving ALH here is to determine a more realistic aerosol altitude value less dependent on actual aerosol properties by opposition to the effective cp. However, any inaccurate prior assumption such as an inadequate aerosol model will obviously bias ALH.

The particle properties in this layer are homogeneous. Among all the input parameters, ALH retrieval requires a prior τ(550 nm) knowledge as ALH and τ(550 nm) simultaneously affect $N^s_{O_2-O_2}$ and need to be separated (Chimot et al., 2017). In theory, this source of information may be diverse (e.g. atmospheric models, prior guess, or observations). In practice, MODIS τ product has systematically been preferred due to its good spatio and temporal collocation with OMI and its recognized high quality. Retrieved OMI τ may also be used as they come from a same spectral measurement (same instrument). However due to their higher uncertainty compared to MODIS, they impact the quality of OMI ALH (Chimot et al., 2017). For OMI τ retrieval, R_c(475 nm) is considered instead of τ as prior input. Note that in the next sections we, define NN_{MODIS} when prior MODIS τ is considered, NN_{OMI} based on the retrieved OMI τ, and NN_{True} when true τ value is considered for the synthetic cases (see Sect. 6.3.1).

The training data was generated by full-physical spectral simulations, assuming explicit aerosol particles without clouds, generated by he Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval (DISAMAR) software of KNMI (de Haan, 2011). Aerosol scattering phase function $\Phi(\Theta)$ was simulated by the Henyey–Greenstein (HG) function parameterized by the asymmetry parameter g, the average of the cosine of the scattering angle (Hovenier and Hage, 1989). Aerosols were specified as standard fine particles with a unique value of the extinction Ångström exponent $\alpha = 1.5$ and $g = 0.7$. They are assumed to fully cover the OMI pixel. To take into account the inaccuracies of the assumed aerosol single scattering albedo ω_0 properties, two training data sets were generated with a different typical value: one with $\omega_0 = 0.95$ and one with $\omega_0 = 0.9$ in the visible spectral domain. Therefore, two OMI ALH NN algorithms have been created, one for each aerosol ω_0 values.

The HG function is known to be smooth and reproduce the Mie scattering functions reasonably well with $g = 0.7$ for most of aerosol types (Dubovik et al., 2002). A similar approach is considered for the operational ALH retrieval algorithms for Sentinel-4 and Sentinel-5 Precursor (Leitão et al., 2010; Sanders et al., 2015; Colosimo et al., 2016; Nanda et al., 2017), and when applying various explicit aerosol corrections in the tropospheric NO_2 AMF calculation over urban and industrial areas dominated by anthropogenic pollution, for instance in east China (Spada et al., 2006; Wagner et al., 2007; Castellanos et al., 2015; Vlemmix et al., 2010).

Similarly to the high effective cp inaccuracy in case of low effective cf, high ALH bias is expected below a minimum particle load (i.e. threshold of $\tau(550\text{ nm}) = 0.5$). This is directly due to the nature of the O_2-O_2 spectral band. Below this threshold, low amounts of aerosols have negligible impacts on $N^{s}_{O_2-O_2}$ signal.

Over cloud-free scenes, OMI ALH has shown consistent spatial patterns with CALIOP level 2 (L2) aerosol heights over urban and industrial areas in east China, with an uncertainty in the range of 500-700 m and for collocated MODIS scenes with $\tau(550\text{ nm}) \geq 0.5$ (Chimot et al., 2018b). Additional comparisons with the LIdar climatology of vertical Aerosol Structure for space-based lidar simulation (LIVAS) and 3-year OMI ALH with MODIS $\tau(550m) \geq 1.0$ were in the range of 180–800 m (Amiridis et al., 2015; Chimot et al., 2017). Finally, Chimot et al. (2018b) showed the potential of OMI visible measurements to observe the height of thick and absorbing aerosol layers released by widespread fire episodes such as in South-America. The aerosol model assumptions, in particular ω_0, are the most critical as they may affect ALH retrieval uncertainty up to 660 m. Potential aerosol residuals in the considered surface albedo may impact up to 200 m. An accuracy of 0.2 is necessary on prior $\tau(550\text{ nm})$ information to limit ALH bias close to zero over scenes with $\tau(550\text{ nm})$ ≥ 1.0, and below 500 m when $\tau(550\text{ nm})$ values below 1.0.

A summary of all the OMI NN aerosol algorithms as well as related input and output parameters is given in Tab. 6.1.

6.3 From aerosol impacts to aerosol correction – Methodology

6.3.1 General methodology

OMI tropospheric NO_2 data are taken from the DOMINO-v2 dataset (see Sect. 6.3.3). They are reprocessed by recomputing the AMF (see Sect. 6.3.2) using the DISAMAR radiative transfer model over cloud-free scenes contaminated by aerosols. This AMF replaces then the original AMF of DOMINO and is applied to the available NO_2 SCD.

The computation of the tropospheric NO_2 AMF follows the formulation detailed in Sect. 6.3.2 and applies either an implicit or an explicit aerosol correction: the implicit correction considers the effective cloud retrievals obtained from OMCLDO2 (cf. Sect. 6.2.2); the explicit aerosol correction employs aerosol parameters: either ALH and OMI τ from the OMI aerosol NN (cf. Sect. 6.2.3), or ALH and MODIS τ. The complementary aerosol parameters (i.e. ω_o, g, α) follow those specified in the associated training dataset.

The surface albedo is based on the OMI Lambertian Equivalent Reflectivity (LER) climatology (Kleipool et al., 2008). In DOMINO-v2, this climatology is based on a 3-year OMI time series measurements. However, it evolved since with an extended 5-year OMI time series (Veefkind et al., 2016). This evolved OMI LER is considered for all the tropospheric NO_2 VCD reprocessing performed in this study. All the other geophysical parameters associated with DOMINO-v2, such as the NO_2 vertical profile, remain identical.

To identify cloud-free OMI observation pixels with aerosols, a similar strategy as Chimot et al. (2016, 2017) is considered. The DOMINO-v2 NO_2 scenes are collected together with the MODIS-Aqua aerosol τ(550 nm) from the combined Dark Target (DT) and Deep Blue (DB) product of COllection 6 available at the resolution of 10 km (Levy et al., 2013). They are collocated within a distance of 15 km. The probability of cloud-free OMI scene is a priori ensured by the availability of the MODIS aerosol product with the highest quality assurance flag. MODIS Aqua τ is then exclusively retrieved when a sufficient high amount of cloud-free sub-pixels is available (i.e. at the MODIS measurement resolution of 1 km) (Levy et al., 2013). However, it is well recognized this may be not completely representative of the atmospheric situation of the OMI pixel. Therefore, we added two thresholds for each collocated OMI-MODIS pixel: the geometric MODIS cloud fraction to be smaller than 0.1, and the effective OMI cloud fraction lower than 0.1. Past experiences OMI cf values higher than 0.2 may still contain clouds (or both clouds and aerosols) (Boersma et al., 2011; Chimot et al., 2016).

Additional synthetic cases analysed in Sect. 6.4 and 6.5 are also based on the DIMSAR model, specified in a similar way than the NN training dataset in Sect. 6.2.3. Then either OMCLDO2 or the OMI NN aerosol algorithms are used to determine the expected tropospheric NO_2 VCD biases.

6.3.2 Air mass factor computations

The computation of tropospheric NO_2 AMF A_{NO_2} is a key step for converting NO_2 SCD $N^{s}_{NO_2}$ into tropospheric NO_2 VCD $N^{v}_{NO_2}$, which represents the number of NO_2 molecules cm^{-2} integrated along the vertical atmospheric layers (by opposite of along the average light path) from the surface P_0 to the tropopause P_{trop} pressure. Its has generally been recognized as the principal source of errors in the tropospheric NO_2 vertical column density (VCD) determination in areas with a high level of air pollution (Boersma et al., 2007). This was even more emphasized by Lorente et al. (2017) who discussed how AMF structural uncertainty significantly increases because of assumed prior information, and cloud and aerosol correction strategies: up to 42% over polluted regions, and 31% over unpolluted regions.

In the context of OMI visible spectral measurements, a fifth-degree polynomial is considered over the 405-465 nm spectral band (Boersma et al., 2011). The application of $A^{\mathrm{v}}_{\mathrm{NO}_2}$ is crucial to correct of the average light path variability contained in $N^{\mathrm{s}}_{\mathrm{NO}_2}$. A_{NO_2} is defined as the ratio of the atmospheric SCD and VCD (Boersma et al., 2011):

$$A^{\mathrm{v}}_{\mathrm{NO}_2}(\Psi,\lambda) = N^{\mathrm{s}}_{\mathrm{NO}_2}(\Psi,\lambda)/N^{\mathrm{v}}_{\mathrm{NO}_2}, \tag{6.3}$$

with Ψ the list of input parameters prerequisite for the radiative transfer model. Note that before performing this conversion, the stratospheric and tropospheric contributions to $N^{\mathrm{s}}_{\mathrm{NO}_2}$ must be separated. Therefore, $A^{\mathrm{v}}_{\mathrm{NO}_2}(\Psi,\lambda)$ is only applied to the tropospheric NO_2 SCD.

The OMI tropospheric NO_2 formulation follows Palmer et al. (2001); Boersma et al. (2004) and the concept of altitude-resolved AMF a (also named block AMF or BAMF) introduced by Wagner et al. (2007); Rozanov and Rozanov (2010); Richter and Wagner (2011). The ratio of a to the total air mass factor A_{NO_2} (deduced from the NO_2 shape profile) gives the vertical averaging kernel AK: i.e. the sensitivity of the satellite measurement to each vertical atmospheric layer (Eskes and Boersma, 2003; Richter and Wagner, 2011). Overall, $A^{\mathrm{v}}_{\mathrm{NO}_2}$ can then be seen as a unitless number representative of the length of the average light path followed by the detected photons in the troposphere. It includes then an indication about the sensitivity to the amount of NO_2 in the troposphere, larger values indicating a higher sensitivity assuming no change in vertical NO_2 profile. Indeed, in those cases, a change in $A^{\mathrm{v}}_{\mathrm{NO}_2}$ is directly associated with a change of a at the atmospheric levels where the trace gas is present. a is in theory spectrally dependent, but the spectral dependence in case of tropospheric NO_2 retrieval is very small. The reference wavelength considered in this paper is 439 nm, following the OMI NO_2 product (see Sect. 6.3.3) (Boersma et al., 2011).

Aerosols may apply either a shielding or an enhancement effect. A shielding effect occurs when the length of the average light path is reduced leading then to a decrease of $A^{\mathrm{v}}_{\mathrm{NO}_2}$. Reciprocally, an enhancement affect results in an increase of $A^{\mathrm{v}}_{\mathrm{NO}_2}$ (Leitão et al., 2010; Chimot et al., 2016). Following Eq. 6.3, any bias in $A^{\mathrm{v}}_{\mathrm{NO}_2}$ calculation leads to a direct bias into $N^{\mathrm{v}}_{\mathrm{NO}_2}$: same value but opposite sign. For example, an over-shielding applied in $A^{\mathrm{v}}_{\mathrm{NO}_2}$, due to a too high ALH or cp for instance, leads to an underestimation of $A^{\mathrm{v}}_{\mathrm{NO}_2}$ and consequently an overestimation of $N^{\mathrm{v}}_{\mathrm{NO}_2}$.

Note that in the case of real OMI tropospheric NO_2 retrievals, a temperature correction must be applied as the temperature of the assumed NO_2 absorption cross section, fixed at 221 K, can differ from the actual temperature when deriving $N^{\mathrm{s}}_{\mathrm{NO}_2}$. The correction term is thus implemented in the computation of $A^{\mathrm{v}}_{\mathrm{NO}_2}$ such that it represents the ratio of $N^{\mathrm{s}}_{\mathrm{NO}_2}$ derived with a NO_2 cross section at the real temperature T to the column derived at 221 K. European Centre for Medium-Range Weather Forecasts (ECMWF) temperature fields are used for this correction (Boersma et al., 2004, 2011).

Computation of $A^{\mathrm{v}}_{\mathrm{NO}_2}$ requires accurate knowledge about all the parameters Ψ affecting the optical properties of the atmosphere and the length of the average light path. For an aerosol and cloud-free scene, Ψ generally includes the satellite and solar geometries, ground-pressure and the Lambertian surface albedo. $A^{v}_{NO_2}$ computation in presence of cloud and aerosol particles is known to be a complex problem. The main reason is due to the complex light paths resulting from particle scattering effects, and therefore the complex methodologies to determine a. Effective cloud parameters derived from OMCLDO2 (see Sect. 6.2.2) allow to correct of cloud effects.

Ideally, an explicit aerosol correction requires comprehensive multiple scattering radiative transfer calculations combined with an assumption on the vertical distribution of the absorber. Currently only the effective cloud parameters (and no aerosol parameters) are considered when computing a in the OMI tropospheric NO_2 product, even over cloud-free scenes contaminated by aerosol particles. Since aerosols and clouds cannot be easily distinguished in the frame of the OMI operational chain, the modified effective clouds (cf. see Sect. 6.2.2) apply then a so-called "implicit aerosol correction" in the tropospheric NO_2 retrieval (Boersma et al., 2011; Chimot et al., 2016).

To apply an explicit aerosol correction, ALH combined with τ are the most crucial necessary variables to control the quality of a calculation (Leitão et al., 2010; Bousserez, 2014; Chimot et al., 2016). However, an aerosol model must be selected to be able to use these parameters. The retrievals performed by the OMI NN aerosol algorithms in Sect. 6.2.3 can then be used by combining their results consistently with the assumed prior information (i.e. prior τ and aerosol model considered during the training phase).

6.3.3 OMI tropospheric NO_2 dataset – DOMINO v2

The Derivation of OMI tropospheric NO_2 (DOMINO) version 2 (Boersma et al., 2011) is a reference worldwide tropospheric NO_2 product derived from the OMI visible measurements and can be downloaded from the Tropospheric Emissions Monitoring Internet Service (TEMIS) website (`http:www.temis.nl`).

The calculation of a is achieved though a linear interpolation within a LUT represented as a function of the required input geophysical parameters Ψ.

Chimot et al. (2016) demonstrated that the implicit aerosol correction in DOMINO-v2 is better than the clear-sky assumption (Richter and Burrows, 2002)), with remaining biases between -20% and -40% on tropospheric NO_2 VCD, especially in presence of absorbing particles and for $\tau(550 \text{ nm}) \geq 0.5$. One the main identified limitation was the coarse sampling of the OMI cloud LUT nodes used in OMCLDO2 (see Sect. 6.2.2). The effect of OMCLDO2-New version on the implicit aerosol correction have not yet been analyzed.

At our knowledge, no reprocessing has yet been done by applying an explicit aerosol correction based on (nearly) explicit aerosol parameters that are retrieved from the OMI 477 nm O_2-O_2 spectral band. Thus, the use of OMI ALH and τ parameters from Sect. 6.2.3 is a first attempt to apply a (nearly) explicit aerosol correction in A_{NO_2} computation by using visible spectral measurements acquired by a same sensor.

DOMINO has recently evolved through the Quality Assurance for Essential Climate Variables (QA4ECV) project (`www.qa4ecv.eu`) which aims to address reliable and fully tracable quality infomration on some of the "essential climate variables" (ECVs) , such as tropospheric NO_2, as defined by the Global Climate observing System (GCOS) (Lorente et al., 2017). This reprocessing contains numerous changes in the complete chain of retrieval, from the calibrated spectrum, spectral fitting with DOAS, to the AMF computation and all the ancillary dataset. This new generation of product is expected to represent one of the best NO_2 dataset. Since the reprocessing products of QA4ECV is still under thorough validation at the time of this paper (and its technical work), and given our specific objective focused on the aerosol radiation correction by using information from the 477 nm, DOMINO has been preferred to avoid interference from other modified parameters.

At the time of the DOMINO development, the employed OMI surface albedo climatology (Kleipool et al., 2008) was based on a 3-year OMI UV-visible spectra time series for both the derivation of the effective cloud parameters and the AMF computation (Boersma et al., 2011). Since then, this climatology has been upgraded by extending this time series from 3 to 5 years. Such changes may have some great impacts on both the NO_2 and O_2-O_2 algorithms. As discussed by Boersma et al. (2011); Lorente et al. (2017), AMFs over clear-sky (and potentially also in presence of aerosols) are very sensitive to the assumed surface properties. These aspects are further discussed in Sect. 6.4.1.

The OMI cloud algorithm configuration used at the time of DOMINO-v2, and its comparison with the other algorithms are summarized in Tab. 6.1.

6.4 Reprocessed OMI NO_2 and O_2-O_2 products – Results

All the OMI tropospheric NO_2 reprocessing achieved here are based on the OMI cloud and aerosol algorithms discussed in the previous sections, and summarized in Tab. 6.1. The differences between the different reprocessings are synthesized in Tab. 6.2 for all collocated OMI-MODIS aerosol scenes (i.e. $\tau(550\text{ nm}) \geq 0.$) However, some analyses below zoom over scenes with high aerosol load.

6.4.1 Impact of OMI cloud algorithm changes

The main changes in the OMCLDO2 algorithm that can overall impact $N^{\mathrm{v}}_{\mathrm{NO_2}}$ in DOMINO are of two types:

- the extension of UV-visible time series observations considered for the OMI LER climatology from 3 years (as implemented in DOMINO) to 5 years (cf. QA4ECV);
- the increased sampling of the OMI cloud LUT from the OMCLDO2-Old to the OMCLDO2-New version.

Table 6.2: Summary of the changes in the diverse reprocessing OMI tropospheric NO_2 VCD $N^{v}_{NO_2}$ depending on the applied aerosol correction strategy (see Tab. 6.1) over all collocated MODIS aerosol scenes (MODIS $\tau(550\text{ nm}) \geq 0.$). See more analyses in Sect. 6.4.1-6.4.3:

Focus on:	Aerosol correction strategies comparison	Region - Season	Changes in $N^{v}_{NO_2}$ in [%] Average ± Standard Deviation
Implicit correction	OMCLDO2-Old − DOMINO	China - summer	−1.0 ± 9.0
		China - winter	−15.6 ± 29.8
		South-America - biomass burning	−6.2 ± 16.0
	OMCLDO2-New − OMCLDO2-Old	China - summer	1.3 ± 6.9
		China - winter	7.9 ± 19.3
		South-America - biomass burning	7.9 ± 14.4
	OMCLDO2-New − DOMINO	China - summer	0.4 ± 10.6
		China - winter	−4.0 ± 26.8
		South-America - biomass burning	3.1 ± 17.3
Explicit correction	$NN_{MODIS,\omega_0=0.95}$ − OMCLDO2-New	China - summer	−2.9 ± 12.5
		China - winter	6.8 ± 26.1
		South-America - biomass burning	−8.1± 16.8
	$NN_{MODIS,\omega_0=0.9} - NN_{MODIS,\omega_0=0.95}$	China - summer	−0.2 ± 7.8
		China - winter	−8.2 ± 22.3
		South-America - biomass burning	1.3 ± 8.7
	$NN_{OMI,\omega_0=0.95}$ − OMCLDO2-New	China - summer	6.5 ± 11.9
		China - winter	11.2 ± 18.4
		South-America - biomass burning	−3.0 ± 14.0
	$NN_{OMI,\omega_0=0.9} - NN_{OMI,\omega_0=0.95}$	China - summer	8.5 ± 13.7
		China - winter	−1.9 ± 24.3
		South-America - biomass burning	1.0 ± 14.6

Figure 6.1 show overall an increase of OMI Lambertian surface albedo. These increases mostly occur in the north-east part of China, at the edge of the desert zones (e.g. Gobi desert), and in most of the parts of South-America.

They can reach up to 0.03, but are actually quite variable depending on the areas and seasons: from 0.97 ± 22.1 (%) in summertime over east-China to 27.2 ± to 37.8 (%) over South-America during the biomass burning season. Such changes might be due to the increase of used observations to update the climatology, although more analyses should be performed to understand the exact causes and the reliability.

Enhanced surface albedos primarily impact the retrieved effective cf since both parameters are strongly connected through R_c(475 nm). This last variable may be seen as an indicator of the scene brightness resulting from surface reflection combined with additional scattering effects induced by particles. Areas with a larger (lower) surface albedo will result in a lower (larger) effective cf to balance the overall interpretation of the scene brightness (cf. Figs. 6.1). This leads to new spatial patterns of the effective cf. The magnitudes may however greatly vary depending on the actual presence of aerosols in the scene and any potential aerosol or cloud residuals in the new OMI climatology. Although average cf changes are low (less than -0.008), standard deviation exhibit values up to 0.016.

Effective c_p is subject to major changes primarily because of the higher sampling nodes implemented in the OMI cloud LUT of OMCLDO2-New (cf. see Sect. 6.2.2). As depicted by Fig. 6.2, differences are quite significant at low τ. On average, c_p values from OMCLDO2-New are lower of about 200 hPa than those from OMCLDO2-Old (with large standard deviation) over scenes with MODIS $\tau(550\text{ nm}) \leq 0.5$. Indeed,

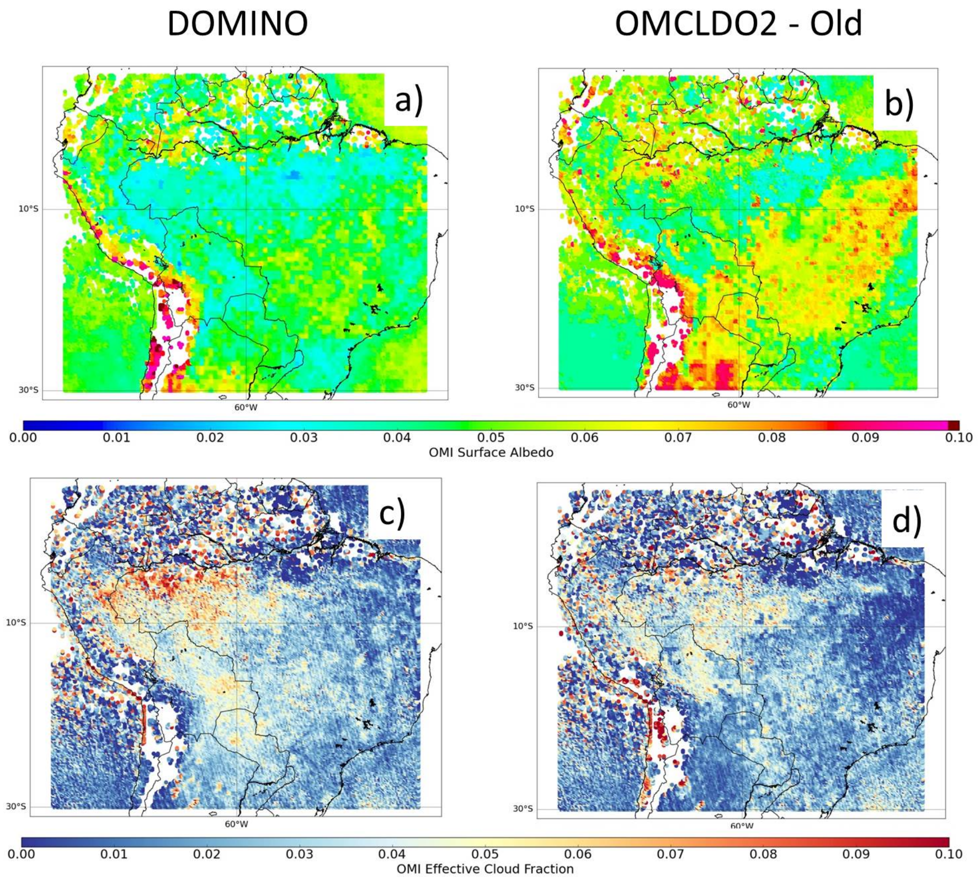


Figure 6.1: Changes in OMI climatology Lambertian surface albedo (Kleipool et al., 2008) and effective cf between DOMINO and OMCLDO2-old datasets over South-America during the biomass burning season (August-September) in 2006-2007 (see Sect. 6.3.3 and Sect. 6.4.1): , **(a)** DOMI surface albedo, **(b)** OMCLDO-Old surface albedo, **(c)** DOMINO effective cf , **(d)** OMCLDO2-old effective cf.

low aerosol load has very limited effects on $N^{s}_{O_2-O_2}$ and does not dominate the measured radiance signal. This result in large uncertainties on the retrieved effective cp and then a large sensitivity of the resolution at which the LUT interpolation is performed for these cases. Over scenes with high aerosol load (MODIS τ(550 nm) $\geq$ 1.0), differences are relatively minor and may even reverse sign. We attribute the small reverse sign to the application of the temperature correction on $N^{s}_{O_2-O_2}$ (see Sect. 6.2.1) which, depending on the temperature difference compared to the assumed mid-latitude summer atmosphere, may apply a positive or negative small modification of c_p in case of high τ. Overall, all these changes are consistent with those analyzed by Veefkind et al. (2016) over cloudy scenes, with low and high effective cf. The new OMI surface albedo climatology also governs changes in c_p, but with a second order of magnitude (although not negligible) compared to the upgraded LUT: between 8.8 ± 83.6 hPa in South-America and 44.1 ± 77.2 hPa in China summertime.

