GAIA-CLIM deliverable D1.11

Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:

Scientific Assessment of Gaps Based on Forward, Inverse and Data Assimilation Modelling

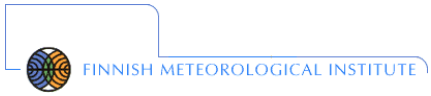
A Horizon 2020 project; Grant agreement: 640276

Date: 31/12/17

Lead Beneficiary: KIT

Nature: R

Dissemination level: PU



Executive Summary

Gaps in the ground-based observational networks for characterization and validation of satellite observations of atmospheric essential climate variables (ECVs) are assessed with the help of forward, inverse and data assimilation models. This analysis addresses primarily, but not exclusively, geographical gaps in the current observational networks. Different modelling approaches are exploited to assess the impacts of geographical gaps in the networks for the greenhouse gases methane and carbon dioxide, for the short-lived climate forcing gas carbon monoxide (CO), for aerosols, ozone and reference networks for temperature and humidity.

Specific findings and corresponding recommendations of this work are:

Greenhouse gases:

- The coverage of the TCCON network in geographic and parameter space was assessed using OCO-2 measurements for the years 2015-2016
- A lack of geographic coverage in the tropics is clear: while this will be somewhat remedied with the 2017 addition of the station in Burgos, Philippines, this is offset by the removal of the Amazon station at Manaus
- Ocean glint measurements are poorly constrained polewards of 50° south and in the northern hemisphere in winter, despite the large geographic match criteria used.
- The processing time and/or embargo time of the TCCON data can create a temporary temporal gap in the validation network, of particular importance for young missions or when investigating anomalous satellite results.
- In terms of parameters space, gaps were identified at high-albedo and high-aerosol-load sites.

Carbon monoxide:

- The added value of profile information of carbon monoxide (CO) from Fourier Transform Infrared (FTIR) measurements of the Network for the Detection of Atmospheric Composition Change (NDACC) for providing additional constraints on CO sources and OH concentrations has been explored in an inverse modeling study.
- The FTIR observations however, do not provide a clear picture: the FTIR profiles currently do not provide sufficient information to discriminate between different CO budgets, suggesting that model/data discrepancies for the most part cannot be traced to uncertainties in CO sources and sinks.
- Evaluation of FTIR profiles using independent high-quality vertically resolved CO measurements shows that the model/data discrepancies can be attributed for a substantial part to potentially erroneous FTIR profiles.
- Priority should be given to strengthening the quality and robustness of the FTIR profile retrievals. Additional measurements (either more frequent measurements at existing sites or FTIR retrievals at additional sites) without improvements in the FTIR

measurement understanding and processing will most likely not contribute to narrow the vertical CO information gap.

- We suggest profiles with AIRCORE (which performs in-situ balloon based profile information capture) co-located with FTIR measurements would help to better characterize FTIR profiles.

Aerosols:

- Correlation lengths of aerosol optical depths (AOD) from the MODIS satellite instrument vary from 100-200 km over populated continental areas to 1500 km in remote ocean or desert areas
- Global climate-aerosol models are not yet good enough to reproduce measured diurnal cycles or spatial correlations
- We recommend to take local correlation structure into account when filling gaps between AERONET stations with satellite data.
- Further, we recommend to carry out long-term, in-depth analysis and development of aerosol transport, chemistry, and removal in models beyond simply validating weather parameters.

Ozone:

- Projected ozone changes are in the range of a few per cent for total ozone, with changes in the stratospheric column and the tropopspheric column of similar absolute size. Changes in tropospheric ozone thus partly mask changes in stratospheric ozone, which is an important climate variable. This demonstrates the challenges for a ground-based network for the purpose of validating long-term satellite ozone records.
- From the model simulations, we find that the long-term changes in tropical stratospheric ozone (which has very long correlation scales) are sufficiently well represented by the existing stations of the reference network. However, for tropical tropospheric ozone, the network is inadequate as much shorter correlation scales mean that single stations are less representative for the tropics as a whole.
- A statistical analysis from the model simulations supports the observationally driven statistical analysis of small scale ozone variations (GAIA-CLIM report D1.10) and offers the potential to extend this analysis to regions were there are not enough observational data available for a statistical analysis.
- Maintaining a ground-based network in the tropics to observe vertically resolved ozone is essential to characterize long-term ozone changes, in particular for the validation of nadir viewing satellite data that either provide only total columns or have only broad weighting functions/averaging kernels.

