

Exploring TROPOMI NO₂ observations over low clouds and fog

Understanding
the impact of the air-mass factor
on cloudy scenes over the North Sea

Nathan Van der Borgh

MSc Thesis, Applied Earth Sciences

Exploring
TROPOMI NO₂
observations
over low clouds and fog

Understanding
the impact of the air-mass factor
on cloudy scenes over the North Sea

by

Nathan Van der Borgh

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 11, 2024 at 14:00 AM.

Student number: 5075815
Project duration: February 12, 2024 – June 28, 2024
Thesis committee: Dr. J. P. Veefkind, TU Delft & KNMI, supervisor
Dr. Ir. J. Timmermans, TU Delft
Dr. J. de Laat, KNMI, daily supervisor

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.
Cover: NPP-VIIRS image taken on 14/02/2023.

Preface

Before you lies the master's thesis "Exploring TROPOMI NO_2 observations over low clouds and fog". This marks the formal end of my education. This started as a general fascination by the natural and human processes around us, which led me to pursue a bachelor in Geography. At the end of this 3-year journey, I was twisted whether to go further down the human geography road, being fascinated by urbanism and the way in which we organise the space around us, or down the physical geography road, with a strong awe for the science and technology which allows us to better understand the most complex system of all, Earth. As you might notice from the title, I chose the latter. Quickly after starting my Master at TU Delft in Applied Earth Sciences, I decided that I would focus on the atmosphere, rather than the subsurface. Everyone knows the atmosphere is the most fun of the Earth's subsystems. Why? Because it is constantly interacting with other subsystems in a highly dynamic and unpredictable way. When the opportunity to work with TROPOMI arised, I immediately grasped it. It provided me with the opportunity to better understand an omnipresent issue, nitrogen. I greatly enjoyed immersing myself in the world of TROPOMI, and I want to contribute this thesis to all of the researchers who, through a collective effort, have made this mission a succes, and something to be proud of. I feel honored to have been a small part of this innovative project.

This thesis was written in the light of a research internship at the Royal Meteorological Institute of the Netherlands (KNMI). It was a privilege to have been a small part of this community of passionate scientists over the past few months. The experience only further established my respect for the practice of science, and the role KNMI plays in leveraging this. I am grateful to have met inspiring people, with a drive to constantly find out more about this world. This experience has certainly motivated me to seek out a similar sense of purpose in my own future.

I would like to explicitly thank Jos, for the day-to-day guidance. The engaging conversations and critical way of thinking, really pushed me to explore, think independently and never simply assume something. I would also like to thank Pepijn, for the expert insights and steering towards the right direction. And also a thanks to Joris, who had an interesting external perspective, which really helped me to zoom out a bit as well and understand things in a broader way. I am also grateful for your efforts in checking how the process was going and ensuring that I was not overwhelmed by the workload. And I would also like to thank some people at KNMI for all the help. Thank you Henk, for explaining me everything TM5-MP and CAMS related, Jos van Geffen for the help with the spectral testing and Benjamin for allowing me to work with your "processor-code". Lastly, I am very grateful for the entire RDSW team for always making me feel welcome and a part of the team!

And finally, a heartfelt thanks to my family and friends. Thanks to Thibaud and Zoé, for being a great mental support during this process, and providing me with much-needed distraction at times. Also thanks to my parents, for listening and always being interested in what I was doing.

Nathan Van der Borcht
De Bilt, June 2024

Abstract

This thesis investigates TROPOMI NO_2 observations over low clouds and fog. Such scenes present unique opportunities due to the increased reflectivity of clouds in the lower troposphere, which enhances sensor sensitivity. However, in order to accurately estimate vertical NO_2 density in these scenarios, a better understanding of the retrieval process in such conditions is imperative. Two NO_2 retrieval case studies, both exhibiting a spurious retrieval in the presence of low clouds and fog above the North Sea, are examined: February 14th, 2023, and April 9th, 2023. The main focus was on the influence of the air-mass factor on the retrieval. This factor converts measured slant column data along the satellite path into vertical tropospheric columns. Spurious spatial patterns are strongly amplified after the application of the air-mass factor. Key dependencies influencing the conversion from slant to vertical column are identified, being the the a-priori NO_2 profile, air mass factor calculation, and cloud characterization. Notably, the low spatial resolution of the a-priori model and its pronounced peak near the surface contribute to inflated NO_2 values over clouds. The majority of the a-priori NO_2 profile is simulated underneath the cloud field, leading the retrieval algorithm to assume it did not capture a significant portion of the concentration, thus strongly increasing retrieved NO_2 densities over clouds. Furthermore, spatial reflectivity patterns in the North Sea are not always accounted for in the surface reflectivity climatology. This results in misinterpreted clouds with a low cloud fraction. Misinterpreted clouds over open sea areas, combined with low cloud pressures, result in downward corrections over clear-sky seas. These effects culminate in an exacerbated cloudy-clear contrast in NO_2 densities, aligning with the outlines of low cloud fields. Additionally, the heterogeneous cloud field challenges the assumption of a fixed-height, fixed-albedo Lambertian reflector. This inflation effect is modulated by the cloud height. The uncertain height of the Lambertian cloud in the lowest troposphere, combined with an assumed NO_2 profile at the same height, dominates the air-mass factor calculation. Four recommendations to enhance NO_2 retrieval accuracy in the presence of low clouds and fog over the North Sea are presented, which include a Cloud-as-Layers characterization, high-resolution simulations and in-situ measurements of vertical NO_2 profiles in the presence of low clouds, an a-priori correction of the NO_2 profile in the presence of low clouds and implementing auxiliary data relating to spatial patterns in sea-colour of the North Sea.

Contents

1	Introduction	1
2	Methodology	5
2.1	Retrieval algorithm	5
2.2	Data	6
2.2.1	Initial identification case studies	6
2.2.2	Data retrieval	7
2.2.3	Quality flag and error calculation	8
2.2.4	Wind fields	9
2.3	Study area	9
2.4	Models	10
2.5	Set-up vertical profiles	11
2.6	Air-mass factor calculation	13
2.7	Intersection gradient	17
2.8	Surface albedo	19
3	Results	21
3.1	Selected case studies	21
3.2	NO ₂ field	23
3.3	Cloud field	28
3.4	A-priori	29
3.4.1	TM5-MP	29
3.4.2	CAMS-Europe	32
3.4.3	Vertical profiles	34
3.4.4	CAMS-Europe as a-priori	36
3.5	Air-mass factor calculation	38
3.5.1	Step-wise analysis	38
3.5.2	Alternative air-mass factors	41
3.5.3	Contribution analysis	42
3.6	Intersection gradient	44
3.6.1	Introducing a cloud pressure bias	44
3.6.2	Pure clear-sky air-mass factor	47
3.6.3	Spatial analysis of introduced changes	49
3.7	Reflectance test	51
4	Discussion	53
4.1	A-priori	53
4.1.1	TM5-MP	53
4.1.2	CAMS-Europe	54
4.2	Air-mass factor calculation	56
4.3	Cloud characterization	57
4.4	Synthesis	59
5	Conclusion	61
6	Recommendations	63
A	Appendix A: Retrieval Algorithm	65
A.1	General overview	65
A.2	Box-AMF	67
A.3	A-priori	68
A.4	Temperature correction	69

A.5	Clouds	69
A.6	Ghost column	71
A.7	Averaging kernels	71
A.8	Surface albedo	72
B	Appendix B: Methodology	75
B.1	Proof air-mass factor derivation	76
B.2	Temperature correction term	77
C	Appendix C: Results	79

Introduction

This thesis focuses on nitrogen dioxide (NO_2), an atmospheric trace gas which is released in the atmosphere primarily through high temperature burning of fossil fuels. Depending on atmospheric conditions, NO_2 has a lifetime of hours to days. Its primary sources are located near or at the surface and its short lifetime means NO_2 is concentrated within the boundary layer. The spatial distribution of NO_2 is for the same reasons spatially inhomogeneous [1]. Satellite measurements clearly show that high NO_2 tropospheric vertical column densities are present in urban aggregations with dense populations and near large industrial facilities (figure 1.1) [2].

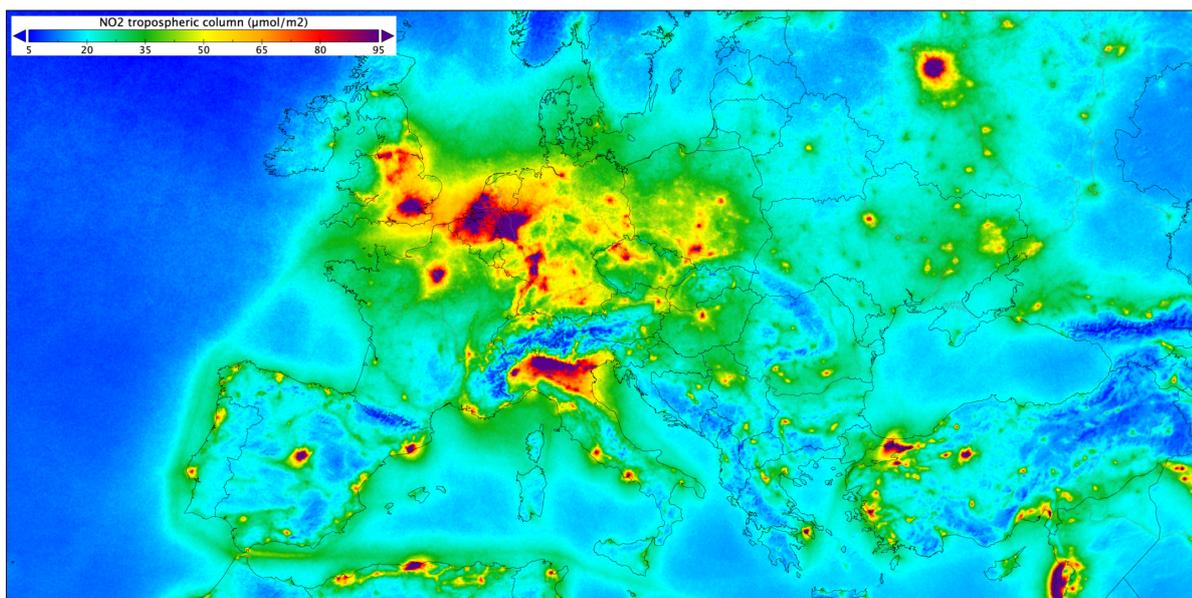


Figure 1.1: Average NO_2 concentration based on measurements gathered by the Copernicus Sentinel-5P mission (TROPOMI) between April 2018 and March 2019 [3].

Under the influence of sunlight, NO_2 dissociates into an NO and an O -atom. This oxygen atom then reacts with O_2 in the troposphere and creates tropospheric ozone (O_3), which is harmful both to human health and the environment, and is a key component of haze and photochemical smog [2]. NO_2 can also lead to the formation of particular matter which is also harmful to human health. Nitric acid from NO_x and from NH_3 are important contributors to the acidification of soil and water [4]. The deposition of nitrogen on the surface also causes eutrophication. The chemical balance of ecosystems is significantly disrupted through these two processes [4]. The effect of NO_2 on global climate is complex, in most cases a net cooling effect is presumed driven by oxidation-fueled aerosols [5]. Nitrogen oxides reduce OH concentrations - via formation of O_3 [6]. Reduced OH concentrations subsequently lead to

an enhanced residence time of greenhouse gases and pollutants [6].

The TROPospheric Monitoring Instrument mission, also known as TROPOMI, is the instrument on board the Sentinel-5 Precursor mission, launched in October 2017 as a part of the European Copernicus Earth Observation programme. The Sentinel-5 Precursor mission is a single-payload mission, the multi-spectral spectrometer aboard observes crucial atmospheric constituents such as ozone, nitrogen dioxide, carbon monoxide, sulfur dioxide, methane, formaldehyde, aerosols, and clouds [1]. TROPOMI measures reflected solar radiation and operates in the UV-VIS-NIR-SWIR frequencies. NO_2 has a number of useful absorption lines in the visible light spectrum 405-465 nm (figure 1.2). The observations are made in the early afternoon with an equator crossing of approximately 13:30 local solar time. The spatial resolution of TROPOMI at nadir is $5.5 \times 3.5 \text{ km}$ since 2019. The swath is 2600 km wide. The TROPOMI mission plays a pivotal role in monitoring trace gases globally, contributing to air quality monitoring, climate research, and ozone layer protection, and aligns with future initiatives for enhanced temporal resolution in European observations [1]. TROPOMI observations contribute to regional policies by monitoring concentrations of trace gases and by determining sources and sinks of atmospheric pollutants, thanks to its high spatial resolution. TROPOMI data is also used as input for air quality models that are further used to issue alerts in case of poor air quality forecasts [7].

In comparison with predecesing satellites from the past 20+ years - like GOME, SCIAMACHY and OMI - TROPOMI is able to observe NO_2 amounts with a higher spatial resolution than ever before. Some key publications are *Veefkind et al. (2012)* [1] describing the mission objectives, *van Geffen et al. (2020)* [5] describing the DOAS slant column retrieval, *van Geffen et al (2022)* [8] describing the tropospheric vertical column retrieval and *Riess et al. (2022)* [9], which focuses on recent progress in the cloud pressure product and its effects on NO_2 retrieval. Milestone applications of TROPOMI include *Goldberg et al. (2019)* [10], which demonstrates the use of TROPOMI to estimate NO_x emissions from North American cities and power plants. Similarly, *Miyazaki et al. (2020)* [11] and *Liu et al. (2020)* [12] highlight TROPOMI's potential in detecting a significant reduction in NO_x emissions due to the pandemic-induced lockdown in China.

A major source of error is cloudiness [1]. Because of the preference to observe NO_2 down to the surface generally NO_2 data is only used for a certain cloud fraction. Observations with a cloud fraction above this defined threshold, the pixel is flagged as not qualitative enough. The aim of this thesis is to explore the observations over these *too* cloudy pixels. A better understanding of the current difficulties, can then contribute to the improvement of the retrieval algorithm so it can measure NO_2 in the presence of clouds. TROPOMI derives cloud height from the O_2 A-band in the NIR (760 nm) and cloud fraction from the VIS band, as shown in figure 1.2. For the OMI satellite - the predecessor that laid out the grounds for TROPOMI - clouds were characterized using the $O_2 - O_2$ absorption band around 477 nm (OMI does not observe around 760 nm). By measuring at three different wavelengths around the O_2 A-band, from strong to moderate to no absorption, clouds can be detected [13]. Total tropospheric NO_2 column "cloud free" retrievals with an acceptable error (10%) and variation ($10^{15} \text{ mol cm}^{-2}$) can only be achieved up to a certain cloud fraction, which is theoretically limited at 0.2 [1]. The TROPOMI overpass time in the early afternoon also allows for a comparison with the cloud fraction from the VIIRS instrument on the NPP satellite, which has an overpass time of 13:30 pm.

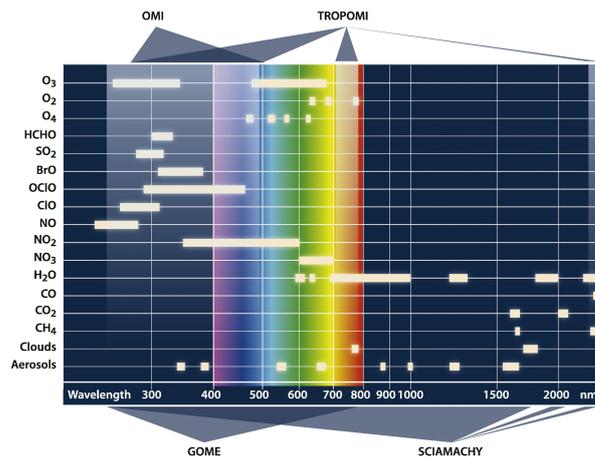


Figure 1.2: Spectral range TROPOMI and its predecessors [1].

Over the last few years the boundaries in using TROPOMI's measurements has been pushed - in no small part due to the success of TROPOMI. By continuously improving the retrieval algorithm and incorporating auxiliary data, the spatial coverage and quality of TROPOMI NO_2 retrievals has constantly progressed. For example, by making use of the frequent overpass time above $70^\circ N$ and combining with an auxiliary data set on lightning, the production of NO_2 related to lightning events in the Arctic was proven [14]. Another example is the phenomenon of sunglint, where sunlight is directly reflected by the ocean surface waves to the satellite. This has been used to infer information that had not been available before. The brighter ocean surface leads to a higher signal-to-noise ratio for TROPOMI NO_2 measurements. In combination with an auxiliary dataset on ship traffic, this type of scenes allowed for the distinction of NO_2 pollution plumes from individual ships in this case [15]. Another example is the use of cloud height information to discriminate clouds from snow covered surfaces. Snow- or ice- covered pixels were previously flagged as not useful due to the reflectance similarities of snow (bright) with clouds (also bright). The updated retrieval algorithm led to an increase of 6-39 % of data, depending on the month [16]. In a case study in Siberia, NO_2 pollution could be detected from individual compressor stations along natural gas pipelines, also because the bright snow surface led to significantly higher sensitivity up to a factor of eight [16]. All of these examples show how challenging scenes, mostly due to their albedo properties, could be transformed into new sources of data thanks to innovative processing and auxiliary data sources.

An obvious question is then whether TROPOMI NO_2 measurements over other bright surfaces could be useful. A logical scenario to first focus on are low clouds and fog. Such clouds are bright and often spatially rather homogeneous, always close to the surface, and thin, which in principle could make interpretation of enhanced NO_2 over low clouds and fog relatively straightforward. A similar approach has been made over a decade ago developed for Carbon Monoxide for which it is now a standard approach [17]. A first exploration of NO_2 fields over low clouds and fog is made in this thesis in support of continued ongoing efforts at KNMI of improving the TROPOMI NO_2 product and expand its use. The aim is to explore how a retrieval algorithm can be developed to measure TROPOMI NO_2 under such conditions. The bright surface of the low clouds and fog offers a possibility a priori not dissimilar from snow covered surfaces. Higher reflectivity leads to a higher signal-to-noise ratio under such conditions. However the situation is complex, because a significant fraction of the NO_2 pollution can also be present underneath and within the clouds - which is for example much less the case for Carbon Monoxide [17]. An important first step is to properly understand and characterize how the operational TROPOMI retrieval algorithm currently handles the presence of clouds and identify any potential constraints. This then could be developed into a workflow on how to extract useful information about NO_2 above low clouds.

Currently, TROPOMI NO_2 fields over low clouds are mostly flagged as not useful enough for user purposes. Peculiar patterns and unrealistic contrasts above low clouds show up in the vertical tropospheric NO_2 column. The retrieval process, explained in chapter 2, is strongly influenced by the calculation of

the so-called air-mass factor and cloud characteristics. The focus of this thesis will be mostly on that step and its implications.

The main research question and its supporting sub-questions are:

How does the air mass factor impact TROPOMI's observations of NO_2 in scenes with low clouds and fog?

- a) Which biases or errors are introduced during air mass factor derivation under these specific atmospheric conditions?
- b) Which assumptions are made in the derivation of an NO_2 column from a reflectance spectrum?
- c) How are local processes influencing the retrieval of NO_2 over low clouds and fog?
- d) Are tropospheric NO_2 columns from TROPOMI more accurate over (low) clouds, and to what extent?
- e) What additional information and insights do tropospheric NO_2 columns over (low) clouds provide compared to cloud-free observations?

The findings of this research can be applied specifically to the Dutch context, where NO_2 is a trace gas of special importance, since it plays a key role in the pressing environmental issue of nitrogen (N) deposition in the environment seriously disrupting soil and water quality [18]. If observations in the presence of low clouds and fog would be usable, this would contribute to enhanced monitoring capabilities of TROPOMI. In July 2021, the Dutch *Stikstofwet* came into action, which aims to significantly reduce the deposition of nitrogen in Dutch nature [19]. This law aims to tackle the nitrogen-issue by reducing emissions from different sectors, like agriculture, industry and traffic. To quantify how certain emission reduction measures are having an effect, TROPOMI NO_2 monitoring plays an important role, as it provides data on a high spatial and temporal scale. A report of the Netherlands Organisation for Applied Scientific Research (TNO) notes the current value of satellite data as a complementary source of information, with the ability to unlock a better understanding of emissions over large areas and how they are transported. They also plea for future missions focused on nitrogen emission on a more detailed scale [20] which will materialize in the ESA CO2M mission (in preparation; 2 km spatial resolution) and the recently approved ESA TANGO mission (sub 500 meter spatial resolution). However, the European action to reduce nitrogen has also resulted in a rise in - sometimes violent - farmers' protests around Europe [21]. One aspect of their frustration is unclear guidelines, which are perceived as not fair and not proportional [21]. This shows how there is need for fair, clear and effective guidelines, which are supported by high-quality data. Data with a high uncertainty and untrustworthy results are not a solid base to build sound policy.

The different data and methods used to assess the effect of the air-mass factor on the retrieval algorithm are described in chapter 2 (Methodology). A concise description of the retrieval algorithm is provided in appendix A (Retrieval Algorithm). The results of the analysis of the case studies is presented in chapter 3 (Results). Chapter 4 (Discussion) discusses the lessons and implications of the results. Key findings are then summarized in chapter 5 (Conclusion). Finally, in chapter 6 (Recommendations), recommendations and ideas for further steps on how to deal with these situations are laid out.

2

Methodology

This chapter describes the methods used to explore TROPOMI NO_2 observations over low clouds and fog. To properly analyse the retrieved NO_2 densities in these scenes, it is essential to understand how the retrieval algorithm approaches these situations. A short summary of the retrieval algorithm is given below, with a complete description available in appendix A. Following this, a concise overview of the selected data and study area is presented. The selection and conversion of the two a-priori models, a global and regional, are described, along with an alternative retrieval performed using the high-resolution regional a-priori model. Then, the set-up of the vertical profiles is explained. This is necessary to allow for a vertical comparison of the averaging kernels (a measure of sensor sensitivity) with the two a-priori's. Apart from the a-priori effect, the air-mass factor calculation process in situations with a low cloud field is also analysed. This analysis includes defining different intermediary factors and performing alternative retrievals using adapted air-mass factors. The mathematical derivation to define different unit-less contributors to the air-mass factor is given. Then, to properly assess the effect of cloud height, an intersection gradient is defined. This intersection explores the clear-cloud gradient and how it is affected by cloud height. Finally, some brief remarks on surface reflectivity are provided.

2.1. Retrieval algorithm

This thesis focuses on the final step of the retrieval algorithm: converting the slant column to the vertical column through the air-mass factor. A comprehensive explanation of the key terms related to the calculation of the air-mass factor, as well as the associated formulas, is given in appendix A. In short, the air-mass factor is calculated for every TROPOMI grid cell, as shown in figure 2.1. The air-mass factor is dependent on a look-up table through the so-called *AMF LUT* (depicted in blue in figure 2.1), which is a measure of vertical sensor sensitivity.

The effect of clouds on sensor sensitivity is parameterized through this *AMF LUT* variable. The cloud pressure affects the vertical sensitivity profile through the measured cloud pressure p_{cloud} . The cloud radiance fraction crf is then used to combine a cloudy and clear-sky air-mass factor. A cloud is represented as a Lambertian reflector with the associated cloud pressure representing the middle of the cloud. More information on the cloud characterization is given in appendix A, section A.5.

The a-priori NO_2 concentration ($[NO_2]_{TM5}$ shown in brown) weighs this look-up table air-mass factor. The relative shape of the profile is what determines the assigned weights, rather than the absolute values. The *AMF LUT* will have very low values underneath a cloud, representing low sensitivity underneath a cloud. These values will obtain high weights through the high a-priori values in the lowest troposphere. This means the total air-mass factor is lower in the presence of clouds, which leads to a more upwards correction of the measured slant column. This is related to the ghost column correction, where the algorithm implicitly accounts for *invisible* NO_2 concentration underneath the cloud.

Additionally, a vertical temperature correction is applied to account for the assumed constant vertical temperature in the Differential Optical Absorption Spectroscopy (DOAS), which is the initial step of the retrieval algorithm. This first step fits measured reflectance spectra with simulated spectra corresponding with a certain NO_2 density.

The slant column is divided by the air-mass factor to obtain the vertical column. This means the air-mass factor has an inverse effect, where a smaller air-mass factor leads to a larger slant-vertical correction and vice versa.

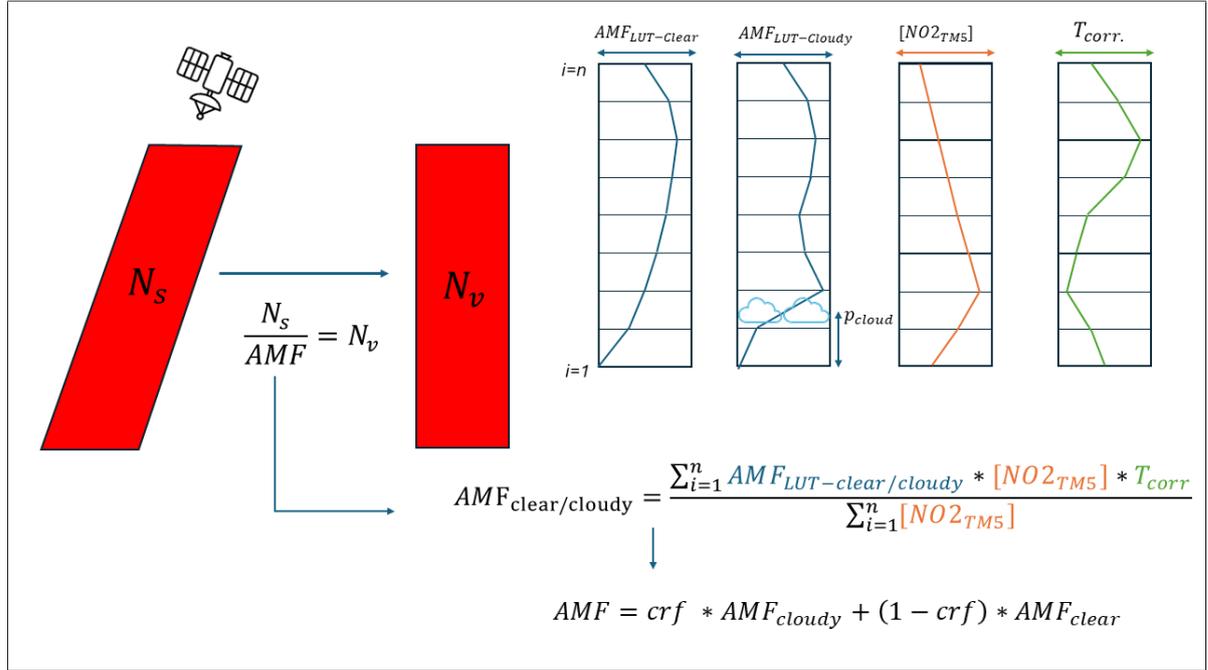


Figure 2.1: Scheme depicting transformation of slant column to vertical column through an air-mass factor. N_s refers to the measured slant column, N_v refers to the vertical column. AMF is the air-mass factor which is calculated through the look-up table derived factors for every layer, $AMF_{LUT\ clear/cloudy}$, both for the clear and cloudy contribution, the a-priori profile $[NO2_{TMS}]$ and the temperature correction T_{corr} . These last two are also defined for every vertical layer. The air-mass factor is calculated through the sum of all the n layers, which depicts the amount of layers in this case. p_{cloud} refers to the FRESKO-CRB cloud pressure. crf refers to the cloud radiance fraction measured in the NO_2 window.

2.2. Data

2.2.1. Initial identification case studies

Initial exploration to identify an adequate amount of cases with a low cloud or fog field and spurious patterns is conducted through a quick visualisation tool: the *S5P Mission Performance Center Level-2 Quality Control Data Portal* [22]. This publicly accessible tool has been developed by KNMI. It allows for fast visualization of multiple data layers, including NPP-VIIRS imagery, FRESKO cloud height, cloud fraction and the tropospheric NO_2 column. Since this tool only had complete data coverage starting from July 2023 at the time of the start of the data exploration, an older visualisation tool is used for the earlier months. This older tool, which is developed in light of the *S5P Mission Performance Center* is not publicly accessible (only through an internal KNMI connection). Unlike the publicly accessible tool, it does not display VIIRS imagery. To still compare with VIIRS imagery the NASA data-view tool *EOSDIS Worldview* is used [23]. This NASA-tool is also used to download the Suomi NPP-VIIRS and Terra-MODIS imagery. The identified potential study cases are all stored in an online folder.

2.2.2. Data retrieval

In order to further analyse the selected study cases, TROPOMI data is locally downloaded in an Anaconda Spyder environment. This is done through the Amazon Web Services S3 Bucket. This commercial storage service contains the Copernicus Data Space Ecosystem collection [24]. Through this tool, TROPOMI Level-2 and Auxiliary files are downloaded. For the two case studies discussed in this thesis, TROPOMI orbits 27667 and 28433 were downloaded. For both cases the area of interest lies within the ascending swath. Three processing streams are defined for every file: *near real-time*, *offline* or *reprocessing* [25]. For this thesis, only the *offline* processing stream is considered. This offline processing contains an assimilation step of measured NO_2 to improve data quality [25]. The *offline* data product usually has a delay of about two weeks compared to the true measurement time (note: near real-time data is available within 3 hours after the measurement). More information on the internal structure of the L2 NO_2 and auxiliary file can be found in their respective Product User Manuals [25] [26]. The variables in these files that were used for the analysis are listed in table 2.1.

In order to obtain a more in-depth understanding of the operational processing taking place between the L1 and L2 product, an extensive Python "step-by-step-algorithm-following" tool developed by Benjamin Leune (KNMI) was retrieved through Gitlab. This tool is only available within the KNMI community. In order to run the tool, additional data is required. The look-up table for AMF calculation, the look-up table for cloud characterization and the DLER-database are all downloaded from internal KNMI servers. A KNMI code to assimilate TM5-MP model profiles to the TROPOMI L2 NO_2 grid is also used. This tool is developed by the KNMI RDSW department. All of the developed code and resulting data, is made available to KNMI. This can also be shared on request to users outside of KNMI. All of the results shown in this thesis are reproducible through the code. The code and data follows the FAIR principles: Findable, Accessible, Interoperable and Reusable. A replication package will also be handed over to KNMI.

To achieve a comparison with a high-resolution chemical transport model, the Copernicus Atmosphere Monitoring Service Europe ensemble forecast model is downloaded. The NO_2 -product of this model is downloaded for the days of the selected study cases. The download is done through the online *climate data store* [27]. A recent KNMI-led research project also processed TROPOMI data using this CAMS Europe ensemble model as a-priori [28]. Complementary to this research, the reprocessed TROPOMI L2 NO_2 product is also publicly available at the *Tropospheric Emission Monitoring Internet Service*, where it was subsequently downloaded [29].

Table 2.1 gives an overview of all the variables of the TROPOMI NO_2 Level-2 (*S5P_OFFL_L2_NO2*) and auxiliary file (*S5P_OPER_AUX_CTMANA*) that are used in the final data analysis. The X and Y dimensions are defined by respectively the *ground_pixel* and the *scanline* variable. The Z dimension is defined by the TM5-MP *levels* variable. The TM5-MP model also stipulates the *time* dimension, which is defined in fractional number of days since the 1st of January, 1950 [26]. It has a size of 48 instances, representing 30-minute steps in the 24-hour day of the measurement, which is 14/02/2023 and 09/04/2023 in this case.

FILE	VARIABLE	UNIT	DIMENSION
<u>S5P_OFFL_L2_NO2</u>	air_mass_factor_troposphere		(X,Y)
	averaging_kernel		(X,Y,Z)
	scanline		Y
	ground_pixel		X
	latitude	degrees_north	(X,Y)
	longitude	degrees_east	(X,Y)
	nitrogen_dioxide_tropospheric_column	mol m-2	(X,Y)
	nitrogen_dioxide_tropospheric_column_precision	mol m-2	(X,Y)
	nitrogen_dioxide_tropospheric_column_precision_kernel	mol m-2	(X,Y)
	tm5_constant_a, hybrid A coefficient at upper and lower interface levels		(Z,2)
	tm5_constant_b, hybrid B coefficient at upper and lower interface levels		(Z,2)
	tm5_tropopause_layer_index		(X,Y)
	air_mass_factor_clear		(X,Y)
	air_mass_factor_cloudy		(X,Y)
	cloud_fraction_crb_nitrogen_dioxide_window		(X,Y)
	cloud_radiance_fraction_nitrogen_dioxide_window		(X,Y)
	nitrogen_dioxide_ghost_column	mol m-2	(X,Y)
	nitrogen_dioxide_slant_column_density	mol m-2	(X,Y)
	nitrogen_dioxide_slant_column_density_precision	mol m-2	(X,Y)
	multiplication_factor_to_convert_to_molecules_per_cm2 (6.022141E19)		(X,Y)
	fresco_cloud_pressure_crb	Pa	(X,Y)
	fresco_cloud_albedo_crb		(X,Y)
	fresco_cloud_fraction_crb		(X,Y)
	fresco_scene_albedo		(X,Y)
	fresco_surface_albedo		(X,Y)
	eastward_wind	m s-1	(X,Y)
	northward_wind	m s-1	(X,Y)
	surface_albedo_nitrogen_dioxide_window		(X,Y)
<u>S5P_OPER_AUX_CTMANA</u>	hyam (hybrid A coefficient at layer midpoints)	Pa	Z
	hybm (hybrid B coefficient at layer midpoints)		Z
	lat	degrees_north	lat
	lon	degrees_east	lon
	no2, volume mixing ratio of NO2 in humid air		(time, Z, lat, lon)
	ps (surface pressure)	Pa	(time, lat, lon)
	t (temperature)	K	(time, Z, lat, lon)
	tropopause_layer_index		(time, lat, lon)

Table 2.1: Table describing the variables downloaded from the NO2 and auxiliary data file.

2.2.3. Quality flag and error calculation

The NO_2 concentrations in the Level-2 (L2) data file are given in $mol m^{-2}$. Through the *multiplication_factor_to_convert_to_molecules_per_cm2* (table 2.1) these values can be converted to $molecules cm^{-2}$. Subsequently, they are also divided by 10^{15} to be visualised in $10^{15} molecules cm^{-2}$. This unit is chosen because the corresponding order of magnitude, which typically runs between 0 and 20 for tropospheric concentrations, is easier to interpret. The associated tropospheric vertical column precision is also retrieved from the Level-2 data file (table 2.1). This precision represents the propagated effect of different errors throughout the retrieval algorithm: the slant column error, the error associated with the separation of the stratospheric and tropospheric signal (see appendix A, section A.1 for more information on this separation) and the error associated with the air-mass factor calculation [30]. Both the precision with and without the kernel setting is retrieved. This kernel setting allows for a distinction of the uncertainty caused by the a-priori profile, also referred to as the smoothing error [30]. The precision *with* kernel then refers to the uncertainty *with* the a-priori effect taken into account and vice versa.

The quality selection flag (qa-value) threshold in all considered TROPOMI data products is 0.5. Pixels with a qa-value above 0.5 are deemed suitable for use in data assimilation and comparison against models [30]. In the case studies considered here this roughly corresponds to values with a radiance weighted cloud fraction in the NO_2 measurement window lower than 0.5 [30]. The effective cloud fraction f_{eff} is the cloud fraction that results in the same top-of-atmosphere reflectance as the real cloud and assumes an optically thick Lambertian cloud with a fixed cloud albedo of 0.8. The *cloud radiance fraction* or *crf* variable is a radiance weighted cloud fraction depending on the effective cloud fraction through formula A.6 [30]. Unless stated differently, the cloud (radiance) fraction used in this thesis is the one measured in the NO_2 window, as opposed to the cloud (radiance) fraction measured in the O_2 A-band for the cloud retrieval.

2.2.4. Wind fields

The *northward_wind* and *eastward_wind* variables are used to calculate the wind direction W_d and magnitude W_m using a vector approach, as shown in Formula 2.1 and 2.2.

$$W_d = \arctan\left(\frac{W_{north}}{W_{east}}\right) \times \frac{180}{\pi} \quad (2.1)$$

$$W_m = \sqrt{W_{north}^2 + W_{east}^2} \quad (2.2)$$

2.3. Study area

The area of interest is the North Sea. The reasons for this region are multiple since a lot of conditions coexist here. Over an ocean surface, the albedo is generally very low. This means that the effect of bright low clouds and fog will be clearly visible. Secondly, this region is close to many strong NO_2 emission regions (see also figure 1.1) whose emissions under favorable conditions can be rapidly transported onto the North Sea. The North Sea also has dense shipping lanes, which causes significant NO_x emissions. Furthermore, the North Sea is frequently covered by low clouds and fog.

For both study cases, an extent of longitude and latitude is defined. Within this spatial mask, all of the calculations and visualisations are done, in order to keep computation and visualisation times manageable. On the other hand, the mask can also not be too restrictive, since a wider area allows for a more holistic insight of local dynamics. The retrieval process has to be re-run a substantial amount of times to calculate alternative data products. For these iterative processor-runs, a smaller swath area inside the area of interest is defined. All maps are shown in the *plate carrée* projection. The spatial masks are defined as shown in figure 2.2:

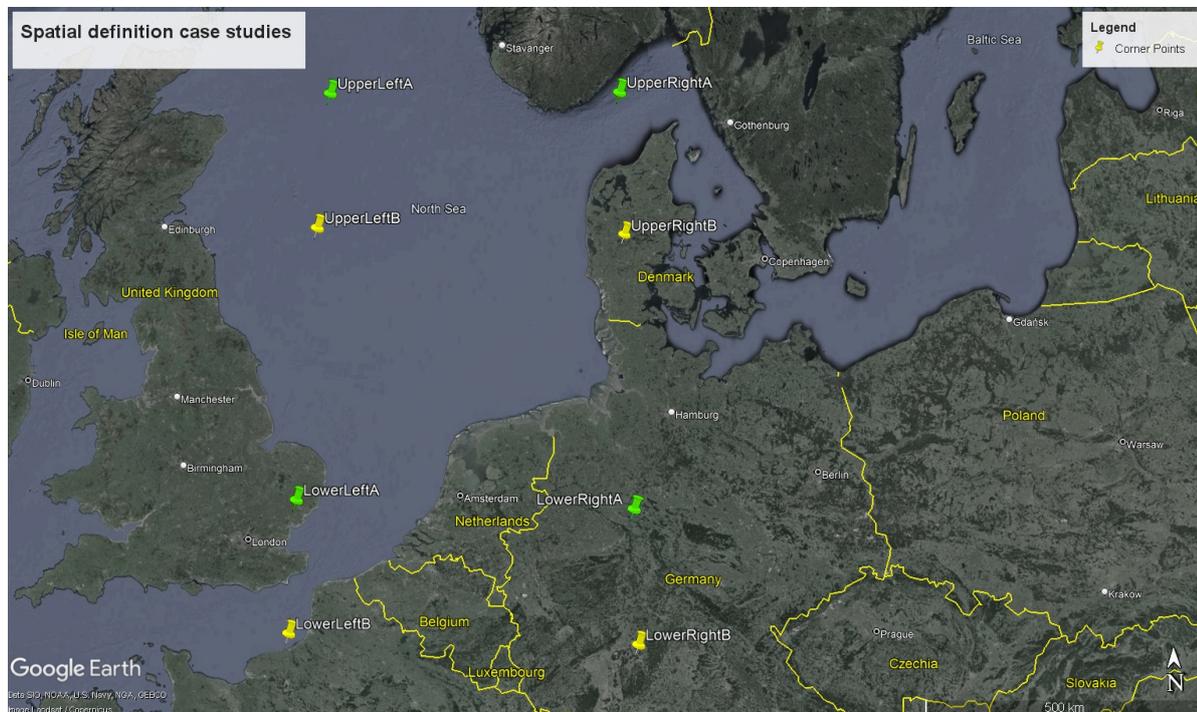


Figure 2.2: Spatial extent of case studies A and B as shown on a map of Northwestern Europe [31]. Bounds mask case A: (52°N, 1°E), (52°N, 9°E), (58°N, 1°E), (58°N, 9°E). Bounds mask case B: (50°N, 1°E), (50°N, 9°E), (56°N, 1°E), (56°N, 9°E)

2.4. Models

The TM5-MP model uses 34 terrain-following dynamic layers with dynamic coefficients A_i^{TM5} and B_i^{TM5} converting these layers to the corresponding pressure levels, as shown in formula 2.3 [26]. The auxiliary data product also contains information on the tropopause height, *tropopause_layer_height* in table 2.1. This index contains the layer $n_{tropopause}$ at which the tropopause is defined. This can then be converted to the corresponding pressure level through formula 2.3. The tropospheric TM5-MP column is subsequently calculated by integrating the partial columns from the surface level until the tropopause level.

$$p_i^{TM5} = A_i^{TM5} + B_i^{TM5} * p_s \quad (2.3)$$

with p_s the surface pressure from the EMWF ERA Interim-5 model, A_i^{TM5} and B_i^{TM5} the dynamic coefficients from TM5-MP.

The Copernicus Europe Atmospheric Monitoring Service (CAMS) Ensemble forecast model is also considered. The entire day of the observation is retrieved from the online data repository, within the bounds of the area of interest (figure 2.2). To gain better insight into the dynamics before and after the overpass time, an animation is made, with each frame corresponding to a two-hour interval of the NO_2 concentrations at different height levels. The 13th forecast hour after midnight is selected to compare with the TROPOMI retrieval (overpass of 13:30 local solar time [30]).