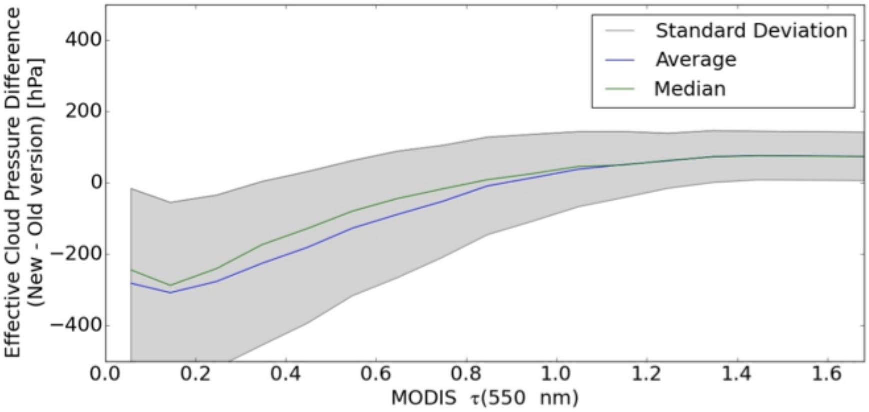


Figure 6.2: Statistics of effective cloud pressure differences from OMCLDO2 (new - old) in [hPa] in 2006-2007 as a function of MODIS aerosol optical thickness (AOT) τ(550 nm):. An example over China in summertime (June-July-August).

Figures 6.3a, c and e illustrate the changes in reprocessed $N^{\mathrm{v}}_{\mathrm{NO_2}}$ from DOMINO to OMCLDO2-Old. As indicated in Tab. 6.1, these changes result from two consequences: 1) the modified effective cloud parameters retrieved with the new OMI surface albedo, and 2) the application of this new albedo when computing $A^{\mathrm{v}}_{\mathrm{NO_2}}$. A higher surface albedo should result in an increased length of the average light path, and therefore an enhanced $A^{\mathrm{v}}_{\mathrm{NO_2}}$. However, this can become more complex when combined with the new effective cloud parameters as they may either enhance, attenuate or even counterbalance this effect. On average, $N^{\mathrm{v}}_{\mathrm{NO_2}}$ is lower (i.e. higher $A^{\mathrm{v}}_{\mathrm{NO_2}}$), between -1 ± 9 (%) in China summertime and -15.6 ± 29.8 (%) in China wintertime. The quality of these changes depends on the accuracy of the new surface albedo climatology.

Figures 6.3b, d and f) depict the impacts of the the implicit aerosol correction evolution from OMCLDO2-Old to OMCLDO2-New. They are directly driven by the improved effective cp (cf. Tab. 6.1). Over scenes with MODIS τ(550 nm) in the range of 0.0–0.5, a decreased c_p results in a stronger shielding (or reduced enhancement) effect from particles: $N^{\mathrm{v}}_{\mathrm{NO_2}}$ generally increases. By opposition, larger c_p over scenes with MODIS τ(550 nm) $\geq$ 1.0 leads to a lower shielding (or stronger enhancement) effect: $N^{\mathrm{v}}_{\mathrm{NO_2}}$ decreases. Standard deviation of these changes is between 15 and 20 % in China wintertime and South-America, and lower than 10% in China summertime and South-America. Averages are in the range of 1.3-7.9 %. Regional and seasonal differences may reflect the implicit dependencies on the aerosol types, the combined effects on c_f-c_p and the impacts of seasonal NO_2 vertical profile shape. Interestingly, these overall changes seem to be in line with the average AMF uncertainty of 11% evaluated by Lorente et al. (2017) due to different cloud correction scheme in polluted conditions and assuming low effective cloud fraction, $cf \leq 0.2$.

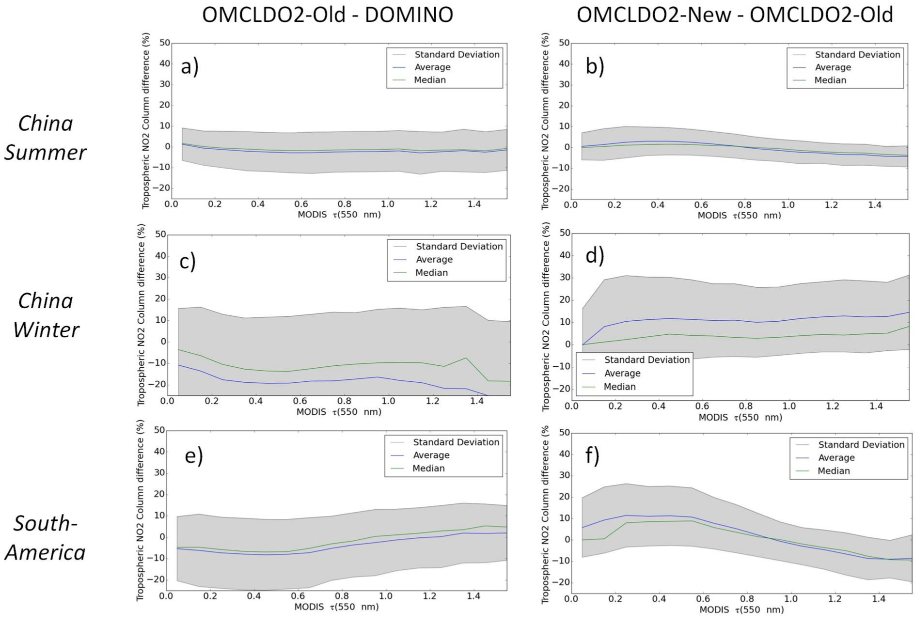


Figure 6.3: Statistics of relative $N^{v}_{NO_2}$ differences in (%) as a function of MODIS τ(550 nm) over China and South-America in 2006-2007 due to changes in the applied implicit aerosol correction (cf. Sect. 6.4.1) from DOMINO to OMCLDO2-Old, and from OMCLDO2-Old to OMCLDO2-New: **(a)** China Summer (June-July-August), **(b)** China Winter (December-January-February), **(c)** South-America (August-September).

Overall, maps in Fig. 6.4c-6.6c show that the total changes in $N^{v}_{NO_2}$, from DOMINO to OMCLDO2-New, mostly occur in the eastern part of China, where the NO_2 pollution is higher. Its spatial pattern mostly results from a complex combination with MODIS aerosol horizontal distribution as suggested by Fig. 6.3: a decrease over Bejing areas in summertime, and an increase in the same area in wintertime.

Figure 6.7 illustrates the expected improvements of the implicit aerosol correction on $N^{v}_{NO_2}$ due to the higher OMI cloud LUT sampling.

While remaining biases were contained between −20 and −40 (%) with OMCLDO2-Old, they should be now limited to the range of 0-20 (%) with the use of OMCLDO2-New over scenes with relatively scattering aerosol particles (i.e. ω_0=0.95) and assuming a typical NO_2 summer vertical profile shape. Such improvements are particularly good in case of aerosols located at elevated altitude (i.e. more than 1 km). However, although improved, biases can be higher in case of more absorbing particles: i.e. in the range of −10-20 (%). Additional geophysical parameters, in particular the NO_2 profile shape, may affect these biases and are therefore of high importance (see further discussions in Sect. 6.5.2). Finally, the observed changes in these changes are inline with those observed in Fig. 6.3b, d and f, confirming then our main analyses.

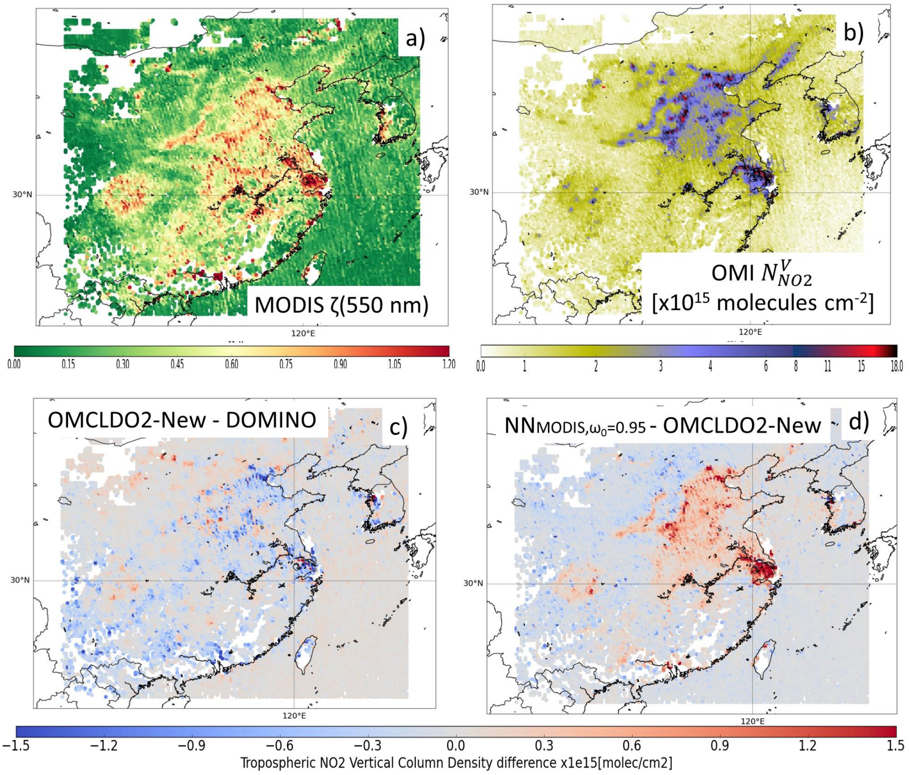


Figure 6.4: Average maps of MODIS τ(550 nm), OMI DOMINO $N^{\mathrm{v}}_{\mathrm{NO_2}}$ and differences after applying the implicit (with OMCLDO2-New) or explicit (with $NN_{MODIS,\omega_0=0.95}$) aerosol correction over China in summertime (June-July-August) 2006-2007: **(a)** MODIS τ(550 nm), **(b)** OMI DOMINO $N^{\mathrm{v}}_{\mathrm{NO_2}}$, **(c)** OMI $N^{\mathrm{v}}_{\mathrm{NO_2}}$ differences due to changes between OMCLDO2-New and DOMINO implicit aerosol corrections, **(d)** $N^{\mathrm{v}}_{\mathrm{NO_2}}$ differences between explicit aerosol correction based on the $NN_{MODIS,\omega_0=0.95}$ aerosol parameters (i.e. aerosol forward model assuming $\omega_0 = 0.95$, MODIS τ(550 nm), and retrieved ALH) and implicit aerosol correction from OMCLDO2-New.

6.4.2 OMI effective cloud and aerosol layer height differences

As discussed in Sect. 6.3.2, ALH is the first crucial parameter for the computation of $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm). Therefore, at a first assumption, it is expected that the accuracy of the OMI ALH retrieval, and its difference with the effective cp, will be the first causes (although not unique) of the difference between the applied implicit and explicit aerosol correction. The first priority is then to retrieve OMI ALH with MODIS τ(550 nm) in order to maximize its accuracy and mitigate the impacts of aerosol model (see Sect. 6.2.3).

Figure 6.8 compares the average OMI ALH (retrieved with the NN trained with aerosol $\omega_0 = 0.95$) and c_p from OMCLDO2-New, both converted in metric unit (km) over cloud-free scenes and for collocated MODIS scenes with $\tau(550\text{ nm}) \geq 0.5$. Overall, both variables are quite well correlated, presenting similar spatial and seasonal

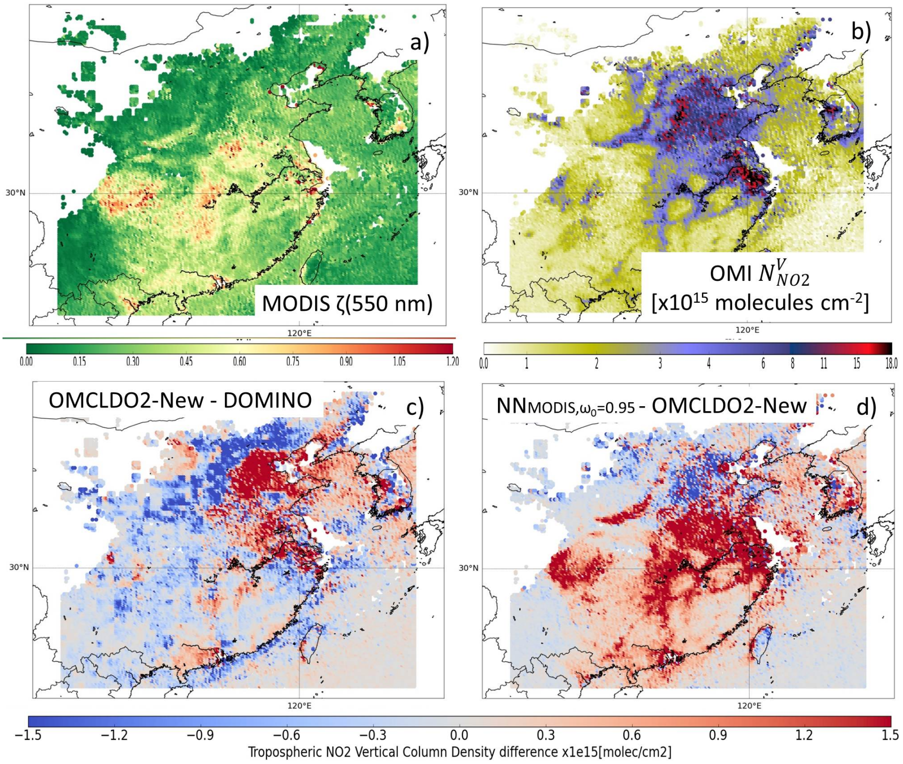


Figure 6.5: Same as Fig. 6.4 but over China in wintertime (December-January-February) 2006-2007

distributions. Values are higher in China summertime and over South-America and lie in the range of 1.5-5.0 km. They are lower in China wintertime, between 0.4 and 2.0 km.

Overall, ALH derived from $NN_{MODIS,\omega_0=0.95}$ are located higher than the effective clouds (i.e. aerosol pressure smaller than cp) with average differences lying in the range of -0.49--50.3 hPa. Standard deviation of the differences are in the order of 120 hPa. The sign of the differences is reversed when employing $NN_{MODIS,\omega_0=0.9}$ (average differences 12.9-59.3 hPa). This may suggest then that, in terms of optical property for layer height retrieval purpose, the opaque Lambertian reflector forward model, with an albedo of 0.8 and covering only a fractional part of the observation pixel, is somehow between (perhaps even at mid-level) the aerosol model fully covering the pixel with $\omega_0 = 0.9$ and $\omega_0 = 0.95$. Of course, both methods also include numerical errors, such as potential LUT interpolation artefact for OMCLDO2-New, and NN training uncertainty for the OMI ALH.

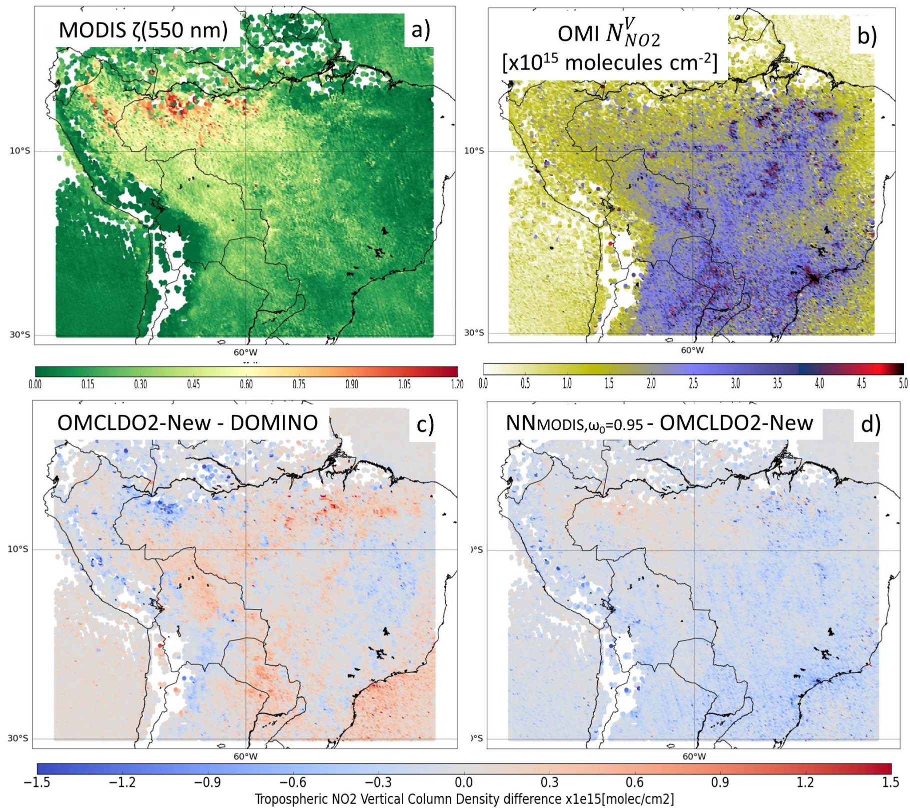


Figure 6.6: Same as Fig. 6.4 but over over South-America during the biomass burning season (August-September) in 2006-2007

6.4.3 Explicit vs. implicit aerosol correction

Over China OMI pixels with MODIS $\tau(550\text{ nm}) \geq 0.5$, Figs 6.9a, c, and e show that $N^{\mathrm{v}}_{\mathrm{NO_2}}$ is generally higher with the explicit aerosol correction from $NN_{MODIS,\omega_0=0.95}$ (i.e. NN based on prior MODIS τ and trained with aerosol $\omega_0 = 0.95$) than with the implicit aerosol correction from OMCLDO2-New. This then suggests a stronger shielding effect generally applied by this aerosol correction strategy leading to lower $A^{\mathrm{v}}_{\mathrm{NO_2}}(439\text{ nm})$. The differences increase with increasing MODIS τ as aerosol effects consequently amplify along the average light path: increases are about 12 ± 12.5 % in summer and 40 ± 26.1 % in winter over scenes with MODIS $\tau(550\text{ nm}) \geq 0.5$. The larger increase in wintertime is likely due to different NO_2 profile shape conditions, with NO_2 molecules being closer to the surface (see further discussions in Sect. 6.5.2). Figs. 6.4-6.6 shows that most of the changes in $N^{\mathrm{v}}_{\mathrm{NO_2}}$ are located on the eastern part, and over areas dominated by heavy NO_2 pollutions such as the mega-cities and the Pearl River Delta. Horizontal distribution of aerosol adds some complex patterns as summarized by Fig. 6.9.

As a first assumption, this stronger shielding effect is likely the result of using

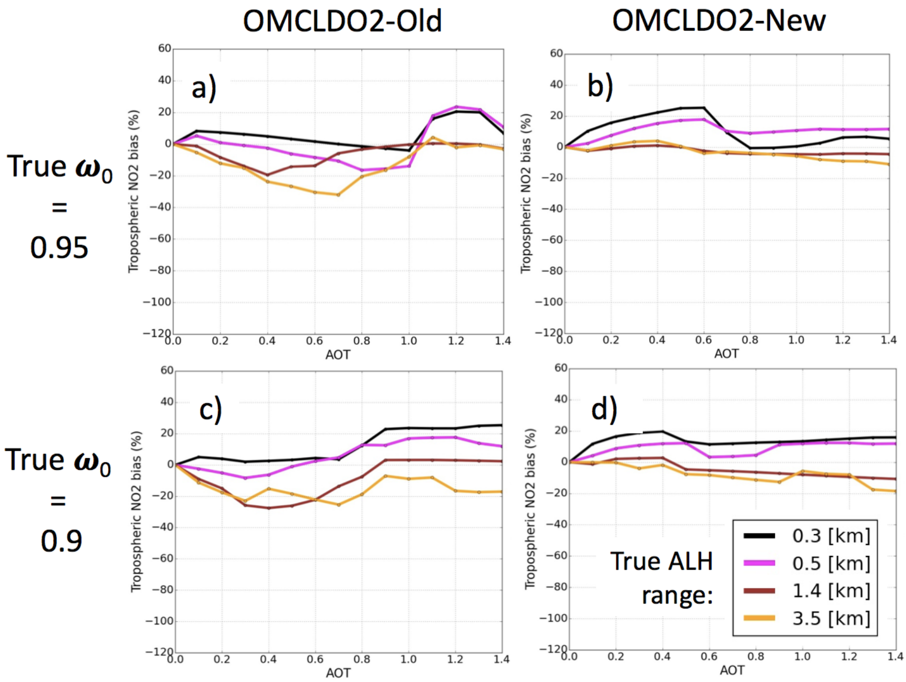


Figure 6.7: Relative $N^{v}_{NO_2}$ biases after application of the implicit aerosol correction as a function of true aerosol optical thickness (AOT) τ(550 nm) and based on synthetic cases including different true ALH, surface albedo = 0.05, $\mu_0 = 25°$, $\mu = 25°$, and a typical TM5 NO_2 vertical profile for 1st of July, 2006 at 12:00 pm over China (van Noije et al., 2014; Chimot et al., 2016). True aerosol properties are defined by $\alpha = 1.5$, $\omega_0 = 0.95$ or 0.9 and $g = 0.7$. Implicit aerosol correction is derived from the retrievals given by OMCLDO2-Old or OMCLDO2-New (cf. Sect. 6.2.2): **(a)** Relative $N^{v}_{NO_2}$ bias resulting from OMCLDO2-Old, true $\omega_0 = 0.95$, **(b)** Relative $N^{v}_{NO_2}$ resulting from OMCLDO2-New, true $\omega_0 = 0.95$, **(c)** Relative $N^{v}_{NO_2}$ bias resulting from OMCLDO2-Old, true $\omega_0 = 0.9$, **(d)** Relative $N^{v}_{NO_2}$ resulting from OMCLDO2-New, true $\omega_0 = 0.9$.

height value of particles (i.e. here ALH) higher than the altitude location of the effective cloud layer as analysed in Sect. 6.4.2 and in Fig. 6.8. Similarly, Lorente et al. (2017) indicated that the choice of an aerosol correction leads to an average uncertainty of 45% over highly polluted scenes and with large aerosol loading (τ(550 nm) $\geq$0.5). In particular, it was found that $N^{v}_{NO_2}$ from POMINO dataset over China (Lin et al., 2015) is 55% higher (A_{NO_2} smaller) than if no explicit aerosol correction was considered when aerosol layer is located above the tropospheric NO_2 bulk. The main identified reason was a lower screening applied when only the effective cloud parameters were used as the Lambertian reflector was defined at a lower altitude (effective cloud pressure higher of 350 hPa). This may of course be attenuated when aerosols are mixed with NO_2. as their multiple scattering effects increase then the average light path length and thus the NO_2 absorption.