Temperature and humidity:

- The current GRUAN site locations are able to help diagnose the sources of some types of O - B bias for some instruments including: global MHS biases, global ATMS biases and the Northern Hemisphere biases of AMSU-A
- The planned expansion of the GRUAN network is needed for diagnosing the source of geographical bias in AMSU-A and MHS O - B values and this study supports that expansion. Sites in Antarctica are of particular importance for diagnosing O - B biases in AMSU-A, since there are strong biases over the South Pole for AMSU-A channels 10 - 11 (500 - 20 hPa) in the Southern Hemisphere winter.
- Further expansion of the GRUAN network to sites in the Southern Hemisphere, at latitudes below 50S (capturing temperature at heights of 500 - 20 hPa), and sites just south of North-Eastern Russia and Alaska in the Northern Hemisphere (capturing temperature at heights of 1000 - 50 hPa), would be desirable

It is intended that this report's relevant findings be reported back to the assessed networks to inform their future strategic planning although this will largely occur subsequently to formal completion of the GAIA-CLIM project.

Table of Contents

1. Introduction and Background

Gaps in the ground-based observational networks for characterization and validation of satellite observations of atmospheric essential climate variables (ECVs) are assessed with the help of forward, inverse and data assimilation models. This report, summarising the findings from GAIA-CLIM Task 1.5, builds upon two previous reports, published as GAIA-CLIM Deliverable Reports D1.1 and D1.5. These two reports presented intermediate steps during the project; thus the current report in many cases encompasses results from these previous reports.

The presentation and discussion of results in this report are organized by the topics addressed, together with the specific model systems that are used for:

(1) Greenhouse gases CO_2 and CH_4,
(2) air quality and short-lived climate forcing gases (CO),
(3) aerosols,
(4) ozone and related trace gases, and
(5) assessing geographical gaps for reference temperature and humidity.

Table 1 summarizes the topics addressed and models involved.

Table 1. Updated topics addressed in Task 1.5 by specific models, the respective ground-based observational networks addressed, and the institutions responsible for this part of the work.

	Model	***Technique***	***Network addressed***	***Institution***
Greenhouse gases, CO_2, CH_4	**TM3**	**Global tracer transport model**	**TCCON**	**MPI BGG**
Air quality, CO	**IMAGESv2**	**Inverse modelling**	**NDACC**	**BIRA**
Aerosols	**ECHAM-HAMMOZ**	**Aerosol Climate model**	**AERONET**	**FMI**
Ozone and related trace gases	**EMAC (ECHAM5-MESSy2)**	**Chemistry Climate model**	**NDACC**	**KIT**
Temperature and humidity	**ECMWF IFS**	**Numerical Weather Prediction**	**GRUAN**	**ECMWF**

2. Results from model simulations

2.1 Greenhouse gases, CO_2 and CH_4 (MPI BGC)

2.1.1 Scope of the assessment

Spaceborne measurements of the greenhouse gases CO_2 and CH_4 in the shortwave infrared, such as those from GOSAT (CO_2 and CH_4), OCO-2 (CO_2) and most recently Sentinel-5P (CH_4) are reliant upon more precise and accurate ground-based total column measurements for calibration and validation. Profile measurements, available from both aircraft and more recently AirCore (Karion et al., 2010), provide an extra dimension for comparison with their vertical resolution, but these tend to be quite limited in space and time due to cost, and exclude at least part of the stratosphere. At present the main source of calibration and validation measurements for column-integrated CO_2 and CH_4 (hereafter XCO_2 and XCH_4, respectively) is the Total Carbon Column Observing Network, TCCON (Wunch et al., 2011). TCCON is a distributed network of surface-based Fourier Transform Infrared (FTIR) spectrometers, operated independently by different institutions, but following common standards and practices to ensure comparability of the measurements.