The CAMS Europe product is defined in a longitude and latitude grid. The NO_2 mass concentration is defined in $\mu g/m^3$. The different vertical levels are as follows: 0, 50, 100, 250, 500, 750, 1000, 2000, 3000 and 5000 meters. To compare vertical variability between CAMS and TM5-MP, NO_2 concentrations at specific height levels have to be defined instead of the default layer concentrations in CAMS. This is achieved by integrating the concentration around the middle-points of the different layers. The integration is performed from the middle point of the layer below each height level, to the middle point of the layer above each height level. The level height is defined in cm to stay consistent with the units of other data products. The concentration of CAMS is defined in the volume dimension cm^3 , whereas concentrations are defined at a specific pressure level in cm^2 in TROPOMI and TM5-MP. So the volumetric concentration needs to be converted to an integrated surface concentration. This is achieved by multiplying the volumetric concentration by height for each level, which effectively integrates the concentration over that level. This concept is shown in formula 2.4

$$C_{level} = \int_{MP_{anterior}}^{MP_{superior}} C_{volumetric} dz \quad (2.4)$$

where C_{level} is the NO_2 concentration at a certain pressure level in $\mu g cm^{-2}$, $MP_{superior}$ and $MP_{anterior}$ represents the middle point of respectively the anterior and superior layer relative to the specific pressure level, $C_{volumetric}$ represents the NO_2 mass concentration in $\mu g cm^{-3}$ and dz represents the change in altitude around the level height in cm

This leads to concentrations defined in $\mu g cm^{-2}$. These are then further converted to $10^{15} molecules cm^{-3}$ to match TROPOMI and TM5-MP values. This conversion is shown in formula 2.5.

$$M_{NO_2} = 46.0055 g mol^{-1}$$

$$[NO_2]_{10^{15} molecules cm^{-3}} = [NO_2]_{\mu g cm^{-2}} * N_{Avogadro} * (10^6 * M_{NO_2}) \quad (2.5)$$

A 2023 KNMI-led research analysed the CAMS-Europe NO_2 field and its potential use regarding TROPOMI [28]. In light of this project, the CAMS-Europe product has been rescaled following the averaging kernels, as shown in formula 2.6. This means information on TROPOMI viewing geometry, cloud properties, aerosols and surface albedo is taken into account in the CAMS field, with the goal of making the

CAMS model output and TROPOMI measured NO_2 field [28] one-to-one comparable. Directly comparing the TROPOMI NO_2 vertical column with the model output would introduce additional uncertainties, because the TROPOMI NO_2 field is dependent on both the way the sensor sensitivity is parameterized, i.e. the averaging kernels, and an a-priori. By applying these TROPOMI averaging kernels on the model output and thus formulating a *CAMS model simulation of the measured NO_2 column*, the uncertainties in the comparison are reduced to only the a-priori driven uncertainty. This allows for a comparison of the a-priori with the measured TROPOMI column independent of the averaging kernels used in TROPOMI [28]. This application of the averaging kernels follows from the rationale in formula A.8.

$$C^A = \sum_i^{n_{trop}} A_i^{trop} * x_i^{cams} \quad (2.6)$$

with C^A the CAMS model simulation of the measured NO_2 column, n_{trop} the layer corresponding with the tropopause, A_i^{trop} the tropospheric averaging kernel and x_i^{cams} the CAMS model profile.

The CAMS NO_2 simulation corrected with the TROPOMI averaging kernels can also be understood as shown in formula 2.7. This shows how the correction from the initial model field is dependent on the proportion of the air-mass factor using CAMS as a-priori to the original air-mass factor (using TM5-MP as a-priori). This ratio is a measure of how much the CAMS simulated NO_2 field is changed to take into account the TROPOMI measurement sensitivity.

$$\sum_i^{n_{trop}} A_i^{trop} * x_i^{cams} = \frac{AMF_{CAMSapriori}}{AMF_{TM5apriori}} * \sum_i^{n_{trop}} x_i^{cams} \quad (2.7)$$

with $AMF_{CAMSapriori}(AMF_{TM5apriori})^{-1}$ the ratio between the newly formulated air-mass factor based on the CAMS-Europe a-priori and the original air-mass factor based on the TM5-MP a-priori. This ratio is an indicator of how strong the CAMS-Europe a-priori is shifted to represent the measurement sensitivity. A value of the ratio above 1 means the initial CAMS field is increased to better simulate the slant TROPOMI tropospheric column, if the ratio is lower than 1 the CAMS field is lowered to simulate the slant tropospheric column.

Then, the regional CAMS-Europe model is used as the a-priori profile in the TROPOMI retrieval algorithm. KNMI has already re-processed the TROPOMI data from 2018 on using CAMS-Europe as the a-priori. This so-called European TROPOMI NO_2 product is available through the Tropospheric Emission Monitoring Service [29]. This European TROPOMI NO_2 product is calculated by replacing the TM5-MP profile with the CAMS profile. The CAMS-Europe profile is linearly interpolated to match the height of the averaging kernels, which are defined in the 34 TM5-MP dynamic pressure layers [28]. The exact method of vertical and horizontal interpolation of the CAMS observations to match the TROPOMI data structure is defined in *Douros et al. (2023)* [28].

2.5. Set-up vertical profiles

Vertical profiles are also set up to allow for a vertical comparison of the different a-priori's and their relation with the averaging kernels in the air-mass factor calculation. This profile is defined for two smaller areas, as shown in figure 2.3. These areas are chosen because they contain both a cloudy and clear-sky contribution. They also contain a strong contrast in tropospheric NO_2 concentration. The vertical profiles show several elements: the averaging kernels, the cloud pressures, the TM5-MP profile, and the CAMS vertical profile. The first two elements are defined within the TROPOMI grid (*ground_pixel*

× *scanline*). The KNMI-tool *TM5_profile* allows for a downscaling of the TM5-MP temperature and NO_2 profiles on the TROPOMI grid in the L2 file [26]. This interpolation is based on the temperature and NO_2 profiles in the auxiliary data file. In order to compare the previously mentioned variables with the CAMS profile, the CAMS cells overlapping with the TROPOMI pixels in the area of interest are defined first. Then, the CAMS height levels are converted to pressure levels. This is done through a simple pressure-height relationship, which follows from the hydrostatic balance equation (2.8). This relationship neglects small-scale acceleration, which is a reasonable assumption in for the application and scale considered here. If one wants to very accurately solve small-scale systems like a local convection system or a thunderstorm, this assumption does not hold [32]. Another assumption made is the ideal gas law (formula 2.9), which is also considered to be acceptable in this case. The derivation which starts from the hydrostatic balance, formula 2.8 and leads to a pressure-height relation, formula 2.10, is shown below.

Starting from the hydrostatic balance equation:

$$\frac{\delta p}{\delta z} = -\rho g \quad (2.8)$$

Then ρ is replaced through the ideal gas law:

$$\rho = \frac{p}{RT} \quad (2.9)$$

$$\begin{aligned} \frac{\delta p}{\delta z} &= -g \frac{p}{RT} \\ p(z) &= p_s e^{\frac{-z_i g}{RT_i}} \end{aligned} \quad (2.10)$$

with $p(z)$ the pressure at every level, z is expressed in pressure levels ranging from the surface to the tropopause, p_s the surface pressure from EMWF ERA Interim-5, R the gas constant for dry air ($J \text{ kg}^{-1} \text{ K}^{-1}$), T_i the temperature at every layer, which comes from the temperature profile downscaled by the *TM5 profile* tool and g the gravitational acceleration in $N \text{ kg}^{-1}$.

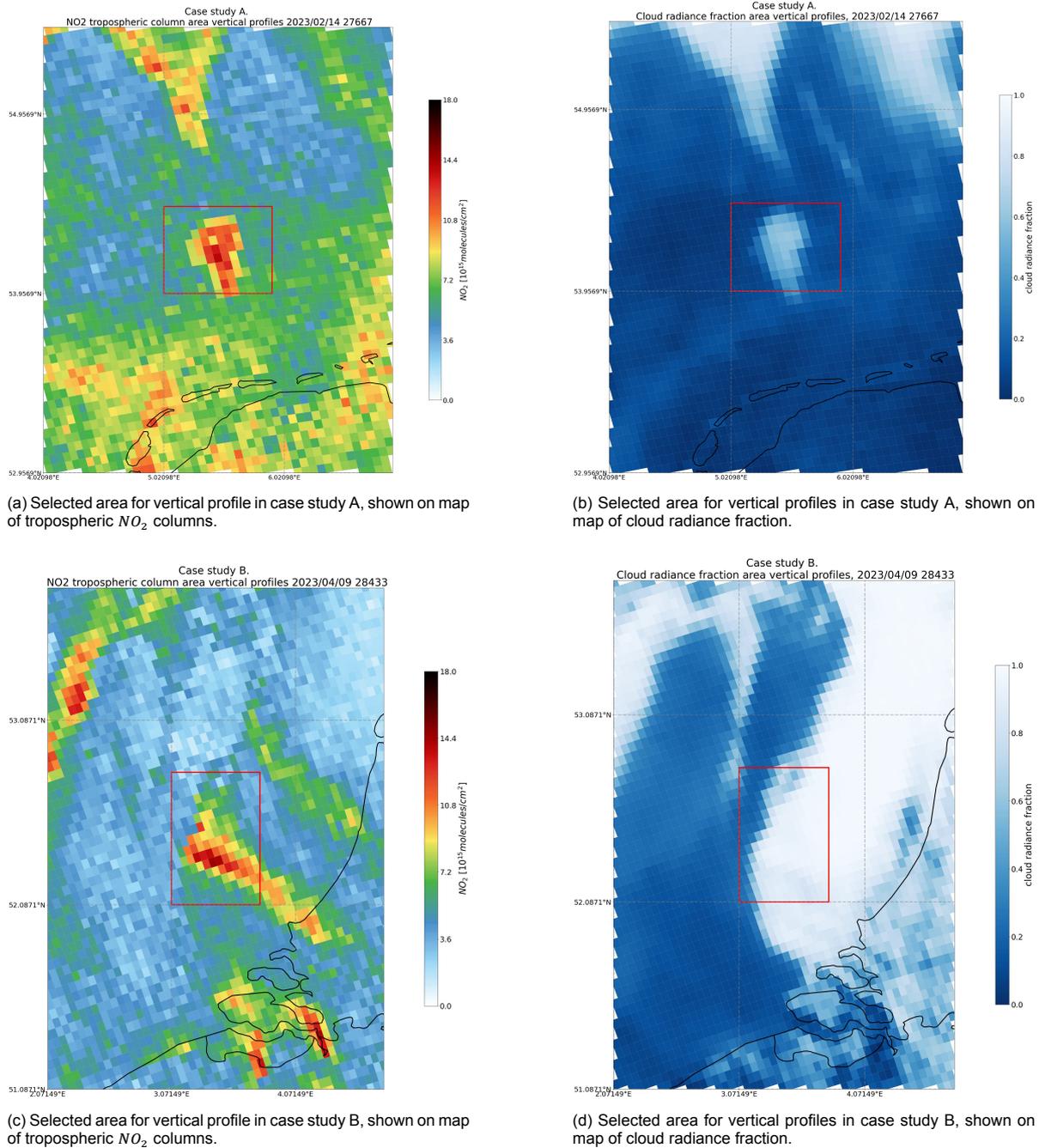


Figure 2.3: Selected area in study case A and B for vertical profiles, shown on top of maps of the NO_2 vertical tropospheric columns and cloud radiance fraction.

2.6. Air-mass factor calculation

To better understand the calculation of the air-mass factor (formula A.3), a step-wise analysis of the air-mass factor calculation is done. Each subsequent factor is defined by adding one more variable to the calculation of the air-mass factor. In practice this means the first factor is purely a geometrical conversion from the slant column, based on the satellite viewing and sun zenith angle. This geometrical factor is defined as F_g in formula 2.11.

$$\begin{aligned}
\mu &= \cos\left(\frac{vza}{180 * \pi}\right) \\
\mu_0 &= \cos\left(\frac{sza}{*180 * \pi}\right) \\
F_g &= \frac{1}{\mu} + \frac{1}{\mu_0}
\end{aligned} \tag{2.11}$$

where F_g is a geometric correction term, vza is the viewing zenith angle and sza is the sun zenith angle.

The next step in the air-mass factor calculation is deriving the box-AMF's from the look-up table. From this step on, only the tropospheric part of the AMF is considered, unless mentioned differently. As mentioned in chapter 2, the look-up table is dependent on six dimensions and the resulting box-AMF is the interpolated result for all six of these variables: viewing zenith angle, sun zenith angle, surface pressure and albedo, cloud fraction and pressure. The viewing zenith angle and sun zenith angle are read from the Level-2 file (table 2.1). The surface albedo is interpolated from the DLER climatology. The surface pressure is derived from the TM5-MP model, which in its turn is derived from pressure data by ECMWF [30]. The cloud fraction is the measured cloud radiance fraction in the NO_2 window and the cloud pressure is interpolated from the FRESCO cloud product. The method of interpolation has not been considered in this thesis. The code of this interpolation can be found in the processor-code by Benjamin Leune (KNMI). After interpolation, the box-AMF's are multiplied with the geometric factor F_g . This product for every level is then summed over the troposphere. This leads to the second factor AMF_{GeoLUT} , defined as follows:

$$AMF_{GeoLUT} = \sum_{i=1}^n (AMF_{box_i} * F_{g_i}) \tag{2.12}$$

with AMF_{GeoLUT} the air-mass factor including the geometric correction and the look-up table, n the amount of atmospheric layers in the troposphere and AMF_{box} depicting the value resulting from the look-up table.

The next factor includes the temperature correction term. This term is calculated for every height level Z . The relationship between vertical temperatures and the temperature correction is shown in formula A.4. The subsequent factor $AMF_{GeoLUTcorr}$ is defined in formula 2.13. A map showing the resulting total T_{corr} correction term for both case A and B is shown in appendix B, figure B.1.

$$AMF_{GeoLUTcorr} = \sum_{i=1}^n (F_{g_i} * AMF_{box_i} * T_{corr_i}) \tag{2.13}$$

with $AMF_{GeoLUTcorr}$ the air-mass factor including the geometric correction term, the look-up table and the temperature correction and T_{corr_i} the temperature correction term defined for every vertical layer i in a TROPOMI grid cell.

The final step of the AMF calculation is weighting with the a-priori. This a-priori comes from the TM5-MP model. TM5-MP defines NO_2 concentration as volume mixing ratio of the trace gas in humid air, in $mol\ mol^{-1}$. These are converted to partial columns in $mol\ m^{-2}$ for every layer, as shown in formula 2.14. These partial columns are then integrated into the calculation of the air mass factor, as shown in formula 2.15. The resulting formula in formula 2.15 is in fact the operational air-mass factor used in the L2 NO_2 retrieval. All of the intermediate air-mass factors, as shown above in formula 2.11, 2.12 and 2.13, are rescaled to match the range and magnitude of the operational air-mass factor (formula 2.15). This allows to distinguish spatial patterns in every step. The different intermediate factors result

in varying magnitudes, which cannot be physically interpreted because the partial calculations lack certain variables necessary for a physically meaningful unit.

$$\begin{aligned}
 R_{air} &= 28.94/1000 \\
 g &= 9.80665 \\
 p_{diff} &= p_{i-1} - p_i \\
 NO_{2partialcolumn} &= p_{diff} * \frac{NO_{2mixingratio}}{g * R_{air}}
 \end{aligned} \tag{2.14}$$

$$AMF_{GeoLUTcorrAP} = \frac{\sum_{i=1}^n (F_{g_i} * AMF_{box_i} * T_{corr_i} * NO_{2partialcolumn_i})}{\sum_{i=1}^n NO_{2partialcolumn_i}} \tag{2.15}$$

with R_{air} the molar mass of air in $kg\ mol^{-1}$, g the gravitational constant in $kg\ s^{-2}$, p_{diff} the pressure drop between two layers i and $i-1$ in *Pascal*, $NO_{2partialcolumn}$ the partial column of layer i in $mol\ m^{-2}$ and $NO_{2mixingratio}$ the volume mixing ratio from the TM5-MP model in $mol\ mol^{-1}$. $AMF_{GeoLUTcorrAP}$ corresponds with the operational air-mass factor calculation.

To better grasp how the different steps in the air-mass factor calculation introduce different spatial fingerprints, three new air-mass factors are introduced. All three contributors (box-AMF, temperature correction and a-priori) are varying in the three-dimensional plane (X,Y,Z). In these alternative air-mass factors, one of the three aspects is averaged out in the vertical plane and the variation of the other two contributors is retained (formula 2.16). In other words, the vertical profile of one contributor is flattened out and the variability is reduced to a two-dimensional plane. This is done to assess the effect of vertical variability in the different factors. By flattening out the profile of a different contributor each time, better insight into how the vertical profile of that contributor is affecting the air-mass factor can be obtained. The slant column is then divided by these alternative factors to calculate the alternative vertical column densities. The geometric contributor F_g is not considered in this analysis. The geometric factor is encapsulated within the box-AMF for simplicity. Only the tropospheric profile is considered when calculating the vertical average.

$$\begin{aligned}
 AMF_{gbox} &= F_g * AMF_{box} \\
 AMF_A &= \frac{\sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i})}{\sum_{i=1}^n NO_{2partialcolumn_i}} \\
 AMF_B &= \frac{\sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i})}{\sum_{i=1}^n NO_{2partialcolumn_i}} \\
 AMF_C &= \frac{\sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i})}{\sum_{i=1}^n NO_{2partialcolumn_i}}
 \end{aligned} \tag{2.16}$$

with AMF_A , AMF_B and AMF_C representing the air-mass factor with only the variation of respectively the box-AMF, the temperature correction and the a-priori. The overbar refers to an averaging out the vertical dimension.

These alternative air-mass factors vary quite strongly in magnitude compared to the original air-mass factor, which subsequently leads to an overall downwards bias in the retrieved NO_2 vertical columns. The resulting NO_2 fields are subsequently normalised and rescaled to match the range of the original tropospheric vertical column, as shown in formula 2.17.

$$[NO_2]_{rescaled} = \left(\frac{[NO_2]_{alternative} - [NO_2]_{original_{min}}}{[NO_2]_{original_{max}} - [NO_2]_{original_{min}}} \right) \cdot ([NO_2]_{original_{max}} - [NO_2]_{original_{min}}) + [NO_2]_{original_{min}} \quad (2.17)$$

with $[NO_2]_{original}$ referring to the tropospheric vertical column retrieved with the original air-mass factor and $[NO_2]_{alternative}$ the newly formulated field using one of the three alternative air-mass factors as shown in formula 2.16

Although the above-mentioned methodology allowed to better distinguish the contribution of the different steps in the AMF calculation, it still does not lead to a conclusive comparison. The different magnitudes and units of the three contributors call for a comparison using a unit-less, relative contribution. The cumulative contribution of each step has to be represented in a one-to-one comparable and non-dimensional manner. These three comparable factors are derived from the air-mass factor calculation as shown below. The derivation starts from two formulas, the independent pixel approximation, which is a combination of the cloudy and clear contribution stipulated by the radiance weighted cloud fraction crf , and the general air-mass factor formula.

$$AMF = crf * AMF_{cloud} + (1 - crf) * AMF_{clear}$$

$$AMF = \frac{\sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i})}{\sum_{i=1}^n NO_{2partialcolumn_i}} \quad (2.18)$$

Then, the denominator in the air-mass factor (2.18) is brought to the other side. The product within the summation expression is rewritten as a sum. Proof that expression 2.19 is equal to expression 2.20 is detailed in appendix B.

$$AMF * \sum_{i=1}^n NO_{2partialcolumn_i} =$$

$$\sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i}) \quad (2.19)$$

$$AMF * \sum_{i=1}^n NO_{2partialcolumn_i} =$$

$$\sum_{i=1}^n \frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{AMF_{gbox_i}}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}$$

$$+ \sum_{i=1}^n \frac{T_{corr_i}}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}$$

$$+ \sum_{i=1}^n \frac{NO_{2partialcolumn_i}}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}} \quad (2.20)$$

The summated $NO_{2partialcolumn_i}$ is then brought back to the right hand side:

$$\begin{aligned}
& AMF = \\
& \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{AMF_{gbox_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
& + \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{T_{corr_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
& + \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{NO_{2partialcolumn_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}}
\end{aligned} \tag{2.21}$$

Subsequently, the sum of three factors in equation 2.21 can be rewritten as a sum of three components A_{gbox} , B_{Tcorr} and $C_{apriori}$:

$$\begin{aligned}
A_{gbox} &= \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{AMF_{gbox_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
B_{Tcorr} &= \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{T_{corr_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
C_{apriori} &= \frac{1}{\sum_{i=1}^n NO_{2partialcolumn_i}} * \sum_{i=1}^n \frac{NO_{2partialcolumn_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}}
\end{aligned} \tag{2.22}$$

which leads to the AMF now being formulated as a sum of three contributors. This derivation is done separately for the clear and cloudy air-mass factor, which leads to both being defined as a sum of three contributors (formula 2.23 and 2.24). This can then be combined through the independent pixel approximation, as shown in formula 2.25. Note that only the AMF_{gbox_i} (i.e. the look-up table) term is different in the clear and cloudy air-mass factor calculation. Because this term is present in the denominator of every factor, this leads to a separate definition of the cloudy and clear contributors. This also leads to the important condition not to compare the cloudy contributors with the clear contributors. Only the contributors within one air-mass factor are comparable. The final expression is given in formula 2.26, with the total air-mass factor being written as the sum of three non-dimensional contributors.

$$AMF_{clear} = A_{gbox_{clear}} + B_{Tcorr_{clear}} + C_{apriori_{clear}} \tag{2.23}$$

$$AMF_{cloudy} = A_{gbox_{cloudy}} + B_{Tcorr_{cloudy}} + C_{apriori_{cloudy}} \tag{2.24}$$

$$\begin{aligned}
A_{gbox} &= (1 - crf)A_{gbox_{clear}} + crfA_{gbox_{cloudy}} \\
B_{Tcorr} &= (1 - crf)B_{Tcorr_{clear}} + crfB_{Tcorr_{cloudy}} \\
C_{apriori} &= (1 - crf)C_{apriori_{clear}} + crfC_{apriori_{cloudy}}
\end{aligned} \tag{2.25}$$

$$\begin{aligned}
AMF &= (1 - crf)AMF_{clear} + crfAMF_{cloudy} \\
AMF &= A_{gbox} + B_{Tcorr} + C_{apriori}
\end{aligned} \tag{2.26}$$

2.7. Intersection gradient

Next, the impact of cloud properties is explored. To better understand the effect of different retrieval settings on the NO_2 vertical column densities in and around clouds, an intersection is defined. This

intersection is defined along the main wind direction and in the seawards direction for both study cases. The gradient intersection is also defined such that it contains a clear and cloudy sky part as well as a contrast in high and low concentration. This is done to best capture the contrast between different types of pixels (polluted vs. non-polluted and clear sky vs. cloudy sky). The spatial definition of this intersection is shown in figure 2.4.

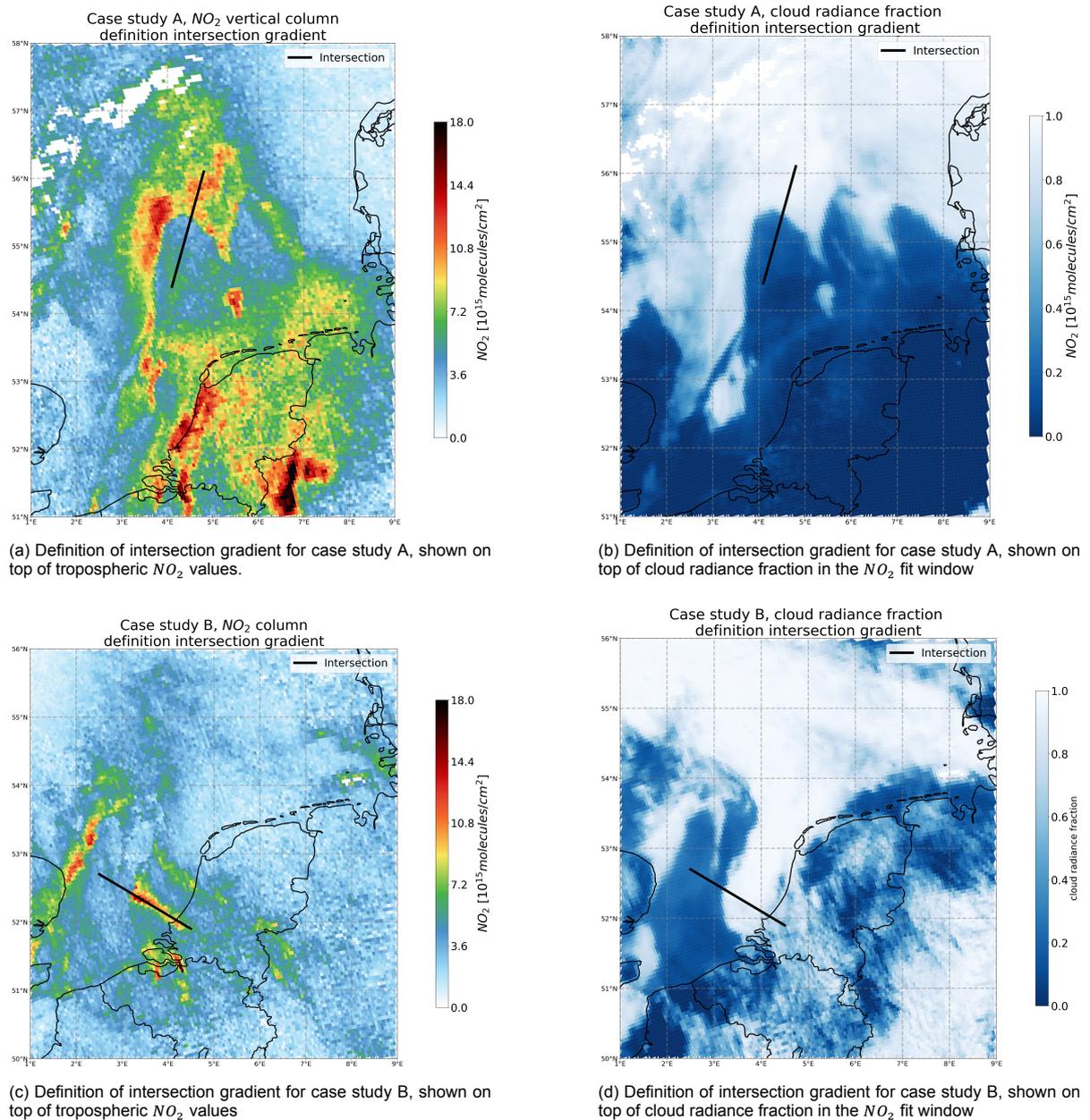
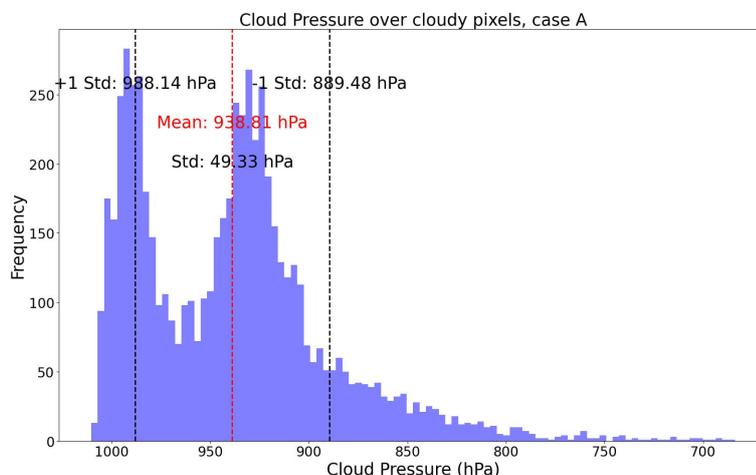


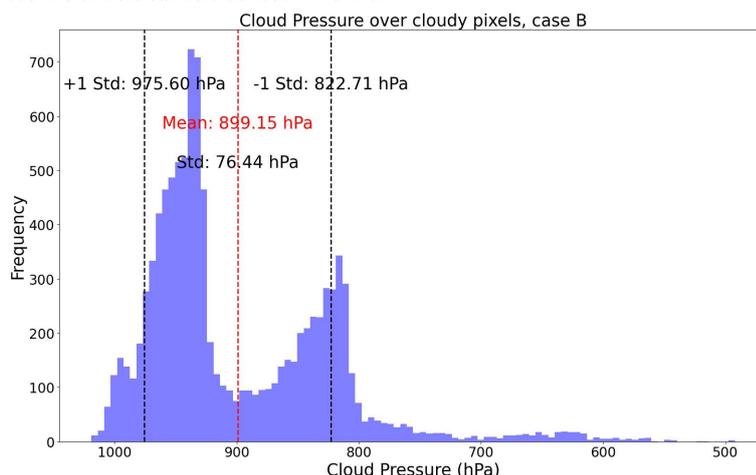
Figure 2.4: Intersection gradient for study case A and B, shown in black on maps of tropospheric NO_2 and cloud radiance fraction. The longitude-latitude grid is set to a 1.5 proportion.

To assess sensitivity to cloud pressure, the input cloud pressures from FRESKO are changed to alternative values. The order of magnitude of this change stems from the standard deviation of cloud pressures over both case studies, as shown in figure 2.5. Since the standard deviation for both cases is around 50 hPa, the first introduced bias is ± 50 hPa. This value is also in line with the error found in a previous FRESKO assessment [33]. Starting from this initial bias, other intermediary values are defined that flatten the contrast between the cloudy and clear sky NO_2 column.

To compare the effect of changing cloud pressures to the effect of a changing air-mass factor, another gradient is defined in the same intersection. This gradient depicts the transition between clear and cloudy sky when applying only the pure clear-sky air-mass factor, contrary to the cloud radiance fraction weighted combination of the cloudy and clear air-mass factor. This implies that the effect of clouds is not considered anymore and the whole scene is regarded as a clear scene.



(a) Histogram of FRESCO cloud pressures over cloudy pixels in the area of interest for case A. The mean cloud pressure is 938 hPa and the standard deviation is 49 hPa.



(b) Histogram of FRESCO cloud pressures over cloudy pixels in the area of interest for case B. The mean cloud pressure is 915 hPa and the standard deviation is 55 hPa.

Figure 2.5: Statistics of cloud pressure over cloudy pixels

2.8. Surface albedo

The effect of the surface albedo is also assessed. In the operational retrieval algorithm, the interpolated DLER (directionally dependent Lambertian-equivalent reflectivity) surface albedo at 440 nm is used as an input. The surface albedo in the NO_2 fit window in the entire swath is simulated for that specific day, based on the past years of surface albedo. This is then used in the look-up table to define sensitivity of the sensor through the box-AMF's. Surface albedo is also used to determine surface radiance, which is then used in the cloud radiance fraction calculation (in the NO_2 fit window). A bright deviation from the DLER surface albedo climatology in the measured reflectance is seen as a cloud, on the condition that the EMWF snow/ice mask has not flagged that pixel. This DLER climatology is known to introduce errors in the algorithm [34]. Deviations on that specific day from the expected climatology can influence both the cloud retrieval and air-mass factor calculation. A strong effect of the DLER on cloud radiance is only to be expected in (partly) clear pixels. In the case of a fully clouded pixel, the DLER surface

albedo climatology does not influence the radiance parameterization.

A check to rule out inconsistencies related to the spectral fit is also done. In some rare cases, the fit of the measured to the modelled absorption spectrum (step 1 in the retrieval algorithm, formula A.1), can show spatial inconsistencies, with large errors over contiguous areas. The deviation between the measured and modelled spectrum should be random and not spatially coherent. The cause of this lies in an improper simulation of the expected absorption spectrum. This mismatch in the DOAS fit is caused by an absorber which is not accounted for in the simulated spectrum. This could be e.g. sediment, chlorophyll or organic matter [35]. The poor DOAS fit could lead to higher NO_2 retrievals, and thus a potential misinterpretation of high emissions [35].

The above-mentioned phenomenon is tested for using the the Wald-Wolfowitz test, also known as the runs test [36]. This non-parametrical test is applied on the differences between the modelled and simulated spectrum. Per pixel, the test will retrieve the largest positive or negative sequence of differences. In the case of a good fit, this value should be low since differences are randomly positive and negative. A larger run value means the spectral fit is poor over a longer range of wavelengths. An expected value of runs is also defined, as well as a deviation from this expected value. The test has only been carried for case study A because of computational effort considerations. The testing is performed by Jos van Geffen (KNMI, personal communication, 2023/04/03).

3

Results

First, an overview of the two selected case studies and their associated NO_2 and cloud field is given. Then, the a-priori TM5-MP model is analysed and compared with the higher-resolution CAMS-Europe Ensemble model. For a small area of interest, the vertical profiles of the models and the averaging kernels (chapter A, section A.7) are analysed. An alternative retrieval with CAMS-Europe as a-priori is performed and compared with the original retrieval. Next, the air-mass factor calculation is examined through a step-wise analysis, the formulation of alternative air-mass factors and a contribution analysis. Finally, the effect of cloud height is assessed through an intersection gradient. A test to check for a misfit between the observed and simulated reflectance spectra is also done.

3.1. Selected case studies



(a) NPP-VIIRS image 14/02/2023 above the Netherlands and North Sea



(b) NPP-VIIRS image 09/04/2023 above the Netherlands, south-east England, Northern France and the North Sea

Figure 3.1: VIIRS imagery of case studies

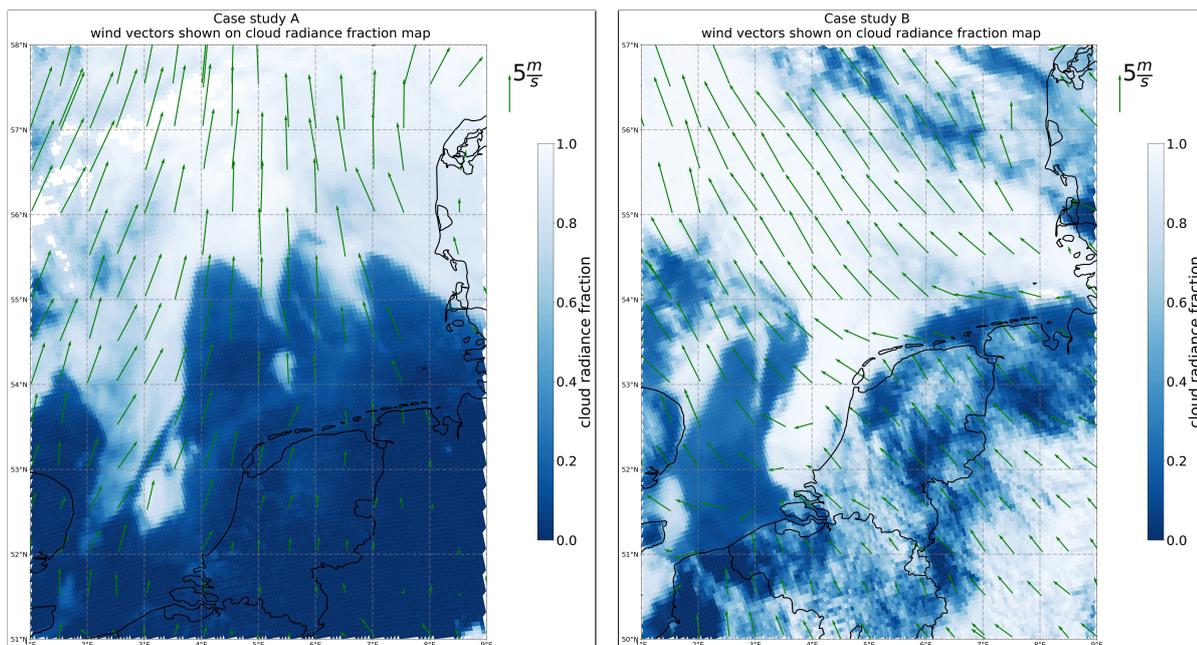
The two selected case studies are presented in figure 3.1a and 3.1. Both cases reflect conditions with an influx of pollution coming from adjacent NO_2 source regions over land, as shown in figure 3.2. The cloud conditions for both case studies are shown in figure 3.1a and 3.1.

Case A is a scene from the 14th of February, 2023. On this winter day, clear-sky conditions exist over the Netherlands and Northwestern Germany. Over the North Sea, we see a low cloud field, where two main parts have been transported northward because of southern offshore winds. In the Terra-MODIS

imagery shown in figure C.1, which has an earlier overpass time of around 10:30 am, the cloud field is more extensive, particularly towards the southeast. VIIRS-NPP imagery shows a persistent field of low clouds and fog over the area the few days before, spanning over Western Europe and the North Sea. This cloud field is being transported northwards on 14 April by a dominant southern wind. Cloud top height detected by VIIRS-NPP ranges between 0 and 800 m. The wind field at 10 meter high at the time of retrieval (ECWMF ERA Interim-5) is shown in figure 3.2a. The tropospheric NO_2 column (figure 3.3c) is filtered to only contain data with a qa-value larger than 0.5, the implications of this filter are explained in chapter 2, section 2.2.3. The NO_2 retrieval shows a contrast over the sea surface between clouded and clear-sky. The tropospheric NO_2 columns follow the outline of the low cloud field with a strong contrast between cloudy and clear sky, which is suspect and may be unrealistic.

Case B is a scene from the 9th of April, 2023. This day shows a field of low clouds stretching out over the Dutch coastline into the North Sea. Wind is blowing from the south-east, dragging the cloud field further seawards. Terra-MODIS imagery (figure C.2) shows a similar situation in the morning, but with a more sharply outlined field of morning fog over the North Sea. As the day progressed, atmospheric vertical and horizontal mixing caused these sharp boundaries to become less distinct. Unlike case A, this low cloud field was not present the days before. Another low cloud field is located around the coastline of southeast England. The height of the cloud top as measured by NPP-VIIRS ranges from 0-1600 meters. As seen in figure 3.3d, a plume from the Rotterdam region is advected towards the northeast, following the prevailing wind direction (figure 3.2b). However the plume appears to suddenly vanish when the clouds disappear (cloud radiance fraction drops in figure 3.2b). The sharp contrast between cloudy and clear-sky in NO_2 seems to be unrealistic, or at least exaggerated by the cloud field. The same qa-filter of 0.5 is used in case B.

Although both cases show some notable differences, both in the cloud and NO_2 field, there are some similar aspects present in both cases. Both scenes have a clearly visible discrepancy in NO_2 pollution, which follows the outline of the low clouds. They also have a clear distinction between parts of the cloudy and clear sky over the North Sea. The contrast in NO_2 concentrations within the scene are also high enough for spatial patterns to be distinguishable, reflecting some real underlying NO_2 variability. Both of these days have the advantage of a partly clear sky above the ocean surface, with a cloud fraction very close to zero. At the same time, the cloudy part in both cases is as good as fully clouded. This sharp contrast eliminates ambiguous interpretations, as it distinctly separates the scene into two clearly defined parts.



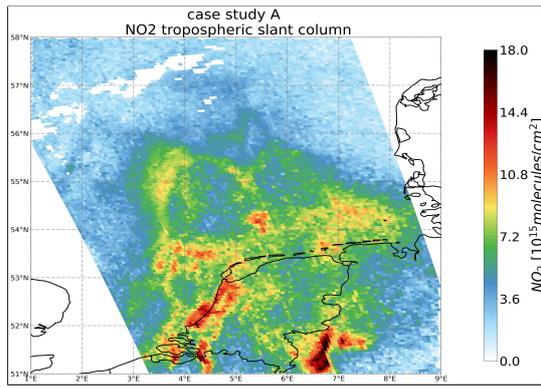
(a) Cloud radiance fraction measured in NO_2 window with wind vector field for case study A

(b) Cloud radiance fraction measured in NO_2 window with wind vector field for case study B

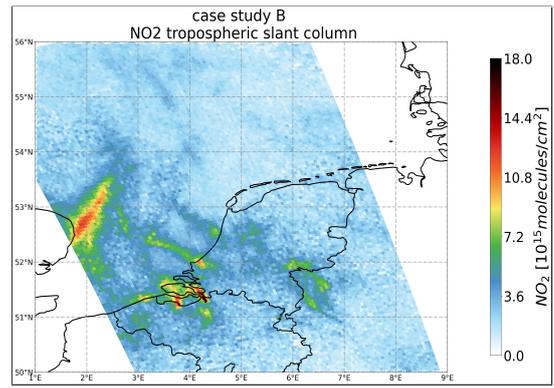
Figure 3.2: Wind vector field (ECMWF) shown on top of cloud radiance fractions for the two case studies analyzed in this thesis. The left panel displays conditions on 14 February 2023, the right panel displays conditions on 9 April 2024.