The difference in $N^{v}_{NO_2}$ between the use of $NN_{MODIS,\omega_0=0.95}$ and OMCLDO2-New

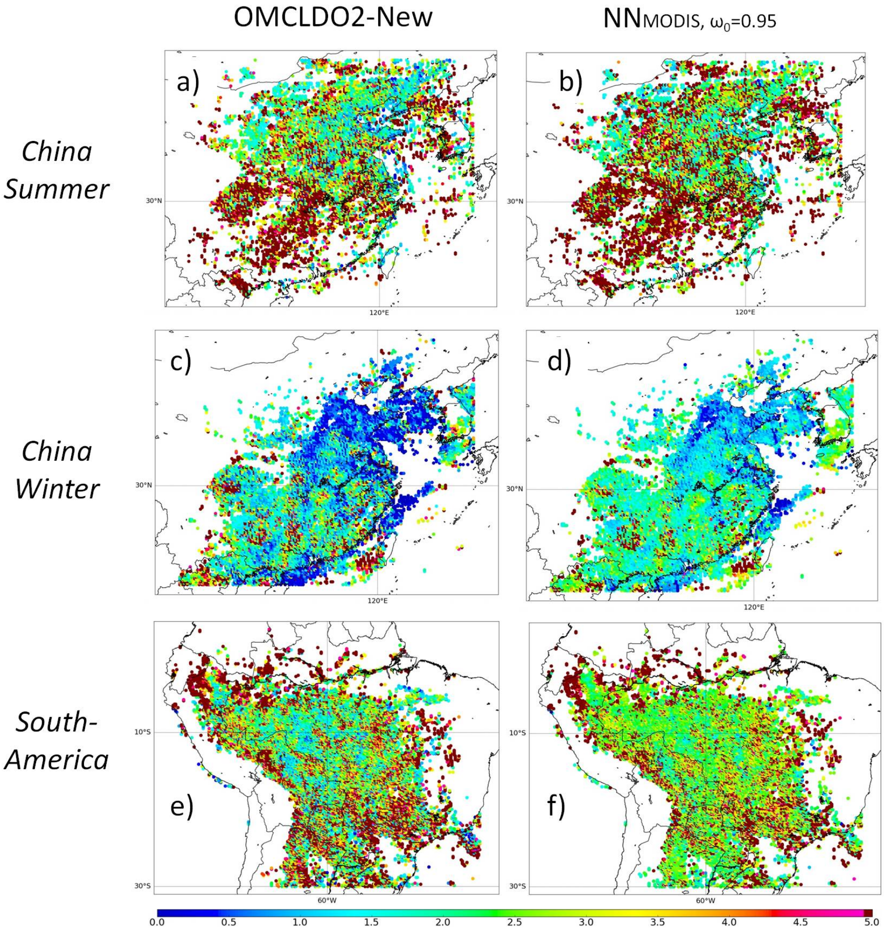


Figure 6.8: Maps of effective cp (converted into cloud height) from OMCLDO2-New and ALH from $NN_{MODIS,\omega_0=0.95}$ in [km] in 2006-2007: **(a)** Effective cp, China Summer, **(b)** ALH, China Summer, **(c)** Effective cp, China Winter, **(d)** ALH, China Winter, **(e)** Effective cp, South-America, **(f)** ALH, South-America.

is somehow reduced when assuming a lower aerosol ω_0: i.e. by using $NN_{MODIS,\omega_0=0.9}$. It has to be noted that in such a configuration, the main differences are: 1) a lower ALH (i.e. aerosol layer located 50-60 hPa closer to the surface - see Sect. 6.4.2) due to an assumed lower ω_0 value, combined with 2) a more absorbing aerosol model used to compute $A_{\mathrm{NO_2}}$(439 nm). In both cases, $NN_{MODIS,\omega_0=0.9}$ and $NN_{MODIS,\omega_0=0.95}$, prior τ (coming from MODIS) remains unchanged. As illustrated in Figs. 6.9b, d, and f, resulting $N^{\mathrm{v}}_{\mathrm{NO_2}}$ using $NN_{MODIS,\omega_0=0.9}$ over China are smaller (i.e. $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm) higher) by about -0.2% ± 7.8% in summer, and -8.2% ± 22.3% in winter compared to $N^{\mathrm{v}}_{\mathrm{NO_2}}$ with $NN_{MODIS,\omega_0=0.95}$. These numbers represent then a first evaluation of the impact of aerosol model uncertainty, assuming one may use a very accurate prior τ information for both ALH retrieval and then $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm) computation.

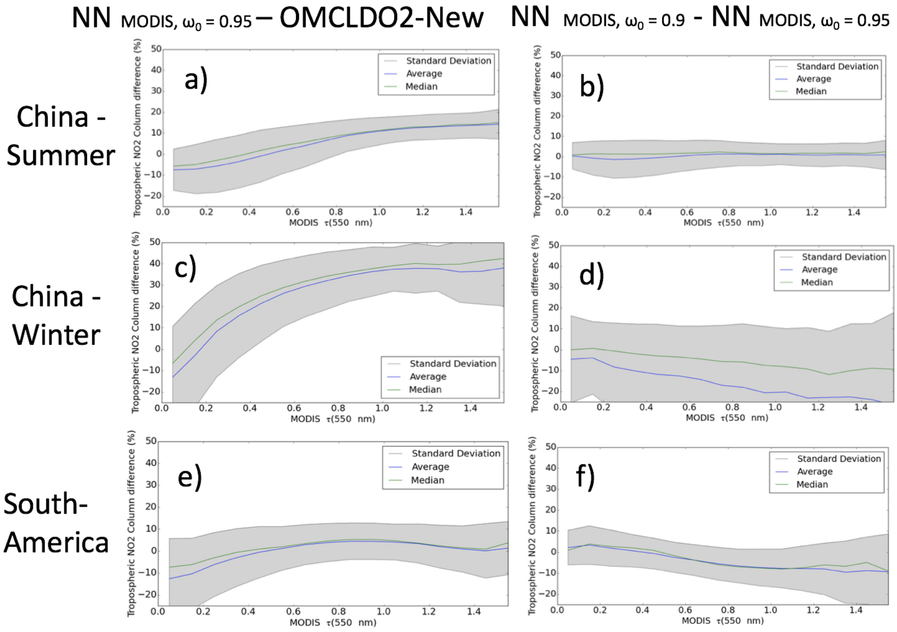


Figure 6.9: Statistics of relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ changes in (%) in 2006-2007, either due to differences between resulting $NN_{MODIS,\omega_0=0.95}$ and OMCLDO2-New aerosol corrections, or $NN_{MODIS,\omega_0=0.9}$ and $NN_{MODIS,\omega_0=0.95}$: **(a)**, and **(b)**: China summertime (June-July-August), **(c)**, and **(d)**: China wintertime (December-January-February), **(e)**, and **(f)**: South-America biomass burning season (August-September).

Interestingly, over scenes in South-America with MODIS $\tau(550\text{ nm}) \geq 0.5$, the difference between $N^{\mathrm{v}}_{\mathrm{NO_2}}$ based on $NN_{MODIS,\omega_0=0.95}$ compared to OMCLDO2-New is on average close to zero with a standard deviation of 16.8%. The use of $NN_{MODIS,\omega_0=0.9}$ reduces $N^{\mathrm{v}}_{\mathrm{NO_2}}$ by about -1.3% ± 8.7%. Interestingly, Castellanos et al. (2015) reported an average change of 0.6% ± 8% on $A^{\mathrm{v}}_{\mathrm{NO_2}}$ after reprocessing DOMINO $N^{\mathrm{v}}_{\mathrm{NO_2}}$ over cloud-free scenes during the biomass burning season in South-America and applying an explicit aerosol correction based on the OMI near-UV aerosol algorithm (OMAERUV) and CALIOP aerosol ALH.

Finally, over scenes with little aerosol amount (i.e. MODIS $\tau(550\text{ nm}) \leq 0.2$), the difference in $N^{\mathrm{v}}_{\mathrm{NO_2}}$ between the explicit and implicit aerosol corrections is systematically lower and non-null compared to scenes with MODIS $\tau(550\text{ nm}) \geq 0.5$: about an average of -10% over all the considered regions, regardless the use of $NN_{MODIS,\omega_0=0.9}$ or $NN_{MODIS,\omega_0=0.95}$. Such a non-zero average difference may seem strange as so little aerosol amounts are expected to have an almost negligible effect on the light path and thus on $A^{\mathrm{v}}_{\mathrm{NO_2}}$. Note that average differences in $N^{\mathrm{v}}_{\mathrm{NO_2}}$ between $NN_{MODIS,\omega_0=0.9}$ and $NN_{MODIS,\omega_0=0.95}$ are, on the contrary, close to zero over those cases. $NN_{MODIS,\omega_0=0.95}$ and $NN_{MODIS,\omega_0=0.9}$ algorithms differ from OMCLDO2 by using an external geophysical parameter (i.e. MODIS τ). Although more accurate than using the retrieved OMI τ, the combination of an external MODIS aerosol paremeter derived from different assumptions about the scattering model and surface reflectance may at the end lead to inconsistencies when combined with the OMI NN model: the 477 nm $O_2 - O_2$ radiance budget is likely not closed. This radiance budget is always closed with OMCLDO2 (apart from the instrument noise) since it simultaneously adjusts both cf and cp to match with the R_c(475 nm)-$N^{s}_{O_2-O_2}$ combination. The topic of radiance closure budget and its impacts on $A^{v}_{NO_2}$ are further discussed in Sect. 6.5.4.

6.5 Specific discussions

In this section, we want to discuss specific elements to evaluate the relevancy of each aerosol correction strategy over cloud-free scenes from OMI, but also in general from all UV-vis satellite measurements. In particular, we wish to draw the reader's attention on the advantage of using an explicit aerosol correction based on the exploitation of the 477 nm O_2-O_2 spectral band, but also the remaining difficulties to implement it in practice. The next subsections focus on the evaluation of $N^{v}_{NO_2}$ biases with implicit and explicit corrections based on synthetic cases, the importance of the NO_2 vertical profile, the significance of model error and the OMI visible radiance closure budget issue.

6.5.1 How to define the best aerosol radiation correction?

In order to define the best aerosol correction scheme, the first criterion should obviously be the lowest remaining bias on $N^{v}_{NO_2}$. Keeping in mind that the assumption of aerosol-free scene leads to the highest biases (Chimot et al., 2016), Fig. 6.10 shows that all reprocessed $N^{v}_{NO_2}$ with implicit or explicit corrections are larger by 10%—50% than if no correction was performed, especially over scenes with MODIS τ(550 nm) ≥ 1.0. This suggests that both strategies converge to the same direction (i.e. same sign) in spite of the different magnitudes of the aerosol correction. Since all the considered strategies attenuate the biases that would result from an aerosol-free scene assumption, it is worth emphasizing that all of them, without distinction, are before hand an aerosol correction, regardless of their implicit or (more) explicit nature.

Figure 6.11 estimates $N^{v}_{NO_2}$ accuracy on synthetic cases based on explicit aerosol correction assuming a NO_2 vertical profile representative of a typical day in summertime over China east (van Noije et al., 2014; Chimot et al., 2016). Provided that the aerosol model is in line with the actual aerosol type present in the observed scene (i.e. ideal scenario), remaining biases are below 20%, and little dependent on aerosol parameters (τ, ω_0 and ALH). In case of absorbing aerosols (i.e. $\omega_0 = 0.9$), biases are even smaller than those with the implicit aerosol correction from OMCLDO2-New. Generally, $N^{v}_{NO_2}$ seems to be higher (i.e. stronger shielding effect applied) with the explicit aerosol correction: i.e. differences up to 20% for τ(550 nm) ≥ 1.0, in line thus with the analyses over China in summer time in Sect. 6.4.3. However, in practice, these results may vary with respect to the NO_2 profile shape and additional errors in the employed aerosol model (see next subsections).

In addition, a second criterion that could be considered is the realism of the phys-

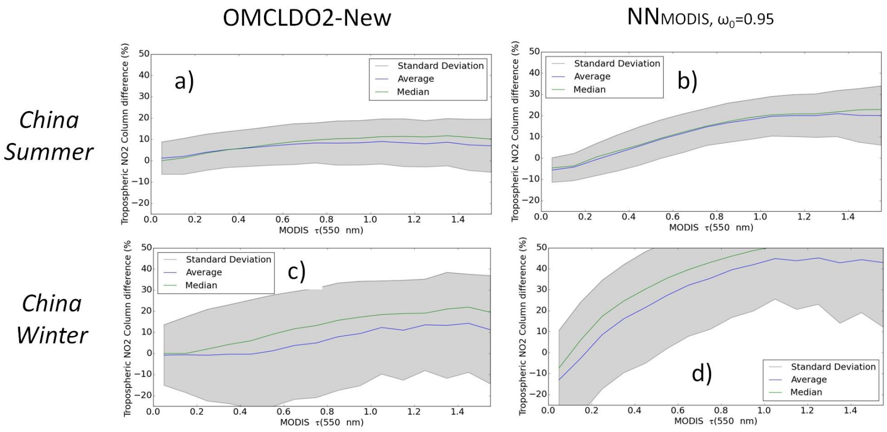


Figure 6.10: Statistic of relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ differences in (%) over China 2006-2007 in Summer (June-July-August) and Winter(December-January-February), after implicit or explicit aerosol correction compared to no aerosol correction (i.e. aerosol-free scene assumption): **(a)** Implicit aerosol correction based on OMCLDO2-New, **(b)** Explicit aerosol correction based on $NN_{MODIS,\omega_0=0.95}$.

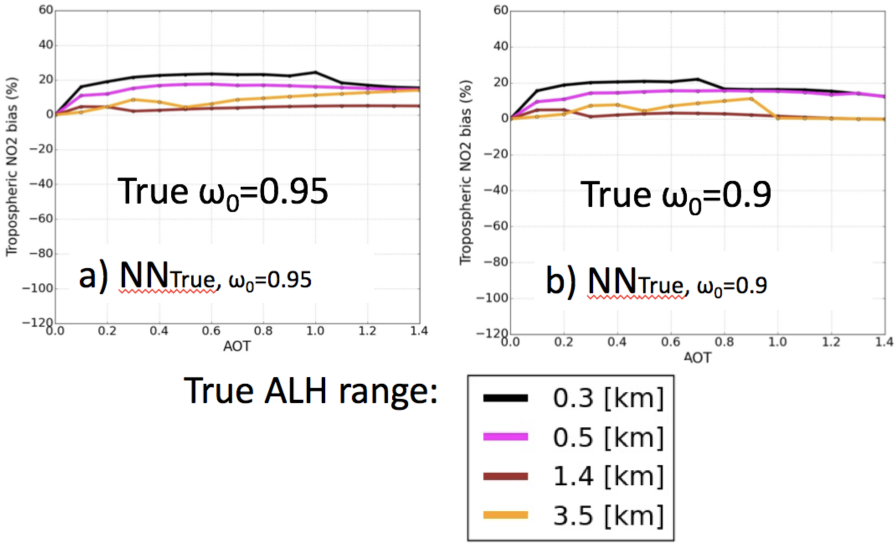


Figure 6.11: Effects of explicit aerosol correction on $N^{\mathrm{v}}_{\mathrm{NO_2}}$ based on synthetic cases. True aerosol ω_0=0.95 and ω_0=0.9 scenario are compared. Other atmospheric conditions are similar to Fig. 6.7. When applied, no biased regarding aerosol model and geophysical conditions are introduced in the ALH retrieval and in the explicit aerosol correction with the OMI aerosol NN (cf. Sect. 6.2.3): **(a)** Relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ bias, ω_0=0.95, explicit aerosol correction from $NN_{MODIS,\omega_0=0.95}$, **(b)** Relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ bias, ω_0=0.9, explicit aerosol correction from $NN_{MODIS,\omega_0=0.9}$.

ical assumptions when computing $A_{NO_2}(439\text{ nm})$. Figure 6.12 illustrates the vertical averaging kernel (AK) (cf. Sect. 6.3.2) assuming $\tau(500\text{ nm}) = 1.0$ and from the application of OMCLDO2-New, $NN_{MODIS,\omega_0=0.95}$ and true aerosol conditions, assuming no bias in prior aerosol assumptions. AKs are very important for estimating the surface NO_x emissions by convoluting the NO_2 vertical profiles from the used atmospheric models to match with the OMI $N^{V}_{NO_2}$ observations (Eskes and Boersma, 2003; Ding et al., 2015). AKs based on OMCLDO2-New display a sharp distinction between the enhanced atmospheric layers located above aerosols and the shielded layers located below. On the contrary, AKs from $NN_{MODIS,\omega_0=0.95}$ depict a smoother transition from then enhanced to the shielded layers. This transition is more in line with the actual AKs (Fig. 6.12a) and results from the multiple scattering effects induced by the more or less wide aerosol particle layer. A bright Lambertian reflector is by nature fully opaque and does not induce multiple scattering effects. This is compensated by the transmission of the clear fraction of the pixel through the IPA assumption. This suggests that applying an explicit aerosol correction leads to the consideration of more realistic physical assumptions and AK productions. However, such a conclusion critically depends on an ensemble of other parameters that contribute to the AK generation: the accuracy of the retrieved ALH which triggers the location of the enhancement / shielding transition, the potential aerosol model biases (e.g. ω_0, scattering phase function, etc...), and the difference between actual and assumed aerosol vertical profile.

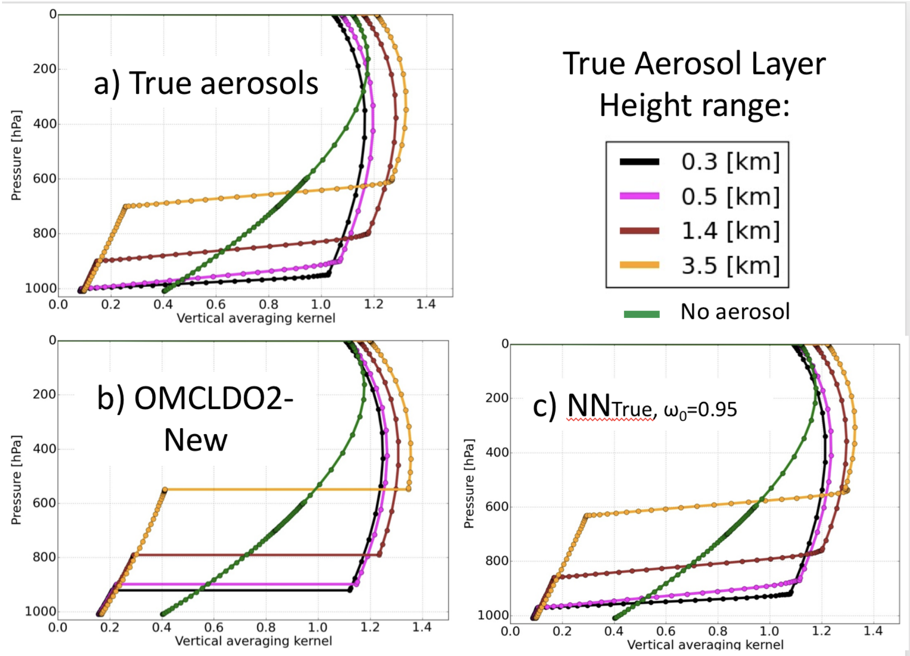


Figure 6.12: Vertical AK (cf. Sect. 6.3.2 and Sect. 6.5.1) based on aerosol $\tau(550\text{ nm}) = 1$, NO_2 vertical profile of 1st of July, 2006 at 12:00 pm over China. Other conditions are similar to Fig. 6.7.

6.5.2 The importance of the relative layer height

A comprehensive aerosol correction for an accurate $A_{\mathrm{NO_2}}$(439 nm) computation requires not only an accurate scheme to mitigate aerosol radiation effects in a (cf. Sect. 6.3.2) but also the actual NO_2 vertical profile. Figure 6.13 shows the accuracy of the aerosol corrections in $A_{\mathrm{NO_2}}$(439 nm), based on synthetic case, assuming the presence of absorbing aerosol particles (ω=0.9) but a NO_2 vertical profile of winter time (1st of January, 12:00 pm) over China. The main difference with Fig. 6.11 is the presence of a more abundant tropospheric NO_2 bulk closer to the surface, and a stronger decrease rate to higher altitudes (Chimot et al., 2016). In such a case, relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ biases with the implicit aerosol correction are strongly degraded from −10-20% (summertime) to −80-40% (winter). Already identified in Sect. 6.5.1 with the summer time NO_2 profile, the insufficient shielding effect applied by the effective cloud parameters from OMCLDO2-New in case of aerosol layers located in elevated altitude is here severely degraded (from −10% to −80%). The insufficient enhancement effect when particles are mixed with the tropospheric NO_2 molecules is also here amplified (from 20% to 40%). For reminder, absorbing particles lead to a reduced effective cp (increasing the shielding effect) but a lower effective cf (lower shielding effect) with then a higher transmittance coming from the clear part of the pixel. This last element likely overall limits the potential of the effective cloud model to apply an adequate shielding effect on $A^{\mathrm{v}}_{\mathrm{NO_2}}$. However, these biases still remain lower than if no aerosol correction was achieved (Chimot et al., 2016).

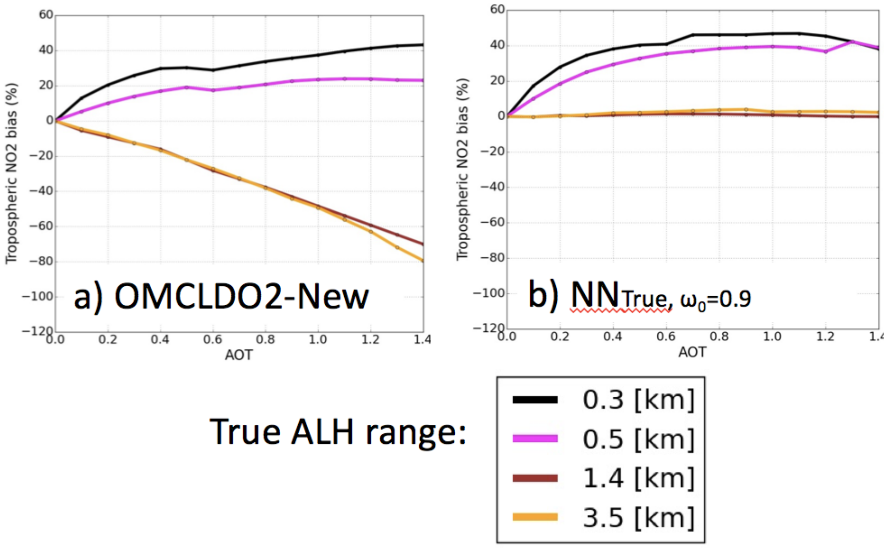


Figure 6.13: Effects of true NO_2 vertical profile shape on $N^{\mathrm{v}}_{\mathrm{NO_2}}$ after implicit or explicit aerosol correction (see Sect. 6.5.2). Atmospheric conditions in simulations as similar to Fig. 6.7 except aerosol ω_0=0.9, and profile TM5 NO_2 vertical profile of 1st of January, 2005 at 12:00 pm over China are considered: **(a)** Relative NO_2 bias, implicit aerosol correction through OMCLDO2-New, **(b)** Relative NO_2 bias, explicit aerosol correction through $NN_{True,\omega_0=0.9}$.

When considering an explicit aerosol correction using $NN_{MODIS,\omega_0=0.9}$, $N^{\mathrm{v}}_{\mathrm{NO_2}}$ are changed to 0-40 (%). Similarly to summertime, they are lower in case of particles at high altitude, suggesting then the strong benefits of such a correction scheme in wintertime and/or in presence of absorbing particles. The cases of aerosols close to the surface (i.e. lower than 0.5 km) remains an issue due to the difficulty to distinguish the scattering effects from the surface and the adjacent aerosol layer when retrieving ALH. The retrieval seems, in such a case, to overestimate the aerosol layer altitude.

The observed positive difference of about 40% between explicit and implicit aerosol corrections is in line with the analyses over China in wintertime (see Sect. 6.4.3 and Fig. 6.15). All these elements strongly remind and emphasize that the quality of the aerosol corrections, and their differences, for $N^{\mathrm{v}}_{\mathrm{NO_2}}$ retrieval is actually more dependent on the relative height between the particles and the tropospheric NO_2 bulk, than ALH or cp themselves. The evolutions of operational aerosol correction schemes in present and future air quality UV-visible space-borne sensors, such as TROPOMI on-board Sentinel-5 Precursor, need then to consider simultaneously a proper characterization of trace gas vertical profiles together with aerosol vertical distribution and related optical properties.

6.5.3 Model error

At a first view, the main error dominating in the different employed aerosol radiation correction schemes for the OMI $N^{\mathrm{v}}_{\mathrm{NO_2}}$ retrieval in this study is the so-called model error. Indeed, at the end, both the implicit and explicit aerosol corrections just represent a different attempt to reproduce aerosol scattering effects on the average light path in the visible light spectrum. The term "model error" actually represents two models with a different purpose:

- the forward model used to derive geophysical parameters from the 477 nm O_2-O_2 spectral band.
- and the model to represent aerosol impacts on $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm).

In theory, both models could be different. Indeed, if accurate and independent information about aerosol parameters would be available, there would be then no need to exploit the O_2-O_2 477 nm band. For example, studies like Lin et al. (2014, 2015); Castellanos et al. (2015) reprocessed OMI NO_2 data based on aerosol simulations given by atmospheric transport simulations, or ancillary observations (e.g. CALIOP aerosol vertical profiles, and OMAERO). The combination of all these data define the aerosol model for computing $A_{\mathrm{NO_2}}$(439 nm). However, in the frame of operational and near-real-time processing, the two models are actually merged together: i.e. the model used within the O_2-O_2 spectral band is (normally) equivalent to compute $A_{\mathrm{NO_2}}$(439 nm). One of the main reasons is the need to have an indication on particle height representative of every single OMI pixel. Such an information is generally not easily and directly available from an external source.