This work aims to assess the representativeness of this network for the validation of satellite measurements, both in terms of spatial coverage and in terms of the parameters that are used for the ad hoc bias corrections currently applied to XCO_2 and XCH_4 retrievals, which are derived from mismatches between satellite retrievals and TCCON measurements. The work previously done in this project, presented in Deliverable 1.5, concentrated on the comparison to total column measurements from the GOSAT satellite for the year 2010. Here the work has been expanded to more recent XCO_2 measurements available from OCO-2. The analysis covers the full calendar years 2015 and 2016. Note that this is not a validation study for OCO-2, and there is no direct comparison of XCO2 values between TCCON and OCO-2: Instead, the emphasis is on the coverage of the validation network, both in geographic space and in the relevant parameter space.

The location of the TCCON stations that made measurements during this time are plotted in Figure 1.1, and the data coverage (in soundings per degree latitude per day) for both OCO-2 and TCCON is presented in Figure 1.2. From this it can be clearly seen that not all TCCON sites are operational over the whole data record. For instance, the instrument at Manaus was only operated for one year at that location and has since been removed, and new sites came online during the study period (East Trout Lake, Paris, Anmyeondo).

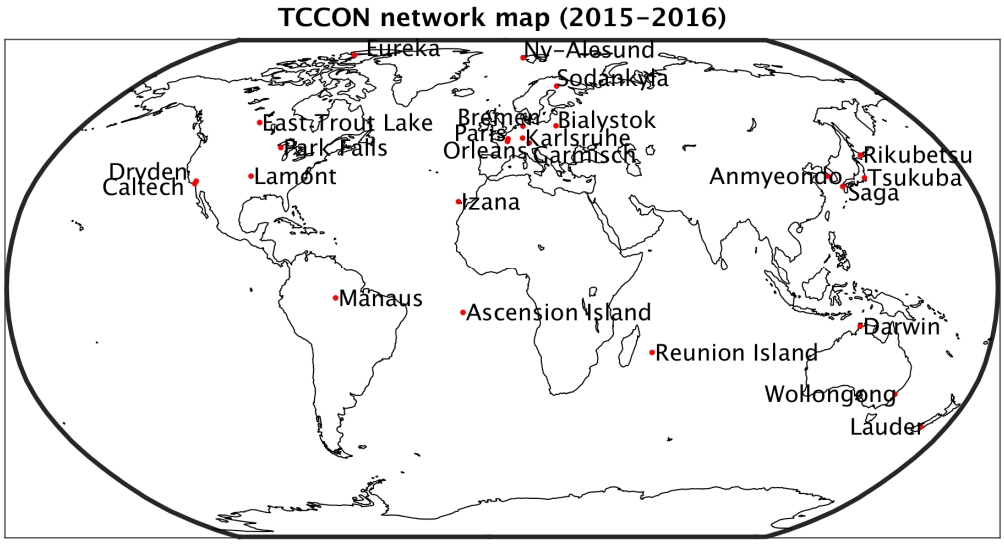


Figure 1.1. Location of TCCON stations operating in 2015 and 2016 which were used for the assessment.