3.2. NO_2 field

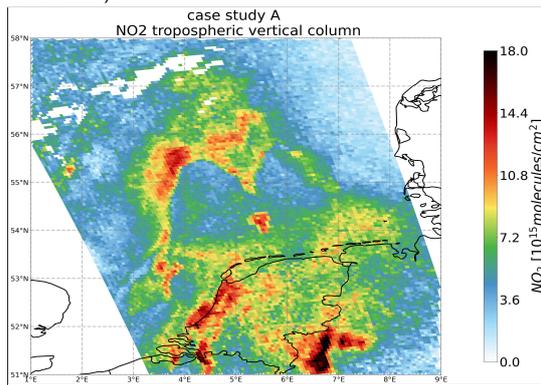
The NO_2 slant column density, the associated vertical column density and the absolute difference between both is shown in figure 3.3. For both cases, a strong increase above the cloudy area is observed. This is in line with expectations, since the part of the tropospheric column underneath the cloud is not captured in the slant column. The cloud fraction and cloud radiance fraction for both cases is shown in figures 3.5 and C.3. Above the clear-sky part, both cases show a decrease in NO_2 densities after the conversion from slant to vertical column through the air-mass factor. For case A, the outlines of the clouds become more clearly distinguishable after applying the air-mass factor. Whereas the slant column measures pollution both in the clear- and cloud-sky, applying the air-mass factor exacerbates this distinction. The plume footprint in case B, starting from the Rotterdam area spreading out northwest above the North Sea, is increased significantly by the air-mass factor. The relative difference between slant and vertical column density is shown in appendix C, figure C.5.



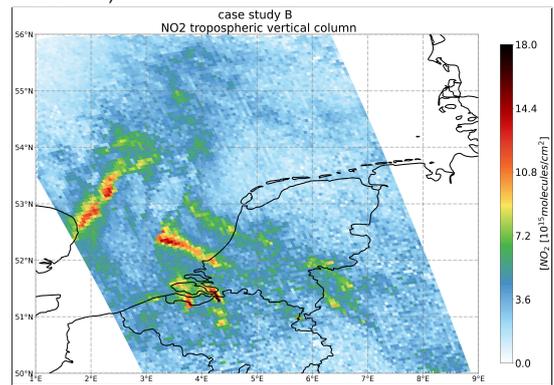
(a) Tropospheric NO_2 slant column, case study A (scaled by geometric AMF).



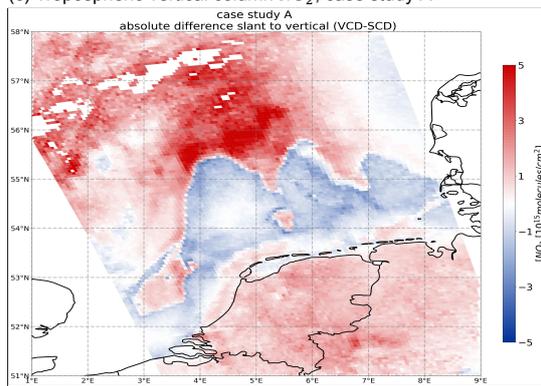
(b) Tropospheric slant column NO_2 , case study B (scaled by geometric AMF).



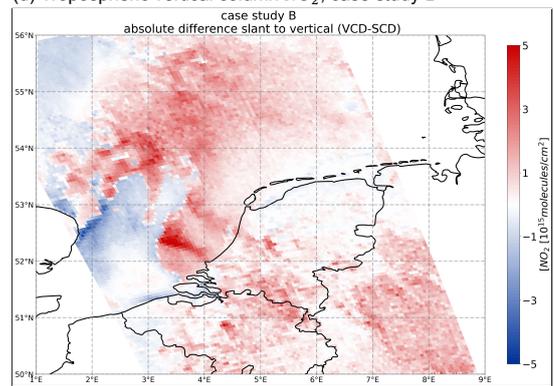
(c) Tropospheric vertical column NO_2 , case study A



(d) Tropospheric vertical column NO_2 , case study B

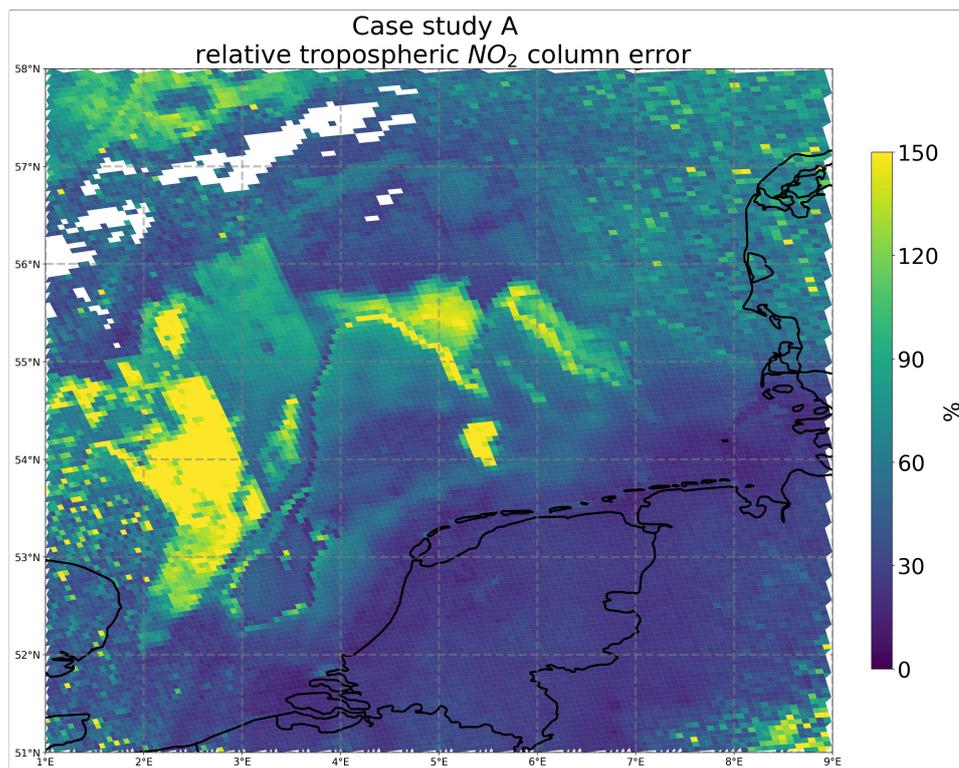


(e) Difference between the slant and vertical column for case study A. A positive value (red) means an increase of the slant column when converted to vertical column. A negative value (blue) means a decrease of the slant column when converted to vertical column.

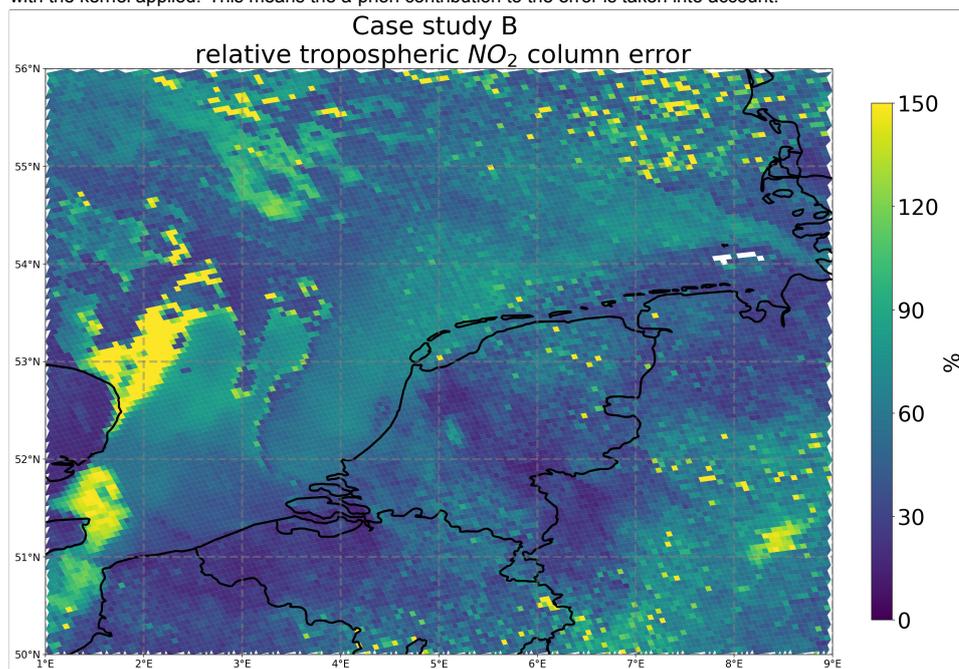


(f) Difference between the slant and vertical column for case study B. A positive value (red) means an increase of the slant column when converted to vertical column. A negative value (blue) means a decrease of the slant column when converted to vertical column.

Figure 3.3: Tropospheric slant and vertical column and the absolute difference for both case studies.



(a) Case study A. Relative error tropospheric vertical column. This relative error is based on the tropospheric precision, with the kernel applied. This means the a-priori contribution to the error is taken into account.



(b) Case study B. Relative error tropospheric vertical column. This relative error is based on the tropospheric precision, with the kernel applied. This means the a-priori contribution to the error is taken into account.

Figure 3.4: Relative error of the tropospheric vertical column for both case studies

A relative error of the vertical tropospheric NO_2 density is also calculated, based on the absolute tropospheric precision *with* the kernel taken into account, as described in chapter 2, section 2.2.3. This absolute precision is then divided by the tropospheric column to obtain the relative error of the vertical column density, as shown in figure 3.4. The relative error is highest over cloudy parts. This effect is mostly visible in case A. Here, the part of the cloudy area, with the highest increase from slant to

vertical column, coincides with the highest relative error, with a relative error ranging up to 150 % of the retrieved densities. In case study B, the borderlines of the cloud field stretching out over the North Sea clearly show a high relative error, with some parts also exceeding 100 %. The relative error is to be interpreted cautiously over areas with a low NO_2 density, because a low absolute error coinciding with a low tropospheric NO_2 density can lead to a high relative error.

The simulated surface albedo in both the NO_2 retrieval window (440 nm) and the FRESCO retrieval window (758 nm), as well as the measured cloud radiance fraction in the NO_2 window and the measured scene albedo in the FRESCO retrieval window are shown in figure 3.5. The same result for case study B is shown in appendix C, figure C.3. The top left panel, the simulated surface albedo in the NO_2 retrieval window, at 440 nm, is the surface albedo used to determine the sensor sensitivity, as an input in the look-up table for the box-AMF's. This surface albedo is also used to determine the cloud (radiance) fraction in the NO_2 window. The top right panel, simulated surface albedo in the O_2 A-band retrieval window, is used in the FRESCO cloud retrieval, where the simulated reflectance at TOA is calculated as a sum of the cloud-free and cloudy parts of the pixel, with the cloud-free reflectance being calculated through the simulated surface albedo [37]. The third panel in figure 3.5 shows the cloud fraction measured in the NO_2 window, this is the effective cloud fraction as shown in formula A.6. Then, the radiance weighted cloud fraction or cloud radiance fraction is shown in the next two panels. The last panel shows the *measured* scene albedo in FRESCO, so in the O_2 A-band.

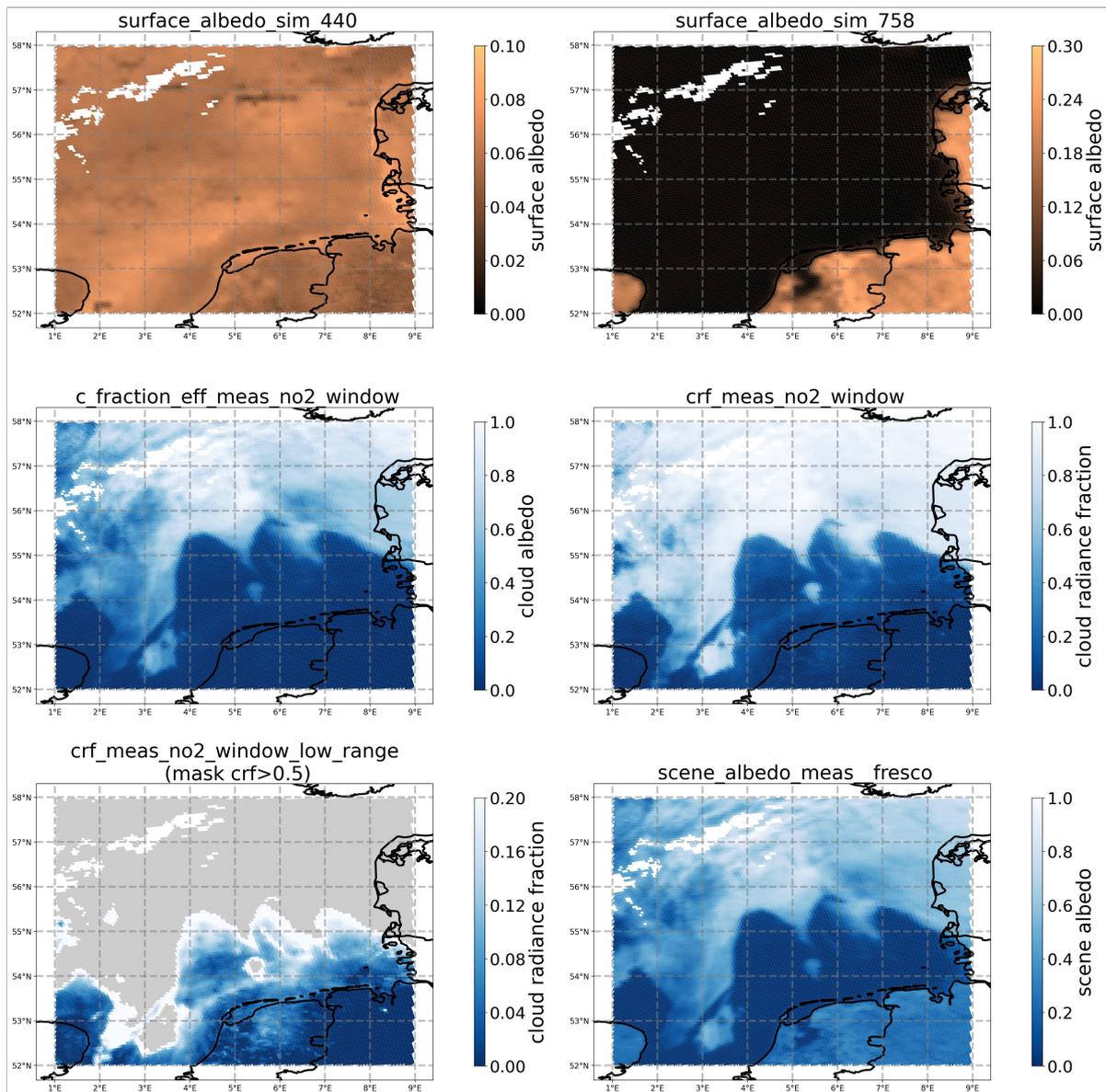
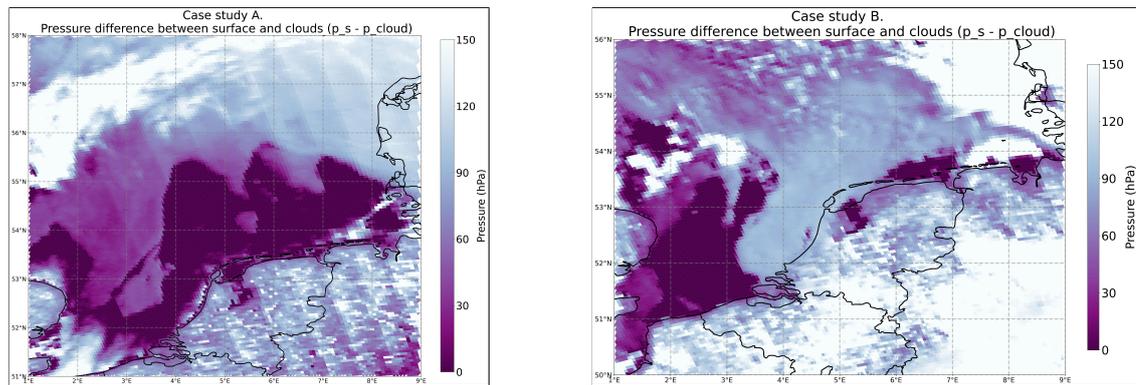


Figure 3.5: Case study A. From bottom left to top right: the simulated surface albedo in the NO_2 retrieval window, the simulated surface albedo in the O_2 A-band retrieval window, the measured cloud fraction in the NO_2 window, the measured cloud radiance fraction in the NO_2 retrieval window, the low range measured cloud radiance fraction in the NO_2 retrieval window with a mask depicting cloudy areas and the measured scene albedo in the FRESKO (O_2 A-band) retrieval window.

3.3. Cloud field



(a) Difference between the surface pressure (ECMWF) and cloud pressure (FRESCO-CRB) for case study A.

(b) Difference between the surface pressure (ECMWF) and cloud pressure (FRESCO-CRB) for case study B.

Figure 3.6: Difference between surface pressure and cloud pressure for both case studies.

Figure 3.6 shows the pressure-height above the surface (pressure) for the FRESCO-CRB clouds, for both case studies ($p_{surface} - p_{cloud}$). A smaller value indicates a cloud closer to the surface. For both cases, low differences, and thus low cloud heights, are measured. In case A, the southern stretches of the cloud field are around 30hPa removed from the surface, which is roughly 150 meters above the sea surface. The clouds seem to gradually increase in cloud height going more north- and seawards. For case B, a more heterogeneous cloud field is observed. Here the gradient is reversed, with high clouds above land that are gradually decreasing in height seawards. The cloud field stretching out from the edge of southeast England has a more pronounced low cloud height.

Figure 3.7a shows both the slant and vertical column densities plotted against cloud pressure. Only points with a cloud radiance fraction above 0.5 are shown, this filters out the more clear pixels, where the effect of cloud pressure is less deterministic. The trend lines through both point clouds indicate that, over cloudy pixels, the vertical column density is consistently higher than the slant column density. This can be explained through the ghost column effect, where the air-mass factor will increase the slant column density to capture the part of the pollution underneath the cloud. This effect is counteracted by the cloud sensitivity effect, where the air-mass factor decreases the slant column density to make up for the heightened reflectivity caused by the presence of a cloud. However, in this case, the ghost column effect is stronger and the vertical column shows an upward correction over all observations with low clouds (clouds higher than the 800hPa pressure level are not considered here). This difference between slant and vertical increases as the clouds are higher. Another important aspect seen in figure 3.7a, is how the NO_2 densities increase with cloud pressure. So lower cloud heights correspond with higher slant and vertical column densities. The same relation is observed for case study B (figure C.4). This observed relation may be attributed to a selection bias, since both case studies are deliberately selected on the condition of high NO_2 concentrations coinciding with low clouds. For both case studies, low clouds correspond to high NO_2 concentrations. Additionally, lower cloud heights correspond with a smaller difference between slant and vertical column densities.

Figure 3.7b shows a scatter plot of the slant column on the x-axis and the vertical column on the y-axis for case A. The colorscale represents the cloud radiance fraction. For both cases, the vertical column is predominantly higher than the slant column over cloudy pixels ($crf > 0.5$). For case A, and to a lesser extent case B (figure C.4), there is also a dependency on cloud radiance fraction. A higher cloud radiance fraction leads to a more positive correction when converting the slant to vertical column. This is related to the effect of clouds in the retrieval. If a pixel has a higher cloud radiance fraction, it means the pixel is more cloudy and the cloudy AMF gains more weight in the total AMF through the independent pixel approximation. The cloudy air-mass factor is lower than the clear air-mass factor, so this leads to a higher vertical column (inverse effect because the slant column is divided by the air-mass

factor in the retrieval algorithm).

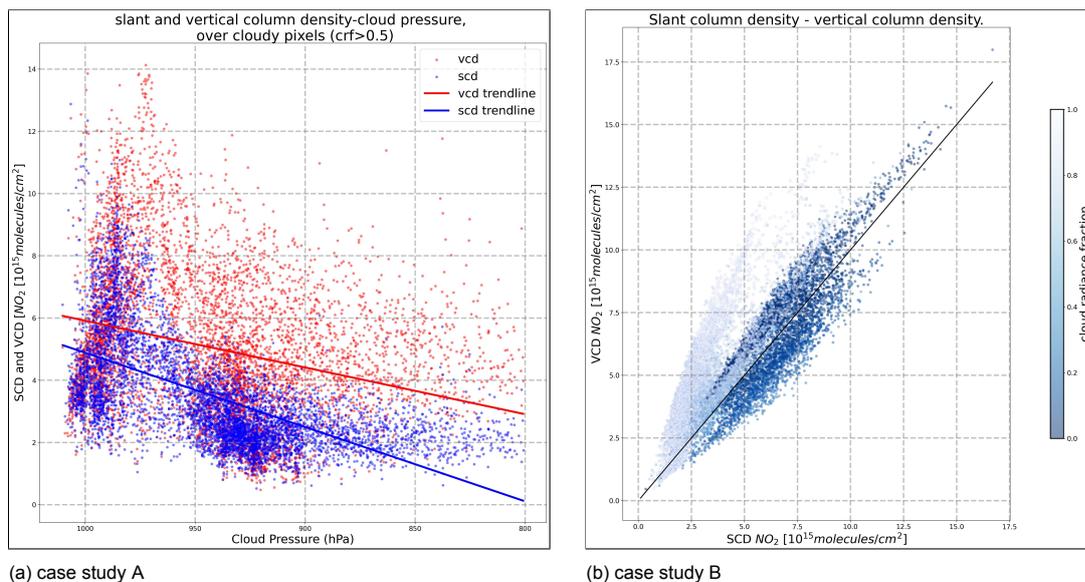


Figure 3.7: case study A. Each point corresponds to a TROPOMI pixel. Left plot shows both the slant and vertical column densities plotted against cloud pressure. The trendline through both scatter clouds is also shown. Only cloudy pixels are plotted with a threshold of $crf > 0.5$ and only cloud pressures above 800 hPa are considered. Right plot shows the scatter plot of slant column density against vertical column density, with colorscale representing cloud radiance fraction in the NO_2 fit window. The 1-1 line is plotted in black.

3.4. A-priori

3.4.1. TM5-MP

The TM5-MP modelled NO_2 field is shown in figure 3.8. The first three layers, covering the first 450 meters of the troposphere, are shown together with the integrated tropospheric column. The horizontal resolution is $1^\circ \times 1^\circ$. This low resolution leads to a spatially *smeared out* effect. For case A, most of the pollution is modelled at the layer between 62 and 220 meters height. The lowest layer linked to the surface, simulates a spot of pollution above the North Sea, this can be from an emission source at the Northern coast of the Netherlands, or perhaps a remnant of pollution from the day before. The tropospheric column field seems to have the Rhine-Ruhr region and the Netherlands as primary emission source. The model simulates a seawards, northeastern transport of this pollution. For case B, significantly less NO_2 is simulated. The pollution in the total tropospheric column seems to be centered around the Antwerp-Rotterdam area, with the highest values at the 65 to 220 meter level. The values in this layer peak above the North Sea. It is unclear which anthropogenic emission source this is related to. The positioning of this centre point could be related to dynamic effects, such as a low-level southeastern wind, rather than the presence of a local source in the North Sea.

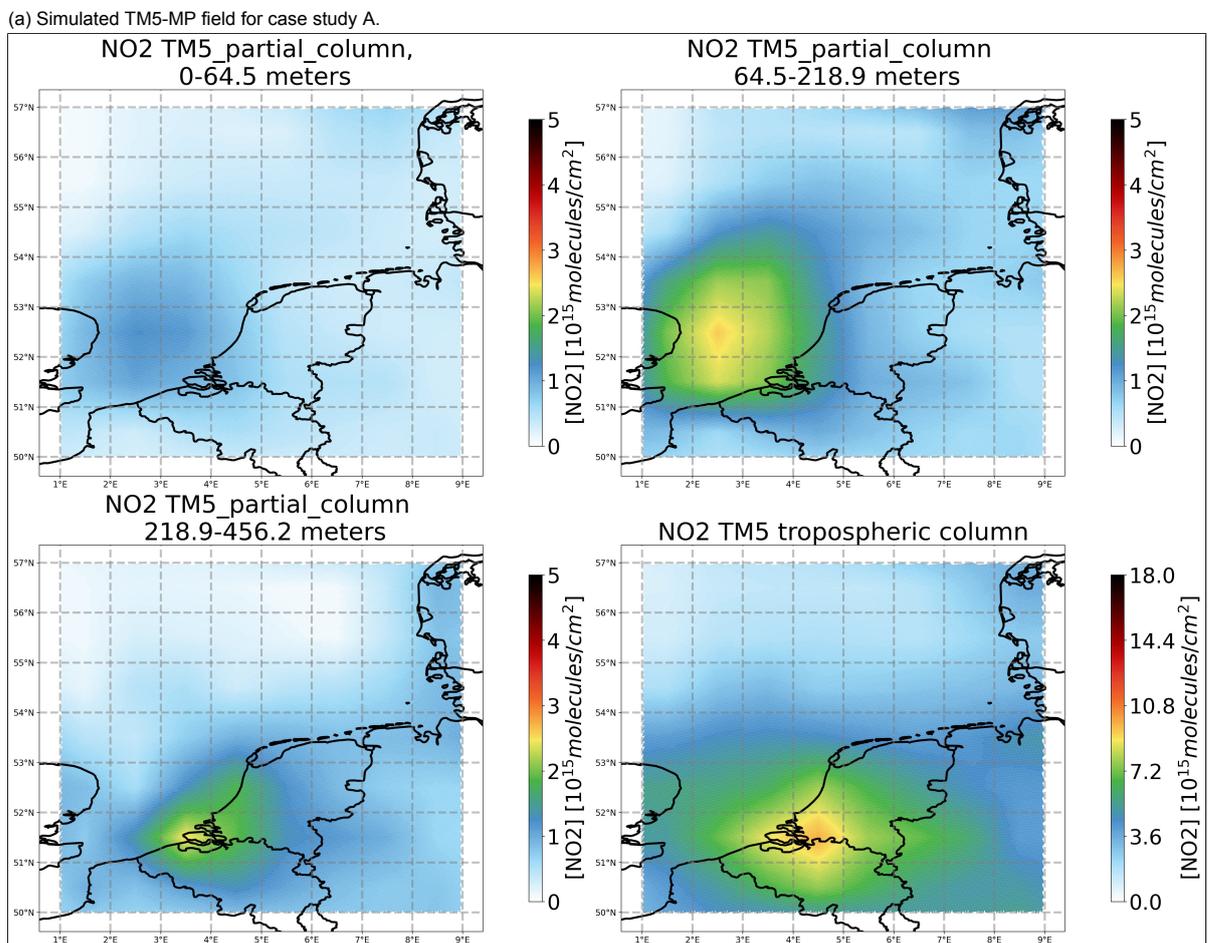
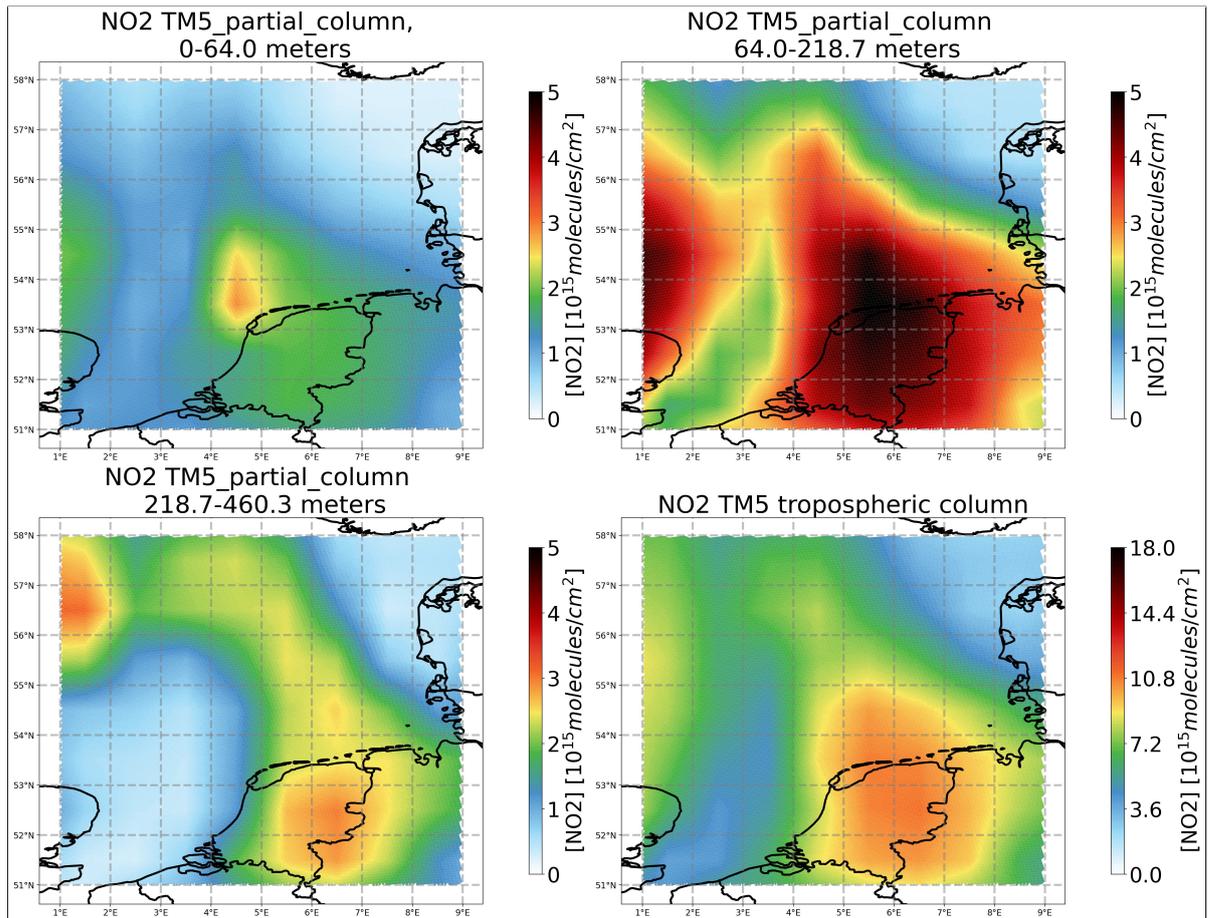


Figure 3.8: Simulated TM5-MP NO_2 concentrations for the surface level on the top left, the 65m-220m layer, the 220m-450m layer and the total tropospheric column on the bottom right, for both case studies.

Figure 3.9 shows the NO_2 ghost column for both cases. As explained in appendix A, this is the integrated TM5-MP profile from the surface until the cloud height. So the ghost column is an indicator both of cloud height and modelled NO_2 field. The ghost column is 0 over areas with a measured cloud radiance fraction in the NO_2 window lower than 0.5. A higher ghost column leads to a more positive correction between the slant and vertical column. For case A, a high ghost column is simulated in the areas with spurious observations. This is likely related to a high simulated NO_2 field (figure 3.8a), since the cloud height is very low over these parts. The spatial pattern seen in 3.9a is closely related to the pattern of cloud height, as seen in figure 3.6a. Since the TM5-MP NO_2 field shows low spatial variations, the spatial variability in the ghost column is dominated by cloud height variability, rather than the TM5-MP field. The ghost column NO_2 densities for case B are comparable in size, although this is also caused by the higher cloud height in this situation. Figure 3.10 shows the same point cloud as the previous scatter plot in figure 3.7b but with the colorscale as a function of the ghost column. This scatter clearly shows how in case A, a higher amount of NO_2 modelled underneath the cloud will lead to a more positive difference when converting the slant to the vertical column. The effect is less deterministic for case B.

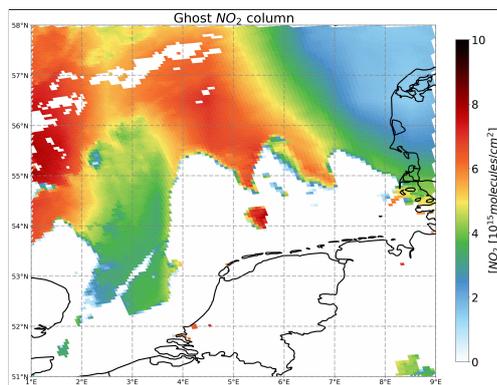
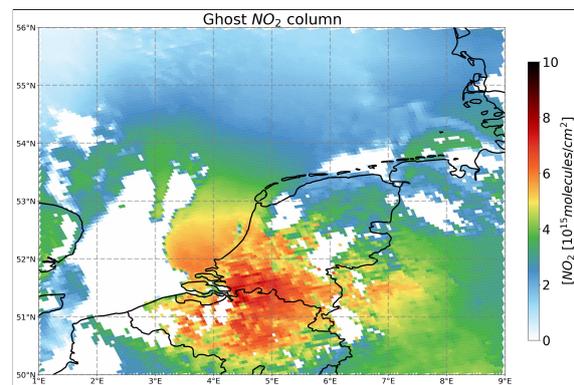
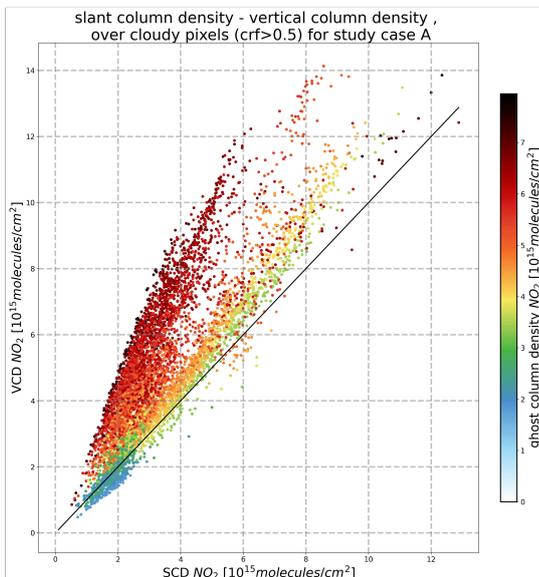
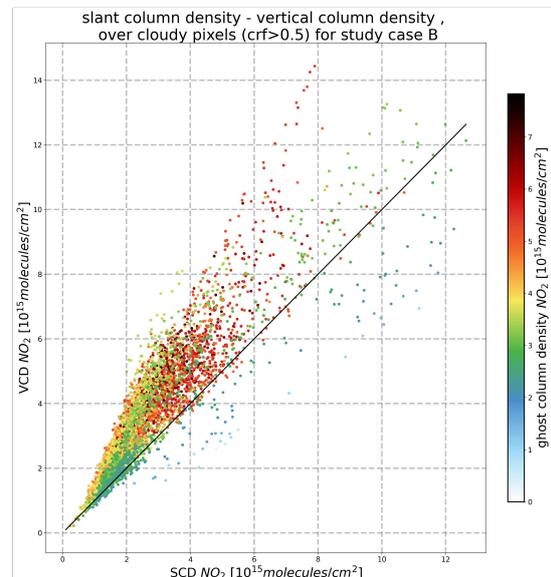
(a) Ghost column NO_2 density for case study A.(b) Ghost column NO_2 density for case study B.

Figure 3.9: Ghost column for both case studies.



(a) Study case A



(b) Study case B. The cloudy pixels are only selected up to a latitude of 54 to focus more on the ghost column near the emission plume.

Figure 3.10: Scatter plot of slant column density against vertical column density, only cloudy pixels ($crf > 0.5$) are considered for both case study A (3.10a) and B (3.10b).

3.4.2. CAMS-Europe

Then, the CAMS-Europe (also referred to as simply CAMS) modelled NO_2 field is shown for case A in figure 3.11. The CAMS-Europe output for case B is shown in figure C.7. CAMS has a resolution of $0.1^\circ \times 0.1^\circ$. This gives the model the ability to better capture emission sources and the associated plume transport. The CAMS field is shown for the layers ranging from the surface to 875 meters height. The total column is integrated from the surface to 5km, which is the highest layer in CAMS, shown on the bottom left panel in figure 3.11. This total integrated column is an approximation of the tropospheric column. In the situation of case A, the lower layers, up to 375 meters, show a well-developed transport of NO_2 quite far onto the North Sea. For the higher layers, a dichotomy appears, with a part of the NO_2 pollution present above the North Sea and a part over some anthropogenic emission sources over land. These pollution spots over sea could be remnants of the previous hours or days. Since case A is in the winter, NO_2 has a long lifetime in the atmosphere. A similar behaviour is seen for case B, albeit the pollution over the North Sea is more concentrated near southeast England here.

The corrected CAMS field after applying the TROPOMI averaging kernels, calculated in the light of a recent KNMI project [28], is shown on the bottom left panel of figure 3.11. This *TROPOMI correction* takes into account the averaging kernels, and thus sensitivity, in the associated TROPOMI overpass. The formula to correct the CAMS-Europe field using the averaging kernel is shown in chapter 2, formula 2.6. This corrected product also uses a lower resolution global model to couple the concentrations in the upper troposphere and troposphere-stratosphere boundary [28]. However, the main effect over a polluted region like the North Sea is to be expected from the averaging kernel correction. The result should be understood as a *CAMS-Europe simulation of what TROPOMI will measure (i.e. the slant column density)*. The averaging kernels will decrease model observations over areas with low sensitivity and vice versa. For the first case, this correction introduces a downwards bias over the North Sea. The values above the North Sea are lowered the most. This is caused by two factors: the darker sea surface and the presence of clouds. The former has a straightforward effect: a darker surface (in the 440nm wavelength) decreases the sensitivity, so TROPOMI is expected to measure lower tropospheric NO_2 columns here. The cloud field has a double effect: if CAMS-Europe simulates the NO_2 plume below the FRESCO-CRB cloud height, which is incapsulated in the averaging kernels, high NO_2 concentrations are multiplied with a low averaging kernel value. So the corrected CAMS concentration will be lower. If the simulated emission plume is situated above the clouds, the higher averaging kernels are multiplied with high CAMS NO_2 concentrations, which would lead to an upwards correction. Mostly the first effect is present, since the field is significantly lower after the averaging kernel correction.

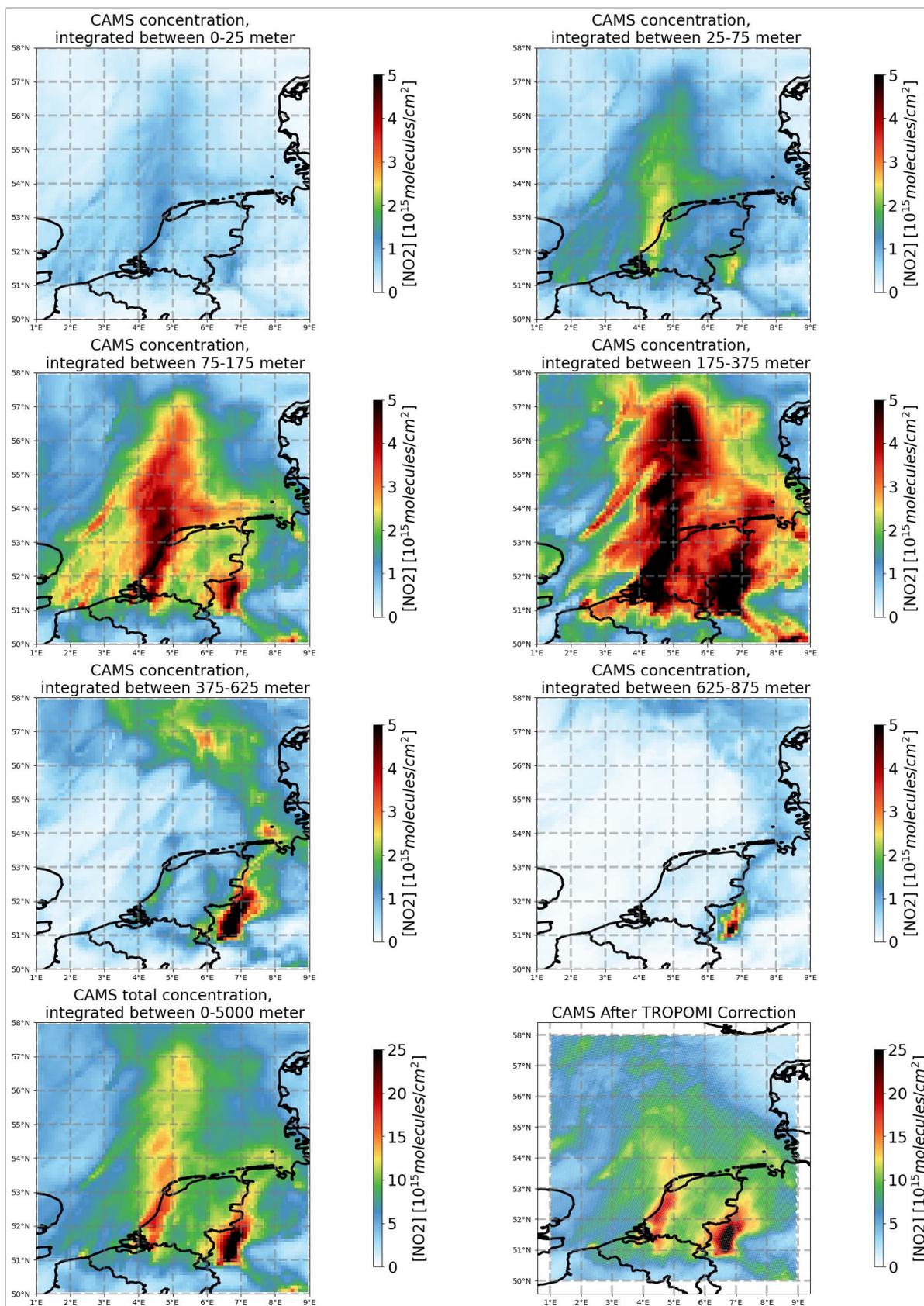


Figure 3.11: case study A. Simulated CAMS-Europe forecast NO_2 fields, integrated around different height levels for the first 7 panels. The bottom left panel shows the summated CAMS-Europe column. Bottom right panel shows the corrected CAMS-Europe tropospheric column [28]. Note the difference in scale between the bottom row and the first three rows.