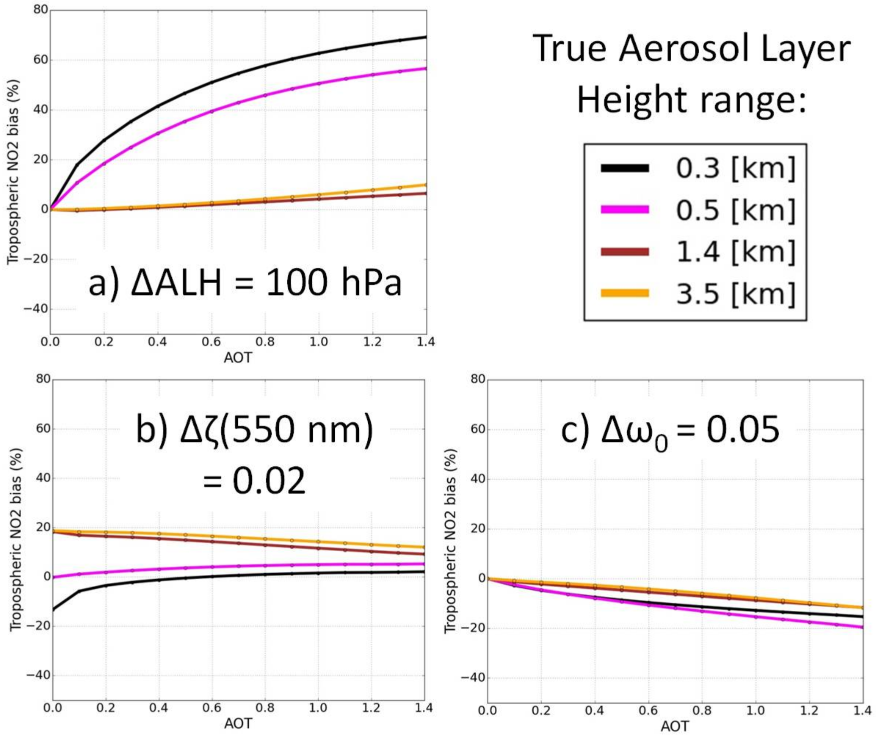


Figure 6.14: Relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ biases due to a direct impact of individual aerosol parameter biases on $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm0 calculation. Synthetic cases are similar to Fig. 6.13: **(a)** ALH bias = 100 hPa, **(a)** τ(550 nm) bias = 0.02, **(c)** ω_0 bias = 0.05.

When applying the explicit aerosol correction strategy, the accuracy of the whole set of parameters defining the aerosol model and their combination, not only ALH alone, is of high importance. Figure 6.14 shows the impacts of individual key variables usually defined as input of a given aerosol model.

A single bias on ALH of 100 hPa directly affects $N^{\mathrm{v}}_{\mathrm{NO_2}}$ within the range of 60-70% when absorbing aerosols (ω_0=0.9) are located below 0.5 km, assuming wintertime NO_2 profile and with τ(550 nm)=1.4. The uncertainties are below 50% for τ(550 nm) $\leq$ 0.5, and overall below 10% when particles are located at elevated altitudes (i.e. true ALH $\geq$ 1.4 km). Such an impact illustrates the sensitivity of the ALH knowledge when particles are mixed with NO_2 molecules due to the complexity to reproduce the enhancement of the average light path caused by scattering effects. A bias of 0.02 on the assumed τ(550 nm) mostly impacts scenes with little aerosol load: while resulting $N^{\mathrm{v}}_{\mathrm{NO_2}}$ uncertainties lie in the range of −20-20% for τ(550 nm) $\leq$ 0.5, they decrease to the range of 0-10% for τ(550 nm)=1.4. Finally, an overestimation of aerosol scattering efficiency (i.e. ω_0 bias of 0.05) leads to an underestimation of $N^{\mathrm{v}}_{\mathrm{NO_2}}$ up to −20% over scenes with high τ as a consequence of an underestimation of aerosol shielding effect and therefore a too large $A^{\mathrm{v}}_{\mathrm{NO_2}}$(439 nm). Overall, ALH is the most crucial parameter to ensure the accuracy of the AMF computation and requires an accuracy better than 50 hPa to limit $N^{\mathrm{v}}_{\mathrm{NO_2}}$ bias below 40%. With τ, it forms the most important set of aerosol parameters prerequisite for $A_{\mathrm{NO_2}}$(439 nm) computation. Although not negligible, model parameters that are more related to the aerosol optical and scattering properties such as ω_0, but also g or α are here of second importance provided that ALH and τ qualities are ensured.

To confirm this, Fig. 6.15a depicts the impacts of a wrong prior aerosol ω_0 (overestimation of 0.05) when applying the explicit aerosol correction through the use of the OMI aerosol NN but with the true prior τ value for both ALH retrieval and $A^{\mathrm{v}}_{\mathrm{NO_2}}$ computation. Compared to Fig. 6.13, the increases of $N^{\mathrm{v}}_{\mathrm{NO_2}}$ biases are limited: from 0-40% (cf. Fig. 6.13b) to 0-60%. This reflects how the quality of prior τ knowledge helps to constrain both ALH retrieval, and then limit the perturbation on $A^{\mathrm{v}}_{\mathrm{NO_2}}$. However, this may be more severely degraded when inaccurate prior τ is considered (see Sect. 6.5.4).

Among other elements, the Lambertian cloud and the aerosol Henyey-Greenstein models differ about the horizontal coverage of the OMI pixel: the effective cloud model only covers part of the pixel (the fraction coverage is fitted through cf, optical properties are fixed), while the aerosol model (and analysed synthetic cases) cover the whole pixels (optical properties can be changed, fraction coverage is fixed). Part of the insufficient shielding effect observed with the implicit correction in Sect. 6.5.1-6.5.2 may be explained by an insufficient coverage of the observation scene. However, true horizontal distributions of aerosols within the observation pixel may actually be quite heterogeneous. Such problems should be further investigated with a focus on areas where absorbing particles are expected such as in winter time in China, or during large biomass burning episodes. Future studies should also determine how often such effect occurs and its overall impact depending on the NO_2 profile shape variability.

6.5.4 Radiance closure budget issue and potential impacts

The discussions about the "model error" in the previous sub-section implicitly made the assumption that the whole set of parameters such as (ecf-ecp-surface reflectance) on the one hand, or (τ-ALH-ω_o-g-α-surface reflectance) are fully consistent in the sense that they form one unique particle model. However, when external data are used to constrain ALH retrieval accuracy, such as MODIS aerosol τ, one may combine inconsistent model assumptions together leading then to complex artefacts such as the issue of OMI closure radiance budget.

Radiance closure budget is not only important in the 477 nm O_2-O_2 band, but also in the NO_2 absorption band at the wavelength where the AMF is computed. Aerosol τ combined with surface reflectance are expected to drive OMI R_c(475 nm). As discussed in Sect. 6.2.2, OMCLDO2 simultaneously adjusts both cf and cp based on the same prior surface albedo such that their combination allows to close the (R_c(475 nm)-$N^{\mathrm{s}}_{\mathrm{O_2}-O_2}$) budget, and thus the OMI 477 nm $O_2 - O_2$ radiance regardless the accuracy of the selected model. On the contrary, by using MODIS τ, only OMI $N^{\mathrm{s}}_{\mathrm{O_2}-O_2}$ is exploited, not R_c(475 nm). If all the prior parameters are accurate and derived from a unique set of aerosol model and surface reflectance parameters, the (R_c(475 nm)-$N^{s}_{O_2-O_2}$) budget should be closed. However, any mismatch between the model employed for MODIS τ determination and the one used in the OMI NN training dataset, between MODIS and OMI instrument radiance, and/ or the surface reflectance hypothesis may leave this budget open. In particularly, it is worth reminding that surface reflectance dataset behind MODIS τ, OMI ecf and $A^{v}_{NO_2}$(439 nm) are not identical: OMI surface albedo assumes Lambertian properties (Kleipool et al., 2008) while MODIS aerosol retrieval uses surface spectral reflectance (Levy et al., 2013).

Therefore, the strategies of implicit vs. explicit aerosol correction analysed in Sect. 6.4.3 do not only differ in terms of assumed particle optical and scattering property model but also on how much the whole OMI radiance budget is eventually fitted. This last difference likely explains the strange systematic difference (about -10%) identified over OMI scenes with MODIS τ(550 nm) ≤ 0.2, where aerosol effects could be assumed as almost insignificant. To confirm this hypothesis, Fig. 6.15 shows the reprocessed OMI $N^{v}_{NO_2}$ over the three regions based on explicit aerosol correction. MODIS τ is now replaced by the retrieved OMI τ (see. Tab. 6.1) which is, like ecf, mostly constrained by R_c(475 nm), prior OMI surface albedo and the same aerosol model used for ALH retrieval and $A^{v}_{NO_2}$(439 nm) calculation. This time, resulting $N^{v}_{NO_2}$ are higher than with the use of MODIS τ (see Tab. 6.2). Moreover, the differences between the implicit and explicit aerosol corrections are on average zero (instead of −10%) everywhere over scenes with small aerosol load confirming the consistency in the employed models for each set of parameters and an almost complete closure of the OMI visible radiance budget (instrument noise apart).

However, using the retrieved OMI τ is not without consequences as its uncertainty is higher than MODIS τ (see Sect. 6.2.3). Firstly, as discussed in Chimot et al. (2017), it degrades the quality of the retrieved OMI ALH. Secondly, the sensitivity of $N^{v}_{NO_2}$ based on the explicit aerosol correction to the selected aerosol ω_0 can be higher: e.g. over China in summer time, 8.5% ± 13.7% on average compared to −0.2% ± 7.8% in Sect. 6.4.3. The main reasons are the separate and also combined impacts on OMI τ, OMI ALH and $A^{v}_{NO_2}$(439 nm) computation. Figure 6.15b shows the impacts of replacing true τ by the retrieved OMI one. $N^{v}_{NO_2}$ degradations are more important, about −40-60%. This is a direct result of a degradation of ALH retrieval due to a too much biased aerosol τ and the resulting impact on the $A^{v}_{NO_2}$. As discussed in Sect. 6.5.3, although ω_0 parameter is of second importance for $A^{v}_{NO_2}$(439 nm) itself compared to the (ALH-τ) combination, its direct impact on τ retrieval consequently affects then ALH determination accuracy and then indirectly $A^{v}_{NO_2}$.

In spite of these drawbacks, $N^{v}_{NO_2}$ biases remain smaller than if the derived effective cloud parameters are employed through the implicit aerosol correction (cf. Fig. 6.13a). Thus, even if not perfect, the explicit aerosol correction based on OMI ALH and τ retrievals seem to remain still advantageous for an efficient aerosol radiation correction in tropospheric NO_2 vcd retrieval from visible satellite radiance.

At the end, one might wonder what the best option is for an optimal aerosol correction: 1) using the best aerosol and surface parameters available for the most accu-

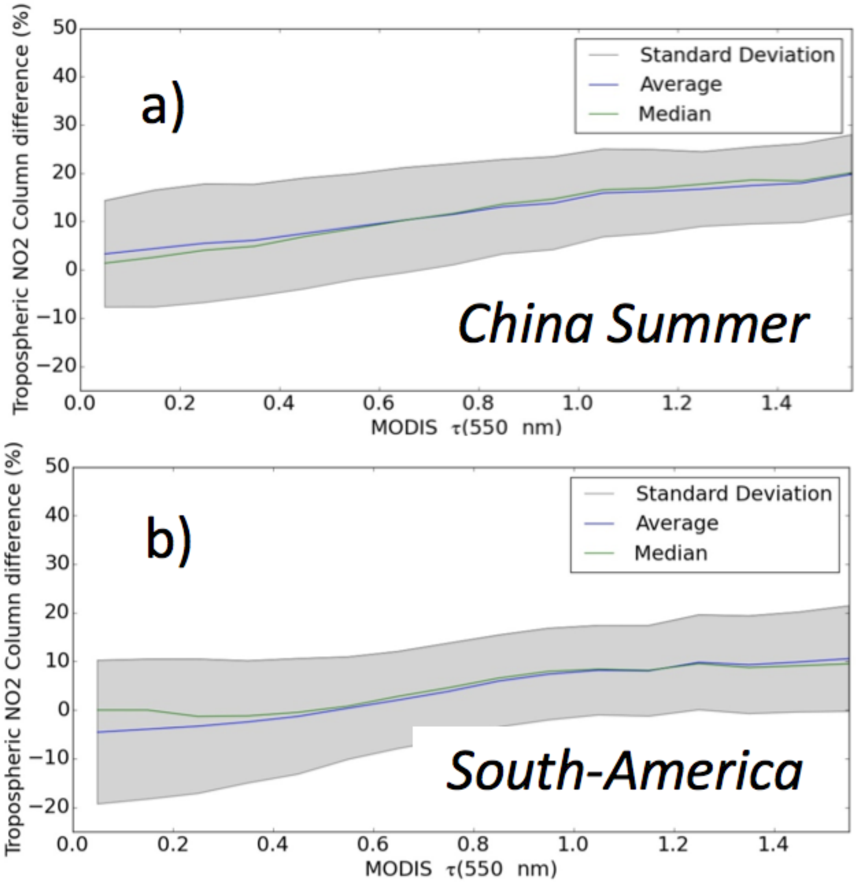


Figure 6.15: Same as Fig. 6.9a, c, e except MODIS τ(550 nm) is replaced by the retrieved OMI τ(550 nm) for both OMI ALH retrieval and the application of the explicit aerosol correction. Assumed ω_0=0.95: **(a)** China summertime (June-July-August), **(a)** South-America biomass burning season (August-September).

rate correction at the cost of not closing the satellite radiance budget, or 2) applying a less accurate correction but with an ensemble of aerosol and surface parameters that eventually comprehensively fit the spectral measurement. The first option gives more weight to the used auxiliary data, while the second option maximise the weights of the amount of information contained in the satellite measurement. The answer to such a problem is, in our opinion, not clear at this moment. But, given the fact that several studies prioritise the application of multiple parameters from very diverse sources (models, ancillary instruments with different techniques, etc..) to satellite spectral measurements, we think the issue of radiance closure budget should be kept in mind by the scientific community and further investigated in future research studies. At the end, an optimal trade off must be found between quality of $N_{NO_2}^{v}$ product and the weights given to the original satellite measurement.

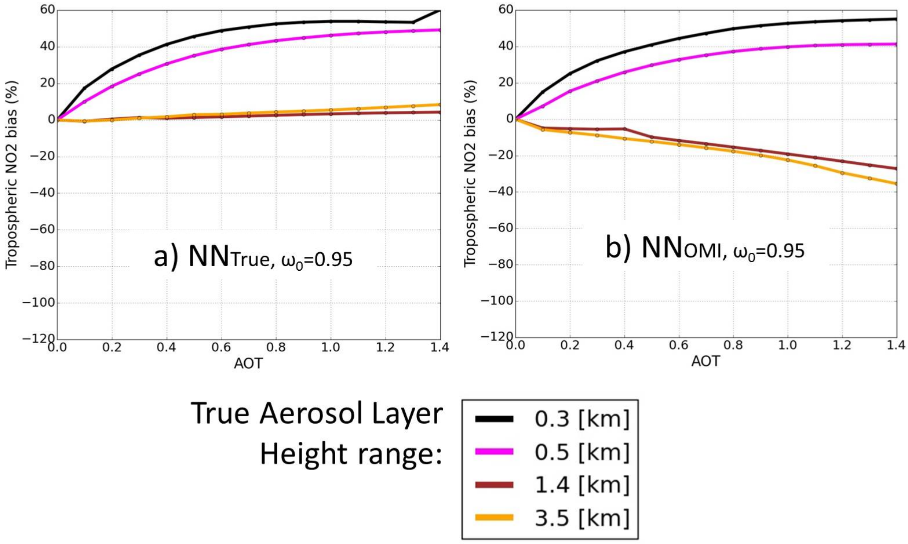


Figure 6.16: Impact of ω_0 bias on the use of OMI NN algorithms: from ALH retrieval to the explicit aerosol correction on $A_{\mathrm{NO_2}}$(439 nm). Simulations are based on true aerosol $\omega_0 = 0.9$. Other atmospheric conditions are similar to Fig. 6.13. Assumed aerosol $\omega_0 = 0.95$ for both ALH retrieval and then $A_{\mathrm{NO_2}}$(439 nm) computation. Relative $N^{\mathrm{v}}_{\mathrm{NO_2}}$ biases due to explicit aerosol correction derived from the use of **(a)** $NN_{True,\omega_0=0.95}$, or **(b)** $NN_{OMI,\omega_0=0.9}$.

6.6 Conclusions

This paper analysed the reprocessing of the reference OMI $N^{\mathrm{v}}_{\mathrm{NO_2}}$ over 2-year (2006-2007) cloud-free scenes in regions dominated by high aerosol particle pollution such as east China and South-America. The goal was to evaluate the benefits of the recent achieved developments during the last years in order to improve the so-called aerosol correction in $A^{\mathrm{v}}_{\mathrm{NO_2}}$, a crucial parameter when retrieving tropospheric NO_2 VCD from visible satellite backscattered measurements. In particular, the tested aerosol correction strategies rely on our last learning experiences with the 477 nm O_2-O_2: a key spectral band present in current and future air quality UV-vis satellite instruments such as OMI, GOME-2, TROPOMI, Sentinel-4-UVN, and Sentinel-5-UVNS. The use of this band is important in view of operational data processing as it allows to derive parameters that, in principle, could reproduce the particle scattering effects on the average light path in the visible spectral window. Regardless of the type of algorithm employed here, these parameters can all be derived very quickly, are directly representative of the OMI pixel in terms of spectral measurement, spatial coverage and temporal acquisition, and can directly be used for the $A^{\mathrm{v}}_{\mathrm{NO_2}}$ computation.

The two tested algorithms were both developed for OMI and represent our improved use of the 477 nm O_2-O_2 spectral band nowadays: the OMI cloud software OMCLDO2 (Veefkind et al., 2016), and the OMI aerosol NN approach (Chimot et al., 2017, 2018b). At high level, they mostly differ in terms of assumed aerosol model not only for the retrieval within the O_2-O_2 spectral window, but also consequently on the way to reproduce and remove aerosol effects in $A^{\mathrm{v}}_{\mathrm{NO_2}}$. OMCLDO2 relies on a Lambertian and opaque reflector model with a high albedo of 0.8 (Stammes et al., 2008). It allows to retrieve the effective cloud parameters cf and cp. The last update from Veefkind et al. (2016) includes, among many elements, a higher cloud LUT resolution and the necessary temperature correction on $N^{\mathrm{s}}_{\mathrm{NO_2}}$. The term "effective" refers to the fact that such a model is generally applied without distinction to both cloudy and cloud-free scenes, aerosol and aerosol-free scenes. Therefore, cf and cp values are adjusted depending on the particle optical properties. The use of OMCLDO2 in $A^{\mathrm{v}}_{\mathrm{NO_2}}$ is generally named "implicit aerosol correction". OMI NN aerosol represents a step ahead by trying to retrieve more explicit aerosol parameters only over cloud-free scenes. The primary goal was to retrieve ALH with an accuracy better than 800 m for $A^{\mathrm{v}}_{\mathrm{NO_2}}$ computation. It requires an accurate prior aerosol τ information, usually taken from the MODIS spectroradiometer flying in the constellation with OMI. However, the OMI NN aerosol algorithm can also retrieve τ from the OMI O_2-O_2 continuum reflectance, although its accuracy is lower than MODIS τ and leads to higher ALH and $A^{\mathrm{v}}_{\mathrm{NO_2}}$ uncertainties. The consideration of OMI NN aerosol in $A^{\mathrm{v}}_{\mathrm{NO_2}}$ computation leads for the first time to an "explicit aerosol correction" by directly using the OMI 477 nm O_2-O_2 spectral band. This differs from several studies where such an explicit correction usually relied on atmospheric model outputs or CALIOP ALH (Lin et al., 2014, 2015; Castellanos et al., 2015).

Overall, both employed strategies show a reduction of OMI $N^{\mathrm{v}}_{\mathrm{NO_2}}$ biases over cloud-free scenes dominated by aerosol pollutions. Previous studies showed an underestimation of $N^{\mathrm{v}}_{\mathrm{NO_2}}$ between -20% and -40% over scenes with MODIS $\tau(550\text{ nm}) \geq 0.6$, assuming scattering aerosol particles with $\omega_0 = 0.95$ and summertime NO_2 vertical profile. In similar conditions, they are expected to be contained in the limit of 0-20%, regardless of the type of aerosol correction applied.

In general, the applied explicit correction based on all the aerosol NN algorithms led to higher $N^{\mathrm{v}}_{\mathrm{NO_2}}$ values compared to the implicit correction from the last OMCLDO2 version: between 20% and 40% depending on the seasons, regions and aerosol pollution episodes. At a first view, these differences are attributed to the differences between the retrieved ALH and the effective cp: a higher ALH would lead to a stronger shielding effect (i.e. smaller $A^{\mathrm{v}}_{\mathrm{NO_2}}$) and therefore a higher $N^{\mathrm{v}}_{\mathrm{NO_2}}$. However, the model differences also play a non negligible role. In particular, all the employed aerosol models assume that particles fully cover the OMI observation pixel. The Lambertian reflector only covers part of the pixel (effective cf is adjusted per scene), the non-covered part leading then to a higher transmittance. Assuming no error in the aerosol model parameters, the simulated physical scattering effects are expected to better represent how aerosols affect the average light path in the visible spectral window, and should lead to more realistic vertical AK.

However, diverse difficulties were identified and recommendations are given for future research studies. In particular:

- In presence of absorbing aerosols (i.e. $\omega_0 \leq 0.9$), the implicit aerosol correction still leads to substantial biases on $N^{\mathrm{v}}_{\mathrm{NO_2}}$: between -80% and 20% assuming a typical NO_2 wintertime. This is mostly severed when particles are located at high altitude (above 1.4 km) and with heavy aerosol load. This is likely due to an insufficient shielding effect by the Lambertian reflector which depicts a reduced coverage of the observation pixel, and thus probably a too high transmittance of the cloud (or aerosol)-free part. In similar conditions, the explicit aerosol correction allows to mitigate the biases in the range of 0-40% if accurate τ and aerosol model are available, or -40-60% with inaccurate τ (e.g. retrieved OMI τ) and aerosol model.

- Biases remain high when particles are located close to the surface, regardless of the aerosol correction methodology. Indeed, the distinction between aerosol scattering and surface relfectance is well known to be difficult. Since particles are close to the tropospheric NO_2 bulk (either mixed or slightly above), any error in ALH or cp dramatically affect how aerosols modify the average light path and the NO_2 absorption, and consequently the $A^{\mathrm{v}}_{\mathrm{NO_2}}$ accuracy

- All aerosol correction methodologies are sensitive to the NO_2 vertical profile shape and the employed surface albedo or reflectance. Future improvements need to address these both parameters together with any evolution that can still be done in the aerosol or effective cloud retrieval from the 477 nm O_2-O_2. Indeed, continuing working only on the improved use of this band with the future space-borne sensors is likely nowadays insufficient. We would like to emphasize the need to invetsigate the observed changes in the new OMI Lambertian albedo climatology and the accuracy of future atmospheric NO_2 models.

- Although the OMI aerosol NN algorithms lead to promising ALH retrieval, its use for an explicit aerosol correction is not straightforward. Its combination with other aerosol parameters (size, ω_0, τ) can lead either to model errors and/or to the risk of not closing the OMI radiance budget if any of these parameters was issued from an external source (model or instrument) not consistent with the OMI ALH retrieval.

Given its estimated performances, we recommend for an operational processing, where no distinction is made between cloudy, cloud-free, and aerosol contaminated scenes, to firstly use the implicit aerosol correction based on OMCLDO2. Indeed, it allows to correct adequately both cloud and aerosol particle effects on the average light path. If, apart of the operational context, OMI cloud-free scenes can be carefully identified and collected in a same way it has been done here, then the explicit aerosol correction based on the OMI NN aerosol algorithms should be considered. We demonstrated in this study, for the first time, its high performance and the realism of the simulated physical effects. However, it is then necessary to pay attention to the ensemble of parameter set defining the complete aerosol model. Moreover, the developed NN approach can ensure a fast cloud-free OMI data processing.

Overall, the considered aerosol corrections can easily be transposed to the future generation of air quality UV-vis satellite sensors, such as TROPOMI, Sentinel-4, and

Sentinel-5-UVNS. Moreover, they can be considered for other trace gases of interest: e.g. SO_2, HCHO. However, they will have to be adapted to the specificities of this new generation of instruments: e.g. the improvement in the spatial resolution. The small pixel sizes will clearly bring additional challenges such as the 3D effects of clouds, and thus the impacts on the adjacent observation pixels.

Chapter 7

Conclusions and recommendations

> "The pessimist complains about the wind; the optimist expects it to change; the realist adjusts the sails."
>
> William Arthur Ward - American writer (1921-1994)

In this chapter, we discuss to which extent the research objectives as presented in Chapter 1 have been reached. Furthermore, recommendations on how this work can be improved and extended are suggested.

7.1 Conclusions

The main motivation for this thesis research is to improve the quality of operational tropospheric NO_2 vertical column density (VCD) retrieval from the Dutch-Finnish satellite Ozone Monitoring Instrument (OMI), over cloud-free scenes and under any pollution conditions. More specifically, we focus on the importance of the aerosol correction. The presence of aerosols affect the average light path followed by the photons that are detected by the instrument. Backscattered solar light measurements acquired by satellite instruments such as OMI are essential for air quality and climate applications. Therefore, the aerosol correction is essential to obtain accurate tropospheric NO_2 retrievals over scenes that have a moderate to strong aerosol loading (and in the absence of clouds). The main objective is to design and evaluate an aerosol layer height (ALH) retrieval approach that can support this correction. This thesis relies on a simultaneous analyses of spectra that contain both the NO_2 (405-465 nm) as well as the O_2-O_2 (460-490 nm) absorption lines. Different aerosol corrections have been applied and evaluated, either based on explicit aerosol variables including ALH or effective cloud parameters.