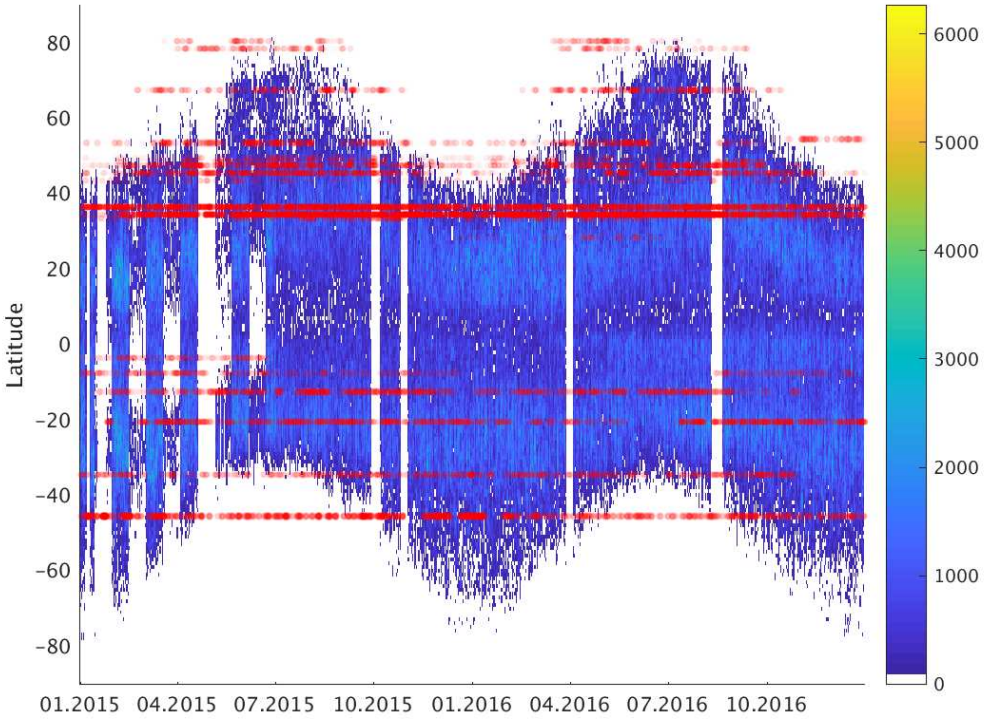


Figure 1.2. An illustration of the data coverage in space and time from both OCO-2 and TCCON. The measurements are summed onto 1 degree latitudinal bins per day. Solid red for TCCON corresponds to 1000 soundings per day, and scales linearly.

In order to quantify at what spatial and temporal scales TCCON measurements are representative of the atmospheric composition, a sophisticated collocation approach was used, following Guerlet et al. (2013). In this method, the spatial criteria are quite loose, but there is an added constraint brought by concentration gradients in modeled fields that take

into account the variability in the total column concentration due to both transport and the influence of local fluxes.

2.1.2 Model and data used and simulations

As described previously in D1.5, the decision was made to use optimized 3D fields of carbon dioxide from the Copernicus Atmosphere Monitoring Service at 1-degree spatial, 3-hourly temporal resolution. The column-integrated CO_2 have been retrieved from CO_2 forecasts constrained with the GOSAT XCO_2 product from the Bremen Optimal Estimation Differential Optical Absorption Spectroscopy (BESD) algorithm (Agustí-Panareda et al., 2014, Massart et al., 2016, Agustí-Panareda et al., 2017). The original 0.25 to 0.5 degree resolution was aggregated to 1 x 1 degree to be consistent with the treatment in Guerlet et al. (2013). This is required as the threshold used to determine collocation is dependent on model resolution.

The OCO-2 data (version B8100r "lite" files) are available from the Goddard Earth Sciences Data and Information Services Center (GES-DISC, OCO-2 Science Team/Michael Gunson (2017)). A documentation of the version B8 lite files can be found in Osterman et al. (2017). The analysis of OCO-2 data from the daily "lite" files has been restricted to data for which the warn level is zero and for which the "xco2 quality flag" is zero. OCO-2 data for two different observation modes are used in this comparison. One is the nadir mode, in which the satellite points straight downwards to the Earth's surface, and one is the glint mode, in which the instruments points just off the glint spot on Earth's surface. The glint data can be divided into glint over land ("land glint") and glint over ocean ("ocean glint"). Because the bias corrections are calculated separately for these three observation modes, the analysis is split into these observation modes as well.
Unlike in the satellite products presented in D1.5, the most recent OCO-2 data version does *not* directly use the TCCON measurements for the parameter-based bias correction. Instead there is a three-step bias correction, described in Osterman et al. (2017), and shown in the following equation:

$$XCO_{2,corrected} = \frac{XCO_{2,raw} - FOOT(fp, mode) - FEATS(mode)}{TCCON\_ADJUST(mode)}$$

The FOOT term is based on the eight across-swath footprints of the OCO-2 measurement, looking at scenes where across-track variability is expected to be negligible, to remove footprint-related systematic features (as a function of cross-track footprint *fp*). The TCCON_ADJUST term *does* directly use collocated TCCON measurements, but only to solve for a global offset as a function of measurement mode. These two terms are not considered in the current assessment. The third term, FEATS, looks at features of unphysical variability within the retrieved XCO2 data that are correlated with various retrieved geophysical parameters. This correlation assessment is done in a variety of ways, but all point to the same three parameters being of significance:

Table 1.1. Retrieved parameters found to be significant in explaining unphysical variability in OCO-2 retrievals.

Parameter name	Description
dP	The difference between the retrieved and the *a priori* surface pressure
DWS	The retrieved abundance of coarse aerosol (DSW = Dust, Water clouds and Sea salt)
co2_grad_del	Very large and unphysical variation in the retrieved vertical profile of CO_2 from that assumed in the prior.