3.4.3. Vertical profiles

Figures 3.12 and 3.13 show the vertical profile of the TM5-MP and CAMS NO_2 partial columns together with the TROPOMI averaging kernels. The cloud pressures in the area of interest are shown as light blue dashed horizontal lines. For these cloud pressures a cloud radiance fraction threshold of 0.5 is used. The averaging kernels clearly show two typologies for both cases: a clear and a cloud type. The averaging kernels relating to cloudy pixels are higher above the cloud and lower below the cloud, whereas the averaging kernels relating to the clear-sky pixels show a near linear decrease in sensitivity towards the surface, which is related to a general decrease in sensitivity more downwards in the atmosphere. These cloudy averaging kernel profiles are the equivalent of the theoretical case shown in chapter A, figure A.2. The averaging kernels show some small variations within these two typologies. This is likely related to small differences in surface albedo, cloud radiance fraction or cloud height. A higher surface albedo will increase the averaging kernel values. A lower cloud means a steeper decrease of sensitivity towards the surface. As mentioned in chapter A, the box-AMF's are also dependent on cloud fraction. A higher cloud fraction leads to an averaging kernel more towards the cloudy typology, whereas a lower cloud fraction in a pixel leads to an averaging kernel profile with a more linear decrease. The averaging kernels are directly proportional to the box-AMF's from the look-up table, as shown in formula A.9. In the vertical profile of case B (figure 3.13), the clouds are generally higher, and exert a higher variation. This more heterogeneous cloud field is reflected in a higher variation of the averaging kernels.

The TM5-MP profile in both cases shows the same general morphology, with a sharp increase from the surface followed by a general decline with height. This is consistent with the different horizontal maps shown in figure 3.8. An important aspect is whether the TM5-MP profile peak is modelled above or below the cloud. In the case of the former, the retrieval algorithm will assume that the slant column is the result of an over-estimated observation of low pollution above a bright surface, and subsequently adjust NO_2 values slightly downward. In the case of the latter, the retrieval algorithm assumes the slant column is missing this below-cloud contribution, and adjust the NO_2 values upward. In both areas of interest, the TM5-MP column is simulated largely below the clouds, which lead to an upward correction from the slant to the vertical column.

This mechanism can also be understood through the air-mass factor calculation (formula A.3). The a-priori is weighed with the box-AMF and the temperature correction in the numerator and the total tropospheric a-priori column is in the denominator. The presence of the a-priori in both numerator and denominator means that the absolute values of the a-priori are not important in driving the air-mass factor but rather the shape of the profile. The variation is driven by the peak of the a-priori profile. When this peak is lower, the high a-priori values are multiplied by very low averaging kernels, especially underneath a cloud, as seen in the vertical profiles. This means the air-mass factor will be lower, which in its turn leads to a higher vertical column. When the peak concentrations of the a-priori are simulated higher, they are multiplied with higher box-AMF's, so the air-mass factor will be higher. The slant column is divided by this higher air-mass factor which results in a less pronounced correction and thus a lower NO_2 density.

The CAMS vertical profile is also shown for both case studies. The vertical levels in CAMS are converted from meters to pressure to compare with the other profiles. Besides from a higher horizontal resolution ($0.1^\circ \times 0.1^\circ$), CAMS-Europe also has a higher vertical resolution, with 7 data-points in the lowest 100 hPa ($\approx 800\text{ m}$), where TM5-MP only defined 5. The resolution is also highest in the area most important for the type of retrievals considered here, which is the lowest part of the troposphere, with mid-layer points at 25, 75, 125 and 275 meter. The partial column values peak higher than TM5-MP for both cases. This is likely caused by the higher total tropospheric density which shows more sharp emission plumes with a lower background signal in the selected area (figure 2.3), as compared with a smeared out field in TM5-MP. Although the highest concentrations are still simulated quite low in the troposphere, the vertical decrease above the cloud is less pronounced. The NO_2 profile in the boundary layer in CAMS is more vertically mixed.

However, the CAMS-Europe model still simulates a significant amount of pollution to be below the cloud

levels, so a similar result as for TM5-MP can be expected when using CAMS-Europe as the a-priori profile. In this case, the retrieval algorithm assumes that the majority of the vertical column is not well represented in the slant column so it strongly amplifies the vertical NO_2 column. Since some of the higher partial column values in the CAMS-Europe profile are simulated more right above or on the same level as the cloud, as seen for case A in figure 3.12b this effect is likely not expected for those pixels. Here, the retrieval algorithm assumes that the measured signal in the slant column already contains most of the vertical NO_2 in that layer. It will subsequently lower the slant column values through the air-mass factor in order to counteract the increased brightness of the clouds. So for case A, a lower vertical column can be expected in the selection area of the vertical profiles. Since both the horizontal and vertical distribution in CAMS-Europe shows greater variability, using CAMS-Europe as a-priori is likely to lead to a more distributed effect in the retrieved vertical NO_2 column, with some cloudy pixels showing a strong upwards correction, where other pixels might be subject to a downwards correction. This downward correction will take place in pixels with an elevated profile, more above the cloud. The upward correction will take place in pixels where a significant amount of NO_2 is still simulated below the cloud height.

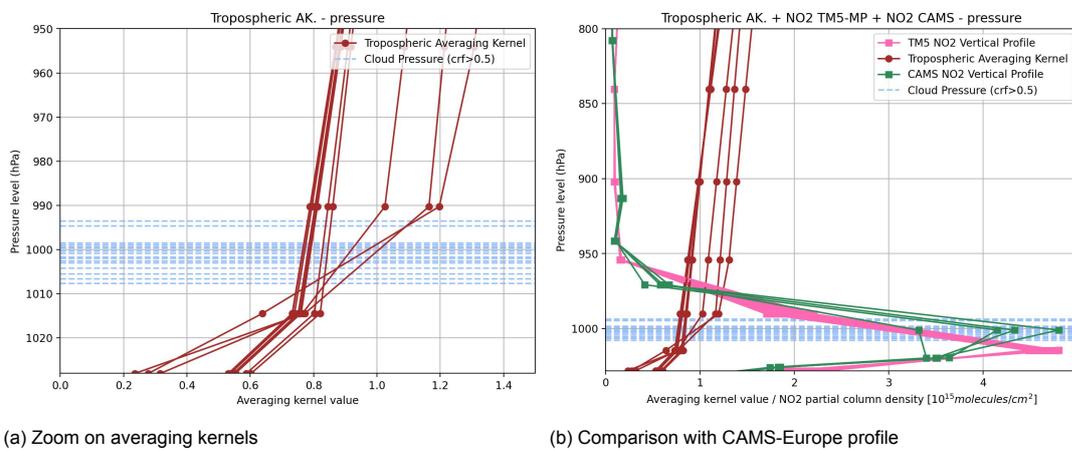


Figure 3.12: Vertical profiles of case study A, with the averaging kernels (brown), the TM5-MP a-priori profile (pink), the cloud pressures (lightblue) and the CAMS-Europe profile (green). The data points for both NO_2 profiles correspond with mid-layer heights.

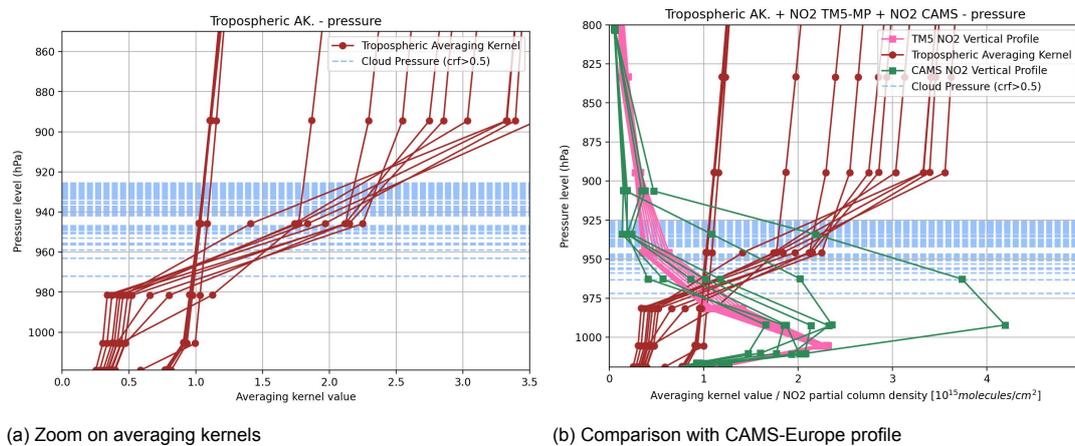


Figure 3.13: Vertical profiles of case study B, with the averaging kernels (brown), the TM5-MP a-priori profile (pink), the cloud pressures (lightblue) and the CAMS-Europe profile (green). The data points for both NO_2 profiles correspond with mid-layer heights.

3.4.4. CAMS-Europe as a-priori

Next, the NO_2 vertical column density using CAMS-Europe as a-priori is shown in figures 3.14 and 3.15. This data product is courtesy of KNMI and has been developed in light of a recent project assessing the use of a high-resolution regional transport model in TROPOMI [28]. For both case studies, CAMS as a-priori does not fundamentally change the vertical distribution of NO_2 . Both cases still show a somewhat similar (and spurious) result compared to the TM5-MP a-priori retrieval, with the vertical NO_2 density seemingly following the outlines of clouds.

However, some noteworthy changes take place. For case study A, the low cloud field has the expected double effect, with both an increase and decrease, compared to the original retrieval. The decrease is mostly seen above the outlines of the low clouds, as well as the small separate cloud around 54 °N and 5 °E, clearly visible as the blue-shaded areas on the difference-plots in the bottom row of figure 3.14. This small clouds corresponds with the selection area of the vertical profiles for case A (figure 3.12). This decrease is related to the CAMS-Europe model simulating pollution transport more *above* instead of below the clouds here. Vice versa an increase is also visible over the more northern cloudy parts of the cloud field, caused by the presence of a pollution plume lower than the cloud heights. Over the entire scene, CAMS-Europe partly succeeds in lowering the strong contrast between clear and cloudy pixels, by simulating a part of the emission plume coming from land above the clouds instead of below. However, the low clouds still lead to a sharply defined outline in the vertical NO_2 column. For case study B, the area with lower clouds sees mostly an increase, which further exacerbates the clear-cloudy contrast in the pollution plume (which is spread out from the Rotterdam area northwest onto the North Sea). This is because the cloud height is still higher than the simulated emission plume height, so the ghost column effect is still strong.

The ratio between the original air-mass factor and the air-mass factor using CAMS as a-priori is shown in figure C.8. This ratio is the inverse of the differences shown below. This follows from the fact that a higher air-mass factor in the CAMS NO_2 product leads to a lower vertical column density and vice versa. This ratio is multiplied with the original CAMS field (formula 2.7) to obtain the *CAMS after TROPOMI correction* field in figure 3.11, as explained in chapter 2, formula 2.7.

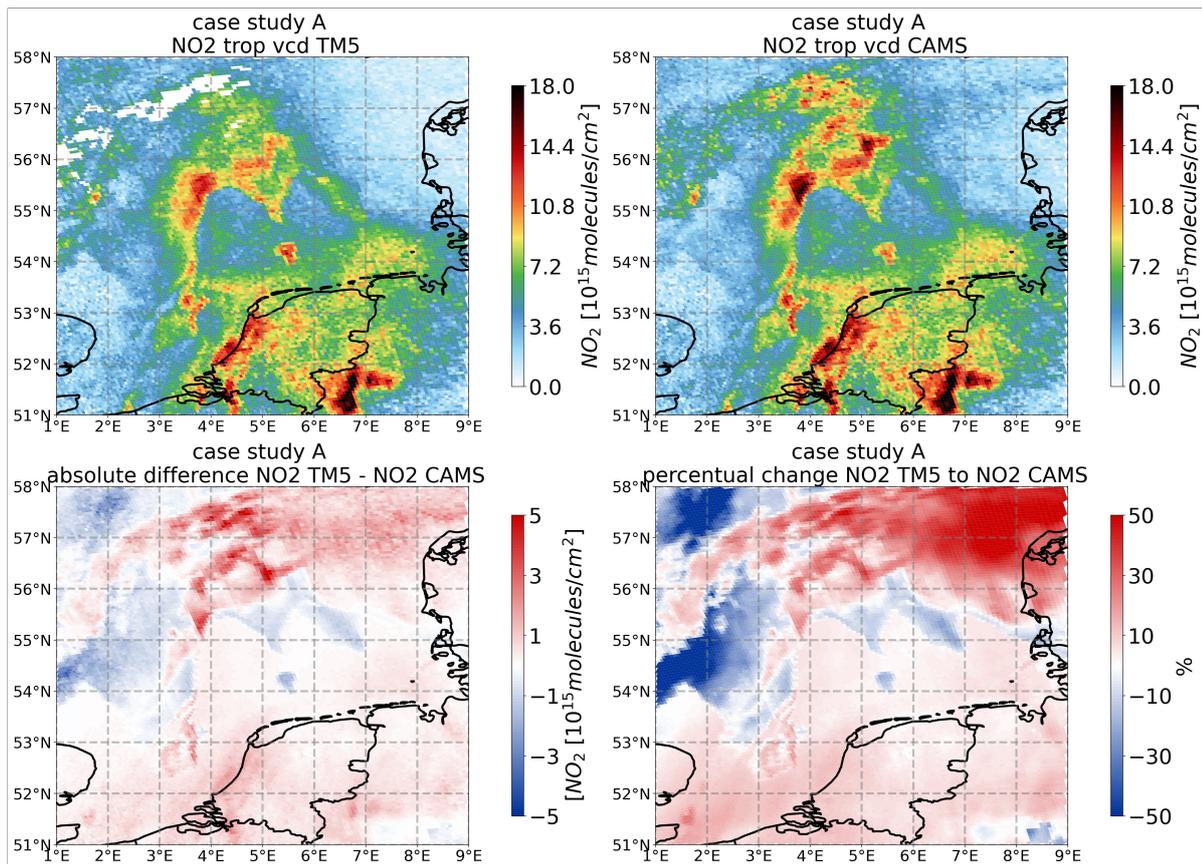


Figure 3.14: case study A: Original tropospheric vertical NO_2 column on top left panel. Tropospheric vertical NO_2 column when using the CAMS-Europe a-priori shown on top right panel. Absolute and relative difference between the original retrieval (TM5-MP a-priori) and CAMS-Europe a-priori retrieval shown on respectively bottom left and bottom right panel.

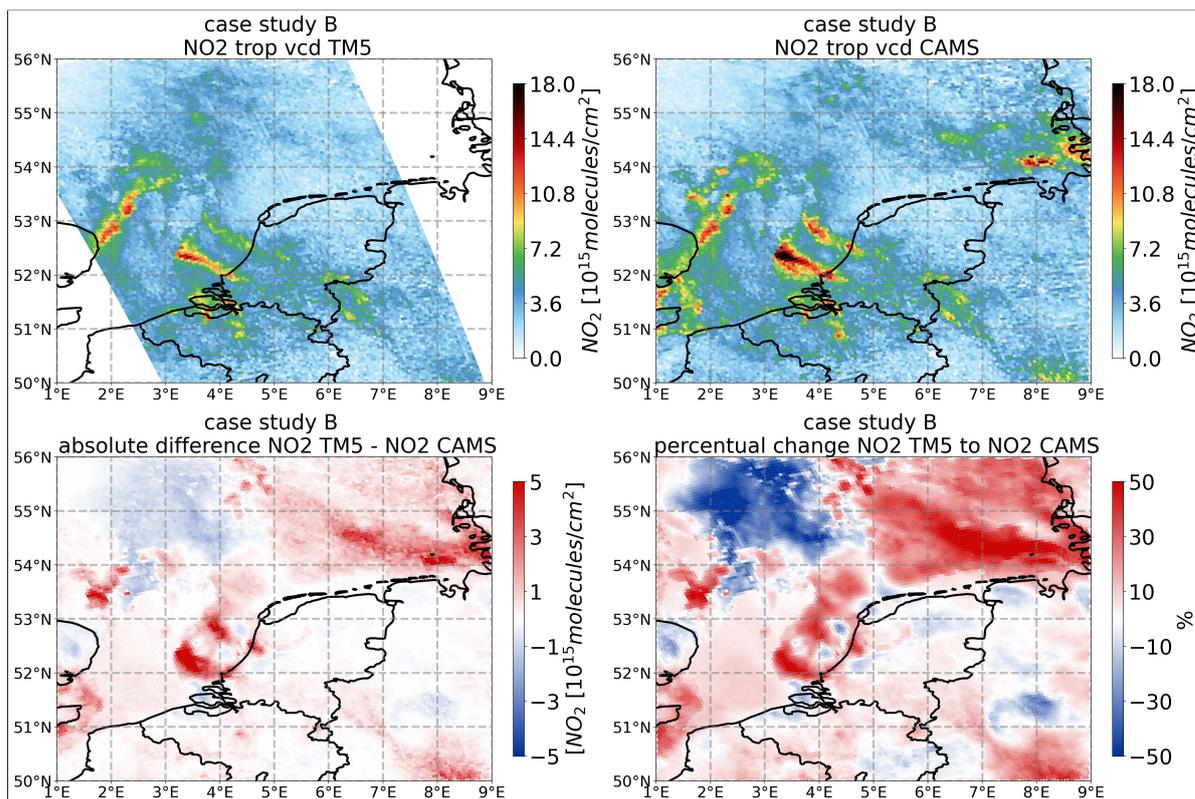


Figure 3.15: case study B: Original tropospheric vertical NO_2 column on top left panel. Tropospheric vertical NO_2 column when using the CAMS-Europe a-priori shown on top right panel. Absolute and relative difference between the original retrieval (TM5-MP a-priori) and CAMS-Europe a-priori retrieval shown on respectively bottom left and bottom right panel.

3.5. Air-mass factor calculation

3.5.1. Step-wise analysis

Next, the results of the step-wise analysis of the air-mass factor calculation are shown for case A. In figure 3.16, each step in the calculation of the air-mass factor is shown spatially, both for the clear, cloudy and total air-mass factor. This calculation process is described extensively in chapter A, section A.2 and chapter 2, section 2.6. The box-AMF, or *AMF LUT*, shows a gridded spatial character. The box-AMF is defined through interpolation of a look-up table with 6 dimensions (input variables). These 6 variables of the look-up table are: viewing zenith angle, sun zenith angle, surface albedo, surface pressure, cloud fraction and cloud pressure. None of these 6 input variables exert a similar spatial behaviour (the surface pressure is shown in C.6, the sun zenith angle and viewing zenith angle are not shown but can be assumed to exert a smooth spatial variation). This implies the gridded spatial behaviour shown is likely an artefact of the interpolation process with the look-up table. However, the interpolation method in itself is not a focus in this thesis, so the exact cause within this interpolation method is not further explored.

The effect of introducing the temperature correction is limited. The temperature correction term does not show large spatial heterogeneity (figure B.1). Since the temperature data is also taken from the $1^\circ \times 1^\circ$ TM5-MP model, low spatial variation on the horizontal scale considered here is expected. The introduction of the TM5-MP a-priori seems to counteract the boxy pattern seen in the *AMF LUT* (fourth panel in figure 3.16). The final clear-sky AMF shows spatial variability over the North Sea. This variability matches with the variability in surface albedo in figure 3.5. This is in line with the theory, since surface albedo is an input parameter for the look-up table defining box-AMF's. The *AMF LUT* for the cloudy (fifth panel) air-mass factor shows a similar gridded behaviour, which seems to coincide with the longitude and latitude field for both the clear and cloudy air-mass factor. The *AMF LUT Cloudy* also shows some patterns in the North Sea, right above the Dutch coastline. As seen on VIIRS imagery

(figure 3.1a), this is likely related to a different water colour, potentially caused by sediment outflow or biological processes.

Unexpected behaviour is observed in the bottom right panel, where both the clear and cloudy part are combined using the cloud radiance fraction crf . The final combined AMF, which is the operational AMF, shows a remarkable result over the North Sea. The patch of clear-sky between the clouds, does not fully correspond with the clear-sky AMF in that region. A part of the cloudy-sky AMF seems to have *leaked* into this area. This is likely related to low measured cloud fractions ($crf < 0.2$), as shown in the cloud radiance fraction (figure 3.5, note the low-range panel). These low cloud fractions affect the air-mass factor through the look-up table.

The cloud radiance fraction is calculated through the radiance from the clouds and surface, as shown in formula A.6. These radiances in its turn are based on a separate cloud look-up table, which takes as input surface and cloud pressure, as well as the albedo of both bounding surfaces [30]. The cloud albedo is set at 0.8, except when a brighter cloud is observed. This upward correction is not observed for both cases in the area of interest. Some upwards correction of cloud albedo in the clear-sky part of the North Sea are observed. Pixels with a cloud radiance fraction lower than 0.5 are not explicitly flagged as *cloudy* pixels. However, low cloud fractions (< 0.5) amplify the AMF over clear sky. These low cloud fractions are measured in the NO_2 window. The VIIRS imagery, figure 3.1, implies that these measured cloud fractions are related to shifts in spectral qualities of the water. Based on the VIIRS imagery, no clouds are present in the area of this low cloud radiance fraction. Only the heightened combined air-mass factor situated at the borders of the cloud field, is the direct effect of the cloud field extending further than the *cloudy* mask based on the $crf > 0.5$ threshold. This leads to higher AMF values at the outlines of the cloud field. Higher AMF values generally imply a lower sensor sensitivity to NO_2 . This lower sensitivity is indeed expected over an unclouded ocean surface, but the peak at the outlines of the clouds does not directly relate to sensor sensitivity. The step-by-step AMF calculation for case study B is shown in appendix C, figure C.9. The same observations hold for this second case study, albeit in a less pronounced manner.

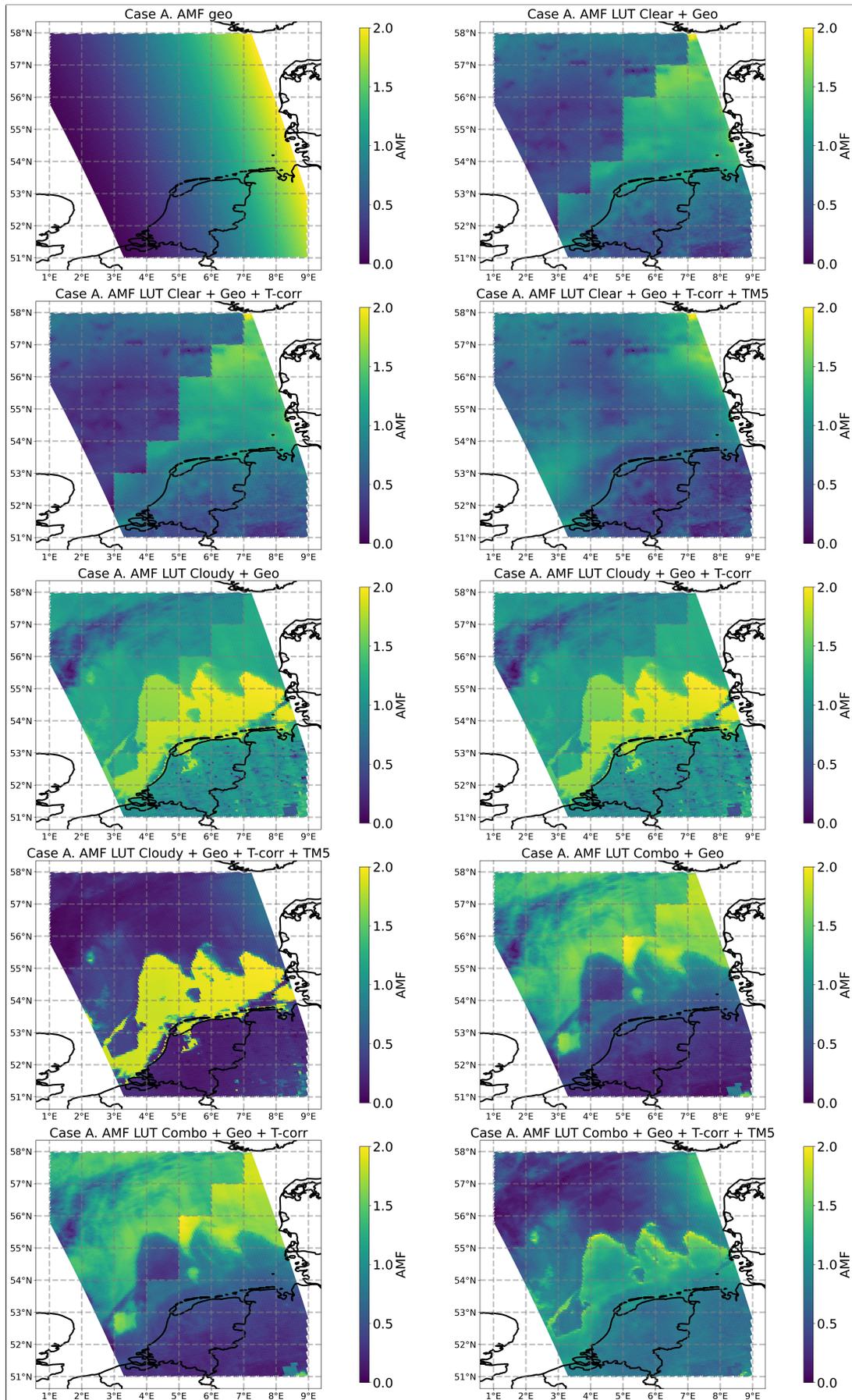


Figure 3.16: case study A. Processor plots showing the geometric conversion factor and the different steps in the clear, cloud and combined air-mass factor retrieval. The top left panel corresponds with formula 2.11. Then, the next three panels correspond with formula 2.12, 2.13 and 2.15. These three last calculation steps are then shown again for the cloudy factor in the next three panels, and again for the combined factor in the last three panels.

3.5.2. Alternative air-mass factors

Next, in figure 3.17, the effect of three alternative air-mass factors on the vertical NO_2 column is shown for case B. For the *averaged out T-corr.*, *averaged out LUT* and *averaged out a-priori* field an air mass-factor with the flattened out vertical variability of respectively the temperature correction, look-up table and a-priori is calculated. Reducing three-dimensional variability of the temperature correction to two-dimensional horizontal variability, results in some differences with the original situation (top left panel). The cloudy and polluted areas show higher vertical column values when the temperature correction variability is reduced. Vertically flattening out this temperature correction term for case B, led to a decrease of the temperature correction term in the first layer above the surface, with the peak in the a-priori value. This lower weight led to a lower air-mass factor, and thus a higher vertical column. This can be seen in the vertical temperature profile of case B, shown in figure C.13, which shows the temperature profile for the selected area (same area as for the other vertical profiles, figure 2.3). Temperature strongly decreases with height, so taking the vertical average over the entire troposphere would mean it increases temperature in the lowest part of the troposphere. However, it should be noted that the temperature is not linearly related to the temperature correction term, as seen in formula A.4. Reducing the variability of the look-up table (bottom left panel), shows a significantly different NO_2 field. Since the LUT contains information on the sensor sensitivity, the reduction of this variable leads to a field that more closely resembles the measured slant column (figure 3.3b). This averaging out assumes that there is no vertical variability in sensor sensitivity, which filters out the strong sensitivity contrast above and below a cloud. The vertical averaging out of the look-up table reduces the correction made for a low sensor sensitivity in the lowest troposphere, which consequentially decreases the vertical column density.

The NO_2 *averaged out a-priori* field looks similar to the *averaged out LUT* field, although related to a different process. By using an averaged out flat tropospheric vertical profile, the partial columns of the a-priori in the lowest few layers will strongly decrease. This means the air-mass factor will be higher over clouded areas, since the low box-AMF's in the low troposphere are no longer weighted with high partial column densities. Clouded areas have a very low box-AMF underneath the clouds, which is amplified in the case of a low peak in the a-priori and thus leads to a low air-mass factor. The flattened out a-priori no longer amplifies this low air-mass factor in the clouded part. This leads to the values over clouds no longer showing a strong upwards correction.

Note that the range of values for the three alternative NO_2 has been rescaled to align with the same range as the original NO_2 vertical fields through a normalization and introduced bias (formula 2.17). The alternative air-mass factors resulted in a significant reduction in overall values due to the averaging process. The primary objective of these alternative formulations is to derive insights into spatial patterns rather than to focus on absolute values. The same figure for case A is shown in the appendix C, figure C.10. The relative differences between the original NO_2 vertical column (top left panel) and the three alternative fields both for case A and B, is shown in appendix C, figures C.11 and C.12.

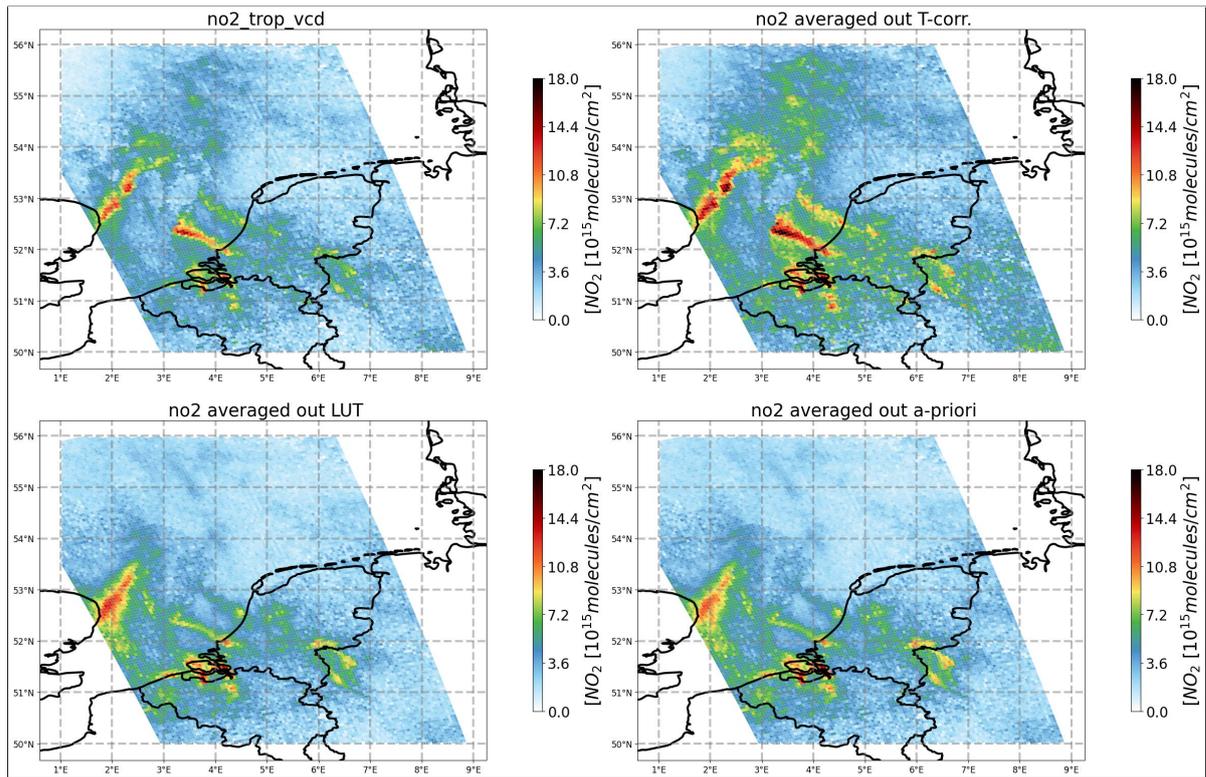


Figure 3.17: case study B. Resulting NO_2 tropospheric vertical column for the original settings, averaging out the temperature correction term, averaging out the look-up table and averaging out the a-priori in the Z-plane (from top left to bottom right).

3.5.3. Contribution analysis

Next, the results of the different calculated contributors to the AMF for case A is shown in figure 3.18 (formula 2.21 and 2.22). The same result for case B is shown in figure 3.19. This figure shows a spatial distribution of the dimensionless contribution of the LUT, temperature correction and a-priori (first three panels) to the final combined AMF, shown in the bottom right panel. This final AMF is the sum of the other three contributors, as shown in formula 2.26. The same principle holds for the cloudy and clear AMF separately (formulas 2.23 and 2.24). The spatial distribution of the three contributors separately for the clear and cloudy air-mass factor for both cases is shown in appendix C, figures C.15a, C.15b, C.16a and C.16b.

The dimensionless contribution of the a-priori on the total air-mass factor is very small (scale on the top left figure). However, based on the earlier results, the a-priori plays an important role in the low troposphere through the ghost column effect. Since the a-priori values are multiplied in the lower troposphere with low averaging kernels, especially underneath a cloud, the contribution to the total tropospheric air-mass factor is low in general. The contribution should also be understood as a counteractive process. When the TM5-MP has an important contribution, which is in the case of a high ghost column, it will lower the air-mass factor (and lead to a higher slant-vertical correction). This means the contribution factor will go towards 0. The effect is highest in the case of large a-priori concentrations and a cloudy sky (and the total TM5 contribution closest to 0). The cloud field is clearly visible in this TM5-contribution. This is because in this case a significant part of the TM5-MP profile is simulated underneath the cloud, which corresponds with a high ghost column. So where the TM5-MP has a larger effect, it will bring the averaging kernel closer to 0 (and the TM5 contribution will be closer to 0), which explains the low values in the contribution factor of the a-priori in the presence of the low cloud field.

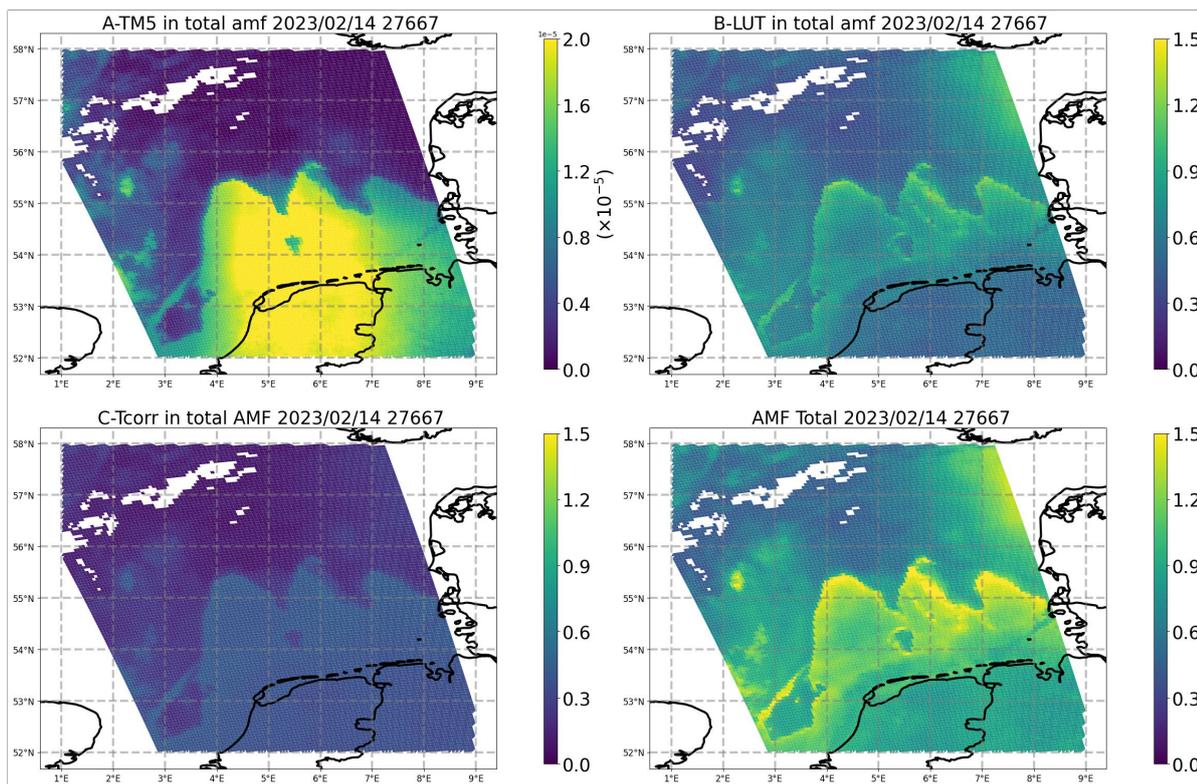


Figure 3.18: Contribution factors to the total air-mass factor for case study A. From top left to bottom right: the TM5-MP contribution (note the scale), the LUT contribution, the temperature correction contribution and the total air-mass factor.

The air-mass factor signal is strongly dominated by the LUT contribution. This contribution analysis highlights the importance of the box-AMF's and the resulting averaging kernels. This contribution of LUT is highest above the North Sea under clear-sky conditions. This high contribution is best understood as a low sensitivity in this area, caused by the dark ocean surface. A reduction in sensitivity leads to a higher air-mass factor. The LUT-contribution peaks near the borders of the clear and cloudy part, indicating more NO_2 -sensitive areas. Since this signal is not present in the LUT clear-sky contribution (figure C.16a in appendix C), it is potentially related to the weighted combination of the cloudy and clear-sky contribution through the independent pixel approximation (formula A.5). The weight in this combination is the cloud radiance fraction. This cloud radiance fraction is related to the effective cloud fraction (formula A.6). The low range cloud radiance fraction panel in figure 3.5 shows indeed similar patterns as observed in the LUT contribution to the air-mass factor, mainly at the borders of the cloud field and north of the Dutch coast. A similar, yet less pronounced, process is seen in case study B (figure 3.19), where the cloud outlines also show a high LUT contribution to the air-mass factor.

The temperature correction contribution can be split between a cloudy and clear part. The reason for this lies in the calculation of the contributions. The temperature correction contribution is calculated separately both for the clear and cloudy air-mass factor. This leads to slightly different values, since they also have a different denominator (formula 2.26). These values are then combined as shown in formula 2.25, which leads to the *crf* clear-cloudy pattern being reflected in the temperature-correction contribution. The temperature variability is low, since the temperature field is downscaled from the $1^\circ \times 1^\circ$ TM5-MP model. The vertical profiles temperature profiles within the two selected areas are shown in figure C.13.

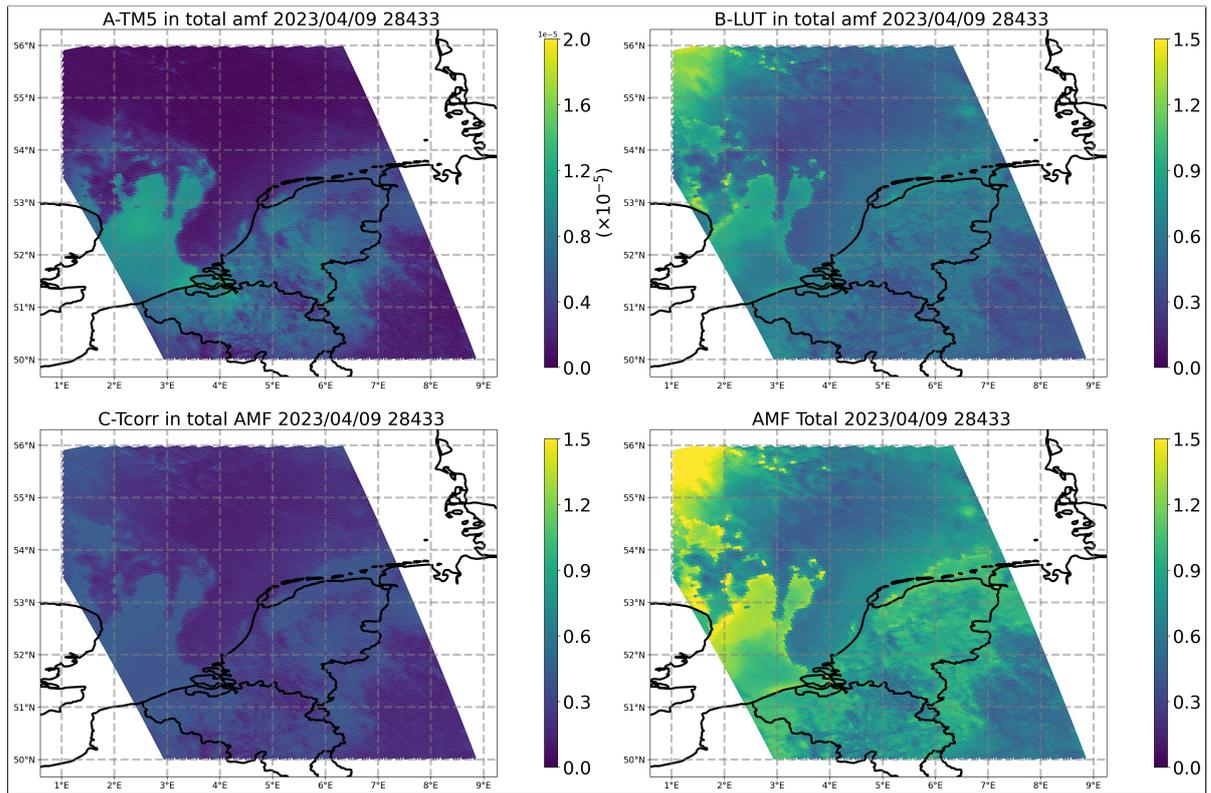


Figure 3.19: Contribution factors to the total air-mass factor for case study B. From top left to bottom right: the TM5-MP contribution (note the scale), the LUT contribution, the temperature correction contribution and the total air-mass factor.