The research questions set out in Sect. 1.7.1 guide this thesis. Below, the main findings related to each of these questions are described.

1. **Is there a quantitative need to correct tropospheric NO_2 retrievals for the effects of aerosols?**

   Chapter 2 clearly demonstrates that aerosol-free assumptions (i.e. no aerosol correction applied) would lead to major biases in the tropospheric NO_2 VCDs: between -60 and 20% for large aerosol optical thickness (AOT) τ values in summertime and cloud-free scenes. These values could even degrade to -150%

in cases of large angles (e.g. solar zenith angle ≥ 50°) or very large vertical separation between aerosols and the tropospheric NO_2 bulk. Such biases are beyond the tropospheric NO_2 accuracy requirements for space-borne sensors, and thus highlight the necessity to correct for aerosol effects.

2. **How do the OMI effective cloud parameters precisely respond to the aerosol properties?**

   The OMI cloud algorithm (OMCLDO2) cannot distinguish aerosol and cloud signals. In the absence of clouds, the effective cloud parameters respond to the presence of aerosols and thus contain information on the aerosol properties (τ, optical properties, particles size, altitude).

   In Chap. 2, we analyse in detail how OMCLDO2 responds to the presence of aerosols within a cloud-free OMI observation. The effective cloud fraction cf linearly increases with increasing τ and can reach values between 0.1 and 0.15 for $\tau(550\text{nm}) = 1$. This represents the enhanced scattering effects on the 460-490 nm continuum reflectance compared to an aerosol-free scene. The slope of the linear regression of τ vs. effective cf depends on the aerosol properties, the surface albedo and Sun-satellite geometry. The response of the effective cloud pressure cp to aerosol scenes represents the O_2-O_2 shielding effect, related to the reduction of photons detected by the satellite instrument within the O_2-O_2 absorption feature. This reduction is caused by optically thicker aerosol layers that shorten the length of the average light path. In case of high aerosol pollution, the retrieved effective cp values correlate with the mean aerosol layer height (ALH), although they are not strictly equivalent. they are weighted by aerosol optical properties (e.g. single scattering albedo ω_0, size, scattering phase function) and the surface reflectance. In cases of low τ or effective cf, aerosols have little effect on O_2-O_2 absorption. In that case, the effective cp cannot be determined accurately.

   We further clearly identify a numerical problem in the OMCLDO2 version 1, caused by too coarse sampling of the look-up-tables (LUTs). For small effective cloud fractions, this caused an increasing effective cp with decreasing effective cf (or aerosol τ), without any physical explanation (Chimot et al., 2016; Veefkind et al., 2016). This has been solved by significantly increasing the sampling of the LUTs, especially in the continuum reflectance dimension, in the OMCLDO2 version 2 (Veefkind et al., 2016).

3. **What is the error budget of the implemented aerosol correction (cf. the so-called implicit aerosol correction) in the OMI tropospheric NO_2, DOMINO-v2, product?**

   For cloud-free scenes, an implicit aerosol correction is applied in the tropospheric NO_2 VCD through the use of the retrieved effective cloud parameters.

   Chapter 2 evaluates the accuracy of the implicit aerosol correction by using OMCLDO2-Old (version 1), as currently achieved in the DOMINO-v2 product: tropospheric NO_2 VCDs are underestimated, in the range of 40% to 20% in case of aerosol particles at elevated altitude (i.e. $\geq$ 1.5 km) and high pollution (τ(550nm) $\geq$ 0.6) in summertime for east China. In case of aerosols located close to the surface or mixed with the tropospheric NO_2 bulk, VCDs are overestimated by 10% to 20%. These values are much smaller than when no aerosol correction would have been applied in the OMI DOMINO products. τ and aerosol altitude are the key drivers of these biases, while aerosol microphysical properties are of secondary importance. The main identified cause of these biases is an underestimation of the aerosol shielding effect by the cloud algorithm, due to the coarse sampling in the employed cloud LUT with OMCLDO2-Old.

Chapter 6 evaluates the new implicit aerosol correction through the use of the updated OMCLDO2-new (version 2) algorithm. With this updated algorithm, OMI tropospheric NO_2 VCD biases over cloud-free scenes are expected to decrease to 0-20% over scenes with τ(550 nm) $\geq$ 0.6, assuming scattering aerosol particles with $\omega_0 = 0.95$ and summertime NO_2 vertical profile. This is mostly thanks to the higher sampling of the OMI cloud LUTs.

However, in presence of absorbing aerosols (i.e. $\omega_0 \leq 0.9$), the implicit aerosol correction still leads to substantial biases: between 80% and 20% assuming a typical NO_2 wintertime profile. This is mostly the case when particles are located at high altitude (above 1.4 km) and with heavy aerosol load. This is likely due to an insufficient shielding effect by the Lambertian reflector model, which leads to a reduced coverage of the observation pixel, and thus to a too high transmittance of the cloud (or aerosol)-free part.

4. **How to design an innovative concept combining global aerosol layer height retrieval from OMI visible spectra and a machine learning approach?**

   Chapter 3 investigates the possibility to develop and apply a machine learning approach, based on the neural network (NN) technique, to retrieve ALH from OMI observations. This technique is developed for cloud-free conditions, over land and water surfaces. ALH is defined as the mid-level of a homogeneous scattering layer with a constant geometric thickness (i.e. 1 km). The key concept is the link between the measured O_2-O_2 slant column density (SCD) at 477 nm and the aerosol altitude. An aerosol layer located at high altitudes applies a large shielding effect on the O_2-O_2 located in the atmospheric layers below. The designed NNs belong to the family of machine learning and artificial intelligence techniques that rely on a multi-layer architecture, also named multilayer perceptron (MLP). The rigorous training task follows the error-back propagation technique and a high-quality training data set. The latter was ensured by full-physical spectral simulations, using the Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval (DISAMAR) software of KNMI (de Haan, 2011).

   Analyses show that a series of precautions are required prior to the retrievals:

   - Cloud-free scenes must be carefully identified and any cloud-residuals should be, ideally, minimized. This is achieved by using MODIS data collocated over OMI pixels. Note that other data from other sensors may theoretically be used for such a filtering.

- An accurate prior τ knowledge is necessary. Both τ and ALH parameters simultaneously contribute to the shielding of O_2-O_2 dimers. The analyses of O_2-O_2 alone lead to an ambiguity since aerosol extinction and aerosol altitude cannot be distinguished from the O_2-O_2 SCD.
- Because low amounts of aerosols have very little impact on the O_2-O_2 SCD changes, large biases are expected over scenes including aerosol particles with $\tau(550nm) \leq 0.5$. This τ(550 nm) value should be considered as a threshold for a good ALH retrieval quality. Moreover, the algorithms are expected to present a very low sensitivity to particles located at an altitude higher than 4 km. This is because of the nature of the O_2-O_2 complex, for which the absorption scales with the pressure-squared instead of being linear with pressure.
- Any prior information on aerosol types (e.g. scattering vs. absorbing) should be reasonably assumed. This can be based on literature knowledge and/or any available external information (e.g. OMI UVAI product (Torres et al., 1998, 2007)).

5. **What is the uncertainty of the retrieved OMI aerosol layer height?**

   Chapter 3 evaluates 3-year ALH retrievals (2005–2007) of OMI cloud-free observations collocated with MODIS Aqua τ over north-east China. Comparisons in ALH between the LIdar climatology of vertical Aerosol Structure for space-based lidar simulation (LIVAS)and OMI retrievals show maximum seasonal differences in the range of 180–800 m for scenes with MODIS $\tau(550m) \geq 1.0$. Additionally, the aerosol model assumptions (i.e. ω_0) affect ALH retrieval uncertainty up to 660 m. Additional analyses based on the MODIS Black Sky Albedo illustrate that potential aerosol residuals in the prior OMI surface albedo may impact up to 200 m. An accuracy of 0.2 is required on prior τ(550 nm) to limit OMI ALH bias close to zero when $\tau(550nm) \geq 1.0$, and below 500 m for τ(550 nm) values close to 0.6.

   Chapter 4 shows consistent spatial patterns between OMI and CALIOP level 2 (L2) ALH over urban and industrial areas in east China with differences in the range of 462-648 m for collocated MODIS scenes with $\tau(550 \text{ nm}) \geq 0.5$. Finally, it is demonstrated that OMI visible measurements can observe the height of thick and absorbing aerosol layers, with fine particle size, caused by widespread fire episodes such as in South-America. In contrast, CALIOP active observations in the visible depict a strong attenuation of the LIDAR beam, limiting the detection to the top of the aerosol layer

   Finally, Chap. 5 shows the feasibility to generate a 1-year (2006) global ALH dataset from OMI, with a low computing time, using the NN approach. After all input data are prepared (cloud filtering, collocation MODIS τ and OMI O_2-O_2), the retrievals are obtained within 12 hours on a single CPU. Additionally, OMI τ(550 nm) can also be derived at the same time.

6. **How could the developed OMI aerosol layer height algorithm be combined with the tropospheric NO_2 retrieval chain?**

   Chapter 6 applies for the first time an explicit aerosol correction on cloud-free OMI tropospheric NO_2 VCD based on the aerosol parameters (ALH primarily, τ secondary) retrieved from the 477 nm O_2-O_2 band. This is performed on 2-year OMI observations (2006-2007), over China in summer and winter times, and South-America during the biomass burning season. Similarly to OMCLDO2-New, all the OMI aerosol NN algorithms show a reduction of tropospheric NO_2 VCD biases compared to those with OMCLDO2-Old.

   In general, the applied explicit corrections lead to higher VCD values compared to the implicit correction from the OMCLDO2-New version: between 20% and 40% depending on the seasons, regions and aerosol pollution episodes. At a first glance, these differences are attributed to the gap between ALH and the effective cp: a higher ALH would lead to a stronger shielding effect (i.e. smaller AMF) and therefore a higher tropospheric NO_2 VCD. However, model differences also play a role that cannot be ignored. The modelled scattering effects are expected to better rep- resent how aerosols affect the average light path. This leads to a more realistic vertical averaging kernel (AK) associated with the retrieved tropospheric NO_2 VCD. Such an ancillary product is of high importance for estimating surface NO_x emissions from OMI.

   Over scenes with absorbing particles (i.e. $\omega_0 \leq 0.9$), the explicit aerosol correction shows lower biases than with OMCLDO2-New: in the range of 0-40% if accurate τ and aerosol model are available, or 40-60% with inaccurate τ (e.g. retrieved OMI τ) and aerosol model.

   However, for all the processed cases, applying an explicit aerosol correction by using the OMI ALH product is not straightforward. Some difficulties are identified and need to be further investigated: 1) Biases remain high when particles are located close to the surface, regardless of the aerosol correction methodology due to the difficulty to separate surface reflectance and aerosol scattering, 2) All aerosol correction methodologies are sensitive to the qualities of NO_2 vertical profile shape and surface albedo, 3) Although the OMI aerosol NN algorithms lead to promising ALH retrieval, its use for an explicit aerosol correction cannot be direct. The dependence on aerosol parameters (size, ω_0, τ) can lead either to model errors and/or to the risk of not closing the OMI radiance budget. The latest may occur if any of these parameters are used from an external source (model or instrument) and thus not consistent with the OMI ALH retrieval.

   For operational processing, it is recommended to apply the implicit aerosol correction based on OMCLDO2-New, in case no distinction can be made between cloudy, cloud-free, and aerosol contaminated scenes. If OMI cloud-free scenes can be carefully identified and collected, then the explicit aerosol correction based on the OMI NN aerosol algorithms should be considered. The developed NN approach can ensure a fast cloud-free OMI data processing. Overall, the considered aerosol corrections can easily be transposed to the future generation of air quality UV-vis satellite sensors, such as TROPOMI, Sentinel-4, and Sentinel-5-UVNS. Moreover, it can be considered for other trace gases of interest: e.g. SO_2, HCHO. However, such an approach have to be adapted to the specificities of this new generation of instruments: e.g. the improvement in the spatial resolution. The small pixel sizes will clearly bring additional challenges such as the 3D effects of clouds, and thus the impacts on the adjacent observation pixels.

7.2 Achievements and contributions

The key achievements and contributions of this thesis are the following:

1. A novel algorithm is developed allowing, for the first time, the retrieval of aerosol layer height (ALH) from the OMI 477 nm (visible) O_2-O_2 spectral band over cloud free land and water scenes. Additionally, AOT can also be derived as a secondary product.
2. A machine learning approach (i.e. neural network) is implemented and tested to face the challenge of big data in view of aerosol retrievals from satellite hyperspectral measurements.
3. The performance of OMI ALH estimation is assessed by comparison with reference datasets for several specific geophysical events, including continental polluted and biomass burning aerosols.
4. The advantage of exploiting the O_2-O_2 spectral band, instead of the traditional O_2-A band, from satellite sensors is demonstrated.
5. A yearly global ALH and τ dataset is generated from OMI measurements over cloud-free scenes in 2006, with a low computing time thanks to the NN approach.
6. The need of an accurate aerosol correction for tropospheric NO_2 retrieval from UV-visible satellite spectrally resolved measurements is quantified, and the key parameters identified.
7. The response of the OMI effective cloud algorithm to aerosol particle presence and related properties is comprehensively analyzed and suggestions for numerical improvements are given.
8. The aerosol correction, as performed in the reference OMI tropospheric NO_2 dataset DOMINO-v2, over cloud-free scenes is evaluated and its benefits, compared to aerosol-free scene assumptions, is quantified.
9. The feasibility of using the OMI ALH product for an explicit aerosol correction for tropospheric NO_2 retrievals is demonstrated. It can also easily be extended to other trace gases (e.g. SO_2, HCHO) and current and future UV-vis satellite

sensors: GOME- 2, TROPOMI on-board Sentinel-5-Precursor, Sentinel-4-UVN, and Sentinel-5- UVNS. Precautions to be considered regarding model errors, and the closure of the measured radiance budget are given. A list of recommendations is given for future studies to investigate these issues.

7.3 Recommendations

This thesis shows encouraging results on improving the tropospheric NO_2 retrievals in the presence of significant aerosol loadings and to derive the ALH from spectrally resolved observations in the visible spectral range. In this section, recommendations on how to apply these results and for future research are given.

1. **Improving ALH retrieval from the visible band**

   ALH retrieval accuracy from the 477 nm satellite O_2-O_2 spectral band mostly suffers from incorrect identification of aerosol type, and related properties (i.e. scattering phase function, size, ω_o, vertical profile, load). The next ALH improvements need to pay attention on methodologies to minimize these inaccuracies, in particular by relying on reference aerosol datasets. Furthermore, there is a high need to have a reliable surface reflectance dataset especially over scenes with aerosol $\tau(550 \text{ nm}) \leq 1$. Therefore, more efforts need to be given on filtering aerosol residuals, and characterizing the angular dependency of the reflectance (e.g. as done in MODIS aerosol products).

2. **UV-visible-NIR synergy for ALH retrieval**

   In addition to the visible, ultra-violet (O_2-O_2 and Ring effect) and near-infrared (O_2-A) also contain some relevant informant on ALH. Each of these bands presents some advantages either due to lower impacts of surface and aerosol types, or enhanced information on aerosol altitude. Future works should consider novel approaches based on the synergy of all these bands, instead of separate exploitation.

3. **Cloud-free, cloudy and cloud-aerosol mix distinction - Still a major issue**

   Distinction of cloud-free and cloudy scenes based on hyperspectral passive satellite sensors remain challenging. Any potential cloud residuals may highly bias the retrievals presented in this thesis research. There is a need to better and consistently characterize the presence of clouds in the individual satellite observations. Strategies like the merged OMI/MODIS cloud product (i.e. the so-called OMCLDRR) seem very promising and should be further deployed for operational cloud-free processing of UV-vis satellite observations.

4. **Consolidated use of machine learning approach**

   The forthcoming future generations of satellite instruments clearly bring the challenge of dealing with big data. The machine learning approach is a promising approach to perform physical analyses on these huge data volumes, while keeping at the same time a fast processing speed. Related strategies on the training process should be further investigated, notably by learning more from the expert scientific community. For example, the following questions could be addressed: How to include instrument noise in the training phase? How to sample the training database? How to better design the algorithm architecture? etc... Moreover, given the large choice of machine learning techniques, it should be more clear how to select the most relevant concept depending on the scientific objective (e.g. neural network, random forest, principal component analysis, etc...).

Finally, variational and statistical approaches should be combined to benefit from advantages of each of them: i.e. fast processing of big data, with at the same time an accurate determination for individual data points or scenes.

5. **Towards a more comprehensive and consistent aerosol correction: NO_2 profile and surface reflectance**

   The effect of aerosols does not only depend on an accurate description of the aerosol profile and related properties., It also requires an accurate NO_2 profile shape. Furthermore, the surface reflectance should be accurate and consistent in the aerosol and the NO_2 retrieval. Consequently, it is recommended to work on joint aerosol-surface-NO_2 retrievals that ensure the consistency in the retrieval processes.

6. **Towards long-time series of ALH and further applications**

   Although the primary motivation of the developed OMI ALH product is here the aerosol correction for the retrieval trace gas from satellite observations, geophysical analyses of aerosol effects in the OMI UV-visible spectra are also of importance. ALH retrieval from visible (and UV) can help to study the effects of aerosols on the atmospheric shortwave radiation field. Several relevant cases of biomass burning, desert dust, and urban and large pollution events can be analysed, e.g. by using a state of the art spectral radiative transfer model (de Haan et al., 1987; de Haan, 2011). The impact of the aerosols could be considered at different vertical levels including the surface, the lower troposphere, the top-of-the atmosphere and inside the aerosol plume.

   Long-time series of OMI observations can lead to a unique 13-year OMIALH dataset that can then be compared to other sensors such as the NASA A-Train constellation with OCO-2 (Geddes and Bösch, 2015), POLDER, and CALIOP, and/or with synthetic global aerosol vertical distributions through the Copernicus Atmosphere Monitoring Service (CAMS). Furthermore, it is worth applying the OMI ALH methodology to other UV-visible satellite sensors such as GOME-2, Sentinel-5-Precursor TROPOMI, Sentinel-4-UVN, and Sentinel-5-UVNS.

Bibliography

Acarreta, J. R., de Haan, J. F., and Stammes, P.: Cloud pressure retrieval using the O_2-O_2 absorption band at 477 nm, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, doi:10.1029/2003JD003915, 2004.

Ahn, C., Torres, O., and Jethva, H.: Assessment of OMI near-UV aerosol optical depth over land, Journal of Geophysical Research: Atmospheres, 119, 2457–2473, doi:10.1002/2013JD020188, 2014.

Amiridis, V., Marinou, E., Tsekeri, A., Wandinger, U., Schwarz, A., Giannakaki, E., Mamouri, R., Kokkalis, P., Binietoglou, I., Solomos, S., Herekakis, T., Kazadzis, S., Gerasopoulos, E., Proestakis, E., Kottas, M., Balis, D., Papayannis, A., Kontoes, C., Kourtidis, K., Papagiannopoulos, N., Mona, L., Pappalardo, G., Le Rille, O., and Ansmann, A.: LIVAS: a 3-D multi-wavelength aerosol/cloud database based on CALIPSO and EARLINET, Atmospheric Chemistry and Physics, 15, 7127–7153, doi:10.5194/acp-15-7127-2015, 2015.

Andreae, M. and Rosenfeld, D.: AerosolâĂŞcloudâĂŞprecipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Science Reviews, 89, 13–41, doi:10.1016/j.earscirev.2008.03.001, 2008.

Arrhenius, S.: On the influence of carbonic acid in the air upon the temperature of the ground, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 41, 237–276, doi:10.1080/14786449608620846, 1896.

Atkinson, P. M. and Tatnall, A. R. L.: Introduction Neural networks in remote sensing, International Journal of Remote Sensing, 18, 699–709, doi:10.1080/014311697218700, 1997.

Barkley, M. P., Kurosu, T. P., Chance, K., De Smedt, I., Van Roozendael, M., Arneth, A., Hagberg, D., and Guenther, A.: Assessing sources of uncertainty in formaldehyde air mass factors over tropical South America: Implications for top-down isoprene emission estimates, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, doi:10.1029/2011JD016827, 2012.

Barkley, M. P., De Smedt, I., VanâĂĽRoozendael, M., Kurosu, T. P., Chance, K., Arneth, A., Hagberg, D., Guenther, A., Paulot, F., Marais, E., and Mao, J.: Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns, Journal of Geophysical Research: Atmospheres, 118, 6849–6868, doi:10.1002/jgrd.50552, 2013.

Bauduin, S., Clarisse, L., Hadji-Lazaro, J., Theys, N., Clerbaux, C., and Coheur, P.-F.: Retrieval of near-surface sulfur dioxide (SO_2) concentrations at a global scale using IASI satellite observations, Atmospheric Measurement Techniques, 9, 721–740, doi:10.5194/amt-9-721-2016, 2016.

Blackwell, W.: Neural Network Applications in High Resolution Atmospheric Remote Sensing, LINCOLN LABORATORY JOURNAL, 15, 9607–9607–7, 2005.

Blackwell, W.: Neural network Jacobian analysis for high-resolution profiling of the atmosphere, EURASIP Journal on Advances in Signal Processing, 2012, 71, doi:10.1186/1687-6180-2012-71, 2012.

Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO2 retrieval from space, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, doi:10.1029/2003JD003962, 2004.

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO_2 from OMI, Atmospheric Chemistry and Physics, 7, 2103–2118, doi:10.5194/acp-7-2103-2007, 2007.

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitao, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmospheric Measurement Techniques, 4, 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.

Bousserez, N.: Space-based retrieval of NO_2 over biomass burning regions: quantifying and reducing uncertainties, Atmospheric Measurement Techniques, 7, 3431–3444, doi:10.5194/amt-7-3431-2014, 2014.

Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V., Chance, K., and Goede, A.:

SCIAMACHY: mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.

Bovensmann, H., Aben, I., Van Roozendael, M., Kühl, S., Gottwald, M., von Savigny, C., Buchwitz, M., Richter, A., Frankenberg, C., Stammes, P., de Graaf, M., Wittrock, F., Sinnhuber, M., Sinnhuber, B. M., Schönhardt, A., Beirle, S., Gloudemans, A., Schrijver, H., Bracher, A., Rozanov, A. V., Weber, M., and Burrows, J.: SCIAMACHY's View of the Changing Earth's Environment, pp. 175–216, Springer Netherlands, Dordrecht, doi:10.1007/978-90-481-9896-2_10, 2011.

Bovensmann, H., Boesch, H., Brunner, D., Ciais, P., Crisp, D., Dolman, H., Hayman, G., houweling, S., and Lichtenberg, G.: Report for Mission Selection: CarbonSat, ESA SP-1330/1 (2 volume series), ESA Communications, 2015.

Buchwitz, M., Rozanov, V., and Burrows, J.: A near-infrared optimized DOAS method for the fast global retrieval of atmospheric CH_4, CO, CO_2, H_2O, and N_2O total column amounts from SCIAMACHY Envisat-1 nadir radiances, Journal of Geophysical Research: Atmospheres, 105, 15 231–15 245, doi:10.1029/2000JD900191, 2000.

Bucsela, E., Celarier, E., Wenig, M., Gleason, J., Veefkind, J., Boersma, K., and Brinksma, E.: Algorithm for NO_2 vertical column retrieval from the ozone monitoring instrument, IEEE Transactions on Geoscience and Remote Sensing, 44, 1245–1258, doi:10.1109/TGRS.2005.863715, 2006.

Burrows, J., Weber, M., Buchwitz, M., Rozanov, V., LadstÃď dter-Weissenmayer, A., Richter, A., de Beek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., Eisinger, M., and Perner, D.: The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results, J. Atmos. Sci., 56, 151–175, 1999.

Burrows, J., Platt, U., and Borrell, P., eds.: Thye Remote Sensing of tropospheric Composition from Space, Springer-Verlag Berlin Heidelberg, doi:10.1007/978-3-642-14791-3, 2011.

Castellanos, P. and Boersma, K. F.: Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession, Scientific Reports, 2, doi:10.1038/srep00265, 2012.

Castellanos, P., Boersma, K. F., Torres, O., and de Haan, J. F.: OMI tropospheric NO_2 air mass factors over South America: effects of biomass burning aerosols, Atmospheric Measurement Techniques, 8, 3831–3849, doi:10.5194/amt-8-3831-2015, 2015.

Chameides, W. L., Yu, H., Liu, S. C., Bergin, M., Zhou, X., Mearns, L., Wang, G., Kiang, C. S., Saylor, R. D., Luo, C., Huang, Y., Steiner, A., and Giorgi, F.: Case study of the effects of atmospheric aerosols and regional haze on agriculture: An opportunity to enhance crop yields in China through emission controls?, Proceedings of the National Academy of Sciences, 96, 13 626–13 633, doi:10.1073/pnas.96.24.13626, 1999.