These retrieved parameters are then applied for the FEATS term in the bias correction as follows:

Table 1.2. The application of the parameters in the FEAT term of the OCO-2 B8r bias correction.

Mode	FEATS term (ppm)
Ocean glint	-0.23 (dP - 3.1) + 0.09 (co2_grad_del + 6.0)
Land (nadir & glint)	-0.36 (dP) - 8.5 (DWS) - 0.029 (co2_grad_del - 15)

Although these parameters are not directly derived using TCCON measurements, the validation of this bias correction approach is still dependent on the representativeness of the TCCON measurements across the parameter space spanned by the retrieved dP, DWS, and co2_grad_del. As such, these, together with the geographic coverage, form the basis of this assessment.

For this study, TCCON data from 25 different stations have been used, as shown in Figures 1.1 and 1.2. The TCCON data were obtained from the TCCON Data Archive, hosted by CaltechDATA, California Institute of Technology, CA (US) (doi:10.14291/tccon.archive/1348407). The data were retrieved on November 1, 2017.

2.1.3 Analysis

The collocation method to determine the region of representativeness around each TCCON site is based on that described in Guerlet et al. (2013). This is done by analyzing the optimized fields of XCO_2 (or XCH_4) at 1° resolution, and determining regions over which the total column value does not vary beyond a given threshold amount (e.g. ±0.5 ppm for XCO_2). What is important to note is that the exact value of the modeled field is not critical, and the value measured and the value modeled are not directly compared. Rather, the variability of the modeled field is assessed. This takes into account variability related to the heterogeneity of the fluxes in the near-field and the impact this has on the total column, as well as synoptic variability through transport.

An example is shown for the TCCON station at Bialystok in Figure 1.3. The upper left panel shows the simulated XCO_2 fields for one three-hourly timestep within the geographic range ±7.5° latitude, ±22.5° longitude. The upper right shows all the regions that match the

simulated XCO_2 value at Bialystok within the ±0.5 ppm threshold, and the lower left shows the effect of filtering such that the matches are contiguous. All OCO-2 soundings that fall within this region are considered collocations. A final step (not shown) involves temporally filtering the colocations to include only those with TCCON measurements within ±2 hours of the OCO-2 overpass. Because the longitudinal extent is rather large, there are sometimes overpasses from two different orbits within a given collocation box, offset in time by approximately 1:40. In this case, the average time of the OCO-2 measurements are used to determine collocation with the TCCON measurements.