3.6. Intersection gradient

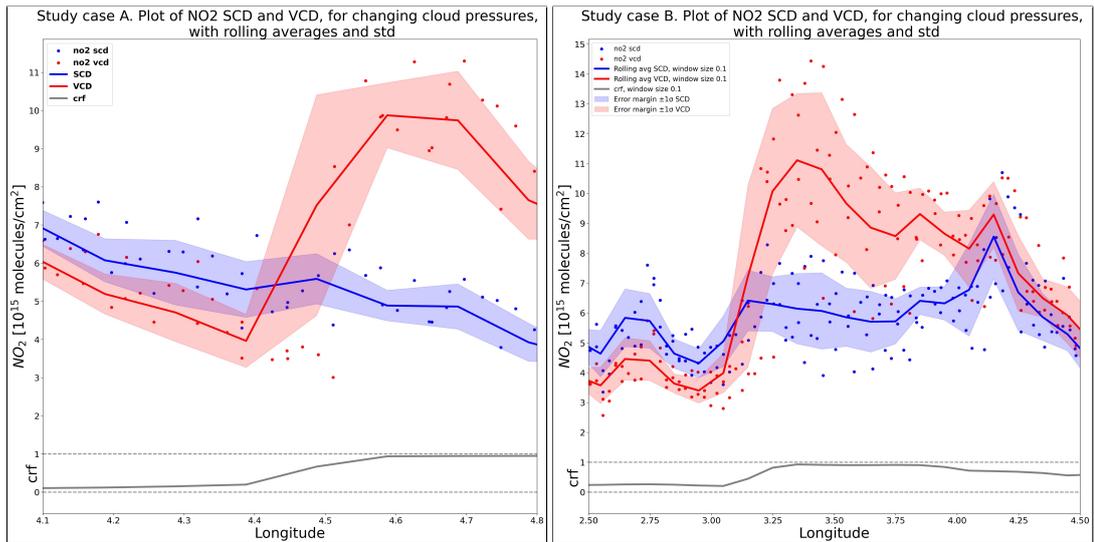
3.6.1. Introducing a cloud pressure bias

Figure 3.20a and 3.20b show the NO_2 slant and vertical column values in the selected intersection for case study A and B, as well as the associated rolling average and standard deviation. The gradients of both case studies show similar behaviour. The vertical column is lower than the slant column over the clear-sky part and vice versa over the cloudy part. Especially this cloudy part shows a strong climb from the slant to the vertical column densities. The vertical column increases strongly over the cloud because of a large assumed ghost column, which causes an upward correction. The vertical column slightly reduces above the clear-sky area, likely because of the *leaking* effect of the cloudy air-mass factor through the effective cloud fraction, which leads to a small increase in total air-mass factor. This is also visible in the average cloud radiance fraction line in the intersection gradients in grey, this line is not exactly 0. Under a fully clear-sky, an increase instead of a reduction is expected when converting the slant to the vertical columns. The physical reason for this expected increase is to counteract the low surface albedo over the sea.

Figures 3.20c and 3.20d show the resulting gradient after adjusting the cloud pressures with ± 50 hPa. An increase in 50 hPa, which for case A corresponds to placing the clouds at the surface level (figure 3.6a), leads to a strong decline in vertical column values. This sets the height over which TM5-MP is integrated to 0. This proves that the ghost column strongly modulates the vertical column densities over the cloud. With the ghost column set to 0, the vertical column densities over the clouded part are lower than the slant column. This is related to the LUT correction due to the enhanced brightness over clouds (correction for the albedo effect of clouds). The box-AMF creates a downward correction here, assuming the sensor overestimated NO_2 density because of the enhanced light reflection paths over a cloud.

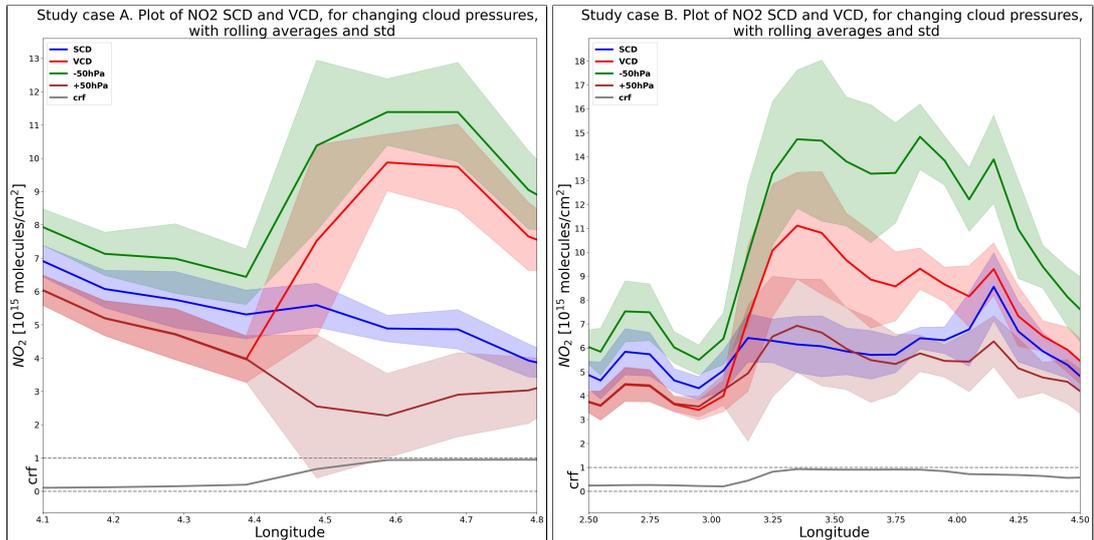
A decrease of 50 hPa, so a higher cloud, leads to a higher vertical column density over the clear sky

part. This can be explained by the small cloud fractions over the clear-sky part leading to a *leaking* effect of the cloudy air-mass factor in the clear-sky part, as shown a few times before in this chapter. Through the small cloud radiance fractions in this clear-sky area, there is also a sensitivity to cloud pressure. The reason a higher cloud leads to a more pronounced upwards correction is likely related to a higher ghost column, meaning the NO_2 field underneath the cloud is larger. This means that the strong decrease in averaging kernels associated with a cloud is transported upwards. This results in a general lower air-mass factor, and thus higher vertical column. The higher vertical column density can also be explained through the averaging kernels. The higher cloud will shift the strong decrease in averaging kernels underneath a cloud, more upwards, which leads to the sum over all the layers being lower, and thus a lower total air-mass factor.



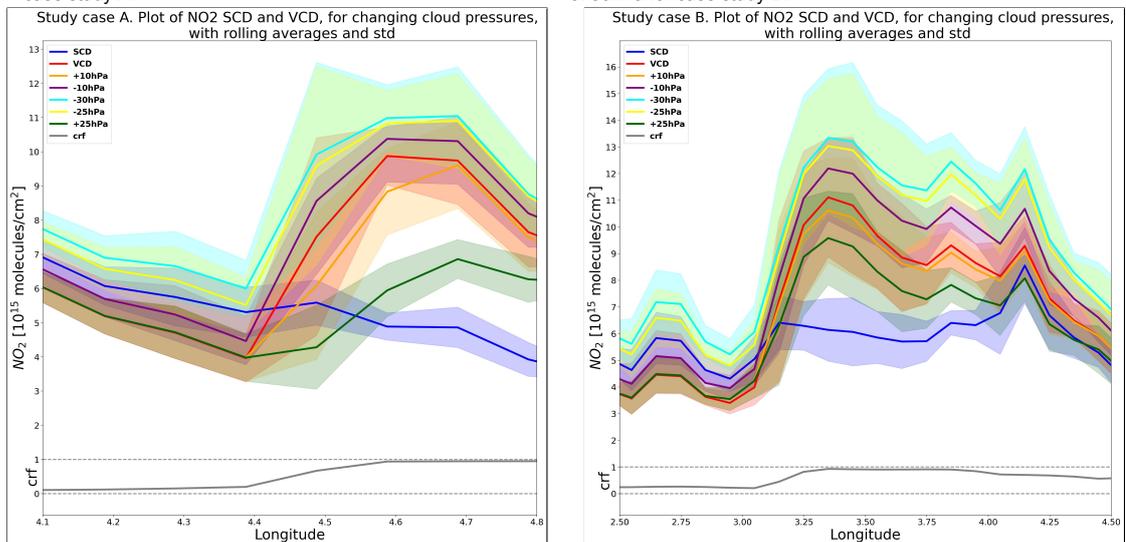
(a) Scatter plots of slant (blue) and vertical (red) column densities in the pixels overlapping with the selected intersection for case study A. The rolling average with a window size of 0.1 in the longitude and latitude dimension is shown in blue and red for respectively the slant and vertical column. The grey line indicates the cloud radiance fraction.

(b) Scatter plots of slant (blue) and vertical (red) column densities in the pixels overlapping with the selected intersection for case study B. The rolling average with a window size of 0.1 in the longitude and latitude dimension is shown in blue and red for respectively the slant and vertical column. The grey line indicates the cloud radiance fraction.



(c) Rolling averages of NO_2 vertical column densities after introducing a positive and negative cloud pressure bias of 50hPa for case study A.

(d) Rolling averages of NO_2 vertical column densities after introducing a positive (brown) and negative (green) cloud pressure bias of 50hPa for case study B.



(e) Rolling averages and standard deviations of NO_2 vertical column densities after introducing positive cloud pressure biases of +10 hPa (orange) and +25 hPa (dark green) and negative cloud pressure biases of -10hPa (purple), -25hPa (yellow) and -30hPa (cyan) for case study A. Rolling average and standard deviation of slant (blue) and vertical (red) column density is also shown.

(f) Rolling averages and standard deviations of NO_2 vertical column densities after introducing positive cloud pressure biases of +10 hPa (orange) and +25 hPa (dark green) and negative cloud pressure biases of -10hPa (purple), -25hPa (yellow) and -30hPa (cyan) for case study B. Rolling average and standard deviation of slant (blue) and vertical (red) column density is also shown.

Figure 3.20: Overview of different intersection gradients under different introduced cloud pressure biases. Note that y-axis is not the same size for every subfigure.

In figures 3.20e and 3.20f, a range of cloud pressure differences is explored. The goal here is to assess whether a certain bias in cloud pressure leads to a more realistic gradient, or at least a gradient which is closer to the slant column densities. Note that a gradient closer to the slant column measurement is not per se more correct since the ground-truth is not known. There are two approaches in creating a less pronounced contrast between the cloudy and clear sky. The first one is by slightly increasing the cloud pressure, as shown in the orange and dark green line in figures 3.20e and 3.20f, corresponding with the rolling average of vertical column densities related to a positive bias of respectively 10 and 25 hPa. This positive bias in cloud pressure leads to a lower ghost column, so a less pronounced correction over the clouds. A different approach is by slightly decreasing the cloud pressure, as shown in the purple, yellow and cyan line in figures 3.20e and 3.20f, corresponding with the rolling average of vertical column densities related to a negative bias of respectively 10, 25 and 30 hPa. This negative bias, so a higher cloud, leads to higher vertical column densities over the clear-sky part. The vertical column densities also slightly increase over the clouded part. Both of these corrections lead to a reduced difference between the clear- and cloudy part. However, they both fail to really tackle the issue. The positive cloud pressure bias reduces both the clear- and clouded part, whereas the negative cloud pressure increases both. A *maximum difference* is defined, which is the difference between the trough and peak of the rolling average. This maximum difference is an indicator of the contrast between the cloudy and clear-sky vertical column. An overview of this peak difference for every proposed change is showed in table 3.1 For case A, the lowest peak-trough distance is achieved with a positive cloud pressure bias of 25 hPa, whereas for case B the more extreme positive bias of 50 hPa leads to the lowest peak-trough distance, followed by the expected FRESKO-CRB cloud pressure error around 50 hPa, the 25 hPa cloud pressure bias lies well within the error margins and leads to the lowest cloudy - clear-sky contrast for both case studies.

3.6.2. Pure clear-sky air-mass factor

Another alternative gradient is introduced in figure 3.21. Here, the clear-sky air-mass factor is applied. This implies neglecting the cloud effects, i.e. ghost column and enhanced sensitivity. The clear-sky air-mass factor is shown in figures 3.16 and C.9 for case A and B, respectively. The clear-sky air-mass factor strongly leans on the simulated DLER surface albedo climatology over the clouded areas, due to the absence of an actual surface albedo measurement. Note that over a clear surface the DLER climatology can be slightly adapted to match the measured surface albedo (chapter A, section A.8). The clear-sky air-mass factor increases the slant column gradient over the entire gradient. The clear air-mass factor assumes a darker ocean surface and thus measurements that are not sensitive to NO_2 . This leads to a slight increase to correct for this darker surface. The a-priori has a weaker effect in the clear-sky air-mass factor, as shown in 3.18, because the averaging kernel profile shows a linear decrease in the case of a clear sky. There is likely no differentiated effect of the a-priori over this intersection, taking into account the low spatial resolution of $1^\circ \times 1^\circ$ in TM5-MP, as shown in 3.8. Applying solely the clear-sky air-mass factor introduces a lower peak-through difference than for the introduced cloud pressure biases, for both cases, as shown in table 3.1. The slant column peak-trough difference is still lower than all of the introduced biases. Note that a lower gradient does not imply a per se more realistic gradient, since the true gradient is not known.

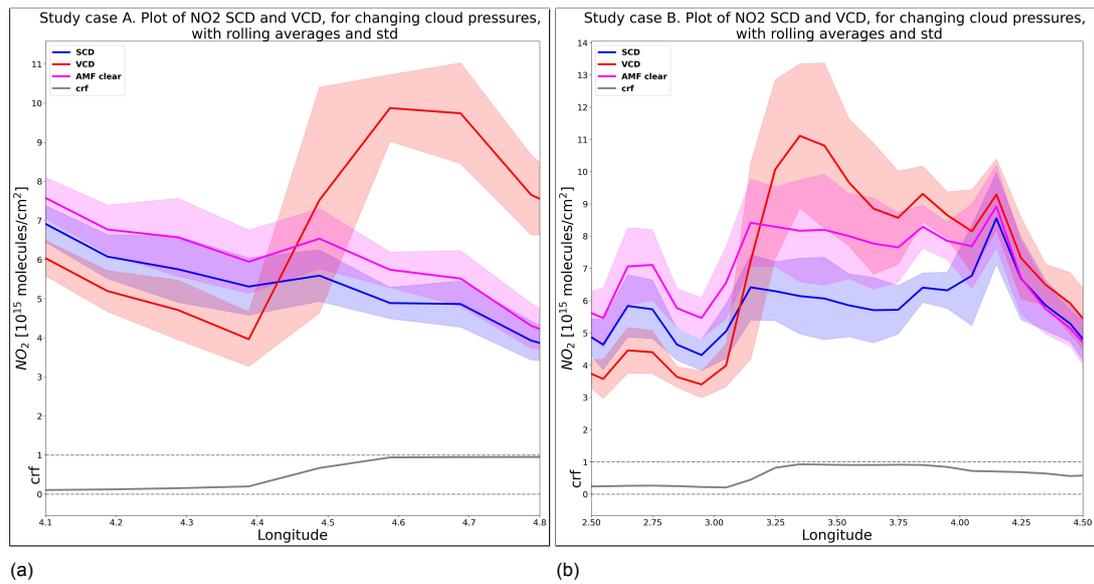


Figure 3.21: Both for case study A and B: rolling average and standard deviation of the vertical column density after applying the pure clear-sky air-mass factor (pink) together with the rolling average and standard deviation of the slant (blue) and vertical (red) column density under the original combined air-mass factor

	Case A Minimum Difference	Case B Minimum Difference
SCD	2.04	3.52
VCD	5.94	7.93
-50hPa	4.98	9.38
+50hPa	3.70	3.66
+10hPa	5.43	7.40
-10hPa	5.74	8.48
-30hPa	4.93	8.21
-25hPa	5.18	8.27
+25hPa	3.21	6.38
AMF Clear	2.13	3.51

Table 3.1: Table showing the maximal peak-trough differences for the different intersection gradients defined.

3.6.3. Spatial analysis of introduced changes

Figure 3.22 shows four maps of NO_2 vertical column densities relating to four introduced changes in the above-mentioned intersection gradients. The resulting tropospheric vertical NO_2 column with an introduced cloud pressure bias of +10 hPa, +25 hPa and -25 hPa is shown (first three panels). The resulting NO_2 field when only considering the clear-sky air-mass factor is also shown. The associated relative differences with the original retrieval are plotted in the bottom 4 panels. Setting the clouds lower, leads to an overall decrease over the clouds and no changes over the clear-sky part, corresponding with the intersection gradients. The relatively small bias of +10hPa already shows a quite different field, with the enhanced values in the presence of clouds much less present. Note the double effect caused by a lower cloud: the reduction of the ghost column and a shift in vertical sensitivity, as the lowering of a very bright surface necessitates a downward adjustment of everything measured above it. The negative bias in cloud pressure, shows a field where almost all of the values are corrected upwards. In the bottom left panel *relativediff -10*, which shows the relative difference after decreasing the cloud pressure with 10hPa, it is visible how the cloud border is not influenced by this correction (white lines following cloud border). The reason for this is not clear. It is also noteworthy how in case B, shown in figure C.17, a positive cloud pressure bias of +10 hPa (the *relative diff +10* panel) decreases the emission plume but increases the background signal. Since TM5-MP does not simulate on the scale of single emission plumes (see figure 3.8), an a-priori effect is not the cause. The +10hPa lower cloud slightly increases the air-mass factor, which in its turn leads to a lower vertical column compared to the original retrieval. In the areas of the highest slant columns (the plume), this leads to a negative difference compared to the original retrieval. The background signal, with a lower measured slant column, still slightly increases compared to the original retrieval. So the +10hPa cloud pressure bias shows an intermediate effect. The same panel also shows a decrease distinctively following the outline of the cloud. This decrease in the emission plume and cloud outline is especially of interest because it decreases the spurious effects in the retrievals. So a cloud pressure bias of -10hPa decreases the spurious spatial pattern for case study B.

Finally, applying only the clear-sky air-mass factor leads to an increase of the vertical column densities over the clear-sky part and a decrease over the cloudy part. The reason for this dichotomy is not entirely clear, as it is only present above the water surface in the North Sea and the IJsselmeer. The decrease is not present above the North Sea in areas with a distinctive spectral shift (as seen on 3.1a). This is seen in the white patches seawards of the Dutch coast on the bottom right panel in figure 3.22. The reason for this is likely related to very low cloud radiance fraction causing a *leaking* effect of the higher cloudy air-mass factor in this area over the North Sea. The pure clear-sky air-mass factor is lower in the clear-sky part in the North Sea than the combined air-mass factor. This effect is present in the entire clear-sky area in the North Sea, except from a small area in the middle (white patch around 54 °N, 4.5 °E in figure 3.22) where the difference is around 0. Above the clouded parts, the clear-sky air-mass factor is slightly higher than the combined air-mass factor (figure 3.16), which leads to a lower vertical column. The higher air-mass factor is caused by the cloud shielding effect being omitted in the clear-sky assumption, so the a-priori is multiplied with higher averaging kernel values in the low troposphere.

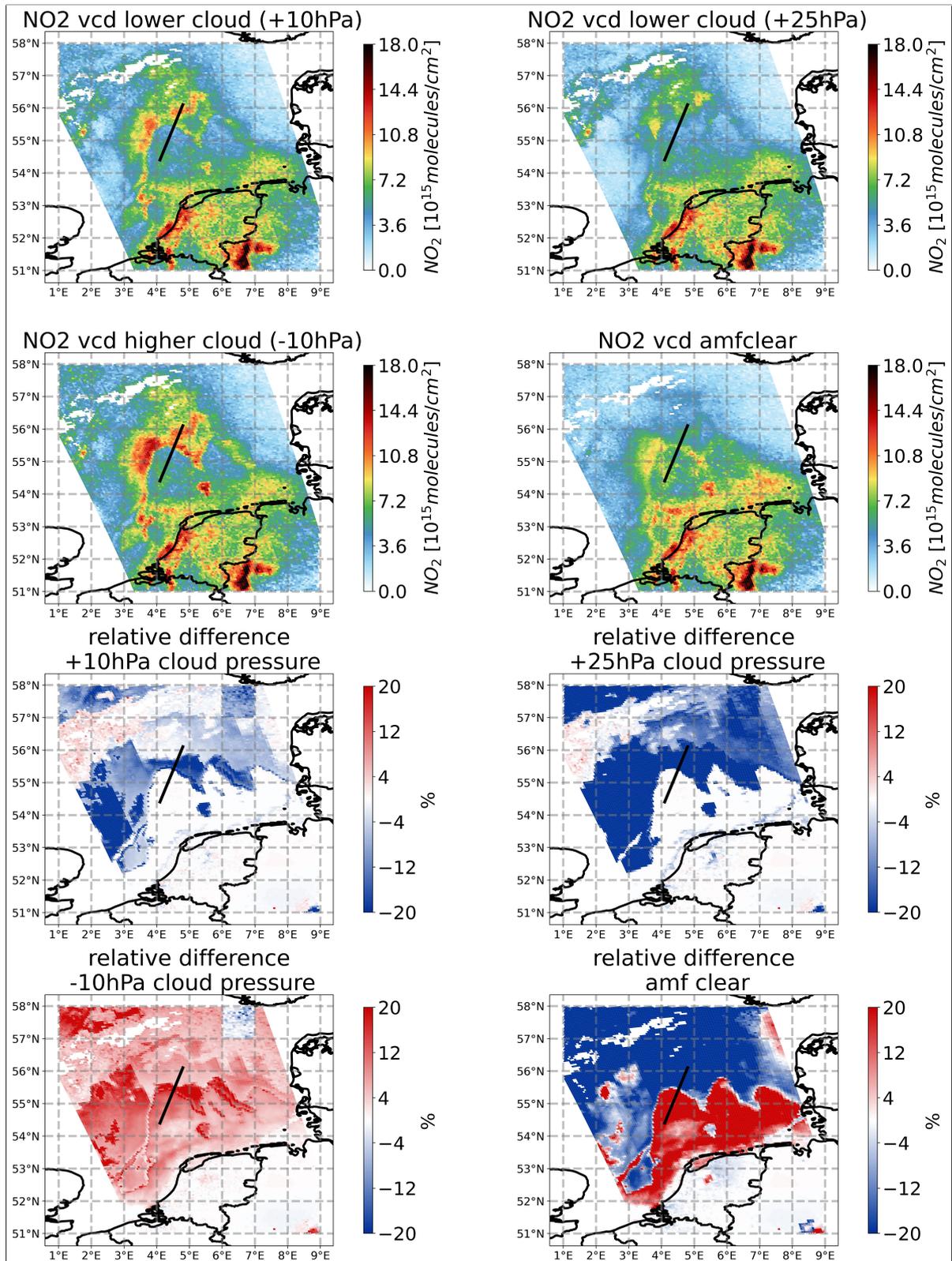


Figure 3.22: Case study A. From top left to bottom right: the NO_2 vertical tropospheric column after applying a positive cloud bias of 10hPa and 25hPa, a negative bias of 25hPa and the pure clear-sky air-mass factor followed by respectively the 4 relative differences under these alternative NO_2 vertical columns. The intersection is shown in black.

3.7. Reflectance test

An important dependency of the air-mass factor is the surface albedo, which drives the ground-based reflectivity response (encapsulated in the box-AMF) and consequently the averaging kernel. As explained in chapter A and 2, the surface albedo in the NO_2 window is simulated through the DLER climatology. The resulting surface albedo is shown on the top left panel in figure 3.5 (and C.3 for case B). The surface reflectivities are for both cases rather low and do not show strong variability. This lies within the expectations of a darker surface (except in the presence of snow or ice) and some small fluctuations relating to e.g. land use, sun and satellite angle. The North Sea is brighter and shows more fluctuations in the NO_2 window, as compared to the FRESCO O_2 A-band (top right panel in figure 3.5). Another important aspect is the cloud radiance fraction, which is the weight to combine the clear and cloudy air-mass factor in formula A.5. The cloud radiance fraction is dependent of the surface and cloud radiance (formula A.6). These radiances are estimated through the cloud look-up table. The low range cloud radiance fraction pattern visible in the fourth panel of figure 3.5, is likely to be effected by the DLER surface albedo, which is one of the input variables to determine surface and cloud radiance through a look-up table. The uncertainty of the simulated DLER surface albedo thus propagates to the cloud radiance fraction, which is a significant factor leading to spurious retrievals as discussed earlier. This effect has not been further explored quantitatively, since this would require an in-depth analysis of measured against simulated surface reflectance, which is related to the Level-1 TROPOMI processing. But a visual comparison of the simulated surface albedo in the top left panel of figure 3.5 with the VIIRS imagery, figure 3.1a, clearly shows how the patterns visible in the North Sea are not reflected in the simulated albedo (at the 440 nm wavelength). This likely plays a significant role in the misinterpretation of the retrieval algorithm of low cloud radiance fractions over the North Sea.

The Wald-Wolfowitz test for consistent deviations of the measured with the simulated reflectance has been done. This test does not focus on a mismatch in surface reflectances, but tries to detect an overall mismatch in the entire DOAS fit. The result of the Wald-Wolfowitz runs test is shown in figure 3.23, this shows a spatially random noise. If important absorbers in the scene would not be taken into account in the DOAS-fit (Differential Optical Absorption Spectroscopy), spatial structures in the longest run deviation could be seen, such as a spatially coherent area of increased test values over the sea or low clouds. The observed random noise implies that this is not the case and the errors in the retrieval scene are not driven by a wrong assumption in the DOAS fit.

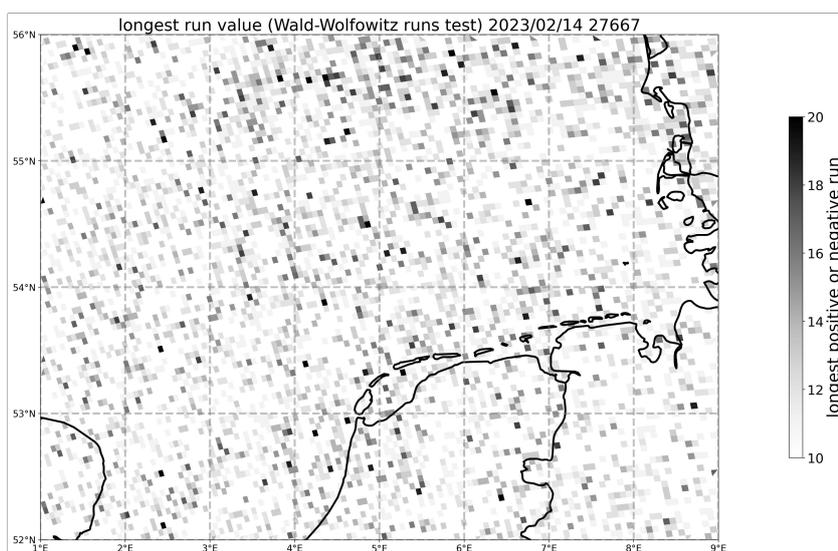


Figure 3.23: case study A. Result of the Wald-Wolfowitz runs test. The plotted value is the longest positive or negative run deviating from the expected absorption spectra, calculated for every pixel.

4

Discussion

The results showed how the a-priori, air-mass factor calculation and the cloud characterization influenced the retrieval in the presence of a low cloud field. The objective of this thesis is to formulate an answer to the question of how the air mass factor affects TROPOMI NO_2 retrievals over low clouds and fog. The air-mass factor takes into account different dependencies. Three main lines of dependency have been explored and will be further discussed here: the a-priori effect, the air-mass factor calculation and the cloud characterization. These lines are not always clearly distinct from each other and influence each other. However, an answer on the research question shall be formulated mainly following these lines. The current effect of each dependency on the retrieval algorithm over low clouds and fog is discussed, as well as its current limitations and implications. Potential next steps to approach these limitations and to work towards an operational retrieval product above low clouds and fog are then laid out in chapter 6 (Recommendations).

4.1. A-priori

4.1.1. TM5-MP

The retrieval algorithm makes use of an a-priori NO_2 field. Both the horizontal and vertical resolution matter. The horizontal resolution of TM5-MP is low, compared to the TROPOMI nominal resolution, and shows a smeared out pollution field. The scale on which low clouds and fog fields are present and interact with the NO_2 field, is not captured in the a-priori. The industrialised areas of the Rhine-Ruhr region and the Rotterdam-Antwerp harbours are the main sources of tropospheric NO_2 emission in TM5-MP. This is then transported following a dominant wind direction, taking into account the NO_2 lifetime in the atmosphere and more complex chemical processes. Sharp outlines and large variations within the scene are not seen in the TM5-MP NO_2 field.

The TM5-MP profiles, interpolated to the different TROPOMI pixels, do not show much variation, which corresponds to the low variation in the horizontal dimension. In the two case studies, TM5-MP defines three layers in the proximity of the clouds. For case study A, one layer is below, one is approximately at the cloud height and one is above the cloud. For the second case study, more layers are situated beneath the cloud. As explained in the previous chapter, the *shape*, rather than the absolute values, of the profile is what drives the response of the air-mass factor. For both case studies, the a-priori profile tends to model a strong peak in the first layer (65m-220m), followed by a gradual decline with height. This peak is much more pronounced for case study A, where a more polluted situation is modelled. Besides from the strength of this peak in the first layer (65m-220m), the profiles of both case studies are similar. The more the a-priori is situated underneath the cloud, the lower the air-mass factor and the higher the vertical column. This highlights the critical role of the interaction between the NO_2 simulated profile and the FRESKO-CRB cloud pressure. However, the first layer ranging from 65 to 220 meters already covers a significant part of the lowest troposphere (155 meters). The type of low clouds and fog of interest here is situated in those lowest few hundred meters of the atmosphere. So there is

usually only one data point in the vertical proximity of the low clouds, and then one or two data points below the cloud. As seen in the vertical profiles, whether a TM5 layer is positioned above or below the cloud height can shift the averaging kernel from about 1.5 (above the cloud) to 0.5 (underneath the cloud), which will strongly affect the air-mass factor. Note that the averaging kernels are defined in the same vertical resolution as the a-priori profile. The combination of the low averaging kernels in the lowest troposphere with a high a-priori weight (because of the strong peak) dominates the resulting total air-mass factor. The stronger their combined effect, the more upwards a correction is given to the vertical column.

Completely averaging out the effect of the a-priori decreases the spurious spatial pattern (figure 3.17), because a vertically flat a-priori means that all the box-AMF's are assigned to same weight. This will lead to little difference being made between the cloudy and clear part, because without the high a-priori weights (caused by the peak) the averaging kernels in the low troposphere no longer dominate the total air-mass factor. The entire troposphere gains the same weight, which leads to a spatially more uniform correction, so the cloud effects in the low troposphere are no longer dominating the tropospheric air-mass factor.

The a-priori dependency is parameterized in the so-called ghost column. This ghost column is defined to visualise what the sensor did not see in a certain retrieval. The ghost column in itself is not directly used in the retrieval algorithm but is integrated in the air-mass factor calculation. As shown in the scatter plot in figure 3.10, a higher ghost column leads to a higher slant-vertical correction. This dependency mostly holds for case study A, caused by a strong peak of the profile shape beneath a very low cloud field. The ghost column throughout the area of interest is high for both cases, with ghost column NO_2 densities ranging up to half of the final tropospheric vertical column, which means that the retrieval algorithm assumes the instrument only measured half of the tropospheric column. The cloud pressure spatial pattern can be distinguished in both ghost column products. This highlights the low spatial variability of the NO_2 field in TM5-MP, with most variations in ghost column being directly related to variations in cloud height, because the integration height changes (rather than changes in a-priori).

The low horizontal resolution, which leads to little spatial variation in the vertical profiles, leads to the a-priori being a near-static dependency throughout the entire retrieval scene. The effect of the a-priori is mostly indirect, in that it amplifies the dynamics of other higher scale assumptions in the retrieval algorithm, like e.g. the averaging kernels underneath and above the clouds.

4.1.2. CAMS-Europe

The CAMS-Europe ensemble model, with a resolution 10 times higher than TM5, was explored and subsequently used as a-priori for an alternative NO_2 tropospheric retrieval. This CAMS-Europe model succeeds in simulating more localized emission sources and their associated plume transport. Atmospheric transport is solved in detail, with figure 3.11 clearly showing a complex transport of the emission plume onto the North Sea, as opposed to only a general direction of transport visible in the TM5-MP field. Another interesting aspect about CAMS-Europe is that it also shows pollution remnants of the previous day above the North Sea. This is interesting for the case of a low cloud field above the North Sea, which is mainly present in the winter time when NO_2 has a longer atmospheric lifetime. The cloud field could also be persistent for multiple days and thus the troposphere above the cloud might also have a remnant of the pollution of the previous hours/days.

The vertical profile of CAMS-Europe has more vertical layers in the lower troposphere than TM5-MP, with about four layers in the first few hundred meters of the troposphere. This is one more layer compared to TM5-MP. The CAMS profiles in the low troposphere have much more inter-variability in shape compared to the TM5-MP profiles. As mentioned above, these lowest layers are the most important in the low-cloud type of cases and weigh the strongest on the AMF. This higher variability in the lower troposphere with distinctively different shapes of vertical NO_2 distributions indeed leads to a NO_2 vertical column that has a much more spatially diverse correction, compared to the vertical column using TM5-MP as a-priori, as can be seen on figures 3.14 and 3.15.

The difference in using CAMS-Europe is mainly that for using TM5-MP the cloud characteristics were driving the correction. The peak of TM5-MP is always at the same height which places it underneath most low clouds. This results in a very dichotomous cloud-clear correction. Using CAMS some vertical profiles peak more above the cloud and some more below, with not just one fixed height of the profile peak but an elevated profile with several layers showing heightened concentrations, which relates directly to the higher horizontal variability, where more polluted grid cells correspond with a more elevated profile. A more elevated profile positively influences the total air-mass factor through the assigned weights, and thus negatively influences the slant-vertical correction. The NO_2 product using CAMS-Europe as a-priori is less dominated by only cloud height and an almost-fixed profile, but has both variations in profile and in cloud height interacting.

This is seen in the small, detached cloud south of the larger cloud field in case study A, where the CAMS-Europe profile models a bulk of the pollution either at the same level or above the cloud. This leads to a lower correction between the slant and vertical column. The algorithm assumes it already sensed a larger proportion of the NO_2 . For case B, the higher cloud field still leads to the bulk of the pollution being simulated underneath the cloud, which could be a realistic approximation. This leads to a less distinct difference when using CAMS as a-priori for case B. The more pronounced peak in CAMS for case B even results in slightly higher vertical densities.

This spatially more diverse correction does not directly solve the case of the spurious observations, as for both cases the NO_2 vertical column still closely follows the cloud outlines. This is because the CAMS-Europe model at $0.1^\circ \times 0.1^\circ$ still does not capture the specific dynamics related to the trace gas behaviour underneath, inside and above a cloud field. The relative height of the plume and the shape of the profile will influence the air-mass factor near a cloud, but this a-priori is not directly linked to the cloud characteristics. It is important to note that CAMS-Europe is also not designed to capture dynamics on a cloud-scale. The retrieval algorithm uses surface and atmospheric characteristics to determine a profile of vertical sensitivity through the box-AMF's, and then it takes an a-priori to weigh the vertical box-AMF's. However, this a-priori is not based on the same atmospheric conditions as the box-AMF's. So a mismatch exists between the averaging kernels on the one hand, which try to parameterize the specific vertical sensitivity of a scene based on local information, and the a-priori, which simulates atmospheric transport of pollutants from emission sources on a large scale. The a-priori might simulate unrealistic NO_2 profiles in the presence of low clouds. The current mismatch in scale between the box-AMF's and the a-priori drives the strong contrast observed between clouds and clear-sky.

The true vertical profile of NO_2 will be influenced by the presence of a low cloud through dynamic and photochemical effects. To succeed in a more realistic retrieval, the vertical behaviour of NO_2 underneath, inside and above a cloud needs to be characterized on the cloud-scale. This allows for the averaging kernels to be weighed in a more appropriate manner. Potential first steps on how to better approach the NO_2 a-priori profile, with the goal of a better NO_2 retrieval over low clouds and fog, are laid out in chapter 6. A better understanding of these case studies can be achieved through a more extensive analysis of the meteorological conditions. The situation in case A seems to be dominated by relatively cold air coming from the land. A moister and warmer air-mass is observed over the North Sea for several days before. A colder and drier air-mass from land pushes the moist air-mass more northwards under influence of the seaward winds. These two different fields interact with the NO_2 concentration in a different way and dictate the trace gas transport in the troposphere. It is likely to assume they will have a different vertical NO_2 contribution. A part of this signal is seen in the slant column of case study A (figure 3.3a). There seems to be a cloud-clear gradient in the measured slant column, which is especially visible in the western part of the cloud-clear border. This could potentially be related to two different air-masses being characterized by different NO_2 concentrations. However, to validate these impressions, an extensive analysis through for example the mesoscale meteorological Harmonie model [38], which has a resolution of 2.5 km, could be an impactful first step.

4.2. Air-mass factor calculation

The analysis of the air-mass factor calculation showed that the calculation process is dominated by the box-AMF's derived from the look-up table. They show a sort of gridded spatial character, which is smoothed out after weighing with the a-priori. This gridded spatiality is likely a consequence of the interpolation process, based on 6 different input variables. These 6 input variables are: viewing zenith angle, sun zenith angle, surface albedo, surface pressure, cloud fraction and cloud pressure. The variables showing strongest variation throughout the scene are the surface albedo, cloud fraction and cloud pressure. These variables are likely to drive variation in the box-AMF. Uncertainties in the surface reflectivity climatology and cloud characterization is through the look-up table which determines the box-AMF's. The cloud radiance fraction is calculated through the radiance from the cloudy part of the pixel and the clear part, which are both determined through a separate look-up table for clouds, based on the viewing geometry, surface albedo and pressure and cloud albedo and pressure. Through this cloud look-up table, the cloud radiance fraction itself is also a function of the simulated surface reflectivity.

The cloud characterization influences the air-mass factor calculation through two variables: the FRESCO-CRB cloud pressure and the cloud radiance fraction in the NO_2 window. The look-up table does not explicitly deal with the presence of a cloud. It calculates both a cloudy and clear-sky air-mass factor and combines them through the independent pixel approximation. The cloud pressure is taken as the new surface pressure in the cloudy air-mass factor, which means the Lambertian reflector, normally the surface, is set higher in the atmosphere at the pressure level where FRESCO-CRB measured the middle of the cloud. Note that this Lambertian reflector has zero transmittance. The cloud height has a strong effect, since the box-AMF's are very sensitive to this cloud. They will strongly increase above (albedo effect) and decrease below (shielding effect) this cloud. The box-AMF's do not decrease horizontally to 0 underneath a cloud because the cloud radiance fraction is (for both of the case studies) never exactly 1 in a single pixel. A more complex cloud characterization, with the cloud no longer being a Lambertian reflector with a fixed albedo, would also lead to the averaging kernels, determined through the box-AMF's, exhibiting more complex vertical behaviour in the presence of a cloud.

The cloud radiance fraction is also not directly used in the look-up table, but serves as a weight to combine the cloudy and clear box-AMF. The results showed this box-AMF being significantly impacted by low cloud radiance fractions over the North Sea, which led to an air-mass factor higher than the pure clear-sky air mass factor over the unclouded part of the North Sea. This higher air-mass factor is caused by the cloudy air-mass factor *leaking* into the combined factor over the clear areas.

The impact is also seen in the intersection gradients, where the introduction of a pure clear-sky air-mass factor instead of the combined one, increased the NO_2 densities over clear-sky. This *leaking* process contributes to the spurious observations, as it lowers the vertical column in the clear-sky area, further exacerbating the contrast between the low clear and high cloudy NO_2 densities. The cloud pressure map in figure 3.6 showed very low FRESCO-CRB cloud pressures over the clear-sky area, this means the ghost column here is 0, as is also seen in figure 3.9. This explains why the *leaking* effect of the cloudy air-mass factor in the clear-sky part exerts a purely negative effect on the vertical density. The algorithm assumes a cloud at the surface here, so it only implements a downward correction to counteract the albedo effect.

These results correspond with earlier results of *Liu et al. (2021)*, as shown in figure 4.1. The top left panel shows an increase in the uncertainty for all three defined uncertainty drivers (surface albedo, cloud pressure and profile shape) for low surface albedo values [39]. The bottom left panel shows how the lower cloud fractions relate to an increase driven in the uncertainty driven by the surface albedo, which also corresponds to the findings here. Note that a threshold of cloud radiance fraction below 0.5 roughly matches a (geometric) cloud fraction threshold of 0.2 [39]. The top right panel shows the highest uncertainty related to profile height (see definition in figure 4.1) at a profile height of around 840 hPa. This means that more polluted situations, so more elevated profiles, lead to higher uncertainties. Even though profiles peaking this high were not present in the case studies regarded here, very high

uncertainties in the AMF are more related to even lower peaks in the cases studied here. Note that the analysis of *Liu et al.* focused on entire Europe. Given the focus of this study on low clouds, it is reasonable to conclude that a lower profile peak beneath the low clouds leads to a peak in uncertainties in the air-mass factor.