Chance, K. V. and Spurr, R. J. D.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, Appl. Opt., 36, 5224–5230, doi:10.1364/AO.36.005224, 1997.

Chapman, S.: On ozone and atomic oxygen in the upper atmosphere, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 10, 369–383, doi:10.1080/14786443009461588, 1930.

Chen, Y., Ebenstein, A., Greenstone, M., and Hongbin, L.: Evidence on the impact of sustained exposure to air pollution on life expectancy from ChinaâĂŹs Huai River policy, Proc Natl Acad Sci USA, 110, 1293–12 941, doi:10.1073/pnas.1300018110, 2013.

Chimot, J., Vlemmix, T., Veefkind, J. P., de Haan, J. F., and Levelt, P. F.: Impact of aerosols on the OMI tropospheric NO_2 retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free scenes via a simple cloud model?, Atmospheric Measurement Techniques, 9, 359–382, doi:10.5194/amt-9-359-2016, 2016.

Chimot, J., Veefkind, J. P., Vlemmix, T., de Haan, J. F., Amiridis, V., Proestakis, E., Marinou, E., and Levelt, P. F.: An exploratory study on the aerosol height retrieval from OMI measurements of the 477 nm O_2-O_2 spectral band using a neural network approach, Atmospheric Measurement Techniques, 10, 783–809, doi:10.5194/amt-10-783-2017, 2017.

Chimot, J., Veefkind, J. P., de Haan, J. F., Stammes, P., and Levelt, P. F.: Minimizing aerosol radiation effects on the OMI tropospheric NO_2 retrieval – An improved use of the 477 nm O_2-O_2 band, Atmospheric Measurement Techniques Discussions - To be submitted, July, 2018a.

Chimot, J., Veefkind, J. P., Vlemmix, T., and Levelt, P. F.: Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations, Atmospheric Measurement Techniques, 11, 2257–2277, doi:10.5194/amt-11-2257-2018, 2018b.

Chu, W. and McCormick, M.: Inversion of stratospheric aerosol and gaseous constituents from spacecraft solar extinction data in the 0.38–1.0-μm wavelength region, Appl. Opt., 18, 1404–1413, doi:10.1364/AO.18.001404, 1979.

Chu, W. P. and McCormick, M. P.: SAGE observations of stratospheric nitrogen dioxide, Journal of Geophysical Research: Atmospheres, 91, 5465–5476, doi:10.1029/JD091iD05p05465, 1986.

Ciais, P., Dolman, A., Dargaville, R., Barrie, L., Bombellie, A., Butler, J., Canadell, P., and Moriyama, T.: Geo Carbon Strategy, p. 48, Geo Secretariat Geneva/FAO, Rome, 2010.

Clarisse, L., R'Honi, Y., Coheur, P.-F., Hurtmans, D., and Clerbaux, C.: Thermal infrared nadir observations of 24 atmospheric gases, Geophysical Research Letters, 38, n/a–n/a, doi:10.1029/2011GL047271, 2011.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, Atmospheric Chemistry and Physics, 9, 6041–6054, doi:10.5194/acp-9-6041-2009, 2009.

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

Connor, B., Bösch, H., McDuffie, J., Taylor, T., Fu, D., Frankenberg, C., O'Dell, C., Payne, V. H., Gunson, M., Pollock, R., Hobbs, J., Oyafuso, F., and Jiang, Y.: Quantification of uncertainties in OCO-2 measurements of XCO_2: simulations and linear error analysis, Atmospheric Measurement Techniques, 9, 5227–5238, doi:10.5194/amt-9-5227-2016, 2016.

Crevoisier, C., Chédin, A., Matsueda, H., Machida, T., Armante, R., and Scott, N. A.: First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations, Atmospheric Chemistry and Physics, 9, 4797–4810, doi:10.5194/acp-9-4797-2009, 2009a.

Crevoisier, C., Nobileau, D., Fiore, A. M., Armante, R., Chédin, A., and Scott, N. A.: Tropospheric methane in the tropics âĂŞ first year from IASI hyperspectral infrared observations, Atmospheric Chemistry and Physics, 9, 6337–6350, doi:10.5194/acp-9-6337-2009, 2009b.

Crisp, D.: Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2), vol. 9607, pp. 9607–9607–7, doi:10.1117/12.2187291, 2015.

Curier, R., Kranenburg, R., Segers, A., Timmermans, R., and Schaap, M.: Synergistic use of OMI NO_2 tropospheric columns and LOTOSâĂŞEUROS to evaluate the NO_x emission trends across Europe, Remote Sensing of Environment, 149, 58 – 69, doi:https://doi.org/10.1016/j.rse.2014.03.032, 2014.

de Graaf, M., Stammes, P., Torres, O., and Koelemeijer, R. B. A.: Absorbing Aerosol Index: Sensitivity analysis, application to GOME and comparison with TOMS, Journal of Geophysical Research: Atmospheres, 110, n/a–n/a, doi:10.1029/2004JD005178, 2005.

de Graaf, M., Stammes, P., and Aben, E. A. A.: Analysis of reflectance spectra of UV-absorbing aerosol scenes measured by SCIAMACHY, Journal of Geophysical Research: Atmospheres, 112, n/a–n/a, doi:10.1029/2006JD007249, 2007.

de Graaf, M., Tilstra, L. G., Wang, P., and Stammes, P.: Retrieval of the aerosol direct radiative effect over clouds from spaceborne spectrometry, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2011JD017160, 2012.

de Haan, J. F.: DISAMAR Algorithm Description and Background Information, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands, 2011.

de Haan, J. F., Bosma, P. B., and Hovenier, J. W.: The adding method for multiple scattering calculations of polarized light, Astron. Astrophys., 183, 371–391, 1987.

de Leeuw, G., Holzer-Popp, T., Bevan, S., Davies, W., Descloitres, J., Grainger, R., Griesfeller, J., Heckel, A., Kinne, S., KlÃijser, Kolmonen, P., Litvinov, P., Martynenko, D., North, P., Ovigneur, B., Pascal, N., Poulsen, C., Ramon, D., Schulz, M., Siddans, R., Sogacheva, L., TanrÃl', D., Thomas, G., Virtanen, T., von Hoyningen Huene, W., Vountas, M., and Pinnock, S.: Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sensing of Environment, 162, 295 – 315, doi:https://doi.org/10.1016/j.rse.2013.04.023, 2015.

de Smedt, I., Müller, J.-F., Stavrakou, T., van der A, R., Eskes, H., and Van Roozendael, M.: Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, Atmospheric Chemistry and Physics, 8, 4947–4963, doi:10.5194/acp-8-4947-2008, 2008.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm Theoretical Baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmospheric Measurement Techniques Discussions, 2017, 1–53, doi:10.5194/amt-2017-393, 2017.

Deutschmann, T. and Wagner, T.: Tracy-II Users Manual, 2008.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmospheric Measurement Techniques, 8, 281–299, doi:10.5194/amt-8-281-2015, 2015.

Ding, J., van der A, R. J., Mijling, B., Levelt, P. F., and Hao, N.: NO_x emission estimates during the 2014 Youth Olympic Games in Nanjing, Atmospheric Chemistry and Physics, 15, 9399–9412, doi:10.5194/acp-15-9399-2015, 2015.

Ding, J., Miyazaki, K., van der A, R. J., Mijling, B., Kurokawa, J.-I., Cho, S., Janssens-Maenhout, G., Zhang, Q., Liu, F., and Levelt, P. F.: Intercomparison of NO_x emission inventories over East Asia, Atmospheric Chemistry and Physics, 17, 10 125–10 141, doi:10.5194/acp-17-10125-2017, 2017a.

Ding, J., van der A, R. J., Mijling, B., and Levelt, P. F.: Space-based NO_x emission estimates over remote regions improved in DECSO, Atmospheric Measurement Techniques, 10, 925–938, doi:10.5194/amt-10-925-2017, 2017b.

Ding, S., Wang, J., and Xu, X.: Polarimetric remote sensing in oxygen A and B bands: sensitivity study and information content analysis for vertical profile of aerosols, Atmospheric Measurement Techniques, 9, 2077–2092, doi:10.5194/amt-9-2077-2016, 2016.

Dirksen, R. J., Folkert Boersma, K., de Laat, J., Stammes, P., van der Werf, G. R., Val Martin, M., and Kelder, H. M.: An aerosol boomerang: Rapid around-the-world transport of smoke from the December 2006 Australian forest fires observed from space, Journal of Geophysical Research: Atmospheres, 114, n/a–n/a, doi:10.1029/2009JD012360, 2009.

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., TanrÃĺ, D., and Slutsker, I.: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations, Journal of the Atmospheric Sciences, 59, 590–608, doi:10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2, 2002.

Dufresne, J.-L., Gautier, C., Ricchiazzi, P., and Fouquart, Y.: Longwave Scattering Effects of Mineral Aerosols, Journal of the Atmospheric Sciences, 59, 1959–1966, doi:10.1175/1520-0469(2002)059<1959:LSEOMA>2.0.CO;2, 2002.

Duncan, B., Prados, A., Lamsal, L., Liu, Y., Streets, D., Gupta, P., Hilsenrath, E., Kahn, R., Nielsen, J., Beyersdorf, A., Burton, S., Fiore, A., Fishman, J., Henze, D., Hostetler, C., Krotkov, N., Lee, P., Lin, M., Pawson, S., Pfister, G., Pickering, K., Pierce, R., Yoshida, Y., and Ziemba, L.: Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid, Atmospheric Environment, 94, 647 – 662, doi:https://doi.org/10.1016/j.atmosenv.2014.05.061, 2014.

Duncan, B., Lamsal, L., Thompson, A., Yoshida, Y., Lu, Z., Streets, D., Hurwitz, M., and Pickering, K.: A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005âĂŞ2014), Journal of Geophysical Research: Atmospheres, 121, 976–996, doi:10.1002/2015JD024121, 2016.

Eck, T. F., Holben, B. N., Dubovik, O., Smirnov, A., Goloub, P., Chen, H. B., Chatenet, B., Gomes, L., Zhang, X.-Y., Tsay, S.-C., Ji, Q., Giles, D., and Slutsker, I.: Columnar aerosol optical properties at AERONET sites in central eastern Asia and aerosol transport to the tropical mid-Pacific, Journal of Geophysical Research: Atmospheres, 110, n/a–n/a, doi:10.1029/2004JD005274, 2005.

Edwards, D. P., Emmons, L. K., Gille, J. C., Chu, A., AttiÃĺ, J.-L., Giglio, L., Wood, S. W., Haywood, J., Deeter, M. N., Massie, S. T., Ziskin, D. C., and Drummond, J. R.: Satellite-observed pollution from Southern Hemisphere biomass burning, Journal of Geophysical Research: Atmospheres, 111, n/a–n/a, doi:10.1029/2005JD006655, 2006.

Einstein, A.: Zur Quantentheorie der Strahlung, Physikalische Zeitschrift, 18, 1917.

Eskes, H. and Boersma, K.: Averaging kernels for DOAS total-column satellite retrievals, Atmospheric Chemistry and Physics, 3, 1285–1291, doi:10.5194/acp-3-1285-2003, 2003.

Eskes, H., Huijnen, V., Arola, A., Benedictow, A., Blechschmidt, A.-M., Botek, E., Boucher, O., Bouarar, I., Chabrillat, S., Cuevas, E., Engelen, R., Flentje, H., Gaudel, A., Griesfeller, J., Jones, L., Kapsomenakis, J., Katragkou, E., Kinne, S., Langerock, B., Razinger, M., Richter, A., Schultz, M., Schulz, M., Sudarchikova, N., Thouret, V., Vrekoussis, M., Wagner, A., and Zerefos, C.: Validation of reactive gases and aerosols in the MACC global analysis and forecast system, Geoscientific Model Development, 8, 3523–3543, doi:10.5194/gmd-8-3523-2015, 2015.

Farman, J., Gardiner, B., and Shanklin, J.: Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, Nature, 315, 207–210, doi:10.1038/315207a0, 1985.

Feingold, G., Cotton, W. R., Kreidenweis, S. M., Janel, and Davis, T.: The impact of giant cloud condensation nuclei on drizzle formation in stratocumulus: Implications for cloud radiative properties, J. Atmos. Sci, 56, 4100–4117, 1999.

Figueras i Ventura, J. and Russchenberg, H.: Towards a better understanding of the impact of anthropogenic aerosols in the hydrological cycle: IDRA, IRCTR drizzle radar, Physics and Chemistry of the Earth, Parts A/B/C, 34, 88 – 92, doi:https://doi.org/10.1016/j.pce.2008.02.038, 2009.

Fioletov, V., McLinden, C. A., Kharol, S. K., Krotkov, N. A., Li, C., Joiner, J., Moran, M. D., Vet, R., Visschedijk, A. J. H., and Denier van der Gon, H. A. C.: Multi-source SO_2 emission retrievals and consistency of satellite and surface measurements with reported emissions, Atmospheric Chemistry and Physics, 17, 12 597–12 616, doi:10.5194/acp-17-12597-2017, 2017.

for Europe, W. H. O. W. R. O.: Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide : report on a WHO working group, Bonn, Germany, 2003.

Frankenberg, C., Bergamaschi, P., Butz, A., Houweling, S., Meirink, J., Notholt, J., Petersen, A., Schrijver, H., Warneke, T., and Aben, I.: Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT, Geophysical Research Letters, 35, n/a–n/a, doi:10.1029/2008GL034300, 2008.

Frederikse, T., Riva, R., Kleinherenbrink, M., Wada, Y., van den Broeke, M., and Marzeion, B.: Closing the sea level budget on a regional scale: Trends and variability on the Northwestern European continental shelf., Geophysical Research Letters, 43, 10 864–10 872, doi:10.1002/2016GL070750, 2016.

Frederikse, T., Jevrejeva, S., Riva, R. E. M., and Dangendorf, S.: A consistent sea-level reconstruction and its budget on basin and global scales over 1958-2014, Journal of Climate, 0, doi:10.1175/JCLI-D-17-0502.1, 2017a.

Frederikse, T., Simon, K., Katsman, C., and Riva, R.: The sea-level budget along the Northwest Atlantic coast: GIA, mass changes and large-scale ocean dynamics, Journal of Geophysical Research: Oceans, doi:10.1002/2017JC012699, 2017b.

Geddes, A. and Bösch, H.: Tropospheric aerosol profile information from high-resolution oxygen A-band measurements from space, Atmospheric Measurement Techniques, 8, 859–874, doi:10.5194/amt-8-859-2015, 2015.

Gordley, L. L., Russell, J. M., Mickley, L. J., Frederick, J. E., Park, J. H., Stone, K. A., Beaver, G. M., McInerney, J. M., Deaver, L. E., Toon, G. C., Murcray, F. J., Blatherwick, R. D., Gunson, M. R., Abbatt, J. P. D., Mauldin, R. L., Mount, G. H., Sen, B., and Blavier, J.-F.: Validation of nitric oxide and nitrogen dioxide measurements made by the Halogen Occultation Experiment for UARS platform, Journal of Geophysical Research: Atmospheres, 101, 10 241–10 266, doi:10.1029/95JD02143, 1996.

Haykin, S.: Neural Networks: A Comprehensive Foundation, International edition, Prentice Hall, 1999.

Heij, G. and Erisman, J.: Acid rain research : do we have enough answers?, Studies in environmental science, Amsterdam ; New York: Elsevier, 1995.

Held, D., Thrun, S., and Savarese, S.: Learning to Track at 100 FPS with Deep Regression Networks, pp. 749–765, Springer International Publishing, Cham, doi:10.1007/978-3-319-46448-0_45, 2016.

Hewson, W., Barkley, M. P., Gonzalez Abad, G., Bösch, H., Kurosu, T., Spurr, R., and Tilstra, L. G.: Development and characterisation of a state-of-the-art GOME-2 formaldehyde air-mass factor algorithm, Atmospheric Measurement Techniques, 8, 4055–4074, doi:10.5194/amt-8-4055-2015, 2015.

Holben, B., Eck, T., Slutsker, I., TanrÃľ, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1–16, 1998.

Holloway, A. M. and Wayne, R. P.: Atmospheric Chemistry, The Royal Society of Chemistry, 2010.

Hovenier, J. W. and Hage, J. I.: Relations involving the spherical albedo and other photometric quantities of planets with thick atmospheres, Astron. Astrophys., 214, 391–401, 1989.

Hsu, N., Herman, J., and Tsay, S.-C.: Radiative impacts from biomass burning in the presence of clouds during boreal spring in southeast Asia, Geophysical Research Letters, 30, doi:10.1029/2002GL016485, 2003.

Hu, H., Hasekamp, O., Butz, A., Galli, A., Landgraf, J., Aan de Brugh, J., Borsdorff, T., Scheepmaker, R., and Aben, I.: The operational methane retrieval algorithm for TROPOMI, Atmospheric Measurement Techniques, 9, 5423–5440, doi:10.5194/amt-9-5423-2016, 2016.

Huijnen, V., Flemming, J., Chabrillat, S., Errera, Q., Christophe, Y., Blechschmidt, A.-M., Richter, A., and Eskes, H.: C-IFS-CB05-BASCOE: stratospheric chemistry in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 9, 3071–3091, doi:10.5194/gmd-9-3071-2016, 2016.

Ingmann, I., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Bazalgette CourrÃĺges-Lacoste, G.: Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, Remote Sensing of Environment, 120, 58 – 69, doi:https://doi.org/10.1016/j.rse.2012.01.023, 2012.

IPCC: The Core Writing Team Pachauri, R. K. and Meyer, L. A., Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva, Switzerland, 2014.

IPCC: Solomon, S., Qin, D., Manning, M., Chen, Z., M., M., Averyt, K. B., Tignor, M., and Miller, H. L.: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Jacob, D., Heikes, B., Fan, S.-M., Logan, J., Mauzerall, D., Bradshaw, J., Singh, H., Gregory, G., Talbot, R., Blake, D., and Sachse, G.: Origin of ozone and NO_x in the tropical troposphere: a photochemical analysis of aircraft observations over the South Atlantic basin, J. Geophys. Res., 101, 24 235âĂŞ24 250, 1996.

Jethva, H., Torres, O., and Ahn, C.: Global assessment of OMI aerosol single-scattering albedo using ground-based AERONET inversion, Journal of Geophysical Research: Atmospheres, 119, 9020–9040, doi:10.1002/2014JD021672, 2014.

Kacenelenbogen, M., Vaughan, M. A., Redemann, J., Hoff, R. M., Rogers, R. R., Ferrare, R. A., Russell, P. B., Hostetler, C. A., Hair, J. W., and Holben, B. N.: An accuracy assessment of the CALIOP/CALIPSO version 2/version 3 daytime aerosol extinction product based on a detailed multi-sensor, multi-platform case study, Atmospheric Chemistry and Physics, 11, 3981–4000, doi:10.5194/acp-11-3981-2011, 2011.

Kanaya, Y., Irie, H., Takashima, H., Iwabuchi, H., Akimoto, H., Sudo, K., Gu, M., Chong, J., Kim, Y. J., Lee, H., Li, A., Si, F., Xu, J., Xie, P.-H., Liu, W.-Q., Dzhola, A., Postylyakov, O., Ivanov, V., Grechko, E., Terpugova, S., and Panchenko, M.: Long-term MAX-DOAS network observations of NO_2 in Russia and Asia (MADRAS) during the period 2007 and 2012: instrumentation, elucidation of climatology, and comparisons with OMI satellite observations and global model simulations, Atmospheric Chemistry and Physics, 14, 7909–7927, doi:10.5194/acp-14-7909-2014, 2014.

Kaufman, Y. J., Tanre, D., and Boucher, O.: A satellite view of aerosols in the climate system, Nature, 419, 215–223, doi:10.1038/nature01091, 2002.

Kerkweg, A., Buchholz, J., Ganzeveld, L., Pozzer, A., Tost, H., and Jöckel, P.: Technical Note: An implementation of the dry removal processes DRY DEPosition and SEDImentation in the Modular Earth Submodel System (MESSy), Atmospheric Chemistry and Physics, 6, 4617–4632, doi:10.5194/acp-6-4617-2006, 2006.

Kim, M.-H., Kim, S.-W., Yoon, S.-C., and Omar, A. H.: Comparison of aerosol optical depth between CALIOP and MODIS-Aqua for CALIOP aerosol subtypes over the ocean, Journal of Geophysical Research: Atmospheres, 118, 13,241–13,252, doi:10.1002/2013JD019527, 2013.

Kipling, Z., Stier, P., Johnson, C. E., Mann, G. W., Bellouin, N., Bauer, S. E., Bergman, T., Chin, M., Diehl, T., Ghan, S. J., Iversen, T., Kirkevåg, A., Kokkola, H., Liu, X., Luo, G., van Noije, T., Pringle, K. J., von Salzen, K., Schulz, M., Seland, Ø., Skeie, R. B., Takemura, T., Tsigaridis, K., and Zhang, K.: What controls the vertical distribution of aerosol? Relationships between process sensitivity in HadGEM3–UKCA and inter-model variation from AeroCom Phase II, Atmospheric Chemistry and Physics, 16, 2221–2241, doi:10.5194/acp-16-2221-2016, 2016.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2008JD010290, 2008.

Koelemeijer, R. B. A., de Haan, J. F., and Stammes, P.: A database of spectral surface reflectivity in the range 335–772 nm derived from 5.5 years of GOME observations, Journal of Geophysical Research: Atmospheres, 108, n/a–n/a, doi:10.1029/2002JD002429, 2003.

Koffi, B., Schulz, M., BrÃľon, F.-M., Griesfeller, J., Winker, D., Balkanski, Y., Bauer, S., Berntsen, T., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Ghan, S., Ginoux, P., Gong, S., Horowitz, L., Iversen, T., KirkevÃěg, A., Koch, D., Krol, M., Myhre, G., Stier, P., and Takemura, T.: Application of the CALIOP layer product to evaluate the vertical distribution of aerosols estimated by global models: AeroCom phase I results, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2011JD016858, 2012.

Krotkov, N. A., McClure, B., Dickerson, R. R., Carn, S. A., Li, C., Bhartia, P. K., Yang, K., Krueger, A. J., Li, Z., Levelt, P. F., Chen, H., Wang, P., and Lu, D.: Validation of SO_2 retrievals from the Ozone Monitoring Instrument over NE China, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2007JD008818, 2008.

Kuhlmann, G., Lam, Y. F., Cheung, H. M., Hartl, A., Fung, J. C. H., Chan, P. W., and Wenig, M. O.: Development of a custom OMI NO_2 data product for evaluating biases in a regional chemistry transport model, Atmospheric Chemistry and Physics, 15, 5627–5644, doi:10.5194/acp-15-5627-2015, 2015.

Kunkel, K., Karl, T., Easterling, D., Redmond, K., Young, J., Yin, X., and Hennon, P.: Probable maximum precipitation and climate change, Geophysical Research Letters, 40, 1402–1408, doi:10.1002/grl.50334, 2013.

Lacis, A. A., Wuebbles, D. J., and Logan, J. A.: Radiative forcing of climate by changes in the vertical distribution of ozone, Journal of Geophysical Research: Atmospheres, 95, 9971–9981, doi:10.1029/JD095iD07p09971, 1990.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.: Evaluation of OMI operational standard NO_2 column retrievals using in situ and surface-based NO_2 observations, Atmospheric Chemistry and Physics, 14, 11 587–11 609, doi:10.5194/acp-14-11587-2014, 2014.

Lee, C., Martin, R. V., van Donkelaar, A., O'Byrne, G., Krotkov, N., Richter, A., Huey, L. G., and Holloway, J. S.: Retrieval of vertical columns of sulfur dioxide from SCIAMACHY and OMI: Air mass factor algorithm development, validation, and error analysis, Journal of Geophysical Research: Atmospheres, 114, n/a–n/a, doi:10.1029/2009JD012123, 2009.

Lee, J., Kim, J., Yang, P., and Hsu, N. C.: Improvement of aerosol optical depth retrieval from MODIS spectral reflectance over the global ocean using new aerosol models archived from AERONET inversion data and tri-axial ellipsoidal dust database, Atmospheric Chemistry and Physics, 12, 7087–7102, doi:10.5194/acp-12-7087-2012, 2012.

Lee, K. H., Li, Z., Wong, M. S., Xin, J., Wang, Y., Hao, W.-M., and Zhao, F.: Aerosol single scattering albedo estimated across China from a combination of ground and satellite measurements, Journal of Geophysical Research (Atmospheres), 112, D22S15, doi:10.1029/2007JD009077, 2007.

Leitão, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q. J., Beekmann, M., and Burrows, J. P.: On the improvement of NO_2 satellite retrievals âĂŞ aerosol impact on the airmass factors, Atmospheric Measurement Techniques, 3, 475–493, doi:10.5194/amt-3-475-2010, 2010.