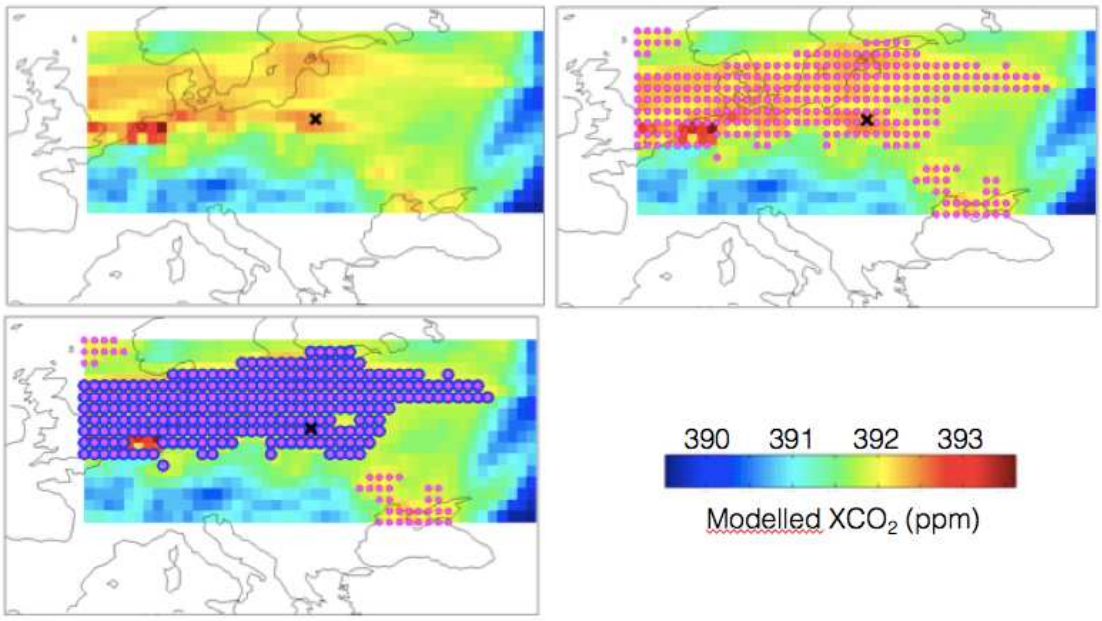


Figure 1.3. An illustration of the applied collocation criteria, based on Guerlet et al. (2013). The upper left panel shows modelled XCO2 fields around the station Bialystok within the geographic area ±22.5° longitude and ±7.5° latitude for one timestep. The upper right panel shows in pink all pixels in this box that fall within the matching threshold ±0.5 ppm, and the lower left panel circles in blue all those that are contiguous. In addition to this, a temporal filter of ±2 hours for satellite/TCCON measurements was applied.

Because the collocation is dependent on both transport and flux fields, the area of collocation varies throughout the year, as was illustrated in D1.5.

2.1.4 Results and Recommendations

Figure 1.4 summarize the findings for the spatial representativeness of the measurements as a function of latitude for the months June, 2015, and December, 2015, for the viewing modes nadir (top panel), land glint (middle panel) and ocean glint (bottom panel). All of the distributions show a distinctive bimodal shape, with a lack of measurements around the equator, due to the persistent cloudiness of the inter tropical convergence zone. As both TCCON and OCO-2 measurements rely upon a clear line of sight between the surface and the sun (and additionally between the surface and the spacecraft for OCO-2), this is to be expected. This dependence on the sun also explains the latitudinal shift towards the summer hemisphere seen when comparing the upper three panels to the lower three panels.

From these plots the most striking absence is any colocations between approximately 0° and 20° north. This is a clear gap in the validation network, affecting all three observation modes

during both seasons shown here, and is typical for the undersampled tropics. Since these measurements were made, a new station has come online: Burgos, Ilocos Norte, Philippines, at 18.5325°N, 120.6496°E, which commenced measurements in March 2017 (Velazco et al., 2017). While not closing the tropical gap entirely, it will certainly be a step in the right direction for better spatial coverage in the tropical region. However, the Amazon site at Manaus is no longer in operation, again leaving a significant gap in spatial coverage.

Another notable gap in the spatial distribution can be seen in the ocean glint colocations polewards of approximately 50° south. Because of the general homogeneity of the southern hemisphere in XCO2, this may not be critical, but certainly any dynamics related to the southern polar vortex are entirely unseen by the current TCCON network.

Interestingly, a partial gap in geographic coverage can be seen in the ocean-glint collocations in the northern hemisphere as well, despite the significantly higher density of stations in the northern mid-latitudes. Only Izaña provides a handful of collocations in this case. Because of the large geographic extent of the oceans, there are many satellite measurements in these latitude bands, despite comparatively high cloud cover and the increased cloud contamination when operating in glint mode. Even when using the large geographic collocation criteria of this approach and despite the fact that more than half of the 25 TCCON stations are near coasts in the northern hemisphere (Eureka, Ny-Ålesund, Sodankylä, Bremen, Paris, Orléans, Rikubetsu, Anmyeondo, Tsukuba, Caltech, Dryden, Saga, and Izaña), the continental influence on the modeled total column abundances restricts the actual collocations significantly, such that the northern hemisphere glint measurements in the winter are very poorly represented by the TCCON network. Here an augmentation with AirCore or aircraft profile measurements, especially throughout the winter, would be of great benefit to data users such as inverse modellers. The added profile information would also be beneficial for the assessment of retrieval algorithms.