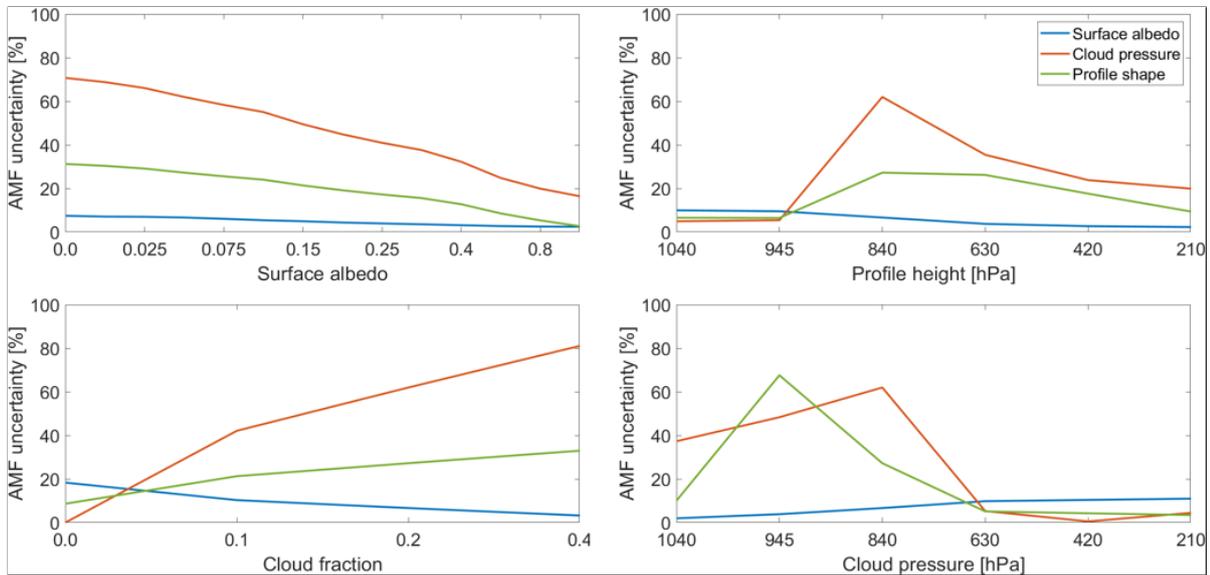


Figure 4.1: Tropospheric AMF uncertainties (Y-axis) related to the surface albedo, cloud pressure, and a-priori profile errors (x-axis) calculated in the study of *Liu et al. (2021)*. The surface pressure is set to 1050 hPa, the surface albedo is 0.05, the profile height is 840 hPa, the cloud pressure is 840 hPa, and the cloud fraction is 0.2 [39]. Profile height is defined as the altitude (pressure) below which 75 % of the integrated a-priori profile [39].

As seen in figure 3.5, the surface albedo in the 440 nm wavelength is not simply a black surface. The surface reflectivity at 440nm shows significant variation. Different spectral characteristics of the sea surface are seen in the VIIRS image, likely caused by sediment fluxes and bio-geo-chemical processes, for example algae or turbulent fluxes in the water. It is unsure whether the DLER climatology succeeds in capturing these dynamics. A comparison of the measured surface reflectance at 440 nm wavelength with the simulated surface reflectance could be a potential next step to assess this. This comparison would require an atmospheric correction on the measured reflectance, in order to properly distinguish the surface contribution from the atmospheric contributors. This was not within the scope of this thesis.

However, a first exploration of inconsistencies between the modelled and measured reflectance spectra has been done. This is done through the Wald-Wolfowitz runs test, which did not show large spatial inconsistencies. This test is set up to focus on an atmospheric spectral mismatch, related to the expected NO_2 absorption spectra. However, it can clearly be seen in both the optical VIIRS pictures and the cloud radiance fraction, that structures in the North Sea, which are not related to clouds but rather to a different colour of the water are present. These structures seem to be interpreted by the retrieval algorithm as clouds through the low cloud radiance fraction. An enhanced understanding of the spectral characteristics of the North Sea could perhaps mitigate the spurious low cloud radiance fractions. This could also be combined with a low cloud radiance fraction threshold encapsulated within the independent pixel approximation, where e.g. cloud fractions lower than 0.2 no longer influence the combined air-mass factor.

4.3. Cloud characterization

Apart from the cloud fraction, cloud height also plays an important role in driving uncertainties. The different intersection gradients, where a range of positive and negative cloud biases are introduced, strongly impacted the NO_2 vertical densities. These changes introduced an up- or downward shift in the

Lambertian reflector, which shifted the sensitivity profile in the lower troposphere. A positive bias of +25 hPa seemed the most reasonable shift in cloud pressure that lead to a lower contrast between the clear and cloudy part of the intersection. Among the different cloud heights investigated, a key observation is the high sensitivity. Cloud height dominates the shape of the averaging kernel profile, which is then weighted by an a-priori profile which barely changes over the wider area. The Lambertian cloud has a large error for low clouds of around 50 hPa, which is comparable to the cloud thickness itself [39]. So the biases lie well within the expected error margin of FRESKO-CRB cloud height. This means a trustworthy retrieval is hard to obtain under the current operational setting. This is also reflected in figure 4.1, where low cloud pressures show an increased AMF-uncertainty driven by cloud pressure, which corresponds to the higher expected error in FRESKO-CRB cloud pressure for low clouds, and an increased uncertainty driven by profile height, which can be explained by uncertain estimations of the a-priori in the presence of a low cloud. Figure 4.1 does show a decrease in uncertainty towards the lowest cloud cloud heights. Low clouds and fog over the North Sea can still be present at these low heights, indicating that, based on the cases examined here, the uncertainty would increase rather than decrease.

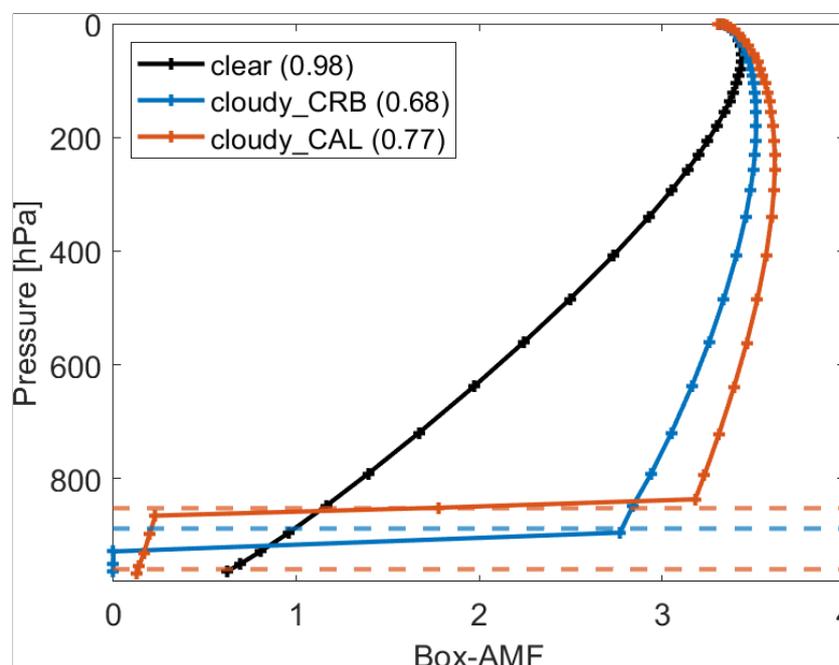


Figure 4.2: box-AMF's for a clear scene and a cloudy scene using both the CRB and CAL assumption. This is an example case retrieved over Munich in Germany on 5 February 2019. The CRB cloud pressure is shown as the horizontal blue dotted line, the CAL cloud top and base pressure is shown as the horizontal dotted brown lines. The total tropospheric air-mass factor is shown in brackets for each situation. [39]

Through the shielding and albedo effect asserted by the Lambertian cloud, the complex light paths inside and on top of the cloud are neglected. In both case studies the majority of the NO_2 profile is at the same height of the cloud, so some degree of cloud-gas interaction is expected. Neglecting this interaction makes the retrievals more uncertain. This also relates to the need for a more intricate understanding of the a-priori profile near clouds. Figure 4.2 shows the averaging kernels defined for both a cloud as a Lambertian reflector and a cloud as optically uniform layers of scattering water droplets (volume scattering cloud - CAL) [39]. The orange averaging kernel profile (CAL), which makes use of the cloud top and base height, leads to a higher air-mass factor. This means the CAL-cloud leads to a lower slant-vertical correction (and a lower contrast between the cloudy and clear-sky vertical densities). Further steps in exploring the implementation of a cloud as a volume scatterer through the FRESKO CAL-approach and how this would relate to more realistic a-priori profiles in the presence of low clouds will contribute to a better retrieval.

4.4. Synthesis

The whole is greater than the sum of its parts. The shape of the a-priori, the sensitivities to surface albedo and cloud radiance fraction in the look-up table and the cloud characterization act as communicating vessels and lead to a highly uncertain vertical column retrieval over low clouds and fog. To create a concise and understandable overview, the most important parameters and their effect on the retrieved vertical column is given in table 4.1. This table gives a general overview of the different variables and their effect. It is assumed that all other variables stay constant if one fluctuates. This may not be true for every individual case, but it does capture the general behaviour, as seen in the two case studies above the North sea. The different parameters exert either a negative or positive effect on the slant-vertical correction. A positive effect (denoted +) means that an increase in this parameter will lead to a higher slant-to-vertical correction. A negative effect (denoted -) means that an increase in this parameter will lead to a lower slant-to-vertical correction. For the a-priori profile, an increase refers to a more elevated profile.

Parameter	Slant-Vertical Correction (VCD-SCD)
Total air-mass factor	-
A-priori profile	-
A-priori column density	-
Total AMF LUT	-
Cloud radiance fraction*	-
Cloud height	+

Table 4.1: Overview of different parameters and their effect on the slant vertical correction. Note that a negative effect does not mean the SCD-VCD difference will be negative. For the a-priori profile, an increase is defined as an increase in profile height, so a more elevated profile with a higher peak. A-priori column density refers to the total tropospheric NO_2 column in that pixel. Cloud height is defined in meters and not in pressure. A higher cloud height is the same as a lower cloud pressure. *:it is assumed the cloud radiance fraction increases over the unclouded North Sea, where the cloud pressure is very close to 0.

5

Conclusion

This thesis explored retrievals of NO_2 over low clouds and fog over the North Sea. Earlier research indicated that the presence of a bright surface could serve as an opportunity for higher-resolution sensing of NO_2 . The main research question was: how does the air-mass factor affect NO_2 retrievals over low clouds and fog. A better understanding of the effect of the air mass factor on retrievals can lead to better retrievals. In order to formulate an answer, two case studies over the North Sea were analysed in depth, 14/02/2023 and 04/09/2023. The main conclusion is that spurious retrievals over low clouds result from a combination of insufficient cloud characterization, an undersampled a-priori with a low peak just above the surface, and an incomplete parameterization of the sea surface's spatial reflectivity patterns.

Initial exploration of the NO_2 field for these two study cases, showed a spurious result, with a high contrast in trace gas density closely following the cloud outlines. This contrast was strongly amplified after the third step of the retrieval algorithm, applying the air-mass factor. To better understand how this air-mass factor influenced the conversion from a slant to a vertical column, three main lines of dependency were defined: the a-priori NO_2 profile, the air-mass factor calculation and the cloud characterization.

The a-priori NO_2 profile plays a crucial role in the accuracy of NO_2 retrievals. Through the profile height, the a-priori weights the vertical averaging kernels. The low profile height leads to a high weight in the lowest layer. Whether this profile is situated beneath or above the Lambertian cloud, significantly impacts the total air-mass factor and, consequently, the vertical NO_2 values. Additionally, the AMF calculation itself, which depends on multiple parameters including the simulated surface albedo and cloud radiance fraction, was shown to introduce spurious spatial patterns over the North Sea, which exacerbated the cloud-clear contrast in retrieved NO_2 densities. Lastly, cloud characterization, parameterized through cloud radiance fraction, measured in the NO_2 retrieval window, and cloud pressure, measured with the FRESKO-CRB algorithm, strongly affects the air-mass factor. Since the cloud is treated as a Lambertian reflector, the height of this surface shifts the entire sensitivity profile of the air-mass factor.

The contrast in NO_2 density following the cloud outlines can be explained as follows: anomalous cloud radiance fractions in clear-sky areas of the North Sea lower the vertical column densities, and the strong a-priori peak beneath the actual low cloud field, increases the vertical column densities. In conclusion, four main causes of the spurious retrievals, which are currently inhibiting an operational use of retrievals over low clouds and fog over the North Sea, are identified:

1. The a-priori profile, characterized by a low spatial resolution, is simulated with a strongly pronounced peak right above the surface in the lower troposphere. In the presence of clouds, this leads to a consistent low air-mass factor and high vertical column densities. This positively amplifies the retrieval over clouds. This also raises questions whether it is realistic to simulate the

majority of the trace gas below the low cloud.

2. The North Sea exhibits spatial patterns in reflectivity, which affects the cloud fraction in the NO_2 window. These patterns are not captured by the surface reflectivity climatology, leading to the algorithm interpreting reflectivity shifts in the sea water as clouds.
3. The cloud field itself is a heterogeneous surface with variability in both reflectivity and height. Unlike previous examples of high-resolution retrievals over very bright surfaces, such as sunglint or snow-ice, the cloud field can not be treated as a bounding surface with a fixed height and albedo.
4. The approximation of a cloud as a Lambertian reflector, as implemented in the FRESCO-Clouds-as-Reflecting-Boundaries algorithm, is insufficient for these type of scenes. The cloud is treated as a bounding surface with a fixed albedo and zero transmittance. The height of the cloud was observed to be a highly sensitive parameter.

6

Recommendations

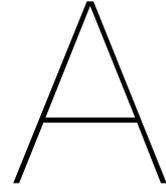
Following on the four main reasons that the current retrieval algorithm fails to robustly retrieve NO_2 vertical densities over low clouds and fog, four potential next steps are set out. These steps aim to improve understanding and contribute to a better retrieval over this type of scenes. Each step is shortly explained in combination with the additional data or measurements that this step would require. The four main steps are a better understanding of NO_2 profiles in the behaviour of low clouds and fog, a correction of the a-priori in the presence of a low cloud field, taking into account variables like ocean color above the North Sea and further work on implementing the Cloud-As-Layers approach in FRESCO.

Besides potentially improving the current TROPOMI retrievals, these insights can also contribute to trace gas retrievals in the presence of clouds for future remote sensing missions. For example, the recently approved ESA TANGO (Twin ANthropogenic Greenhouse Gas Observers) mission [40]. This dual constellation of two agile small-satellites will directly monitor major emission sources with an unprecedented spatial resolution of $300 \times 300 m^2$ [41]. This constellation will measure for example off-source NO_2 emissions and try to estimate fluxes [40]. These high-resolution measurements will be impacted, and hindered, by the presence of local clouds. In order to achieve continuous monitoring on such a small scale, accurately accounting for cloud effects is indispensable. Another example is the future planned CO2M mission, set to have a nominal resolution around $4 km$ [42]. This mission will also carry a NO_2 imaging spectrometer. The recently launched EarthCARE mission can also contribute to a better understanding of smaller scale cloud dynamics and their effect on radiation [43].

The following 4 recommendations are made:

1. Deeper understanding of the dynamic and photo-chemical effects of low clouds on the vertical NO_2 distribution over a sea surface. This can initially be explored by coupling a chemical transport model with a high-resolution atmospheric dynamics model, like the Dutch Atmospheric Large-Eddy Simulation (DALES) model, which succeeds in capturing dynamic behaviour on an eddy scale [44]. This can be validated through vertical NO_2 in-situ measurements over the North Sea. These measurements can be obtained using surface-based optical spectrometers or balloon-borne NO_2 sondes [45]. Vertical profiles can also be obtained using an airplane manned with a sniffer sensor, using a similar methodology as in *Ries et al. (2023)* [46]. This research focused on aircraft-based measurements of NO_2 vertical profiles above the North Sea. This could potentially be extended to further explore these profiles in the presence of low clouds.
2. Through a better understanding of the NO_2 profile near low clouds, a so-called *cloudy NO_2* profile (analogous to the concept of a cloudy AMF) could be introduced, which corrects the original profile for the presence of low clouds or fog. This adjusted profile would account for the effect of low clouds or fog, leading to more accurate estimations of the proportion of pollution beneath the cloud that TROPOMI does not detect, which in its turn would lead to a more realistic ghost column.

3. More variables could be taken into account above the North Sea to better determine the sensitivity through the box-AMF. One example of an additional variable is ocean color, which could be measured through the Ocean Color Instrument aboard the recently launched NASA's Plankton, Aerosol, Cloud and Ocean (PACE) mission [47]. By taking into account a wider range of variables, different processes affecting the absorption could be parameterized on a finer scale. This allows for a better estimation of the surface reflectivity in the NO_2 window over the North Sea. In that case, the look-up table methodology, which showed a low-resolution gridded spatiality, can be expanded to better parameterize horizontal and vertical shifts in sensor sensitivity. A more detailed and spatially more intricate interpolation, taking into account more variables could be achieved through for example neural networks, which will match a vertical sensitivity profile based on a range of variable combinations above the North Sea. This can potentially be combined with a low cloud radiance fraction threshold within the Independent Pixel Approximation, to mitigate the misinterpretation of these spectral patterns as clouds.
4. Explore the implementation of a Cloud-As-Layer (FRESCO-CAL) product, where volume scattering inside the cloud is taken into account. A measurement of both the cloud top and cloud base is required to take into account the cloud optical thickness, properly parameterize cloud-profile dynamics and resolve the vertical a-priori distribution in a more realistic way. This CAL-product does not require additional data or measurements, but a different processing of the reflectance measurements in the O_2 A-band, where clouds are presented as light-scattering layers of water droplets [48].



Appendix A: Retrieval Algorithm

This chapter provides an overview of the retrieval algorithm. The different sections are: the box-AMF, the a-priori, temperature correction, clouds, ghost column, averaging kernels and surface albedo. These terms are frequently mentioned throughout this thesis, the goal of this chapter is to provide the reader with some background information on how these different terms are defined and calculated.

A.1. General overview

The TROPOMI NO_2 retrieval algorithm consists of three steps:

1. Obtain NO_2 slant columns by applying a Differential Optical Absorption Spectroscopy (DOAS) method on the measured reflectance spectra over the different wavelengths, as shown in figure A.1. These slant columns are defined in the line of sight of the satellite.
2. separating the stratospheric and tropospheric contribution to the slant column,
3. converting the tropospheric slant column to a vertical column with the tropospheric air-mass factor (AMF) [6].

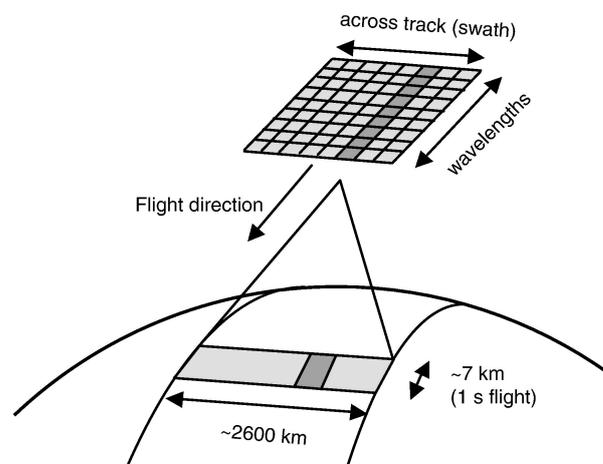


Figure A.1: Conceptual definition TROPOMI measurement set-up. The dark-gray pixels represent the spectrum of wavelengths used for the measurement [1].

The DOAS method applied in the first step fits the measured reflectance spectrum to an associated modelled reflectance spectrum. This is done by minimizing the chi-squared merit function, being the smallest difference between the observed and modelled spectrum [30] (formula A.1). In the second step, the tropospheric column is separated from the stratospheric contribution. The slant columns, the average light paths from the Sun through the atmosphere to the satellite, are assimilated in an atmospheric chemistry model, TM5-MP CTM (Tracer Model 5-Massively Parallel Chemistry Chemistry Transport Model) [49]. The model assimilates stratospheric NO_2 columns based on TROPOMI slant columns over areas with little to no tropospheric contribution. Based on this estimated stratospheric column, the model will estimate global stratospheric contribution through a dynamic transport model [30].

$$\chi^2 = \sum_{i=1}^{n_\lambda} \left(\frac{R_{meas}(\lambda_i) - R_{mod}(\lambda_i)}{\Delta R_{meas}(\lambda_i)} \right)^2 \quad (\text{A.1})$$

with $R_{meas}(\lambda)$ the reflectance spectrum observed by the satellite instrument, R_{mod} the modelled reflectance, n_λ the number of wavelengths in the fit window and $\Delta R_{meas}(\lambda_i)$ the noise on the reflectance, which depends on the radiance and irradiance noise.

This chapter will mainly focus on the final step of the retrieval algorithm, the conversion of the slant column to the vertical column through an air-mass factor (formula A.2). This conversion is currently the dominant source of error for tropospheric NO_2 retrievals over polluted areas [6]. A schematical representation of this step is shown in figure 2.1. The vertical column is defined vertically above a certain pixel in the TROPOMI grid, which has a cell size of $5.5 \times 3.5 \text{ km}$ [6]. The grid relative to the satellite path is shown in figure A.1. As shown in formula A.2, the total vertical column $[NO_2]_{vertical}$ over a pixel is retrieved by dividing the total retrieved slant column density $[NO_2]_{slant}$ with the AMF . Even though the air-mass factor is defined for every atmospheric layer, the entire slant column is divided by the sum of all these layers, so it is not converted per layer. The same principle holds for the tropospheric column, where the tropospheric slant column is divided by a tropospheric air-mass factor.

$$[NO_2]_{vertical} = \frac{[NO_2]_{slant}}{AMF} \quad (\text{A.2})$$

A.2. Box-AMF

AMF calculations require a considerable amount of auxiliary data like atmospheric scattering by air molecules, aerosols, clouds, the shape of the NO_2 vertical distribution and the surface albedo. The spatial sampling of these parameters introduces structural uncertainty [6]. In order to calculate an AMF for a cell, the top-of-atmosphere reflectance is simulated with a Radiative Transfer Model (RTM), in the case of TROPOMI the Doubling-Adding KNMI (DAK) model is used. This RTM calculates the transport of radiation through the atmosphere to the satellite. Different processes affecting this are taken into account, being absorption, scattering, refraction and reflection [50]. These processes then control the calculation of the altitude-dependent AMFs, or box-AMF's. They represent the vertical sensitivity of the measurement to an amount of trace gas in a certain atmospheric layer [50]. The altitude-dependent AMFs are a function of forward model parameters and cloud properties [50] [30]. Based on these dependencies, the box-AMF's are interpolated through a look-up table. This is why they are also referred to as the look-up table air-mass factor (AMF_{LUT} in figure 2.1). The look-up table is defined in the following 6 dimensions, which are the 6 dependencies defining a box-AMF:

1. viewing zenith angle,
2. sun zenith angle,
3. surface albedo in the NO_2 fit window (440 nm)
4. surface pressure,
5. cloud fraction in the NO_2 fit window (440 nm),
6. cloud pressure.

		SOURCE
<i>Dynamic input</i>	slant column density and errors	DOAS
	a-priori profile	TM5-MP
	temperature & pressure profile and tropopause level	TM5-MP
	effective cloud fraction	NO_2 window
	scene albedo and scene pressure	FRESCO
	snow and ice flag	ECMWF
	surface pressure	ECMWF
<i>Static input</i>	viewing geometry and geolocation data	S5P Level-1 product
	surface albedo at 440nm	DLER
	terrain height + land/water classification	Digital Elevation Model
	altitude-dependent AMF look-up table	
	cloud fraction and cloud radiance look-up table	

Table A.1: Overview of the dynamic and static input used in the NO_2 retrieval algorithm [30].

The resulting box-AMF describes how the top-of-atmosphere reflectance is affected by adding a certain trace gas at a certain height. The vertical sum of these altitude-dependent AMFs for a certain grid cell represents the relative length of the mean light path at a certain wavelength interacting with a certain absorber, relative to the vertical light path [50].

$$AMF = \frac{\sum_l m_l v_l c_l}{\sum_l v_l} \quad (\text{A.3})$$

with m_l the altitude-dependent AMFs. v_l represents the column density in layer l . c_l represents the temperature correction.

For the NO_2 product, the AMF is a measure of the sensitivity of the measurement to NO_2 . For example, an increase in surface albedo will lead to an increase in box-AMF's over that grid cell, because of the higher sensitivity due to the stronger reflection of sunlight [50]. Because the AMF is in the denominator (formula A.2), a *larger* AMF leads to a *lower* vertical column density compared to the slant column density. This means that the measurement is more sensitive to NO_2 . If the AMF is smaller, the measurement is less sensitive to NO_2 , and the vertical column will be higher after conversion from the slant column.

A.3. A-priori

The box-AMF's for every vertical layer are weighted with the a-priori NO_2 profile, depicted by v_l in formula A.3. This step aims to adjust the AMF to account for the anticipated vertical distribution of NO_2 . The presence of this trace gas affects measurements. The partial columns of this a-priori are denoted by v_l in formula A.2. In a polluted region with a high a-priori the AMF will be smaller, indicating a reduced sensitivity to NO_2 [6]. The vertical distribution of the a-priori also plays a role. If the NO_2 profile is more elevated, with a higher relative share in the free troposphere, the air-mass factor will be higher and vice versa. Whether the peak of the concentration is modelled below or above a cloud for example, can also significantly impact the air-mass factor [50]. A more elevated profile, or a lower cloud, will generally result in a higher air-mass factor [50]. In a polluted region over land, NO_2 tends to be modelled close to the surface, this will result in a lower AMF because measurement sensitivity decreases towards surface [51]. In a maritime situation, where NO_2 profiles are usually simulated to be more elevated and not so focused near the surface, the AMF will increase in comparison with the land situation.

The a-priori comes from a chemistry transport model. In the case of the TROPOMI NO_2 product, this is the TM5-MP model. This model simulates a three-dimensional field of, among others, NO_2 concentration with a 30-minute time-step ([26]). The horizontal resolution setting used for TROPOMI is

$1^\circ \times 1^\circ$. This resolution is quite low in comparison with the TROPOMI nadir resolution in the order of 5 km . Spatial undersampling of the a-priori can exacerbate relative uncertainties, especially in polluted areas with distinct emission sources [51]. This undersampling effect is more pronounced near coast-lines, because of the large variability within a scene in both atmospheric dynamics and reflectances [51].

Vertically, the model consists of 34 dynamic pressure layers with a higher resolution in the troposphere and upper troposphere-lower stratosphere [49]. The meteorological input data comes from the European Center for Medium range Weather Forecast (ECMWF) ERA-Interim 5 dataset, which is updated every three hours and interpolated for the intermediate time fields. For the OFFL L2 NO_2 product, which is the version considered in this thesis, the processing starts from the DOAS spectral fits and the slant columns [26]. Then, the retrieval is done using a time and space co-located model field. In this retrieval step, the air-mass factors, averaging kernels and vertical columns are computed. TM5-MP also carries out a forecast of the atmospheric chemical state until the time of the orbit the next day. This results in daily fields of TM5-MP simulated trace gas fields with a 30-minute time-step [26].

A.4. Temperature correction

The NO_2 cross sections in the DOAS retrieval are valid for a temperature of 220K [50]. In reality, vertical temperature variability influences the measurement. Thus, a correction term is introduced. To calculate this term, the ECMWF temperature profile is interpolated to the TROPOMI grid. The terms in this correction are an updated version of the version used in OMI [30]. These sensitivity terms are fine-tuned specifically to the NO_2 fit window and TROPOMI's spectral resolution. This temperature correction term, c_t , is integrated in the AMF-calculation (formula A.3). This step is denoted by T_{corr} in figure 2.1.

$$\begin{aligned} t_{sigma} &= 220\text{K} \\ T_{corr} &= 1 - 0.00316 * (t - t_{sigma}) + 3.39^{-6} * (t - t_{sigma})^2 \end{aligned} \quad (\text{A.4})$$

with t_{sigma} the DOAS NO_2 cross-section spectrum temperature and t the temperature in a specific atmospheric layer for a specific TROPOMI pixel.

A.5. Clouds

Cloud-related uncertainties are known to be a significant contributor to the overall uncertainty in NO_2 vertical columns [13]. Within one grid cell of TROPOMI, with a spatial scale less than 5 km , both a cloudy and clear-sky part can be present. In this case, the AMF is expressed a linear combination of a cloudy AMF (AMF_{cloud}) and a clear-sky AMF (AMF_{clear}). This is shown in formula A.5, known as the Independent Pixel Approximation. This holds both for the tropospheric and total AMF. The cloudy and clear AMF are weighted by the radiance weighted cloud fraction crf , which depends on the effective cloud fraction f_{eff} (formula A.6). The Independent Pixel Approximation allows to consider the radiative properties of one grid-cell in isolation from its neighbours [52].

$$M = crf AMF_{cloudy} + (1 - crf) AMF_{clear} \quad (\text{A.5})$$

$$crf = \frac{f_{eff} I_{cl}}{R} = \frac{f_{eff} I_{cl}}{f_{eff} I_{cl} + (1 - f_{eff}) I_{cr}} \quad (\text{A.6})$$

where M is the total air-mass factor of a grid cell, crf is the cloud radiance fraction, AMF_{cloudy} and AMF_{clear} are respectively the clear and cloudy air-mass factor, f_{eff} is the effective cloud fraction, I_{cl} is the radiance from the cloudy part of the pixel, I_{cr} is the radiance from the clear part of the pixel and R is the total scene radiance. For TROPOMI NO_2 retrievals these reflectances are computed from a

look-up table based on the cloud pressure, cloud albedo, surface pressure, surface elevation and surface albedo [30]. This is not to be confused with the look-up table defined for the box-AMF's (chapter A, section A.2).

The cloud retrieval algorithm currently in use for the TROPOMI L2 NO_2 product is named FRESKO, which stands for Fast REtrieval Scheme for Clouds from Oxygen A-band. This method uses multiple absorption bands around the O_2A -band at 760 nm, ranging from strong to moderate to low absorption, in order to determine cloud fraction and cloud pressure [13]. This is combined with the brightness approach in the NIR region. The brightness approach defines a cloud-free background as dark compared to bright clouds. Through these absorption bands, cloud fraction and pressure are defined so that they match the measured top-of-atmosphere reflectance of the real cloud [37]. Recently, the FRESKO algorithm was updated, this included expanding the number of absorption bands in the measurement window ($O_2 A$ -band) [30]. This wider spectrum leads to a generally increased cloud height compared to the previous FRESKO version.

Although the derived cloud parameter referred to "cloud pressure" and thus suggesting it is indicative of the top of the cloud, in reality, the FRESKO cloud pressure is in the middle of a cloud rather than the top [37]. This is an important distinction from cloud detection at longer infrared wavelengths, which are very sensitive to water or ice at the top of clouds. The FRESKO cloud pressure p_{cloud} , and subsequently cloud height, is important in the vertical determination of the box-AMF's, as shown in the conceptual scheme in figure 2.1. Measurement sensitivity is strongly dominated by clouds. The presence of clouds reduces sensitivity to the lower atmosphere but clouds also increase sensitivity for absorbers inside or above the cloud through their higher albedo and multiple scattering within the cloud [13]. This means that above a cloud, box-AMF's will be higher and underneath the cloud lower.

The operational FRESKO version for the TROPOMI NO_2 product is the Cloud-as-Reflecting-Boundaries (FRESKO-CRB). This means that clouds are represented as optically thick Lambertian reflectors centered around the cloud height with a fixed albedo of 0.8 [30]. This version is computationally more efficient and still successfully captures realistic effective cloud fractions [13]. This approach has some limitations like the assumption that a cloud is a reflective boundary with zero transmittance ("a plank"). A more realistic approach would be for a cloud to be considered a volume scatterer and not a reflecting boundary. In reality multiple scattering of the light path also takes place within the cloud. This scattering and absorption inside the cloud is neglected in FRESKO-CRB, which leads to significant errors [53].

The approach of a cloud as a volume scatterer is being explored in the so-called FRESKO-CAL (Clouds-As-Layers) product, where clouds are represented by uniform layers of scattering liquid water particles [39]. For the cell-size of TROPOMI, CAL could potentially lower uncertainty [52]. However, this work is ongoing and a CAL-approach is not yet incorporated in the operational algorithm. The CRB assumption neglects the absorption within a cloud. A smaller top-of-atmosphere reflectance can be misinterpreted as a lower cloud layer [39]. Uncertainties in the NO_2 retrieval due to the combined effect of cloud pressure and a-priori NO_2 profile can lead up to 70 % if the cloud is located below or within NO_2 layer [39]. This effect is strongest for thick clouds at low altitudes and for an elevated a-priori profile, which is the case in more polluted situations. This uncertainty is related to the potential presence of NO_2 inside or below the clouds, which leads to the multiple scattering effect within clouds [37]. The use of a CAL assumption decreases tropospheric NO_2 columns by almost 20% for polluted regions in winter in comparison with CRB, especially when most NO_2 concentrations are located at the surface [39].

FRESKO products have a different spectral footprint because FRESKO measures in a different wavelength window than the NO_2 retrieval window, as shown in figure 1.2. For the NO_2 retrieval, cloud fraction and cloud radiance fraction are measured in the NO_2 spectral window itself and the cloud pressure, after spatial interpolation, is taken from FRESKO [13].

A.6. Ghost column

To get a better idea of what TROPOMI can not sense underneath a cloud, the ghost column concept is introduced. This represents the NO_2 column below the clouds. It is determined by integrating the TM5-MP NO_2 profile from the surface to the cloud pressure level. In other words, it can be defined as the a-priori amount of NO_2 beneath a cloud. This ghost column is added as a variable to the TROPOMI NO_2 product. It allows to get an idea of the NO_2 concentrations underneath the clouds which the sensor is not measuring. The ghost column variable is for visualising and understanding the assumed concentration underneath the cloud, rather than an explicit step in the retrieval algorithm. The ghost column effect is encapsulated within the air-mass factor, through the look-up table. The box-AMF's of the layer underneath the cloud are multiplied with the partial column in that layer, as shown in scheme 2.1. box-AMF's are lower beneath the clouds because of reduced sensitivity [50]. Through the ghost column concept, clouds further strengthen the effect of model under-sampling [51]. Underneath a cloud, there is nearly no measured source of information and the retrieval algorithm needs to infer concentrations from this *dark* part from the chemistry transport model. The photo-chemical and dynamical effects within a cloud, like convection and modified photolysis, influence the vertical profile shape [51]. The model is not likely to capture this on a cloud-scale.

A.7. Averaging kernels

Vertical sensitivity is represented by the averaging kernels. They are defined to better understand the sensitivity of the measurements, without taking into account the assumptions of the modelled a-priori profile. The averaging kernel is an expression of the relationship between the retrieved NO_2 vertical column and the *true* NO_2 distribution [54]. In this case, the true distribution is the modelled TM5-MP field. Another use of averaging kernels is that the TM5-MP model can be replaced by a regional higher resolution model [26]. They are calculated for every layer i as the box-AMF of that layer divided by the total AMF computed *with* the a-priori profile [55] (formula A.9). As shown in formula A.10, the tropospheric averaging kernel A_{trop} is defined by weighing every value of layer i of the total averaging kernel A with AMF/AMF_{trop} and setting all elements above the tropopause to zero.

$$\hat{y} = Ax_m = \sum_{i=1}^n A_i S_i x_{m,i} \quad (\text{A.7})$$

with \hat{y} the retrieved quantity, A the column averaging kernel, which is defined as a vector spanning n pressure levels, x_m the vertical distribution of NO_2 , in this case from a chemistry transport model, S_i is an operator executing a mass-conserving vertical interpolation and a conversion to *molescules cm⁻²*.

The averaging kernels can also be used to compare to in-situ measurements or different model outputs. This relationship is expressed in formula A.8, with the *initial guess*, which is the $X_{apriori}$ or TM5-MP a-priori in this case, relates to the retrieved variable X_{sat} , the vertical NO_2 column, through the averaging kernel vector A . The averaging kernel is a measure of the sensitivity of the retrieved quantity to the true state, X_{true} . This also makes the averaging kernels a useful instrument in validating retrievals with measurements or a high-resolution model, which are approximations of X_{true} .

$$X_{sat} = X_{apriori} + A(X_{true} - X_{apriori}) \quad (\text{A.8})$$

where X are observation vectors spanning over the three dimensional field, X_{sat} is the retrieved NO_2 tropospheric column, $X_{apriori}$ is the a-priori profile and X_{true} can be either a measured or modelled atmospheric profile.

As seen in figure A.2, the averaging kernel generally decreases toward the surface because of the reduced sensitivity further away from the sensor. In the situation of a cloud, the averaging kernel will steeply rise above the cloud because of enhanced brightness (albedo effect), and go quickly towards 0 underneath the cloud (shielding effect). Note that figure A.2 shows a theoretical boundary case. The reality is more an intermediary decline, where the averaging kernels of a (partially) clouded pixel go faster towards 0 in comparison with the clear-sky case. The box-AMF's are directly proportional to the averaging kernel, as shown in formula A.9.

$$A_i^{total} = \frac{AMF_{boxi}}{AMF} \quad (A.9)$$

$$A_i^{trop} = \frac{AMF}{AMF_{trop}} A_i^{total}, i \leq i_{tropopause}$$

$$A_i^{trop} = 0, i > i_{tropopause} \quad (A.10)$$

with A_i^{total} the averaging kernel at every atmospheric layer i for a specific grid cell and $i = 1, \dots, 34$ contains all the layers of the a-priori profile, AMF_{boxi} the box-AMF at every layer i , AMF is the total air-mass factor of that specific grid cell, A_i^{trop} is the tropospheric averaging kernel for every layer i , which is calculated by multiplying every value of total averaging kernel with the ratio of the total AMF divided by the tropospheric AMF , the result is then set to 0 above $i_{tropopause}$, which is the layer at which the tropopause is defined.

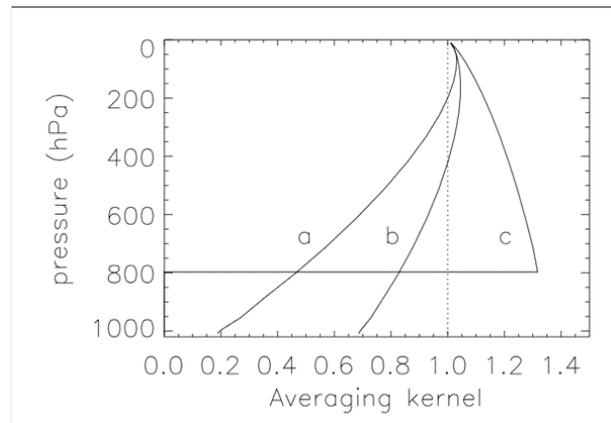


Figure A.2: Example of averaging kernels for three situations: clear-sky with a surface albedo of 0.02 (a), clear-sky with a surface albedo of 0.15 (b), an (optically thick) cloud with a top height at 800 hPa (c) [55]

A.8. Surface albedo

Surface reflectivity, which is also one of the parameters that defines the look-up table for box-AMF's, is estimated by a climatology database, the directionally dependent lambertian-equivalent reflectivity (DLER) for different wavelength bands [34]. This database is based on 5 years of TROPOMI reflectance observations. For every grid-cell and measurement time, the expected surface albedo for the 440-nm band, which is in the NO_2 fitting window, is interpolated through this database. An example of the simulated surface albedo in the NO_2 window is given in figure A.3. The surface albedo is also simulated in the O_2 A-band for FRESCO cloud detection. This FRESCO surface albedo is interpolated using GOME-2 observations at 758 and 772 nm [34]. If the real surface albedo strongly differs from the expected value, for example in the case of snow or ice cover, significant errors in the retrieval algorithm can occur [30]. In the case of snow or ice, a correction based on the daily snow-ice extent maps from ECMWF is applied. Because of the comparable reflection properties of snow and clouds, proper distinction is important. Previously, pixels with snow or ice extent were flagged as not usable because this distinction could not be made. Recently, by checking whether the cloud pressure is close to the

surface pressure, within error margins, cloud-free snow/ice pixels can be filtered out and used [16].

When measured surface reflectivity is lower than the DLER-interpolated reflectivity, the surface albedo in the NO_2 window is adjusted to match the observations and replace the climatology [30]. This lower albedo lowers sensitivity, which subsequently lowers the AMF and increases the NO_2 column density. The same principle holds for a measured reflectivity greater than expected for a cloudy scene. In this case, the cloud albedo is increased to match brighter observations [30]. Normally, cloud albedo is fixed at 0.8 in a clouded scene.