Levelt, P., Joiner, J., Tamminen, J., Veefkind, P., Bhartia, P. K., Stein Zweers, D., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., McLinden, C., Fioletov, V., Carn, S., de Laat, J., DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., Gonzáles Abad, G., Arola, A., Boersma, F., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kleipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suileiman, R., Tilstra, L. G., Torres, O., Wang, H., and Wargan, K.: The Ozone Monitoring Instrument: Overview of twelve years in space, Atmospheric Chemistry and Physics Discussions, 2017, 1–61, doi:10.5194/acp-2017-487, 2017.

Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., van den Oord, G. H. J., Bhartia, P. K., Tamminen, J., de Haan,

J. F., and Veefkind, J. P.: Science Objectives of the Ozone Monitoring Instrument, IEEE Transactions on Geoscience and Remote Sensing, 44, 1199–1208, doi:10.1109/TGRS.2006.872336, 2006.

Levitus, S., Antonov, J. I., Boyer, T. P., Locarnini, R. A., Garcia, H. E., and Mishonov, A. V.: Global ocean heat content 1955âĂŞ2008 in light of recently revealed instrumentation problems, Geophysical Research Letters, 36, n/a–n/a, doi:10.1029/2008GL037155, 2009.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, Journal of Geophysical Research: Atmospheres, 112, n/a–n/a, doi:10.1029/2006JD007811, 2007.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, doi:10.5194/amt-6-2989-2013, 2013.

Li, X., Brauers, T., Hofzumahaus, A., Lu, K., Li, Y. P., Shao, M., Wagner, T., and Wahner, A.: MAX-DOAS measurements of NO_2, HCHO and CHOCHO at a rural site in Southern China, Atmospheric Chemistry and Physics, 13, 2133–2151, doi:10.5194/acp-13-2133-2013, 2013.

Lin, J.-T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendael, M., Clémer, K., and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, Atmospheric Chemistry and Physics, 14, 1441–1461, doi:10.5194/acp-14-1441-2014, 2014.

Lin, J.-T., Liu, M.-Y., Xin, J.-Y., Boersma, K. F., Spurr, R., Martin, R., and Zhang, Q.: Influence of aerosols and surface reflectance on satellite NO_2 retrieval: seasonal and spatial characteristics and implications for NO_x emission constraints, Atmospheric Chemistry and Physics, 15, 11 217–11 241, doi:10.5194/acp-15-11217-2015, 2015.

Liu, Z., Winker, D., Omar, A., Vaughan, M., Trepte, C., Hu, Y., Powell, K., Sun, W., and Lin, B.: Effective lidar ratios of dense dust layers over North Africa derived from the {CALIOP} measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 204 – 213, doi:https://doi.org/10.1016/j.jqsrt.2010.05.006, 2011.

Loeb, N. and Su, W.: Direct Aerosol Radiative Forcing Uncertainty Based on a Radiative Perturbation Analysis, Journal of Climate, 23, 5288–5293, doi:10.1175/2010JCLI3543.1, 2010.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J.-T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J., and Krol, M.: Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals, Atmospheric Measurement Techniques, 10, 759–782, doi:10.5194/amt-10-759-2017, 2017.

Loyola, D.: Automatic cloud analysis from polar-orbiting satellites using neural network and data fusion techniques, in: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, vol. 4, pp. 2530–2533 vol.4, doi:10.1109/IGARSS.2004.1369811, 2004.

Loyola, D., Thomas, W., Livschitz, Y., Ruppert, T., Albert, P., and Hollmann, R.: Cloud Properties Derived From GOME/ERS-2 Backscatter Data for Trace Gas Retrieval, IEEE Transactions on Geoscience and Remote Sensing, 45, 2747–2758, doi:10.1109/TGRS.2007.901043, 2007.

Loyola, D., Thomas, W., Spurr, R., and Mayer, B.: Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements, International Journal of Remote Sensing, 31, 4295–4318, doi:10.1080/01431160903246741, 2010.

Lu, Z. and Streets, D.: Increase in NO_x Emissions from Indian Thermal Power Plants during 1996âĂŞ2010: Unit-Based Inventories and Multisatellite Observations, Environmental Science & Technology, 46, 7463–7470, doi:10.1021/es300831w, 2012.

Lu, Z., Zhang, Q., and Streets, D. G.: Sulfur dioxide and primary carbonaceous aerosol emissions in China and India, 1996âĂŞ2010, Atmospheric Chemistry and Physics, 11, 9839–9864, doi:10.5194/acp-11-9839-2011, 2011.

Luger, G. and Stubblefield, W.: Artificial intelligence : Structures and strategies for complex problem solving, Benjamin/Cummings, 1998.

Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., Hall, F., Sellers, P., Wu, A., and Angal, A.: Scientific impact of MODIS C5 calibration degradation and

C6+ improvements, Atmospheric Measurement Techniques, 7, 4353–4365, doi:10.5194/amt-7-4353-2014, 2014.

Ma, J., Chen, Y., Wang, W., Yan, P., Liu, H., Yang, S., Hu, Z., and Lelieveld, J.: Strong air pollution causes widespread haze-clouds over China, Journal of Geophysical Research: Atmospheres, 115, n/a–n/a, doi:10.1029/2009JD013065, 2010.

Ma, J. Z., Beirle, S., Jin, J. L., Shaiganfar, R., Yan, P., and Wagner, T.: Tropospheric NO_2 vertical column densities over Beijing: results of the first three years of ground-based MAX-DOAS measurements (2008 and 2011) and satellite validation, Atmospheric Chemistry and Physics, 13, 1547–1567, doi:10.5194/acp-13-1547-2013, 2013.

Martin, R.: Satellite remote sensing of surface air quality, Atmospheric Environment, 42, 7823 – 7843, doi:https://doi.org/10.1016/j.atmosenv.2008.07.018, 2008.

Martin, R., Chance, K., Jacob, D., Kurosu, T., Spurr, R., Bucsela, E., Gleason, J., Palmer, P., Bey, I., Fiore, A., Li, Q., Yantosca, R., and Koelemeijer, R.: An improved retrieval of tropospheric nitrogen dioxide from GOME, Journal of Geophysical Research: Atmospheres, 107, ACH 9–1–ACH 9–21, doi:10.1029/2001JD001027, 2002.

Martin, R., Jacob, D., Chance, K., Kurosu, T., Palmer, P., and Evans, M.: Global inventory of nitrogen oxide emissions constrained by space-based observations of NO_2 columns, Journal of Geophysical Research: Atmospheres, 108, n/a–n/a, doi:10.1029/2003JD003453, 2003.

McComiskey, A., Schwartz, S., Schmid, B., Guan, H., Lewis, E., Ricchiazzi, P., and Ogren, J.: Direct aerosol forcing: Calculation from observables and sensitivities to inputs, Journal of Geophysical Research: Atmospheres, 113, doi:10.1029/2007JD009170, 2008.

Mendolia, D., D'Souza, R. J. C., Evans, G. J., and Brook, J.: Comparison of tropospheric NO_2 vertical columns in an urban environment using satellite, multi-axis differential optical absorption spectroscopy, and in situ measurements, Atmospheric Measurement Techniques, 6, 2907–2924, doi:10.5194/amt-6-2907-2013, 2013.

Mijling, B. and van der A, R. J.: Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, doi:10.1029/2012JD017817, 2012.

Mount, G. H., Rusch, D. W., Noxon, J. F., Zawodny, J. M., and Barth, C. A.: Measurements of stratospheric NO2 from the Solar Mesosphere Explorer satellite: 1. An overview of the results, Journal of Geophysical Research: Atmospheres, pp. 1327–1340, doi:10.1029/JD089iD01p01327, 1984.

Nanda, S., de Graaf, M., Sneep, M., de Haan, J. F., Stammes, P., Sanders, A. F. J., Tuinder, O., Veefkind, J. P., and Levelt, P. F.: Error sources in the retrieval of aerosol information over bright surfaces from satellite measurements in the oxygen A-band, Atmospheric Measurement Techniques Discussions, 2017, 1–26, doi:10.5194/amt-2017-323, 2017.

Omar, A., Winker, D., Vaughan, M., Hu, Y., Trepte, C., Ferrare, R., Lee, K.-P., Hostetler, C., Kittaka, C., Rogers, R., Kuehn, R., and Liu, Z.: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, Journal of Atmospheric and Oceanic Technology, 26, 1994–2014, doi:10.1175/2009JTECHA1231.1, 2009.

Palancar, G. G., Lefer, B. L., Hall, S. R., Shaw, W. J., Corr, C. A., Herndon, S. C., Slusser, J. R., and Madronich, S.: Effect of aerosols and NO_2 concentration on ultraviolet actinic flux near Mexico City during MILAGRO: measurements and model calculations, Atmospheric Chemistry and Physics, 13, 1011–1022, doi:10.5194/acp-13-1011-2013, 2013.

Palmer, P., Jacob, D., Chance, K., Martin, R., Spurr, R., Kurosu, T., Bey, I., Yantosca, R., Fiore, A., and Li, Q.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, Journal of Geophysical Research: Atmospheres, 106, 14 539–14 550, doi:10.1029/2000JD900772, 2001.

Park, S. S., Kim, J., Lee, H., Torres, O., Lee, K.-M., and Lee, S. D.: Utilization of O_4 slant column density to derive aerosol layer height from a space-borne UVâĂŞvisible hyperspectral sensor: sensitivity and case study, Atmospheric Chemistry and Physics, 16, 1987–2006, doi:10.5194/acp-16-1987-2016, 2016.

Piters, A. J. M., Boersma, K. F., Kroon, M., Hains, J. C., Van Roozendael, M., Wittrock, F., Abuhassan, N., Adams, C., Akrami, M., Allaart, M. A. F., Apituley, A., Beirle, S., Bergwerff, J. B., Berkhout, A. J. C., Brunner, D., Cede, A., Chong, J., Clémer, K., Fayt, C., Frieß, U., Gast, L. F. L., Gil-Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Großmann, K., Hemerijckx, G., Hendrick, F., Henzing, B., Herman, J., Hermans, C., Hoexum, M., van der Hoff, G. R., Irie, H., Johnston, P. V., Kanaya, Y., Kim, Y. J., Klein Baltink, H., Kreher, K., de Leeuw, G., Leigh, R., Merlaud, A., Moerman, M. M., Monks, P. S., Mount, G. H., Navarro-Comas, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., du Piesanie, A., Pinardi, G., Puentedura, O., Richter, A., Roscoe, H. K., Schönhardt, A., Schwarzenbach, B., Shaiganfar, R., Sluis, W., Spinei, E., Stolk, A. P., Strong, K., Swart, D. P. J., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Whyte, C., Wilson, K. M., Yela, M., Yilmaz, S., Zieger, P., and Zhou, Y.: The Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI): design, execution, and early results, Atmospheric Measurement Techniques, 5, 457–485, doi:10.5194/amt-5-457-2012, 2012.

Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy (DOAS), Principles and Applications, Springer-Verlag Berlin Heidelberg, doi:10.1007/978-3-540-75776-4, 2008.

Proestakis, E., Amiridis, V., Marinou, E., Georgoulias, A. K., Solomos, S., Kazadzis, S., Chimot, J., Che, H., Alexandri, G., Binietoglou, I., Kourtidis, K. A., de Leeuw, G., and van der A, R. J.: 9-year spatial and temporal evolution of desert dust aerosols over South-East Asia as revealed by CALIOP, Atmospheric Chemistry and Physics Discussions, 2017, 1–35, doi:10.5194/acp-2017-797, 2017.

Prospero, J. and Carlson, N.: Saharan Air Outbreaks Over the Tropical North Atlantic, 119, 677–691, 1980.

Prospero, J. M.: Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality, J. Geophys. Res.-Atmos., 104, 15 917âĂŞ15 927, doi:10.1029/1999JD900072, 1999.

Pueschel, R. F. and Livingston, J. M.: Aerosol spectral optical depths: Jet fuel and forest fire smokes, Journal of Geophysical Research: Atmospheres, 95, 22 417–22 422, doi:10.1029/JD095iD13p22417, 1990.

Ramanathan, V., Cicerone, R. J., Singh, H. B., and Kiehl, J. T.: Trace gas trends and their potential role in climate change, Journal of Geophysical Research: Atmospheres, 90, 5547–5566, doi:10.1029/JD090iD03p05547, 1985.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D.: Aerosols, Climate, and the Hydrological Cycle, Science, 294, 2119–2124, doi:10.1126/science.1064034, 2001.

Randall, C. E., Rusch, D. W., Bevilacqua, R. M., Hoppel, K. W., and Lumpe, J. D.: Polar Ozone and Aerosol Measurement (POAM) II stratospheric NO2, 1993âĂŞ1996, Journal of Geophysical Research: Atmospheres, 103, 28 361–28 371, doi:10.1029/98JD02092, 1998.

Remer, L., Kaufman, Y., TanrÃľ, D., Mattoo, S., Chu, D., Martins, J., Li, R.-R., Ichoku, C., Levy, R., Kleidman, R., Eck, T., Vermote, E., and Holben, B.: The MODIS Aerosol Algorithm, Products, and Validation, Journal of the Atmospheric Sciences, 62, 947–973, doi:10.1175/JAS3385.1, 2005.

Remer, L., Kleidman, R., Levy, R., Kaufman, Y., TanrÃľ, D., Mattoo, S., Martins, J., Ichoku, C., Koren, I., Yu, H., and Holben, B.: Global aerosol climatology from the MODIS satellite sensors, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2007JD009661, 2008.

Reuter, M., Buchwitz, M., Hilboll, A., Richter, A., Schneising, O., Hilker, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Decreasing emissions of NO$_x$ relative to CO$_2$ in East Asia inferred from satellite observations, Nat. Geosci. Lett., 7, 792âĂŞ795, doi:10.1038/NGEO2257, 2014.

Richter, A. and Burrows, J.: Tropospheric NO$_2$ from GOME measurements, Advances in Space Research, 29, 1673–1683, doi:https://doi.org/10.1016/S0273-1177(02)00100-X, 2002.

Richter, A. and Wagner, T.: The Use of UV in Visible and Near IR Solar Back Scattered Radiation to Determine Trace Gases, The Remote Sensingof Tropospheric Composition from Space, Physics of Earth and Space Environments, Chapter 2, Springer-Verlag Berlin Heidelberg,, 2011.

Richter, A., Burrows, J., N, H., Granier, C., and Niemeier, U.: Increase in tropospheric nitrogen dioxide over China observed from space, Nature, 437, doi:10.1038/nature04092, 2005.

Rodgers, C.: Inverse methods for atmospheric sounding - theory and practice, Series on Atmospheric, Oceanic and Planetary Physics, World Scientific Publishing, Singapore, 2000.

Rosenfeld, D.: Suppression of Rain and Snow by Urban and Industrial Air Pollution, Science, 287, 1793–1796, doi:10.1126/science.287.5459.1793, 2000.

Rosenfeld, D., Lahav, R., Khain, A., and Pinsky, M.: The Role of Sea Spray in Cleansing Air Pollution over

Ocean via Cloud Processes, Science, 297, 1667–1670, doi:10.1126/science.1073869, 2002.

Rozanov, V. V. and Rozanov, A. V.: Differential optical absorption spectroscopy (DOAS) and air mass factor concept for a multiply scattering vertically inhomogeneous medium: theoretical consideration, Atmospheric Measurement Techniques, 3, 751–780, doi:10.5194/amt-3-751-2010, 2010.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533–536, doi:10.1038/323533a0, 1986.

Sanders, A. F. J., de Haan, J. F., Sneep, M., Apituley, A., Stammes, P., Vieitez, M. O., Tilstra, L. G., Tuinder, O. N. E., Koning, C. E., and Veefkind, J. P.: Evaluation of the operational Aerosol Layer Height retrieval algorithm for Sentinel-5 Precursor: application to O_2 A band observations from GOME-2A, Atmospheric Measurement Techniques, 8, 4947–4977, doi:10.5194/amt-8-4947-2015, 2015.

Sarna, K. and Russchenberg, H. W. J.: Monitoring aerosol–cloud interactions at the CESAR Observatory in the Netherlands, Atmospheric Measurement Techniques, 10, 1987–1997, doi:10.5194/amt-10-1987-2017, 2017.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, Earth System Science Data, 8, 697–751, doi:10.5194/essd-8-697-2016, 2016.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., and Jeong, M.-J.: Validation and uncertainty estimates for MODIS Collection 6 âĂIJDeep BlueâĂİ aerosol data, Journal of Geophysical Research: Atmospheres, 118, 7864–7872, doi:10.1002/jgrd.50600, 2013.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D.: First operational BRDF, albedo nadir reflectance products from {MODIS}, Remote Sensing of Environment, 83, 135 – 148, doi:http://dx.doi.org/10.1016/S0034-4257(02)00091-3, 2002.

Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., Rückstieß, T., and Schmidhuber, J.: PyBrain, Journal of Machine Learning Research, 11, 746–746, 2010.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macatangay, R., and Warneke, T.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite Part 1: Carbon dioxide, Atmospheric Chemistry and Physics, 8, 3827–3853, doi:10.5194/acp-8-3827-2008, 2008.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Bergamaschi, P., and Peters, W.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite Part 2: Methane, Atmospheric Chemistry and Physics, 9, 443–465, doi:10.5194/acp-9-443-2009, 2009.

Seinfeld, J.: Atmospheric chemistry and physics of air pollution, A Wiley-Interscience publication, Wiley, 1986.

Shaiganfar, R., Beirle, S., Sharma, M., Chauhan, A., Singh, R. P., and Wagner, T.: Estimation of NO_x emissions from Delhi using Car MAX-DOAS observations and comparison with OMI satellite data, Atmospheric Chemistry and Physics, 11, 10 871–10 887, doi:10.5194/acp-11-10871-2011, 2011.

Shindell, D., Faluvegi, G., Koch, D., Schmidt, G., Unger, N., and Bauer, S.: Improved attribution of climate forcing to emissions, Science, 326, 716–718, 2009.

Sillman, S., Logan, J., and Wofsy, S.: The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes, Journal of Geophysical Research: Atmospheres, 95, 1837–1851, doi:10.1029/JD095iD02p01837, 1990.

Sneep, M., de Haan, J. F., Stammes, P., Wang, P., Vanbauce, C., Joiner, J., Vasilkov, A. P., and Levelt, P. F.:

Three-way comparison between OMI and PARASOL cloud pressure products, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2007JD008694, 2008.

Sontag, E. D.: Feedback stabilization using two-hidden-layer nets, IEEE Transactions on Neural Networks, 3, 981–990, doi:10.1109/72.165599, 1992.

Spada, F., Krol, M. C., and Stammes, P.: McSCIA: application of the Equivalence Theorem in a Monte Carlo radiative transfer model for spherical shell atmospheres, Atmospheric Chemistry and Physics, 6, 4823–4842, doi:10.5194/acp-6-4823-2006, 2006.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, pp. 229–275, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-540-48546-9_7, 2008.

Stammes, P.: Spectral radiance modelling in the UV-visible range, in Proceedings IRS-2000: Current Problems in Atmospheric Radiation, Studies in geophysical optics and remote sensing, A deepak publishing, 101 Research drive, PO BOX 7390, Hampton, VA 23666 USA, 2001.

Stammes, P., de Haan, J., and Hovenier, J.: The polarized internal radiation field of a planetary atmosphere, Astronomy and Astrophysics, 225, 239–259, 1989.

Stammes, P., Sneep, M., de Haan, J., Veefkind, J., Wang, P., and Levelt, P.: Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2007JD008820, 2008.

Streets, D., Canty, T., Carmichael, G., de Foy, B., Dickerson, R., Duncan, B., Edwards, D., Haynes, J., Henze, D., Houyoux, M., Jacob, D., Krotkov, N., Lamsal, L., Liu, Y., Lu, Z., Martin, R., Pfister, G., Pinder, R., Salawitch, R., and Wecht, K.: Emissions estimation from satellite retrievals: A review of current capability, Atmospheric Environment, 77, 1011–1042, doi:10.1016/j.atmosenv.2013.05.051, 2013.

Sunyer, J., Spix, C., Quenel, P., Ponce-de Leon, A., Ponka, A., Barumandzadeh, T., Touloumi, G., Bacharova, L., Wojtyniak, B., Vonk, J., Bisanti, L., Schwartz, J., and Katsouyanni, K.: Urban air pollution and emergency admissions for asthma in four European cities: the APHEA Project, BMJ Group, 52, 1997.

Teichman, A. and Thrun, S.: Tracking-based semi-supervised learning, The International Journal of Robotics Research, 31, 804–818, doi:10.1177/0278364912442751, 2012.

Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J.: Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis, Journal of Geophysical Research: Atmospheres, 103, 17 099–17 110, doi:10.1029/98JD00900, 1998.

Torres, O., Bhartia, P. K., Herman, J. R., Sinyuk, A., Ginoux, P., and Holben, B.: A Long-Term Record of Aerosol Optical Depth from TOMS Observations and Comparison to AERONET Measurements., Journal of Atmospheric Sciences, 59, 398–413, doi:10.1175/1520-0469(2002)059<0398:ALTROA>2.0.CO;2, 2002.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview, Journal of Geophysical Research: Atmospheres, 112, n/a–n/a, doi:10.1029/2007JD008809, 2007.

Torres, O., Ahn, C., and Chen, Z.: Improvements to the OMI near-UV aerosol algorithm using A-train CALIOP and AIRS observations, Atmospheric Measurement Techniques, 6, 3257–3270, doi:10.5194/amt-6-3257-2013, 2013.

Twomey, S. A., Piepgrass, M., and Wolfe, T. L.: An assessment of the impact of pollution on global cloud albedo, Tellus B, 36B, 356–366, doi:10.1111/j.1600-0889.1984.tb00254.x, 1984.

UNICEF, U. N. C. F.: Clean the air for children – The impact of air pollution on children, 2016.

Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M., and Emmadi, S.: Operational total and tropospheric NO_2 column retrieval for GOME-2, Atmospheric Measurement Techniques, 4, 1491–1514, doi:10.5194/amt-4-1491-2011, 2011.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: effectiveness of air quality policy for SO_2 and NO_x emissions in China, Atmospheric Chemistry and Physics, 17, 1775–1789, doi:10.5194/acp-17-1775-2017, 2017.

van Noije, T., Le Sager, P., Segers, A., van Velthoven, P., Krol, M., Hazeleger, W., Williams, A., and Chambers, S.: Simulation of tropospheric chemistry and aerosols with the climate model EC-Earth, Geoscientific Model Development, 7, 2435–2475, doi:10.5194/gmd-7-2435-2014, 2014.

Vandaele, A., Hermans, C., Simon, P., Carleer, M., Colin, R., Fally, S., MÃľrienne, M., Jenouvrier, A., and Coquart, B.: Measurements of the NO_2 absorption cross-section from 42 000 cm^{1} to 10 000 cm^{1} (238âĂŞ1000 nm) at 220 K and 294 K, Journal of Quantitative Spectroscopy and Radiative Transfer, 59, 171–184, doi:https://doi.org/10.1016/S0022-4073(97)00168-4, 1998.

Vaughan, M., Winker, D., and Powell, K.: CALIOP Algorithm Theoretical Basis Document - Part 2: Feature Detection and Layer Properties Algorithms, 2005.

Veefkind, J. P., Boersma, K. F., Wang, J., Kurosu, T. P., Krotkov, N., Chance, K., and Levelt, P. F.: Global satellite analysis of the relation between aerosols and short-lived trace gases, Atmospheric Chemistry and Physics, 11, 1255–1267, doi:10.5194/acp-11-1255-2011, 2011.

Veefkind, J. P., Aben, I., McMullan, K., FÃűrster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote sensing of environment., 120, 70âĂŞ83, doi:10.1016/j.rse.2011.09.027, 2012.

Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements of the OMI O_2-O_2 Operational Cloud Algorithm and Comparisons with Ground-Based Radar-Lidar Observations, Atmospheric Measurement Techniques Discussions, 2016, 1–28, doi:10.5194/amt-2016-48, 2016.

Veihelmann, B., Levelt, P. F., Stammes, P., and Veefkind, J. P.: Simulation study of the aerosol information content in OMI spectral reflectance measurements, Atmospheric Chemistry and Physics, 7, 3115–3127, doi:10.5194/acp-7-3115-2007, 2007.

Vlemmix, T., Piters, A. J. M., Stammes, P., Wang, P., and Levelt, P. F.: Retrieval of tropospheric NO_2 using the MAX-DOAS method combined with relative intensity measurements for aerosol correction, Atmospheric Measurement Techniques, 3, 1287–1305, doi:10.5194/amt-3-1287-2010, 2010.

Vlemmix, T., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Piters, A., Wang, P., Levelt, P., and Van Roozendael, M.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches, Atmospheric Measurement Techniques, 8, 941–963, doi:10.5194/amt-8-941-2015, 2015.

Volten, H., Brinksma, E. J., Berkhout, A. J. C., Hains, J., Bergwerff, J. B., Van der Hoff, G. R., Apituley, A., Dirksen, R. J., Calabretta-Jongen, S., and Swart, D. P. J.: NO_2 lidar profile measurements for satellite interpretation and validation, Journal of Geophysical Research: Atmospheres, 114, n/a–n/a, doi:10.1029/2009JD012441, 2009.

Wagner, T., Dix, B., Friedeburg, C. v., FrieÃ§, U., Sanghavi, S., Sinreich, R., and Platt, U.: MAX-DOAS O_4 measurements: A new technique to derive information on atmospheric aerosolsâĂŤPrinciples and information content, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, doi:10.1029/2004JD004904, 2004.

Wagner, T., Burrows, J. P., Deutschmann, T., Dix, B., von Friedeburg, C., Frieß, U., Hendrick, F., Heue, K.-P., Irie, H., Iwabuchi, H., Kanaya, Y., Keller, J., McLinden, C. A., Oetjen, H., Palazzi, E., Petritoli, A., Platt, U., Postylyakov, O., Pukite, J., Richter, A., van Roozendael, M., Rozanov, A., Rozanov, V., Sinreich, R., Sanghavi, S., and Wittrock, F.: Comparison of box-air-mass-factors and radiances for Multiple-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) geometries calculated from different UV/visible radiative transfer models, Atmospheric Chemistry and Physics, 7, 1809–1833, doi:10.5194/acp-7-1809-2007, 2007.

Wang, P., Tuinder, O. N. E., Tilstra, L. G., de Graaf, M., and Stammes, P.: Interpretation of FRESCO cloud retrievals in case of absorbing aerosol events, Atmospheric Chemistry and Physics, 12, 9057–9077, doi:10.5194/acp-12-9057-2012, 2012.

Wang, Y., Penning de Vries, M., Xie, P. H., Beirle, S., Dörner, S., Remmers, J., Li, A., and Wagner, T.: Cloud and aerosol classification for 2.5 years of MAX-DOAS observations in Wuxi (China) and comparison to independent data sets, Atmospheric Measurement Techniques, 8, 5133–5156, doi:10.5194/amt-8-5133-2015, 2015a.

Wang, Y., Wagner, T., Xie, P., Li, A., Beirle, S., Theys, N., Stavrakou, T., De Smedt, I., , and Koukouli, M.: MAX-DOAS observations and their application to the validation of satellite and model data in Wuxi, China, in: 12th International Workshop on Lidar Multiple Scattering Experiments, doi:10.13140/RG.2.1.2379.6325, 2015b.

Wang, Y., Lampel, J., Xie, P., Beirle, S., Li, A., Wu, D., and Wagner, T.: Ground-based MAX-DOAS observations of tropospheric aerosols, NO_2, SO_2 and HCHO in Wuxi, China, from 2011 to 2014, Atmospheric Chemistry and Physics, 17, 2189–2215, doi:10.5194/acp-17-2189-2017, 2017.

Wayne, R.: Chemistry of atmospheres (3rd Ed.), Oxford University Press, 2000.

Wilcox, E. M.: Direct and semi-direct radiative forcing of smoke aerosols over clouds, Atmospheric Chemistry and Physics, 12, 139–149, doi:10.5194/acp-12-139-2012, 2012.

Winker, D.: Accounting for multiple scattering in retrievals from space lidar, in: 12th International Workshop on Lidar Multiple Scattering Experiments, vol. 5059, pp. 128–139, doi:10.1117/12.512352, 2003.

Winker, D., Vaughan, M., Omar, A., Hu, Y., Powell, K., Liu, Z., Hunt, W., and Young, S.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, Journal of Atmospheric and Oceanic Technology, 26, 2310–2323, doi:10.1175/2009JTECHA1281.1, 2009.

Winker, D. M., Tackett, J. L., Getzewich, B. J., Liu, Z., Vaughan, M. A., and Rogers, R. R.: The global 3-D distribution of tropospheric aerosols as characterized by CALIOP, Atmospheric Chemistry and Physics, 13, 3345–3361, doi:10.5194/acp-13-3345-2013, 2013.

Worden, H., Bowman, K., Worden, J., Eldering, A., and Beer, R.: Satellite measurements of the clear-sky greenhouse effect from tropospheric ozone, Nature Geoscience, 1, doi:10.1038/ngeo1823, 2008.

Worldbank: World Data Bank of the World Bank, 2015.

Wu, Y., Cordero, L., Gross, B., Moshary, F., and Ahmed, S.: Assessment of {CALIPSO} attenuated backscatter and aerosol retrievals with a combined ground-based multi-wavelength lidar and sunphotometer measurement, Atmospheric Environment, 84, 44 – 53, doi:https://doi.org/10.1016/j.atmosenv.2013.11.016, 2014.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON, Atmospheric Measurement Techniques, 10, 2209–2238, doi:10.5194/amt-10-2209-2017, 2017.

Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R. C., and Liu, Y.: Evaluation of VIIRS, GOCI, and MODIS Collection 6 âĂŕAOD retrievals against ground sunphotometer observations over East Asia, Atmospheric Chemistry and Physics, 16, 1255–1269, doi:10.5194/acp-16-1255-2016, 2016.

Xu, X., Wang, J., Wang, Y., Zeng, J., Torres, O., Yang, Y., Marshak, A., Reid, J., and Miller, S.: Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point, Geophysical Research Letters, 44, 7544–7554, doi:10.1002/2017GL073939, 2017.

Xue, Y., Xu, H., Guang, J., Mei, L., Guo, J., Li, C., Mikusauskas, R., and He, X.: Observation of an agricultural biomass burning in central and east China using merged aerosol optical depth data from multiple satellite missions, International Journal of Remote Sensing, 35, 5971–5983, doi:10.1080/2150704X.2014.943321, 2014.

Young, S. and Vaughan, M.: The Retrieval of Profiles of Particulate Extinction from Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) Data: Algorithm Description, Journal of Atmospheric and Oceanic Technology, 26, 1105–1119, doi:10.1175/2008JTECHA1221.1, 2009.

Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., Balkanski, Y., Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N., Reddy, M. S., Schulz, M., Takemura, T., and Zhou, M.: A review of measurement-based assessments of the aerosol direct radiative effect and forcing, Atmospheric Chemistry and Physics, 6, 613–666, doi:10.5194/acp-6-613-2006, 2006.

Zarzycki, C. M. and Bond, T. C.: How much can the vertical distribution of black carbon affect its global direct radiative forcing?, Geophysical Research Letters, 37, doi:10.1029/2010GL044555, 2010.

"Finally, from so little sleeping and so much reading, his brain dried up and he went completely out of his mind. "

Miquel de Cervantes, Don Quixote de La Mancha, 1615

Merci - Dank u wel - Danke schön - Muchas gracias - Gràcies

> "We are not only scientists; we are men, too. We cannot forget our dependence on our fellow men. I mean not only our material dependence, without which no science would be possible, and without which we could not work; I mean also our deep moral dependence, in that the value of science must lie in the world of men, that all our roots lie there. These are the strongest bonds in the world, stronger than those even that bind us to one another, these are the deepest bonds – that bind us to our fellow men"
>
> Prof. J. Robert Oppenheimer, Speech to the Association of Los Alamos Scientists - November 2, 1945

August 2013, I remember the difficulty I had in announcing my departure to Delft to my manager in Toulouse, who I considered, at that time, my mentor since the end of my studies. It was a hard decision after more than 5 years of successful collaborations together. Nevertheless, some of his words as reply were "You are lucky. You are joining a great project, with good people!". Four years and some months later, here I am, writing this ultimate but not easiest chapter of this book. I look back on my time in Delft, in the Netherlands, and I realize how much he was absolutely right!

This thesis research was like a 4-year sailing journey in an offshore area. Leaving France for the Netherlands was my very first move abroad as an expat citizen, to a country where language difficulties (e.g. the gutturals represented by "ch" or "g", ui, ij, ei, and the single-versus-double "a") may sound like a throat disease, "horeca" (hotel - restaurant - cafe) refers to the Dutch catering industry, you feel almost handicapped when finding a flat tyre on your bike in the morning, and the steep and narrow Dutch staircase looks like a trap for foreigners. In such a situation, you know very well the coast that you are leaving, but you don't know yet how to reach the other side of the ocean or what you will find there. You have only some expectations. During the travel, there were of course a lot of sunny moments (actually mostly!). Nevertheless, stormy challenges with gusts close to 6-8 Beaufort brought their difficulties and stress. Luckily, the Dutch beers were never far away and maintained my hopes at a decent level. And sometimes, I had to face some "petole" moments (cf. French sailing dictionary): no wind at all, stuck on the same problem, I could tack, gybe, deploy a spinnaker, and nothing progressed... However, like Francois Gabard

(Page iii) and many other sailors of the Vendee Globe sailing race, I actually enjoyed all these moments as there was always something new to learn, either about myself, or the talented people surrounding me. Now that the PhD cruise is reaching its destination, I can say that this 4-year period was overall fun and fruitful. And this success would have been very different without all of you on this boat....

I am firstly very grateful to Pepijn Veefkind and Pieternel Levelt. Their help with my research during the entire 4 years has been incredibly diverse and invaluable. I also appreciated their support when facing special challenges (not only scientific but also more personal), their constant enthusiasm, and their full transparency and honesty during key discussions. I very much admire their major success stories, OMI and TROPOMI, and how they have built these unique and long-term projects. I do hope their impacts on our society will continue to be more recognized and concretely used in the future. I was also very honoured to be part of the connection between TU Delft and KNMI. I hope we will be able to continue to collaborate together.

I would also like to thank Tim Vlemmix. Tim, I really appreciated how you welcomed me here in Delft and helped to build my position at GRS, and with KNMI. We also shared great discussions about numerous topics these years. You showed unique teaching skills that have been greatly appreciated by the students. Even if we did not finish this research together, I do not forget that you are also part of this success. I wish you all the best for the next years!

I acknowledge all the experts who very quickly accepted to be members of my evaluation committee. Dr. Folkert Boersma, Prof. dr. Johanna Tamminen, Prof. dr. Ilse Aben, Prof. dr. Thomas Wagner, Prof. dr. ir. Herman Russchenberg, and Prof. dr. Ramon Hanssen. Hope you have enjoyed reading this book as well as the forthcoming debates together.

I arrived in Delft with a background mostly acquired with the French NOVELTIS company. I have quite often thought about my former work colleagues in Toulouse, and tried to apply what we learned together. Great thanks to Pascal Prunet, my manager and colleague during all these years, Olivier Lezeaux, Andrzej Klonecki, Sandrine Bijac, Emmanuel Dufour, Emilien Bernard, Catherine Tine, Jerome Donnadille, Sandrine Olivie, Frederique Ponchaut, and Sylvain Ythier. I also keep in mind the fruitful collaborations we had with Claude Camy-Peyret, and Francois-Marie Breon. Hope to keep seeing you again.

The close collaboration with the Royal Netherlands Meteorological Institute (KNMI) was very important for this work. I remember all the warm welcomes every time I came to visit the 2nd floor full of OMI and TROPOMI images! Notably, I would like to thank Piet Stammes for all the enthusiastic discussions and great support he gave me. This helped me a lot to present my research and to prepare my future. I also thank Johan de Haan for his expertise on the DISAMAR software, and Folkert Boersma for the collaborations. Great thanks to my colleagues Jieying, Marina, Maurits and Nanda. And all the other people from the atmospheric satellite observations department met during diverse events (Maarten, Ronald, Martin, Ankie, Quintus, Deborah, Anche, Henk, etc...).

Presenting at the OMI and NASA Aura conferences was not only important for divulging our findings, but also extending our network. With that respect, I have met people who greatly encouraged my works. I notably think about Johanna Joiner, Bryan Duncan, Michel van Roozendael, and Alexander Vasilkov. And my co-authors Vassilis Amiridis, Emmanouil Proestakis, and Mengyao Liu.

Fortunately, there are not only aerosols, NO_2, and OMI in life. These 4 years would not have been the same without all the international people that I have met through our GRS activities, and who became my friends. I am especially grateful to the first people I met and who supported me until the end of this process: my best office mate Rricky (especially thanks for this last year!), the "Laurel and Hardy" of the TU Delft sea level research group, kings of babbels: Marcel and Thomas, Olga, and Lukas. Quickly after, great people joined the GRS adventure! I especially think about Reenu, Susana, Karen, Monica, Veronica, Gert, Antoon, our social media advertiser Jonathan, Steven, Jianpeng, Dimitra, Francesca, Sarah, Kaixuan, Kevin, Mariska, Beatrice. I do not forget Vincent, Edouard, Pedro, Yann, Manuel, Ruben, Máté, and Siavash who in the meantime have pursued new adventures. Our coffee breaks, social outdoor activities, discussions, disagreements, "babbels", gossips, countless drinks and crazy celebrations were truly revitalizing (even if some headaches persisted the mornings after).

The GRS staff members are also part of this work, and deserve to be acknowledged as well. I cannot mention everyone here. However, I would like to stress the fantastic IT support provided by Erwin de Beus, and the great help from our beloved secretaries Debbie, Lidwien, Suzanne and Rebeca. And all the scientists and researchers who directly or indirectly brought their "two cents" to this project or even myself.

I had the opportunity to contribute to some teaching activities. I have excellent memories of the 4 "Vesc" fieldcampaigns in France. I am grateful to some of the 400 Bsc students that I supervised, and the good teaching collaborations with Jan-Kees, Roderik, Joep, Hemmo, David, Shelley, Koen, Richard, and Rémi. I also acknowledge the warmly welcome from the hosts in France during these campaigns: Tania and Jacquot, Veronique and Alain.

I could also supervise two very competent and motivated students on different projects: Eva Stierman on "Ocean Colour monitoring from Sentinel-3", and Gijs van Ouwerkerk on "Aerosol observation comparisons between CALIOP and MODIS over biomass burning events". Our collaborations were pleasant, and their outcome influenced some conclusions of this work.

I am also grateful to my French (or French speaker) Delft friends Remi (thanks again for the cover!), Emilie, Boris, Dana, Pierre-Olivier, Flo, Edouardo and Baptiste: I am sure my "satellite and radiative transfer" research will remain as mysterious to you as your "geology, licking rocks and fractures" is for me. But that has never stopped

us for all our lively coffee breaks and barbecue dinners. The swimming activities in Den Haag with "My Personal Swim Coach" helped me to deal with the research pressure, take care of my body and improve my swimming skills. I thank Fabian, Prue, Christina, Mandy, and Armando for their instructions and patience!

My sailing hobby has taught me a lot of useful skills for performing scientific activities. It notably illustrates very well how much team spirit is essential. It also provides nice tips to deal with unexpected events and relaxing moments to focus on ourselves and make wise decisions. I would like to thank all the people that I met during this research period: Cyril, Audrey, Ari, Gabriel, Yuanjing, and Laura. And my last captains who encouraged me for the skipper training: Jamie, Jacek, Guillem, and Armel.

My close friends that I met in Toulouse have always had a special place in my heart. It was really not easy to leave all of you who are still in France: Choune, Clem, Carole, Mag, Thom, Paulette, Simon. I remember all the encouragements received by those who had already started their life as expats before me: Vince and Ju. Despite the distance, we have managed to keep in very good contact and share quite amazing adventures (e.g. Norway hiking, parachute jumping, road tripping in the USA, visits in Brittany, Clermond-Ferrand, Paris, the Netherlands, sailing Cherbourg-Alderney)! I am looking forwards to the next ones!

A special thank to my paranymphs Marcel Kleinhereinbrik and Vincent Vionnet, and my Master of Ceremony Ricardo Reinoso Rondinel who agreed to assist me during the ultimate process of this work!

And thanks to you, for reading this book and my work.

And finally, I address my special thoughts and love to my family: my parents, and my sister Marie-Charlotte, of whose achievements and strengths I am very proud. I am looking forward to your future!

For this book, and actually even more: Merci - Dank u wel - Danke schön - Muchas gracias - Gràcies!!!

Julien Chimot
June 9, 2018
Frankfurt am Main

(Many thanks to Karen Simon for the review.)

List of Publications Related to this Research

Peer-reviewed papers

Chimot, J., Veefkind, J. P., de Haan, J. F., Stammes, P., and Levelt, P. F.: Minimizing aerosol effects on the OMI tropospheric NO_2 retrieval – An improved use of the 477 nm O_2-O_2 band, to be submitted in Atmos. Meas. Tech. Disc., July 2018.

Chimot, J., Veefkind, J. P., Vlemmix, T., and Levelt, P. F.: Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations, Atmos. Meas. Tech., 11, 2257-2277, doi:10.5194/amt-11-2257-2018, 2018.

Liu, M., Lin, J., Boersma, K. F., Pinardi, G., Wang, Y., **Chimot, J.**, Wagner, T., Xie, P., Eskes, H., Van Roozendael, M., Hendrick, F., Wang, P., and Yan, Y.: Improved aerosol correction for OMI tropospheric NO_2 retrieval over East Asia: constraint from CALIOP aerosol vertical profile, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-34, in review, 2018.

Proestakis, E., Amiridis, V., Marinou, E., Georgoulias, A. K., Solomos, S., Kazadzis, S., **Chimot, J.**, Che, H., Alexandri, G., Binietoglou, I., Daskalopoulou, V., Kourtidis, K. A., de Leeuw, G., and van der A, R. J.: Nine-year spatial and temporal evolution of desert dust aerosols over South and East Asia as revealed by CALIOP, Atmos. Chem. Phys., 18, 1337-1362, doi:10.5194/acp-18-1337-2018, 2018.

Chimot, J., Veefkind, J. P., Vlemmix, T., de Haan, J. F., Amiridis, V., Proestakis, E., Marinou, E., and Levelt, P. F.: An exploratory study on the aerosol height retrieval from OMI measurements of the 477 nm O_2-O_2 spectral band using a neural network approach, Atmos. Meas. Tech., 10, 783-809, doi:10.5194/amt-10-783-2017, 2017.

Chimot, J., Vlemmix, T., Veefkind, J. P., de Haan, J. F., and Levelt, P. F.: Impact of aerosols on the OMI tropospheric NO_2 retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free scenes via a simple cloud model?, Atmos. Meas. Tech., 9, 359-382, doi:10.5194/amt-9-359-2016, 2016.

Conferences, Seminars, Workshops

Liu M., Lin J., Boersma F., Pinardi G., Wang Y., **Chimot J.**, Wagner T., Xie P., Eskes H., van Roozendael M., Hendrick F., Long-term influence of aerosols on tropospheric NO2 retrieval over China based on OMI and TROPOMI sensors, 15th Annual Meeting Asia Oceania Geosciences Society (AOGS), Honolulu, Hawai, 2018.06.07.

Chimot J., Atmospheric composition remote sensing from OMI and Sentinel-3, Talk invitation by Prof. dr. Ilse Aben, SRON Netherlands Institute for Space Research, Utrecht, the Netherlands, 2018.05.23.

Chimot J., Veefkind P., van Ouwerkerk G., Vlemmix T., de Haan J., Levelt P., Aerosol layer height retrieval from OMI and neural network- Possibility for a 13-year time series?, European Geosciences Union (EGU) 2018, Vienna, Austria, 2018.04.09.

Liu M., Lin J.-T., Boersma K.F., Eskes H., Pinardi G., Wang Y., **Chimot J.**, Explicit and Observation-based Aerosol Treatment in Tropospheric Retrieval of NO2 over East Asia, RDSW-seminar, KNMI, De Bilt, The Netherlands, 2018.03.29.

Chimot J., Veefkind P., van Ouwerkerk G., Vlemmix T., Levelt P., Aerosol layer height retrieval from OMI and neural network- Possibility for a 13-year time series?, 20th OMI Science Team Meeting, NASA Goddard, Greenbelt MD, Virginia, USA, 2017.09.12.

Stierman, E., Lindenbergh, R., **Chimot J.**, Veefkind, P., Levelt, P., Oyen, A., Hartogs, M., Kobbe, J., Knotters, A., Blauw, A., Eleveld, M., Water quality from OLCI for Dutch coastal waters, Highroc science conference, September 2017.

Chimot J., Veefkind P., Vlemmix T., Levelt P., Aerosol layer height retrieval from satellite visible measurements – Application to OMI 477 nm O_2-O_2 spectral band, based on neural networks, European Geosciences Union (EGU) 2017, Vienna, Austria, 2017.04.26.

Chimot J., Veefkind P., Vlemmix T., de Haan J., Levelt P., Aerosol layer height retrieval from OMI+MODIS – A neural network approach, 3rd NASA A-Train symposium, Pasadena, California, USA, 2017.04.20.

Chimot J., How to fight against big (and complex) data set? Neural network vs. conventional methods – An example with atmospheric satellite observations, Second Post-Gradual Teaching Workshop of the Buys Ballot Research School, KNMI, de Bilt, the Netherlands, 2017.03.17.

Chimot J., Veefkind P., Vlemmix T., Levelt P., Where are the aerosols?, Geoscience and Remote Sensing (GRS) – CiTG seminar, TU Delft, the Netherlands, 2017.03.07.

Chimot J., Veefkind P., Vlemmix T., de Haan J., Levelt P., Retrieving aerosol height from space: a neural network approach applied to OMI O_2-O_2 measurements, Buys Ballot research school (BBOS) Autumn Symposium, Oranjewoud, the Netherlands, 2016.10.27.

Chimot J., Veefkind P., Vlemmix T., Levelt P., Aerosol retrievals from the OMI 477 nm O_2-O_2 band – Polluted and cloud-free scenes over North-East Asia – Aerosol height: a new OMI(-MODIS-CALIPSO) product?, NASA EOS Aura Meting, Rotterdam, the Netherlands, 2016.08.31.

Chimot J., Veefkind P., Vlemmix T., Levelt P., Aerosol retrievals from OMI using the 477 nm O_2-O_2 band, O_2–A remote sensing workshop, KNMI, De Bilt, the Netherlands, 2016.07.

Chimot J., Veefkind P., Vlemmix T., Levelt P., OMI tropospheric NO_2, aerosol and cloud retrievals over industrialized regions and cloud-free scenes, Interpretation, evaluation and mitigation, ESA Living Planet, Prague, 2016.05.10.

Chimot J., Vlemmix T., Veefkind P., Levelt P., Impact of aerosols on the OMI tropospheric NO_2 retrievals over industrialized regions – A synthetic approach to understand the crucial role of the OMI O_2-O_2 cloud algorithm, OMI science team meeting nr. 19, KNMI, De Bilt, the Netherlands, 2015.09.02.

Chimot J., Vlemmix T., Veefkind P., Levelt P., Aerosol effects on the OMI tropospheric NO_2 retrievals over industrialized regions: how good is the aerosol correction of cloud-free scenes via a simple cloud model?, ATMOS 2015 conference, Scientific Exploitation of operational missions (SEOM), ESA, Heraklion, Crete, 2015.06.08.

Research proposal

NWO – ALW Open Program, Aerosol vertical distribution – Impacts on climate, atmospheric radiation and satellite remote sensing, with **Julien Chimot**, Prof. dr. Pieternel F. Levelt, Dr. J. Pepijn Veefkind, Dr. Martin de Graaf, and Prof. dr. Sebastian Thrun, TU Delft, KNMI, and Stanford University, 2017.

About Julien

July 16, 1985	Born in Agen, Lot et Garonne, France
2013	Competent Sailor Crew Les Glenans (France)
2016	Short range certificate VHF Agency of frequencies, France
2017	Day skipper qualification Royal Yach Association (RYA), UK
Homepage	https://julien-chimot-research.blog
Twitter	@JChimotScience

Julien Chimot is a remote sensing scientist with specific interest in atmospheric composition (aerosol, trace gas, cloud) and surface satellite observations for air quality and climate purposes.

He received his Master specialized in Atmosphere, Ocean, and Environment monitoring in 2008 at the Paul Sabatier University in Toulouse, France. From 2008 to 2013, as Research and Development engineer and project manager for the NOVELTIS company, in France, he was mostly involved in a variety of projects fund by space agencies (e.g. CNES, ESA, EUMETSAT) related to atmospheric infrared satellite missions (IASI, MTG-IRS, GOSAT, Sentinel-5-UVNS, MicroCarb and CarbonSat), with a focus on trace gas (e.g. CO_2, CH_4 and H_2O) and temperature retrievals. He also contributed to some activities covering altimeter, and ocean colour topics.

From December 2013 to January 2018, he was based at the Geoscience and Remote Sensing department of Delft University of Technology, in the Netherlands. In close collaboration with the Royal Netherlands Meteorological Institute (KNMI), he worked on aerosol, tropospheric NO_2, and cloud retrievals from OMI, in combination with MODIS and CALIOP instruments. His research has led to a machine learning algorithm for determining aerosol height from the OMI O_2-O_2 visible band and an improved correction of aerosol effects in the OMI tropospheric NO_2 dataset.

Since February 2018, he leads the developments of the atmospheric composition products from the Copernicus Sentinel-3 mission at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT, Darmstadt, Germany).

During his career, Julien Chimot has (co-)authored 8 publications in international and peer-reviewed scientific journals and 3 European final project reports. He has given 15 oral and poster presentations at international conferences and final European project meetings.