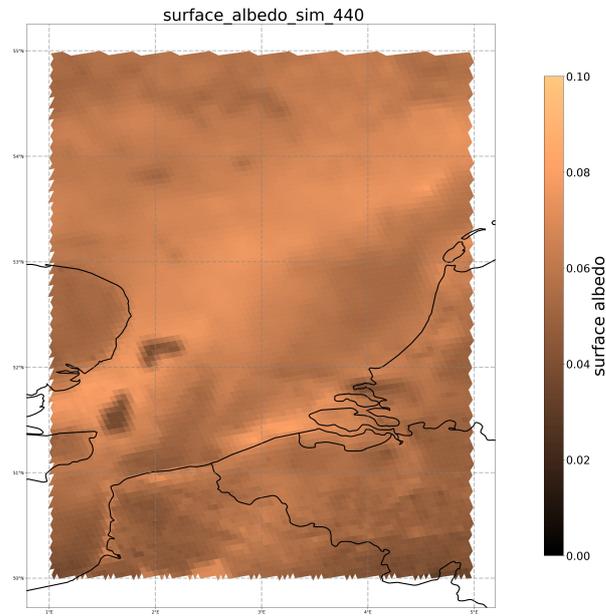


Figure A.3: Example of simulated surface albedo over Belgium and the Netherlands and the North Sea for 14/02/2023.

B

Appendix B: Methodology

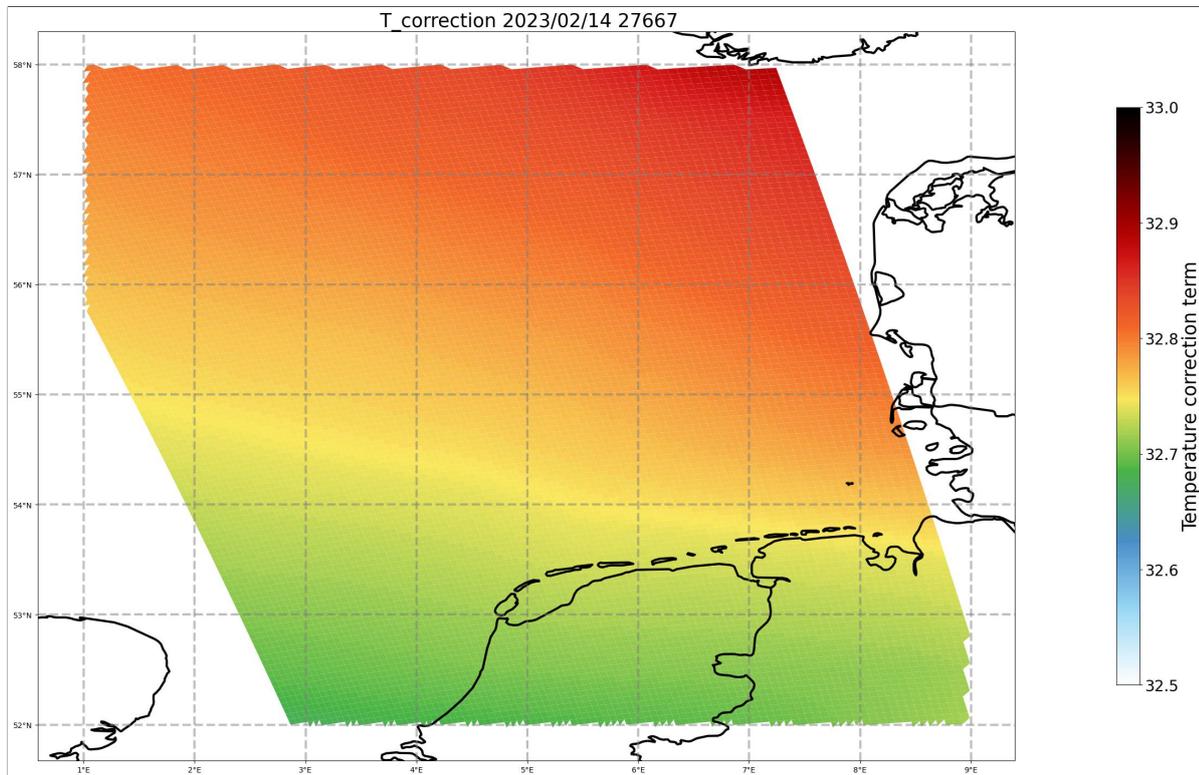
B.1. Proof air-mass factor derivation

The proof that expression 2.19 is mathematically equivalent expression 2.20 is given below.

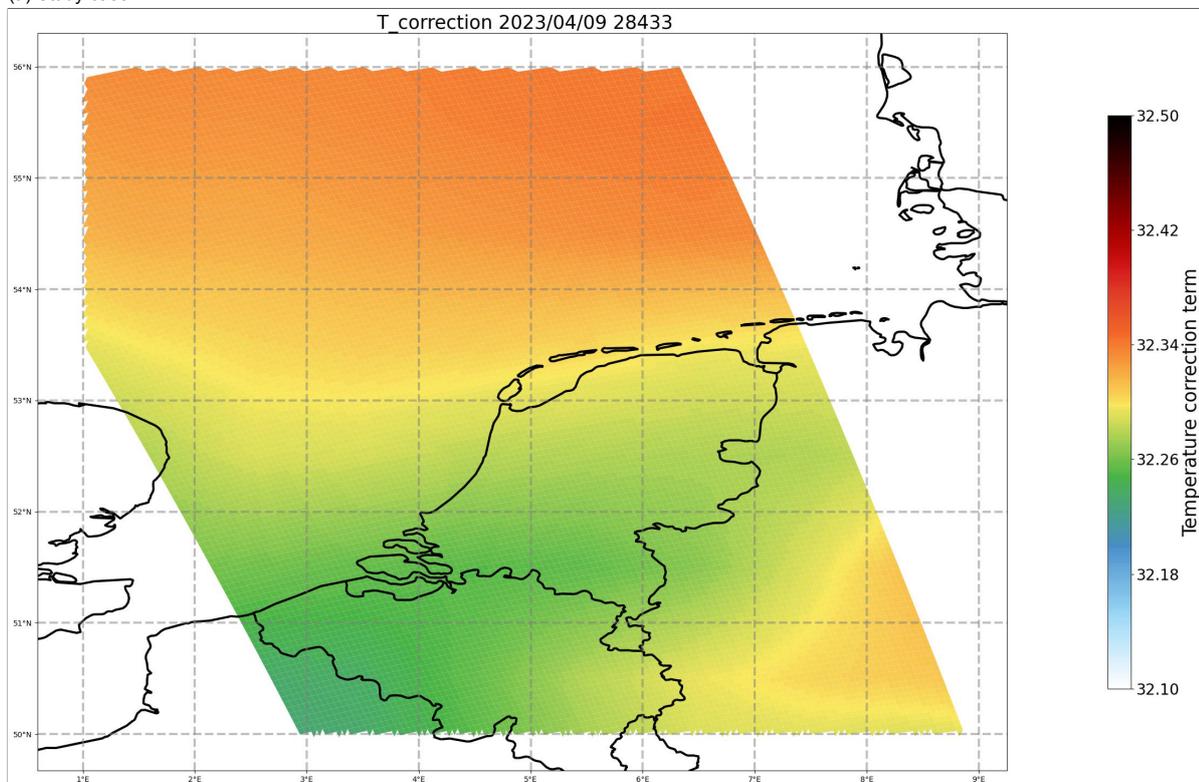
The sum on right-hand side of formula 2.19, corresponding with expression B.1 can be derived to lead back to product on the right-hand of 2.20, expression B.2, as follows:

$$\begin{aligned}
 & \sum_{i=1}^n \frac{AMF_{gbox_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
 & + \sum_{i=1}^n \frac{T_{corr_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \\
 & + \sum_{i=1}^n \frac{NO_{2partialcolumn_i}}{\frac{1}{AMF_{gbox_i} * T_{corr_i}} + \frac{1}{AMF_{gbox_i} * NO_{2partialcolumn_i}} + \frac{1}{T_{corr_i} * NO_{2partialcolumn_i}}} \tag{B.1} \\
 & = \\
 & \sum_{i=1}^n \frac{AMF_{gbox_i}}{\frac{NO_{2partialcolumn}}{AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn}} + \frac{T_{corr_i}}{AMF_{gbox_i} * NO_{2partialcolumn_i} * T_{corr_i}} + \frac{AMF_{gbox_i}}{T_{corr_i} * NO_{2partialcolumn_i} * AMF_{gbox_i}}} \\
 & + \sum_{i=1}^n \frac{T_{corr_i}}{\frac{NO_{2partialcolumn}}{AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn}} + \frac{T_{corr_i}}{AMF_{gbox_i} * NO_{2partialcolumn_i} * T_{corr_i}} + \frac{AMF_{gbox_i}}{T_{corr_i} * NO_{2partialcolumn_i} * AMF_{gbox_i}}} \\
 & + \sum_{i=1}^n \frac{NO_{2partialcolumn}}{\frac{NO_{2partialcolumn}}{AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn}} + \frac{T_{corr_i}}{AMF_{gbox_i} * NO_{2partialcolumn_i} * T_{corr_i}} + \frac{AMF_{gbox_i}}{T_{corr_i} * NO_{2partialcolumn_i} * AMF_{gbox_i}}} = \\
 & \sum_{i=1}^n \frac{(AMF_{gbox_i} + T_{corr_i} + NO_{2partialcolumn}) * (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn})}{NO_{2partialcolumn} + T_{corr_i} + AMF_{gbox_i}} = \\
 & \sum_{i=1}^n (AMF_{gbox_i} * T_{corr_i} * NO_{2partialcolumn_i}) \tag{B.2}
 \end{aligned}$$

B.2. Temperature correction term



(a) Study case A



(b) Study case B

Figure B.1: Temperature correction term, summated over the troposphere. Note the different scale.

C

Appendix C: Results

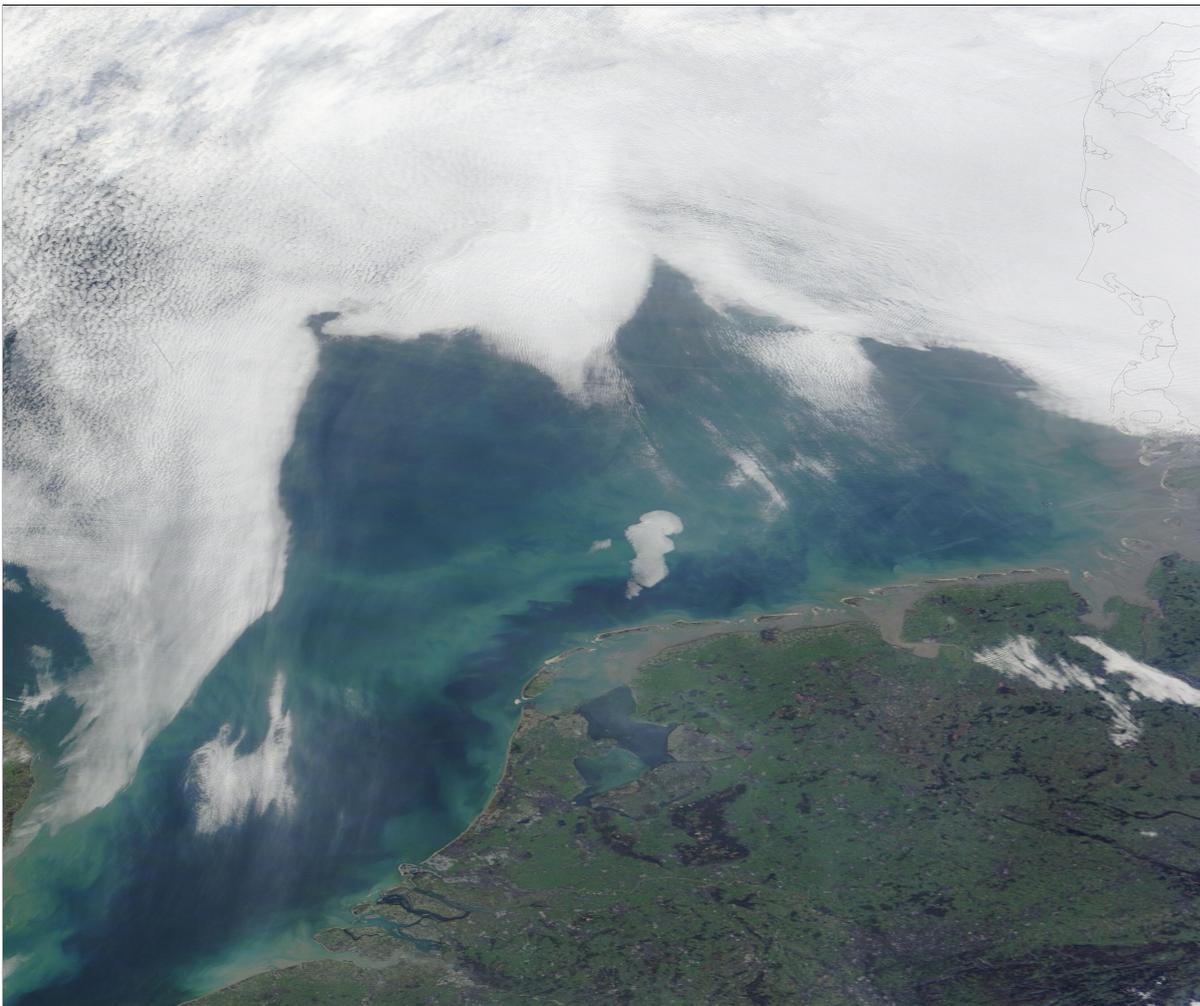


Figure C.1: Terra-MODIS imagery of case study A [23].

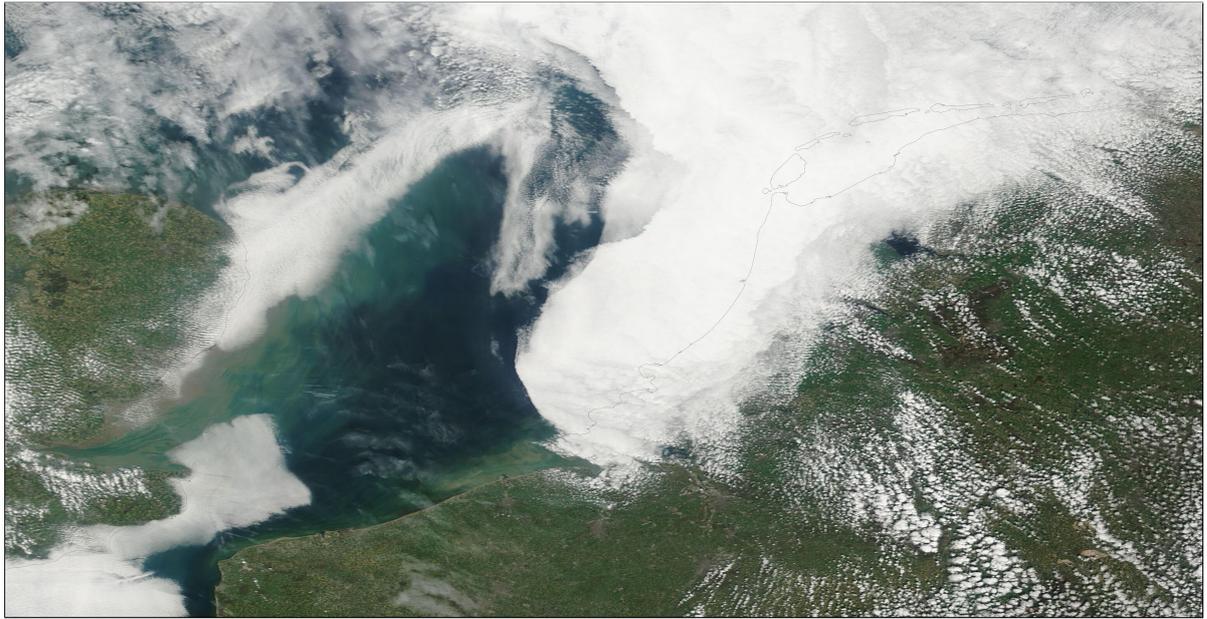


Figure C.2: Terra-MODIS imagery of case study B [23].

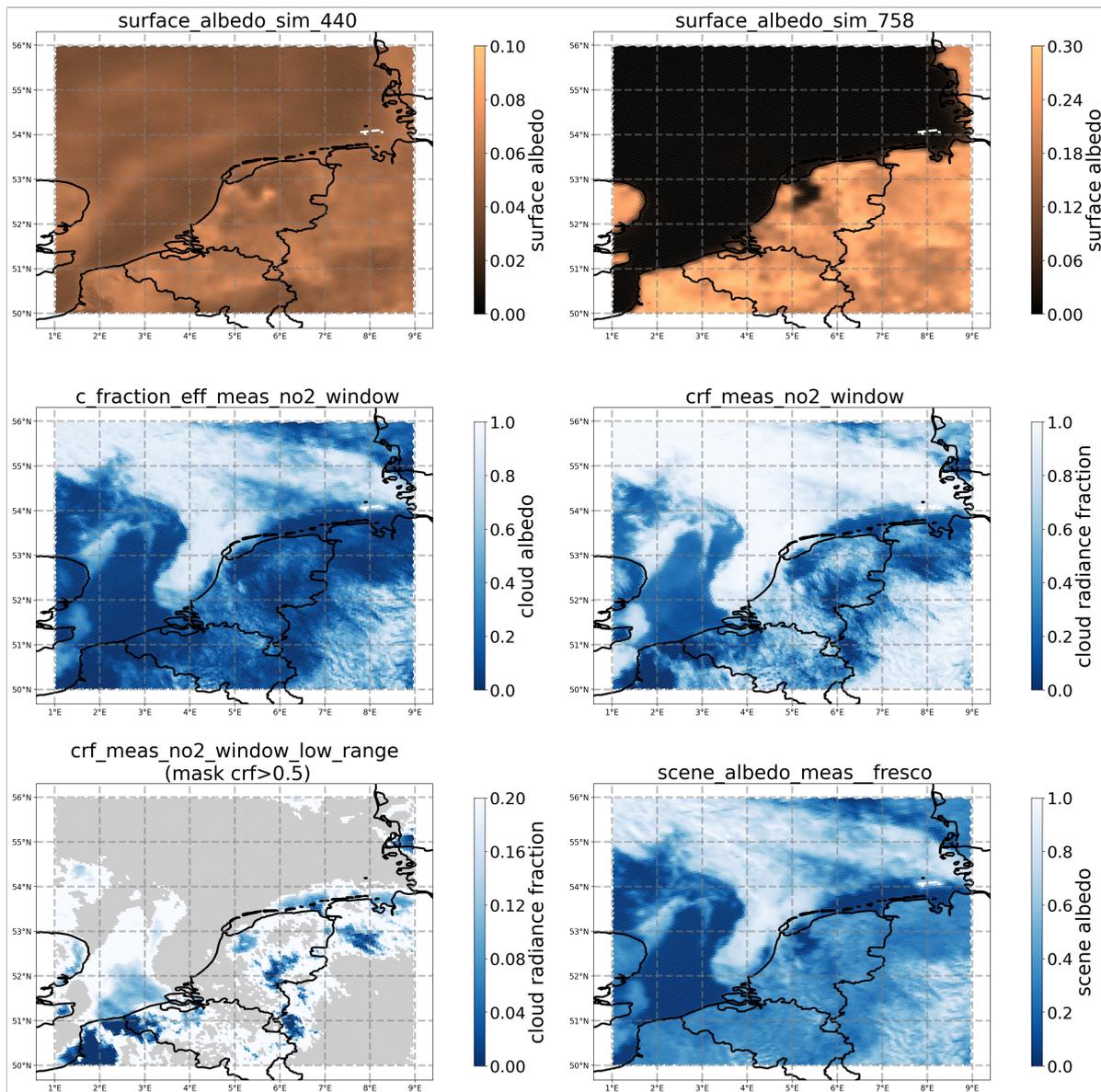


Figure C.3: Measured and simulated reflectances for case study B. From bottom left to top right: the simulated surface albedo in the NO_2 retrieval window, the simulated surface albedo in the O_2 A-band retrieval window, the measured geometric cloud fraction in the NO_2 retrieval window, the measured cloud radiance fraction in the NO_2 retrieval window, the low range measured cloud radiance fraction in the NO_2 retrieval window with a mask depicting cloudy areas, the measured scene albedo in the FRESKO (O_2 A-band) retrieval window.

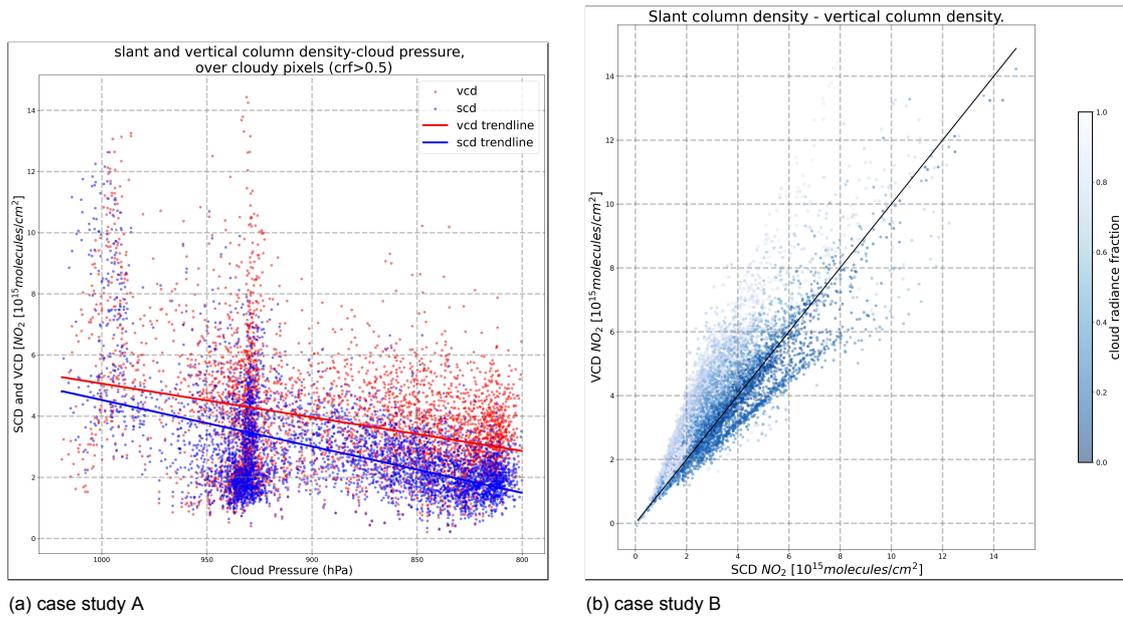


Figure C.4: case study B. Each point corresponds to a TROPOMI pixel. Left plot shows both the slant and vertical column densities plotted against cloud pressure. The trendline through both scatter clouds is also shown. Only cloudy pixels are plotted with a threshold of $crf > 0.5$. Only cloud pressures above 800 hPa are considered. Right plot shows the scatter plot of slant column density against vertical column density, with colorscale representing cloud radiance fraction in the NO_2 fit window. The 1-1 line is plotted in black.

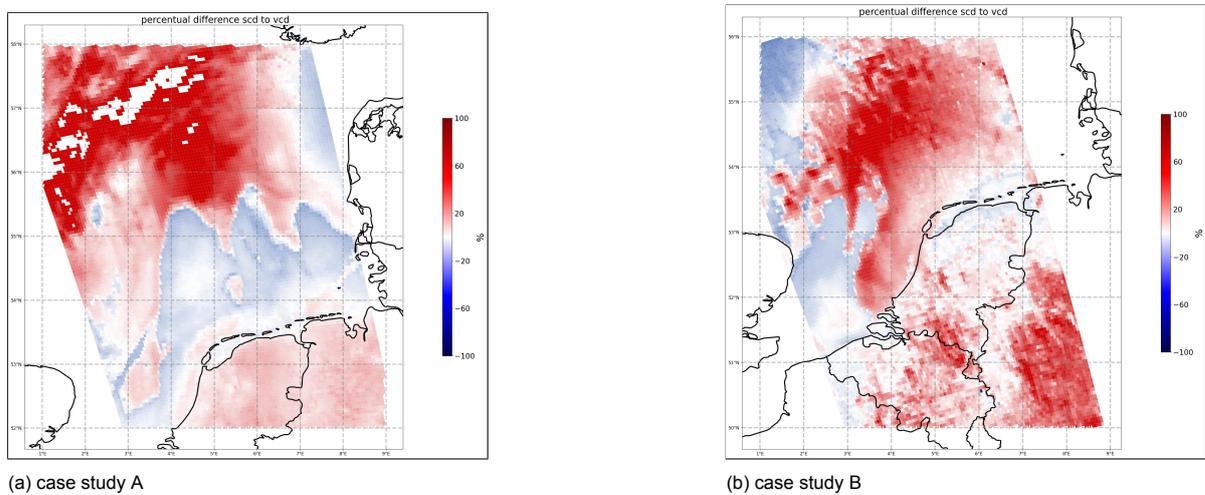
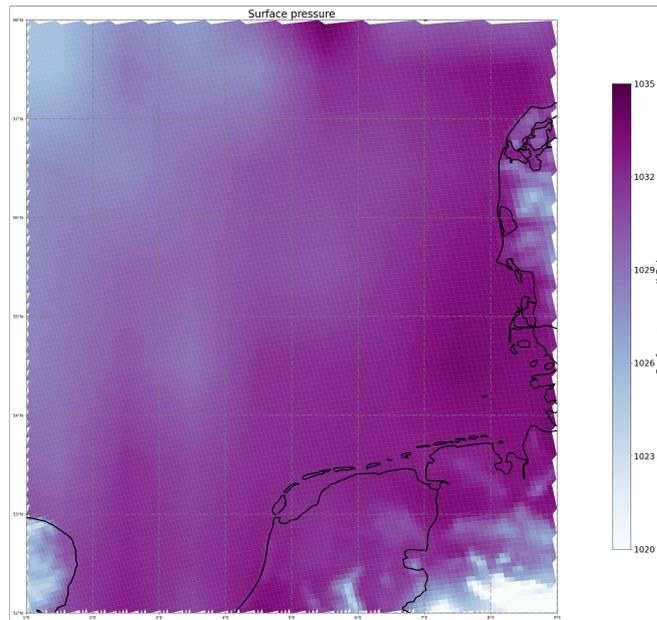
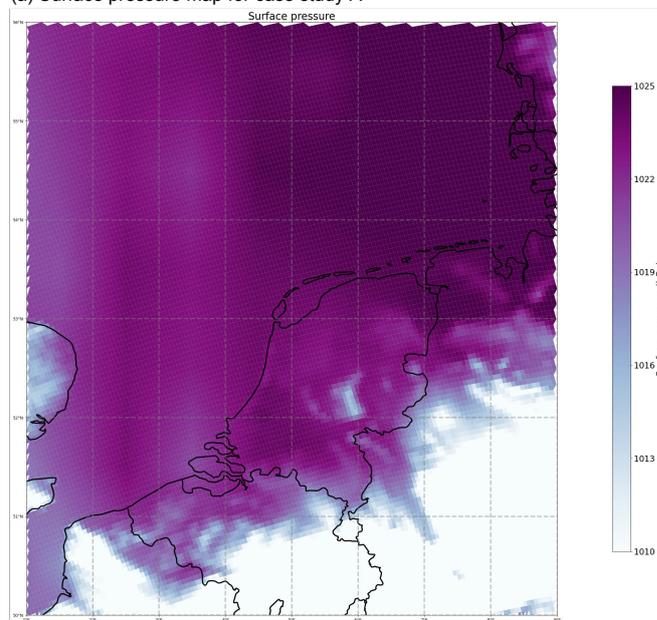


Figure C.5: Relative difference between the slant and vertical column.



(a) Surface pressure map for case study A



(b) Surface pressure map for case study B

Figure C.6: Surface pressure maps for case studies A and B

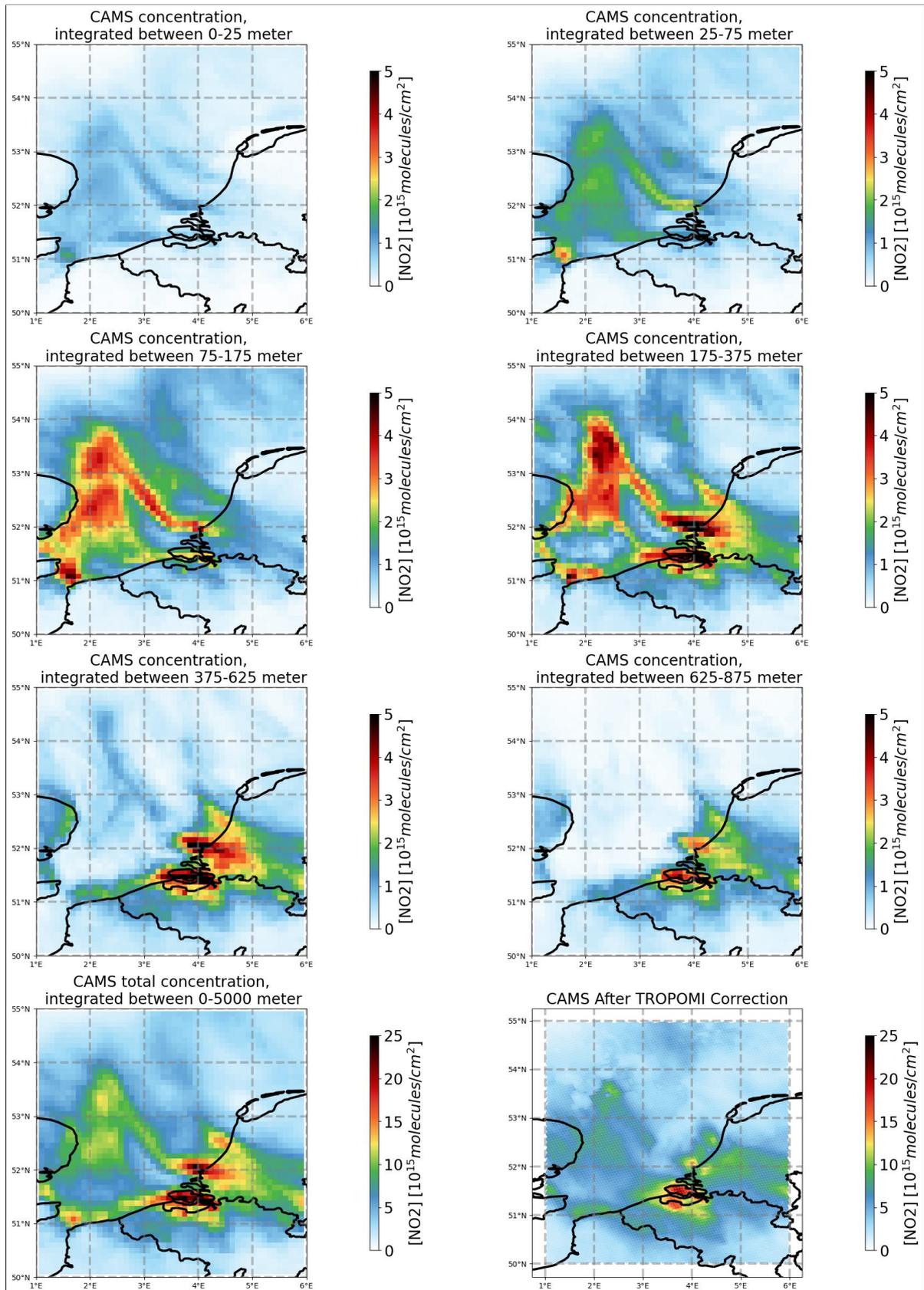


Figure C.7: case study B. Simulated CAMS-Europe forecast NO_2 fields, integrated around different height levels for the first 7 panels. The bottom left panel shows the summated CAMS-Europe column. Bottom right panel shows the corrected CAMS-Europe tropospheric column [28]

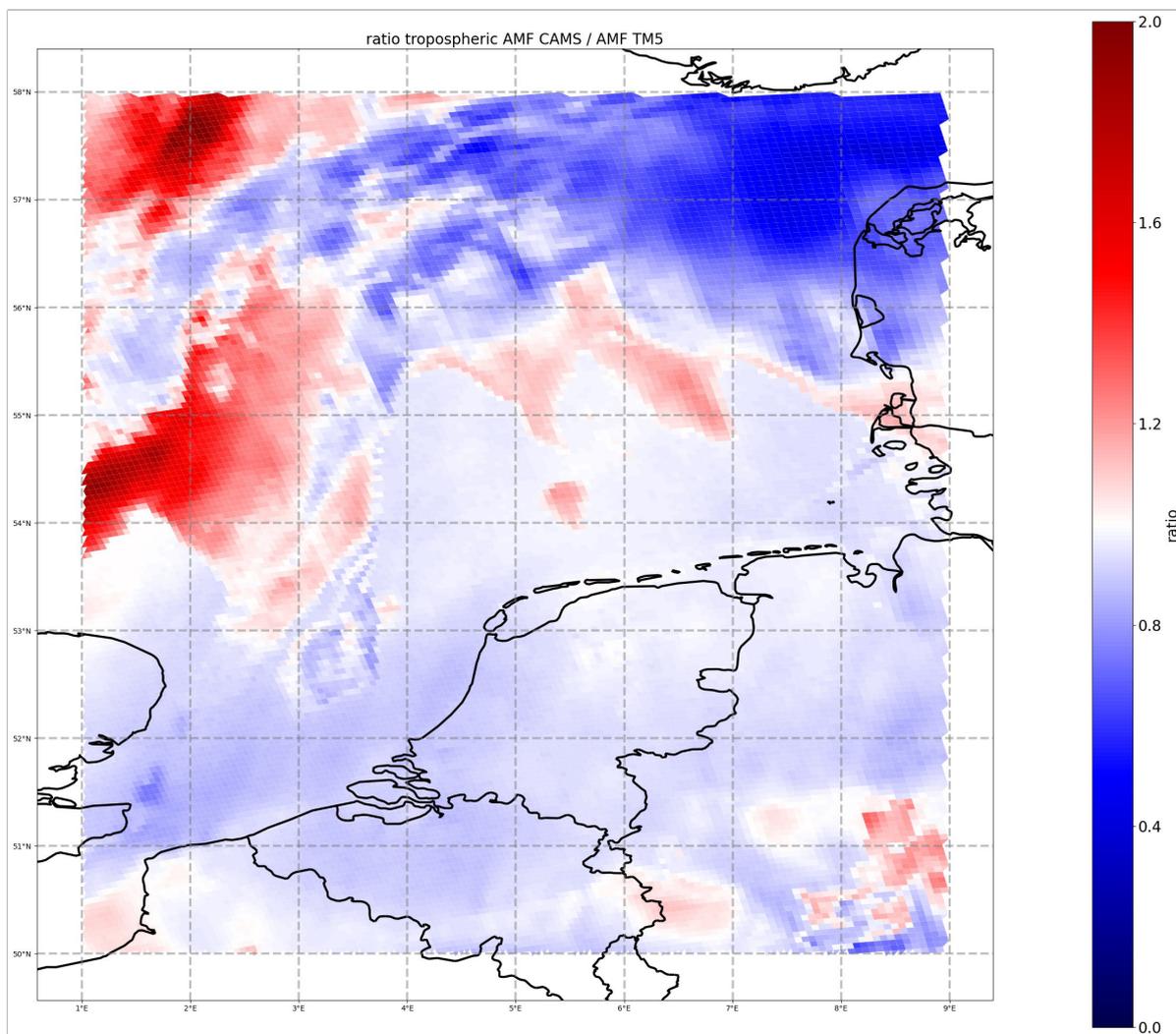


Figure C.8: case study A. Ratio of the air-mass factors using CAMS-Europe as a-priori and TM5-MP as a-priori.

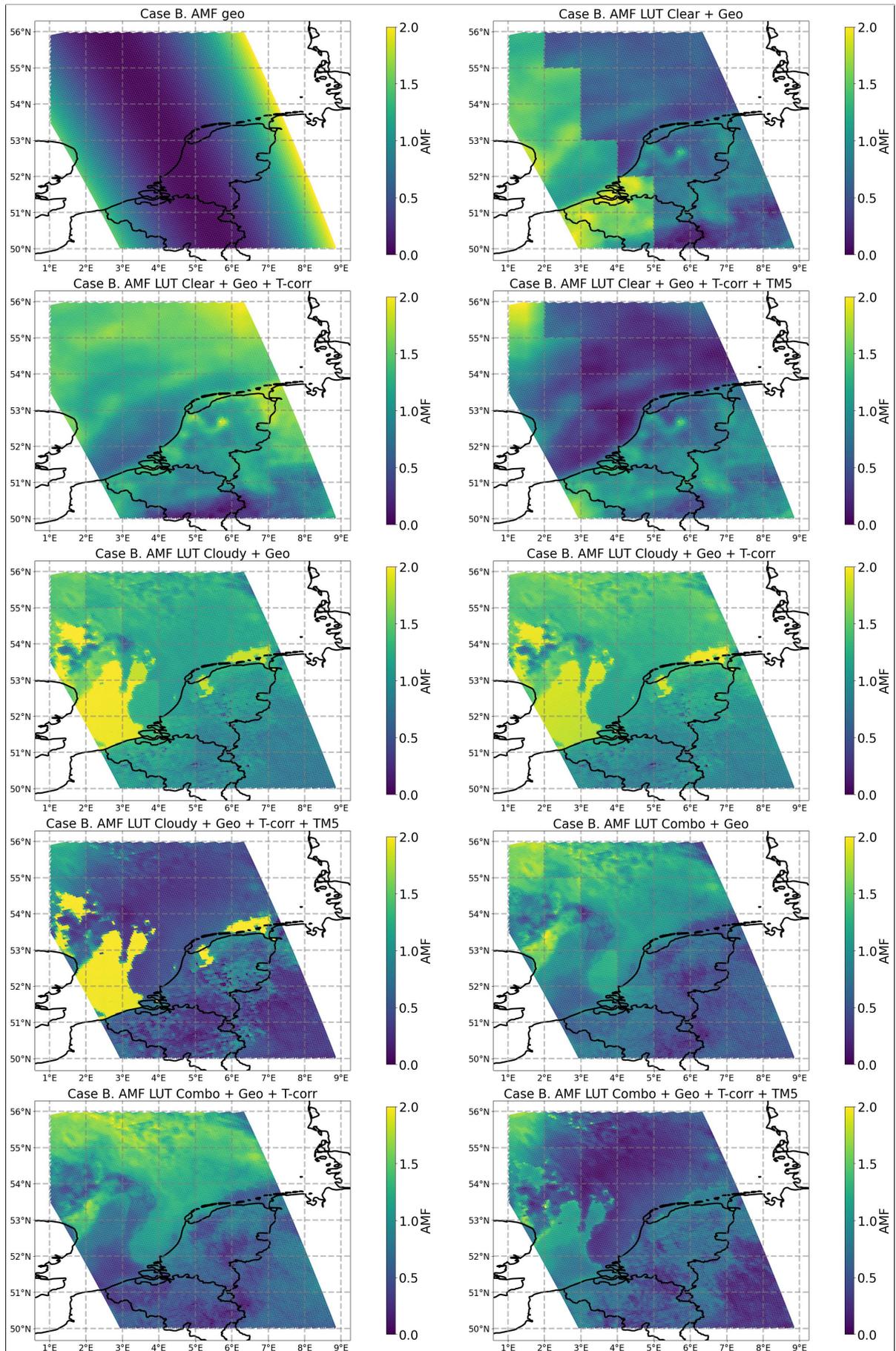


Figure C.9: case study B. Processor plots showing the geometric conversion factor and the different steps in the clear, cloudy and combined air-mass factor retrieval.

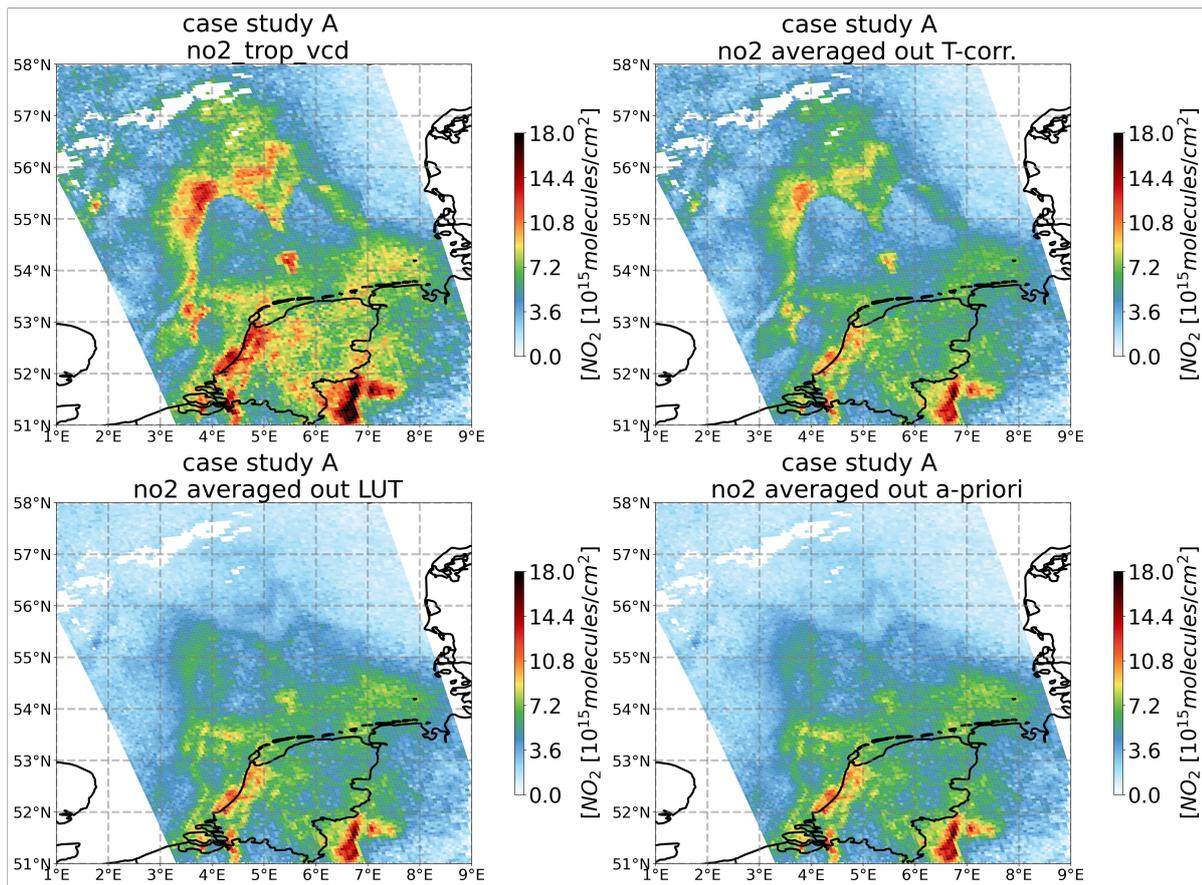


Figure C.10: case study A. Resulting NO_2 tropospheric vertical column for the original settings, averaging out the temperature correction term, averaging out the look-up table and averaging out the a-priori in the Z-plane (from top left to bottom right).

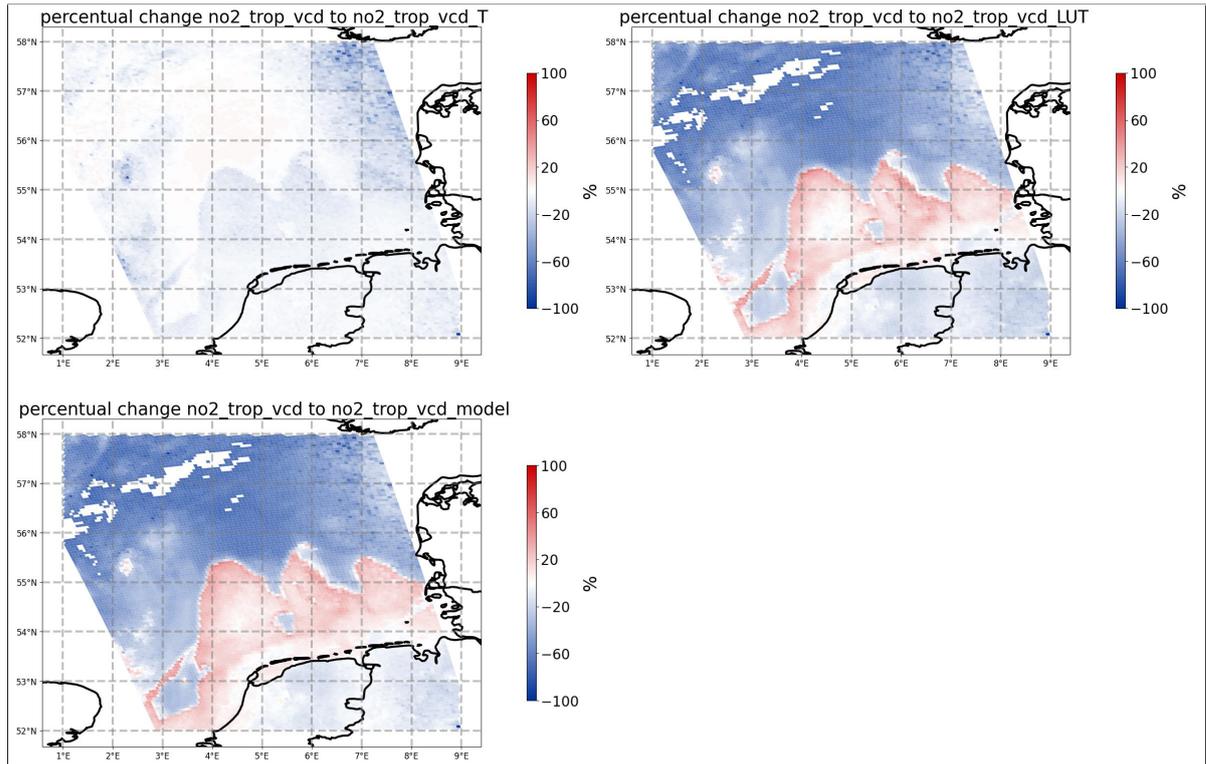


Figure C.11: case study A. Relative difference between the original retrieval and the retrieval after averaging out the temperature correction (*no2 trop vcd T*), the LUT (*no2 trop vcd LUT*) and the a-priori profile (*no2 trop vcd model*).

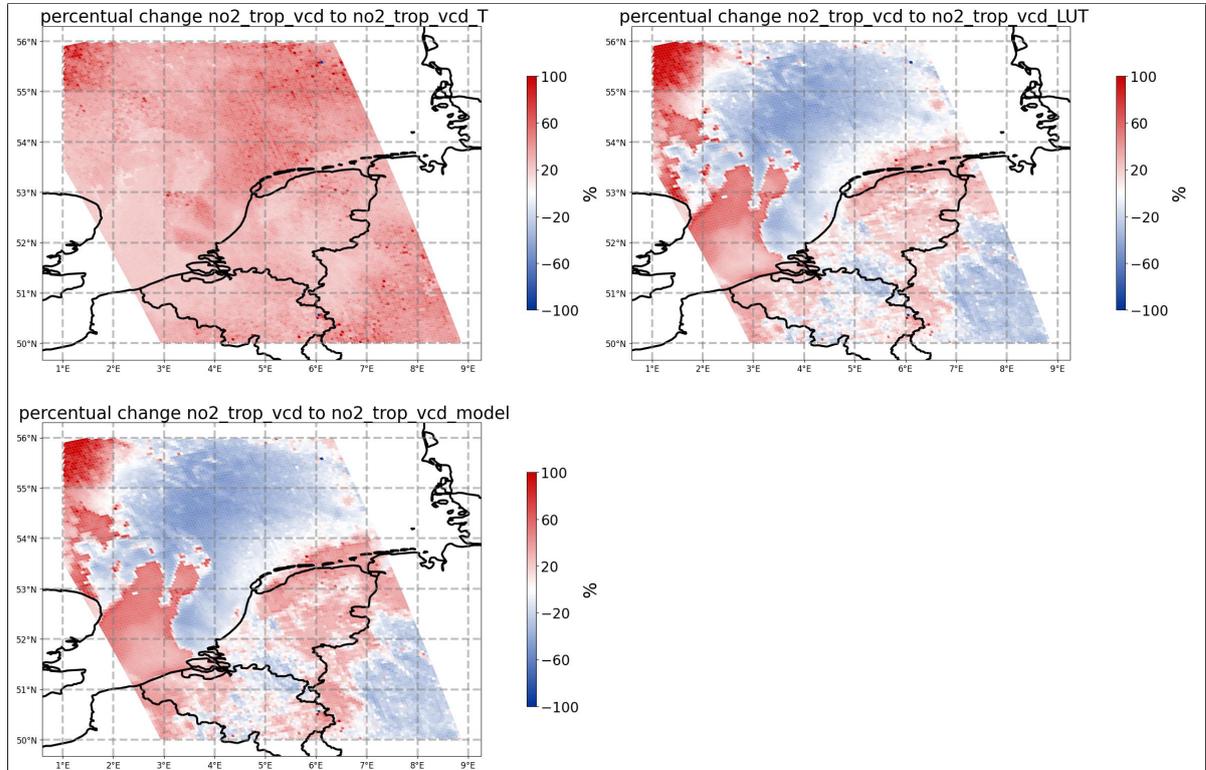
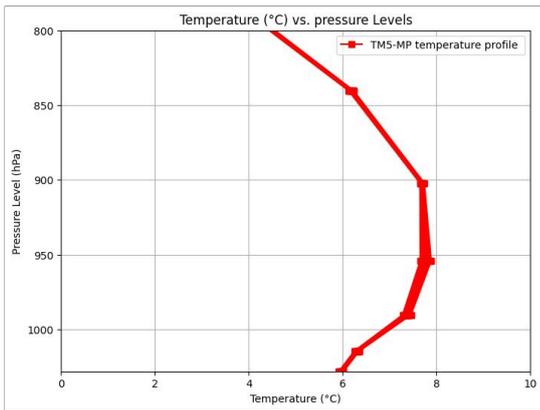
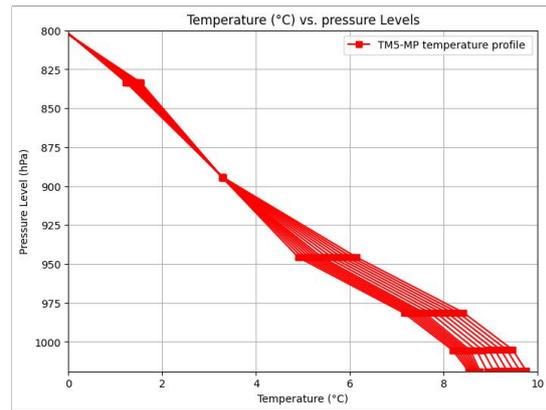


Figure C.12: case study B. Relative difference between the original retrieval and the retrieval after averaging out the temperature correction (*no2 trop vcd T*), the LUT (*no2 trop vcd LUT*) and the a-priori profile (*no2 trop vcd model*).

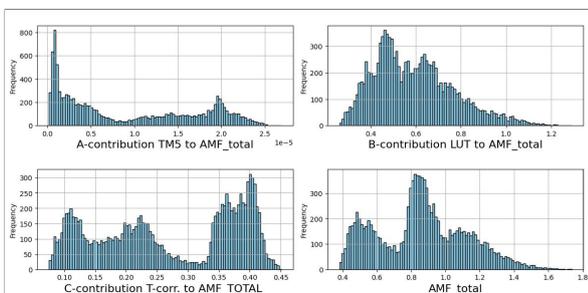


(a) case study A. Temperature profile.

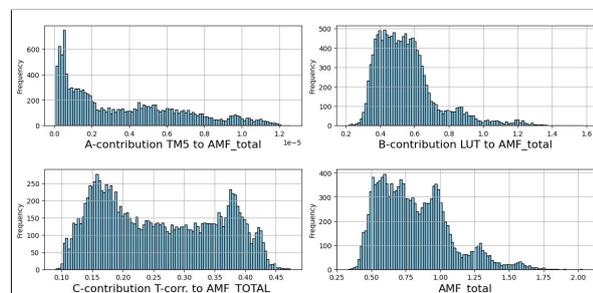


(b) case study B. Temperature profile

Figure C.13: Temperature profiles for both cases from 800 hPa to the surface.

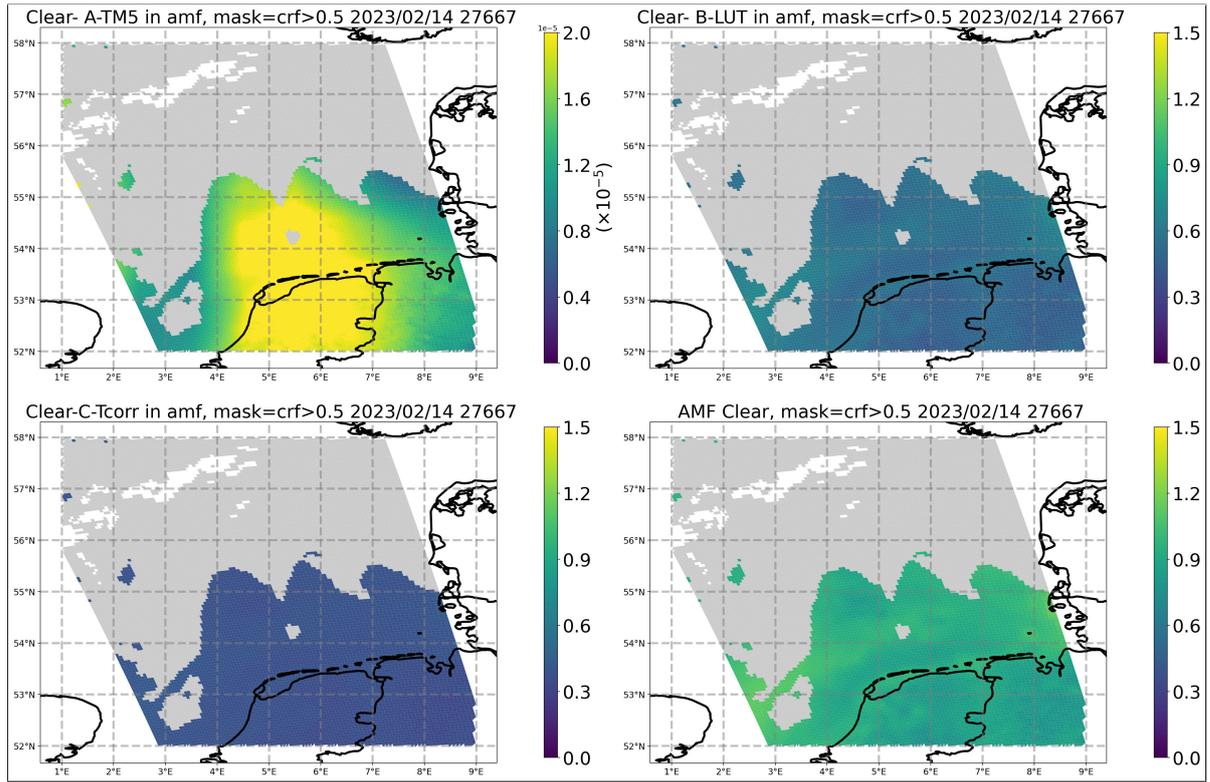


(a) case study A. Histogram of the three contributors (first 3 panels) and the total air-mass factor (bottom right).

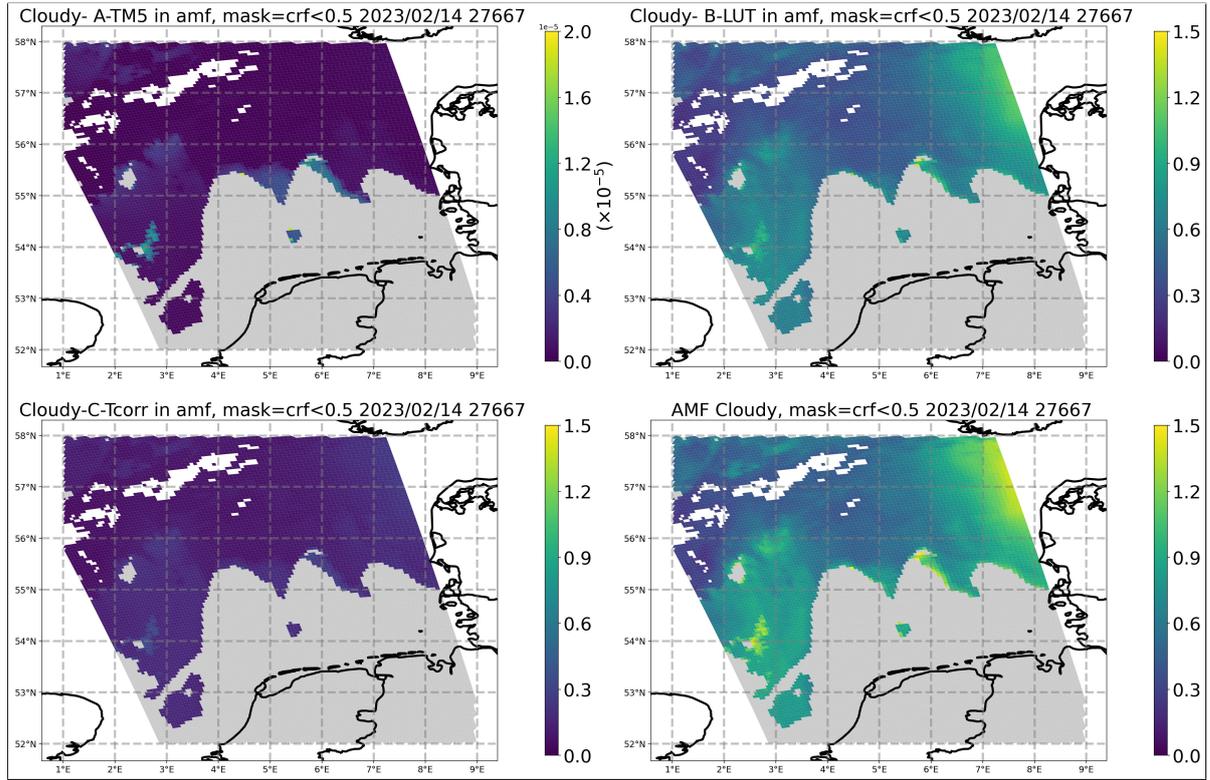


(b) case study B. Histogram of the three contributors (first 3 panels) and the total air-mass factor (bottom right).

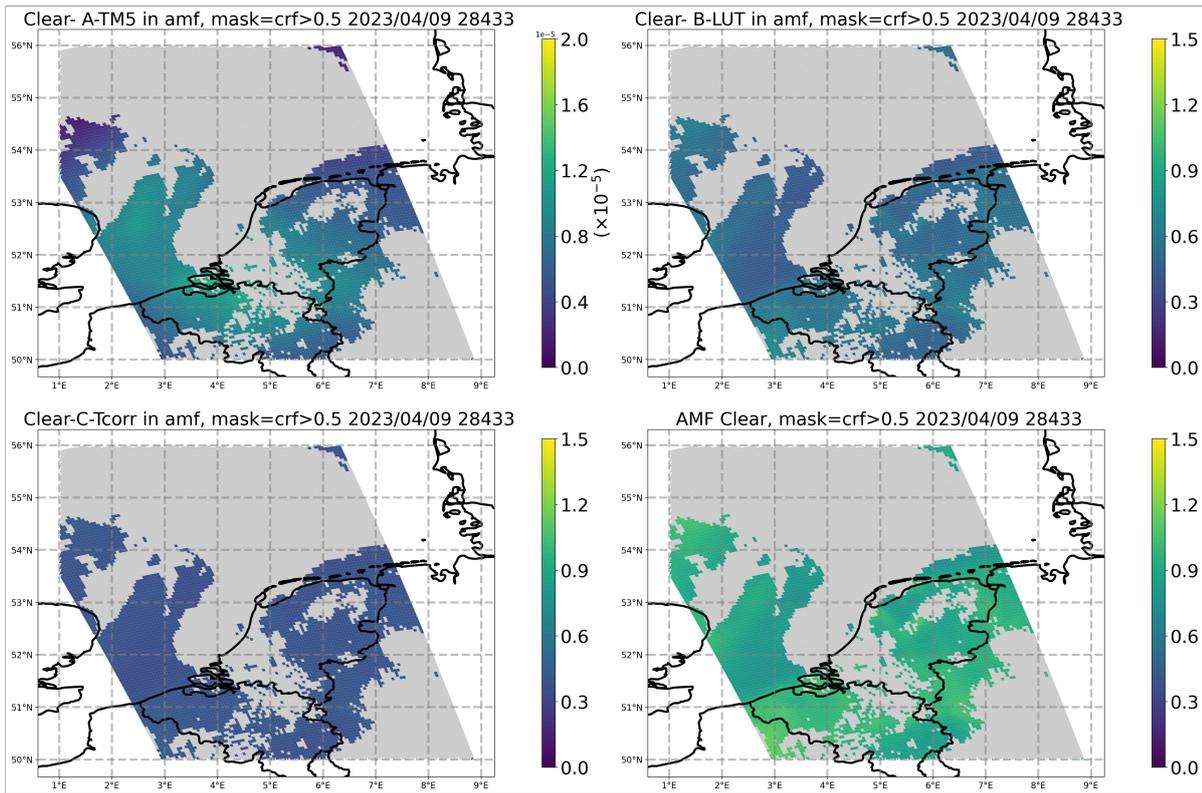
Figure C.14: Histograms of the three contributors and the total air-mass factor for case studies A and B.



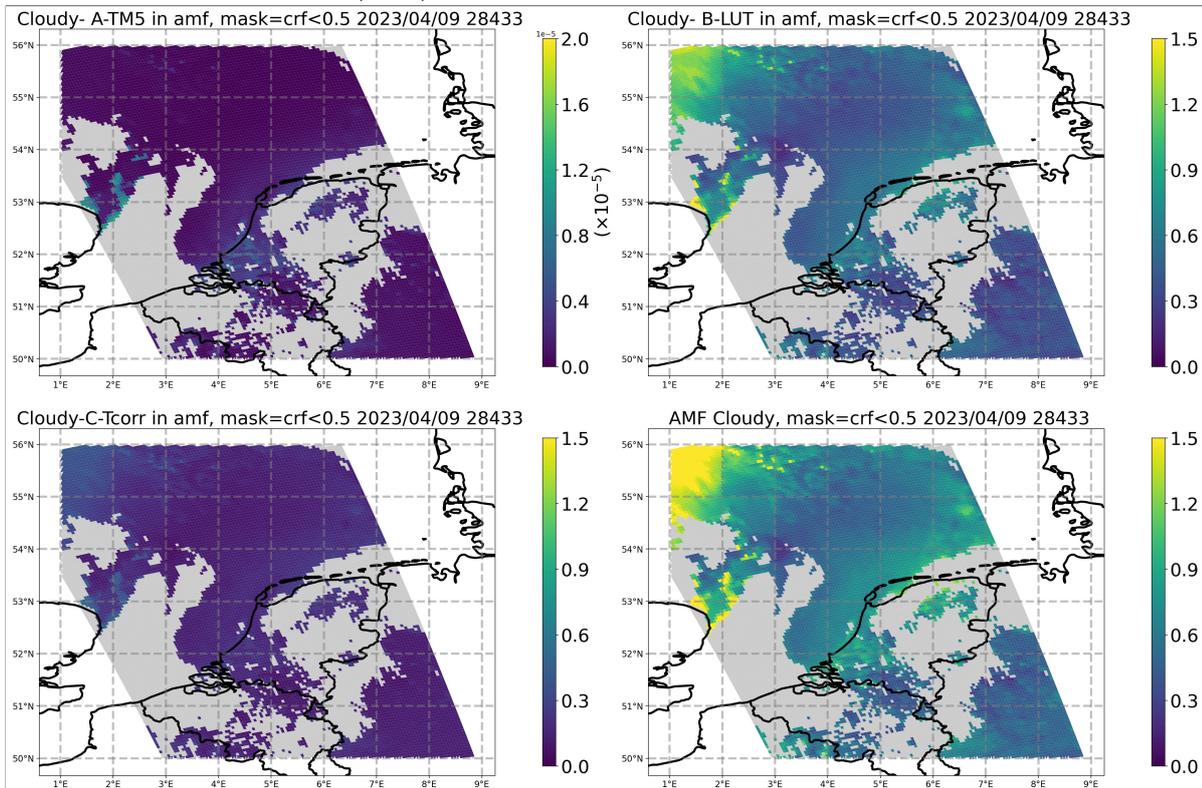
(a) case study A. Relative contribution of the 3 contributors to the clear air-mass factor (first 3 panels) and the clear air-mass factor (bottom left). The masked area indicates the cloud field (crf>0.5)



(b) case study A. Relative contribution of the 3 contributors to the cloudy air-mass factor (first 3 panels) and the cloudy air-mass factor (bottom left). The masked area indicates the clear field (crf<0.5)



(a) case study B. Relative contribution of the 3 contributors to the clear air-mass factor (first 3 panels) and the clear air-mass factor (bottom left). The masked area indicates the cloud field ($crf > 0.5$)



(b) case study B. Relative contribution of the 3 contributors to the cloudy air-mass factor (first 3 panels) and the cloudy air-mass factor (bottom left). The masked area indicates the clear field ($crf < 0.5$)

Figure C.16: Relative contributions to the clear and cloudy air-mass factors for case studies A and B.

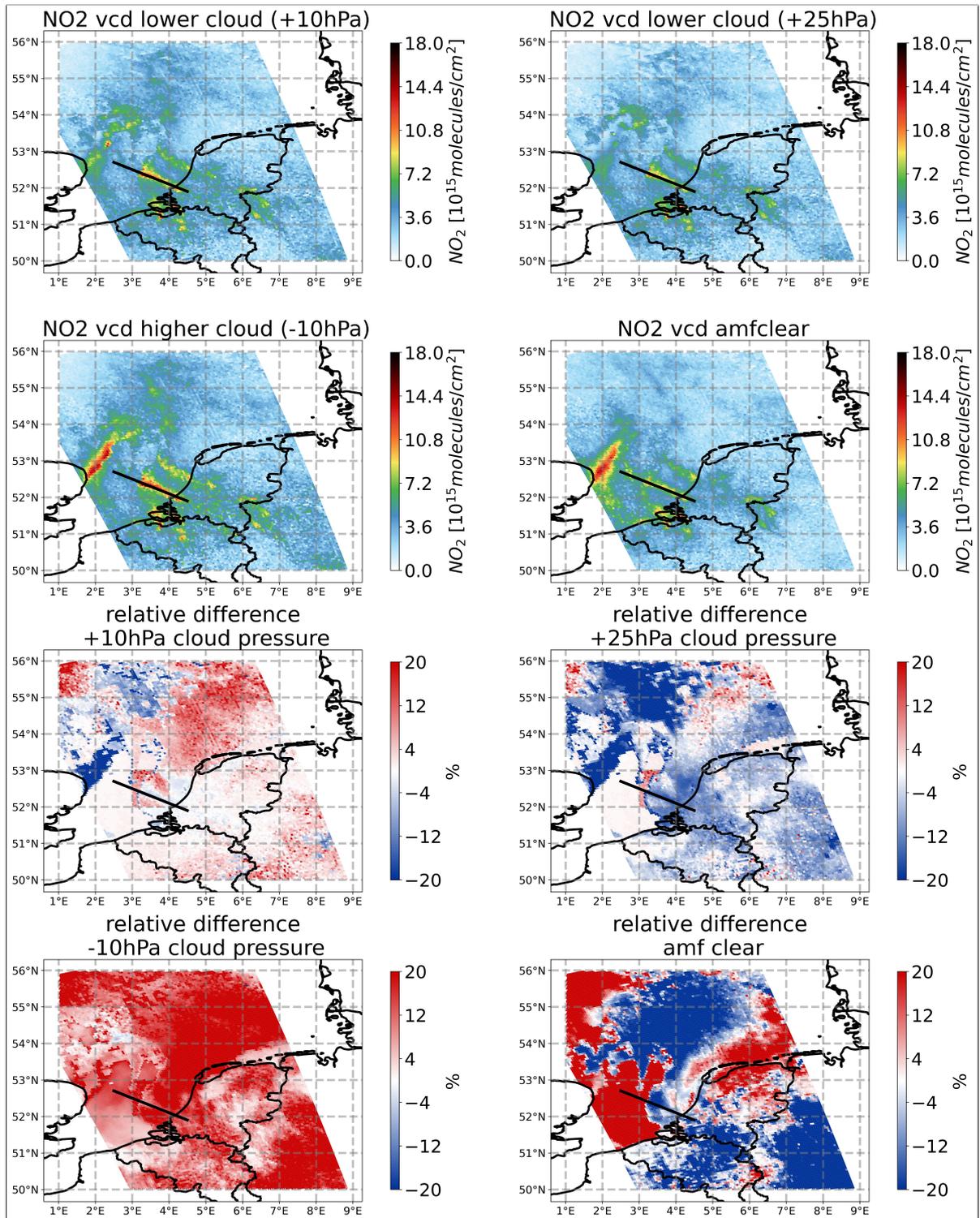


Figure C.17: case study B. From top left to bottom right: the NO_2 vertical tropospheric column after applying a positive cloud bias of 10hPa and 25hPa, a negative bias of 25hPa and the pure clear-sky air-mass factor followed by respectively the 4 relative differences under these alternative NO_2 vertical columns. The intersection gradient is shown in black.

Bibliography

- [1] J.P. Veefkind et al. "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications". In: *Remote Sensing of Environment* 120 (2012). The Sentinel Missions - New Opportunities for Science, pp. 70–83. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2011.09.027>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425712000661>.
- [2] Yuanyun Gao et al. "Assessing the impact of urban form and urbanization process on tropospheric nitrogen dioxide pollution in the Yangtze River Delta, China". In: *Environmental Pollution* 336 (2023), p. 122436. ISSN: 0269-7491. DOI: <https://doi.org/10.1016/j.envpol.2023.122436>. URL: <https://www.sciencedirect.com/science/article/pii/S0269749123014380>.
- [3] *Nitrogen dioxide over Europe*. URL: https://www.esa.int/ESA_Multimedia/Images/2019/05/Nitrogen_dioxide_over_Europe.
- [4] Nico van Breemen, Hans Grinsven, and Wim Vries. "Effects acid atmospheric deposition on soil and groundwater." In: Jan. 1998.
- [5] J. van Geffen et al. "S5P TROPOMI NO_2 slant column retrieval: method, stability, uncertainties and comparisons with OMI". In: *Atmospheric Measurement Techniques* 13.3 (2020), pp. 1315–1335. DOI: 10.5194/amt-13-1315-2020. URL: <https://amt.copernicus.org/articles/13/1315/2020/>.
- [6] K. F. Boersma et al. "An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument". In: *Atmospheric Measurement Techniques* 4.9 (2011), pp. 1905–1928. DOI: 10.5194/amt-4-1905-2011. URL: <https://amt.copernicus.org/articles/4/1905/2011/>.
- [7] ESA. *Thematic areas and services - sentinel-5p mission - sentinel online*. URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-5p/thematic-areas-services>.
- [8] J. van Geffen et al. "Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data". In: *Atmospheric Measurement Techniques* 15.7 (2022), pp. 2037–2060. DOI: 10.5194/amt-15-2037-2022. URL: <https://amt.copernicus.org/articles/15/2037/2022/>.
- [9] T. C. V. W. Riess et al. "Improved monitoring of shipping NO_2 with TROPOMI: decreasing NO_x emissions in European seas during the COVID-19 pandemic". In: *Atmospheric Measurement Techniques* 15.5 (2022), pp. 1415–1438. DOI: 10.5194/amt-15-1415-2022. URL: <https://amt.copernicus.org/articles/15/1415/2022/>.
- [10] Daniel Goldberg et al. "Enhanced capabilities of TROPOMI NO_2 : Estimating NO_x from North American cities and power plants". In: *Environmental Science Technology* 2019 (Oct. 2019). DOI: 10.1021/acs.est.9b04488.
- [11] Kazuyuki Miyazaki et al. "Air Quality Response in China Linked to the 2019 Novel Coronavirus (COVID-19) Lockdown". In: *Geophysical Research Letters* (Oct. 2020). DOI: 10.1029/2020GL089252.
- [12] Fei Liu et al. "Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19". In: *Science Advances* 6 (June 2020), eabc2992. DOI: 10.1126/sciadv.abc2992.
- [13] M. Latsch et al. "Intercomparison of Sentinel-5P TROPOMI cloud products for tropospheric trace gas retrievals". In: *Atmospheric Measurement Techniques* 15.21 (2022), pp. 6257–6283. DOI: 10.5194/amt-15-6257-2022. URL: <https://amt.copernicus.org/articles/15/6257/2022/>.

- [14] Xin Zhang et al. "Spaceborne Observations of Lightning NO₂ in the Arctic". In: *Environmental Science & Technology* 57.6 (2023). PMID: 36724410, pp. 2322–2332. DOI: 10.1021/acs.est.2c07988. eprint: <https://doi.org/10.1021/acs.est.2c07988>. URL: <https://doi.org/10.1021/acs.est.2c07988>.
- [15] Aristeidis K Georgoulis et al. "Detection of NO₂ pollution plumes from individual ships with the TROPOMI/S5P satellite sensor". In: *Environmental Research Letters* 15.12 (Dec. 2020), p. 124037. DOI: 10.1088/1748-9326/abc445. URL: <https://dx.doi.org/10.1088/1748-9326/abc445>.
- [16] Ronald van der A et al. "Connecting the dots: NO_x emissions along a West Siberian natural gas pipeline". In: *npj Climate and Atmospheric Science* 3 (Dec. 2020). DOI: 10.1038/s41612-020-0119-z.
- [17] A. M. S. Gloudemans et al. "SCIAMACHY CO over land and oceans: 2003–2007 interannual variability". In: *Atmospheric Chemistry and Physics* 9.11 (2009), pp. 3799–3813. DOI: 10.5194/acp-9-3799-2009. URL: <https://acp.copernicus.org/articles/9/3799/2009/>.
- [18] Ministerie van Algemene Zaken. *Aanpak stikstofuitstoot verminderen*. Dec. 2023. URL: <https://www.rijksoverheid.nl/onderwerpen/aanpak-stikstof-natuur-water-en-klimaat/aanpak-stikstofuitstoot-verminderen>.
- [19] *Staatsblad 2021, 140 | Overheid.nl > Officiële bekendmakingen*. Mar. 2021. URL: <https://zoek.officielebekendmakingen.nl/stb-2021-140.html>.
- [20] TNO. "FACTSHEET EMISSIES EN DEPOSITIE VAN STIKSTOF IN NEDERLAND". In: (Oct. 2019). URL: https://www.standaard.be/cnt/dmf20240128_97441236.
- [21] Ine Renson. "Dat platteland in lichterlaaie staat, legt betonrot Europees landbouwbeleid bloot". In: (Jan. 2024). URL: https://www.standaard.be/cnt/dmf20240128_97441236.
- [22] *S5P MPC L2 QC Portal*. URL: <https://mpc-l2.ms4.surfsara.nl/>.
- [23] *Worldview: Explore your dynamic planet*. URL: <https://worldview.earthdata.nasa.gov/>.
- [24] *Documentation - Access to EO data via S3*. URL: <https://documentation.dataspace.copernicus.eu/APIs/S3.html>.
- [25] H. Eskes et al. *Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Nitrogen dioxide*. Tech. rep. KNMI, 2023.
- [26] H. Eskes and J. van Geffen. *Product user manual for the TM5 NO₂, SO₂ and HCHO profile auxiliary support product*. Tech. rep. KNMI, 2021.
- [27] *Copernicus Climate Data Store | Copernicus Climate Data Store*. URL: <https://ads.atmosphere.copernicus.eu/>.
- [28] J. Douros et al. "Comparing Sentinel-5P TROPOMI NO₂ column observations with the CAMS regional air quality ensemble". In: *Geoscientific Model Development* 16.2 (2023), pp. 509–534. DOI: 10.5194/gmd-16-509-2023. URL: <https://gmd.copernicus.org/articles/16/509/2023/>.
- [29] *TEMIS – European NO₂ with CAMS a priori profiles*. URL: https://www.temis.nl/airpollution/no2_cams.php.
- [30] J. van Geffen et al. *TROPOMI ATBD of the total and tropospheric NO₂ data products*. Tech. rep. KNMI, 2022.
- [31] *Overview – Google Earth*. URL: <https://earth.google.com/intl/earth/>.
- [32] John Marshall and R. Alan Plumb. *Atmosphere, Ocean and Climate Dynamics: An Introductory Text*. San Diego, UNITED STATES: Elsevier Science & Technology, 2007. ISBN: 9780080556703. URL: <http://ebookcentral.proquest.com/lib/delft/detail.action?docID=534996>.
- [33] Ping Wang, Piet Stammes, and Nicolas Fouriner. "Test and first validation of FRESCO+". In: *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE* (Sept. 2006). DOI: 10.1117/12.687901. URL: <https://doi.org/10.1117/12.687901>.

- [34] L. G. Tilstra et al. "A directional surface reflectance climatology determined from TROPOMI observations". In: *Atmospheric Measurement Techniques* 17.7 (2024), pp. 2235–2256. DOI: 10.5194/amt-17-2235-2024. URL: <https://amt.copernicus.org/articles/17/2235/2024/>.
- [35] L. Labzovskii et al. "What does TROPOMI see over Tibetan lakes: NO_x emissions or unknown absorption in water?" In: *Nature Geoscience* (2024). in review.
- [36] Roger Barlow. *Statistics : a guide to the use of statistical methods in the physical sciences*. English. Manchester physics series. Chichester, England: Wiley, 1989. ISBN: 0471922943; 9780471922940; 0471922951; 9780471922957; 9781118723234; 1118723236.
- [37] P. Wang et al. "FRESCO+: an improved O₂ A-band cloud retrieval algorithm for tropospheric trace gas retrievals". In: *Atmospheric Chemistry and Physics* 8.21 (2008), pp. 6565–6576. DOI: 10.5194/acp-8-6565-2008. URL: <https://acp.copernicus.org/articles/8/6565/2008/>.
- [38] Ministerie van Infrastructuur en Waterstaat. *Documentatie harmonie*. Dec. 2023. URL: <https://www.knmidata.nl/open-data/harmonie>.
- [39] S. Liu et al. "An improved total and tropospheric NO₂ column retrieval for GOME-2". In: *Atmospheric Measurement Techniques* 12.2 (2019), pp. 1029–1057. DOI: 10.5194/amt-12-1029-2019. URL: <https://amt.copernicus.org/articles/12/1029/2019/>.
- [40] J. Landgraf et al. "The TANGO mission: A satellite tandem to measure major sources of anthropogenic greenhouse gas emissions". In: *EGU General Assembly 2020 (2020)*. Online, 4–8 May 2020, EGU2020-19643. DOI: 10.5194/egusphere-egu2020-19643.
- [41] J. Landgraf et al. "It Takes Two to Tango: The Twin Anthropogenic Greenhouse Gas Observers". In: *EGU General Assembly 2024 (2024)*. Vienna, Austria, 14–19 Apr 2024, EGU24-17625. DOI: 10.5194/egusphere-egu24-17625.
- [42] Eumetsat. *CO2M*. June 2021. URL: <https://www.eumetsat.int/co2m>.
- [43] European Space Agency. *EarthCARE*. Dec. 2022. URL: <https://earth.esa.int/eogateway/missions/earthcare>.
- [44] T. Heus et al. "Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications". In: *Geoscientific Model Development* 3.2 (2010), pp. 415–444. DOI: 10.5194/gmd-3-415-2010. URL: <https://gmd.copernicus.org/articles/3/415/2010/>.
- [45] A. J. M. Piters et al. "The Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI): design, execution, and early results". In: *Atmospheric Measurement Techniques* 5.2 (2012), pp. 457–485. DOI: 10.5194/amt-5-457-2012. URL: <https://amt.copernicus.org/articles/5/457/2012/>.
- [46] T. C. V. W. Riess et al. "To new heights by flying low: comparison of aircraft vertical NO₂ profiles to model simulations and implications for TROPOMI NO₂ retrievals". In: *Atmospheric Measurement Techniques* 16.21 (2023), pp. 5287–5304. DOI: 10.5194/amt-16-5287-2023. URL: <https://amt.copernicus.org/articles/16/5287/2023/>.
- [47] Eric T. Gorman et al. "The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission: an emerging era of global, hyperspectral Earth system remote sensing". In: *Sensors, Systems, and Next-Generation Satellites XXIII*. Ed. by Steven P. Neeck, Philippe Martimort, and Toshiyoshi Kimura. Vol. 11151. International Society for Optics and Photonics. SPIE, 2019, 111510G. DOI: 10.1117/12.2537146. URL: <https://doi.org/10.1117/12.2537146>.
- [48] D. G. Loyola et al. "The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor". In: *Atmospheric Measurement Techniques* 11.1 (2018), pp. 409–427. DOI: 10.5194/amt-11-409-2018. URL: <https://amt.copernicus.org/articles/11/409/2018/>.
- [49] J. E. Williams et al. "The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation". In: *Geoscientific Model Development* 10.2 (2017), pp. 721–750. DOI: 10.5194/gmd-10-721-2017. URL: <https://gmd.copernicus.org/articles/10/721/2017/>.

- [50] A. Lorente et al. "Structural uncertainty in air mass factor calculation for NO₂ and HCHO satellite retrievals". In: *Atmospheric Measurement Techniques* 10.3 (2017), pp. 759–782. DOI: 10.5194/amt-10-759-2017. URL: <https://amt.copernicus.org/articles/10/759/2017/>.
- [51] A. Heckel et al. "Influence of low spatial resolution a priori data on tropospheric NO₂ satellite retrievals". In: *Atmospheric Measurement Techniques* 4.9 (2011), pp. 1805–1820. DOI: 10.5194/amt-4-1805-2011. URL: <https://amt.copernicus.org/articles/4/1805/2011/>.
- [52] D. Loyola et al. *S5P/TROPOMI ATBD Cloud Products*. Tech. rep. Deutsches Zentrum für Luft- und Raumfahrt, 2023.
- [53] R. Koelemeijer et al. "Global distributions of effective cloud fraction and cloud top pressure derived from oxygen A band spectra measured by the Global Ozone Monitoring Experiment: Comparison to ISCCP data". In: *Journal of geophysical research* 107.D12 (June 2002). DOI: 10.1029/2001jd000840. URL: <https://doi.org/10.1029/2001jd000840>.
- [54] *QA4ECV – Tropospheric NO2*. URL: <https://www.temis.nl/qa4ecv/no2.html>.
- [55] H. J. Eskes and K. F. Boersma. "Averaging kernels for DOAS total-column satellite retrievals". In: *Atmospheric Chemistry and Physics* 3.5 (2003), pp. 1285–1291. DOI: 10.5194/acp-3-1285-2003. URL: <https://acp.copernicus.org/articles/3/1285/2003/>.