Moving on to the parameters used for the OCO-2 bias correction, Figure 1.5 shows similar histograms for the distribution of dP, the difference between the retrieved and the *a priori* surface pressure, as a function of measurement mode. In order to better see the shape of the distribution of the TCCON collocations (both potential and actual, as in Figure 1.4), these have been multiplied by a factor of five. Ideally these distributions would be quite narrow, and centered on zero, as we expect that our prior knowledge of the surface

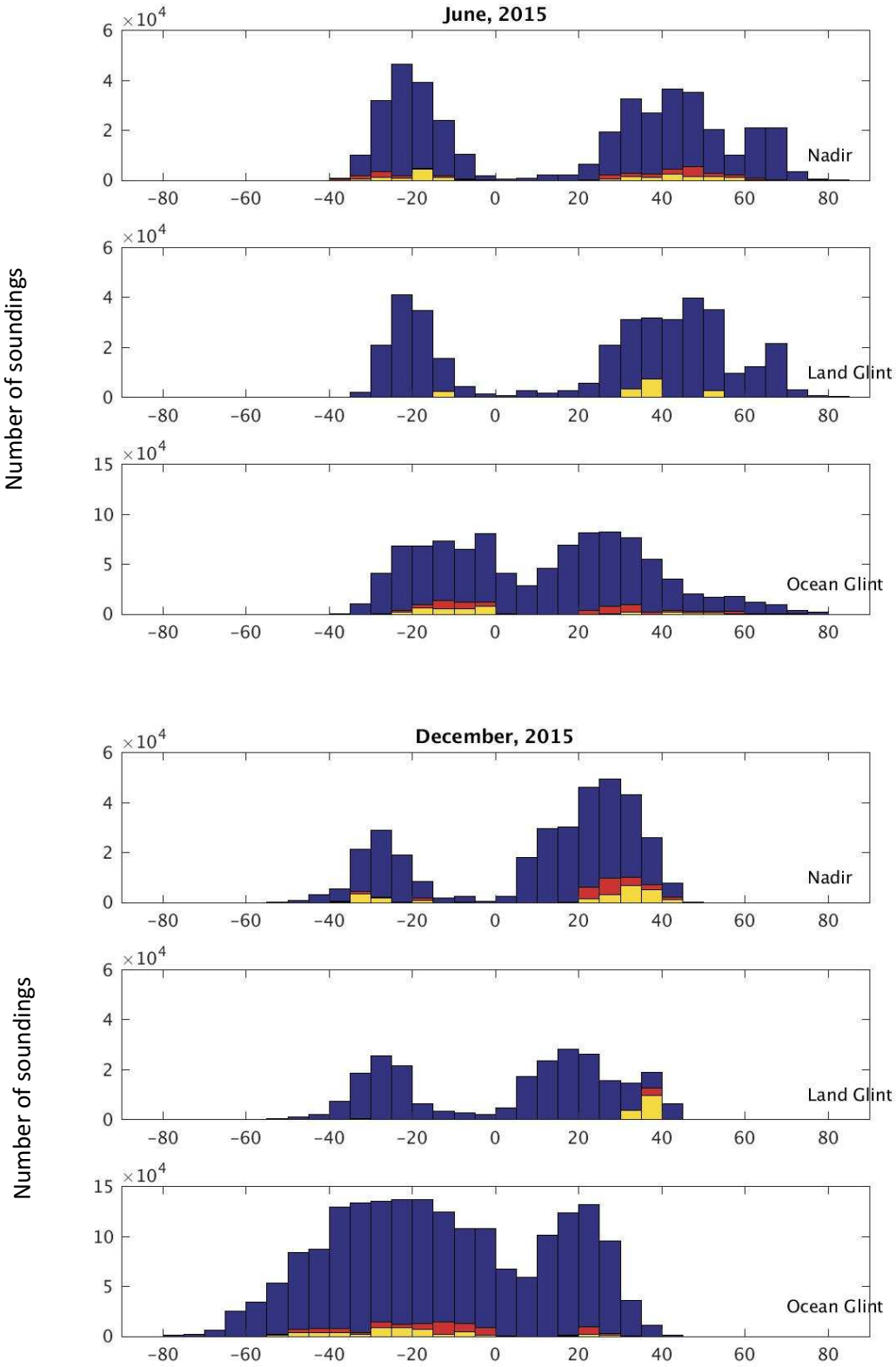


Figure 1.4. In all plots the number of OCO-2 soundings per 5 degree latitude bin per month are shown in blue. The "potential" TCCON collocations using the described collocation criteria and the station locations are shown in red, and the actual collocations are in yellow. The upper three panels are for June, 2015, and the lower three panels are for December, 2015. For each month there is one plot for each of the viewing modes, as indicated on the right.

pressure is rather good, and large deviations may be indicative of general problems in the forward model. A comparison of the distributions shows that the TCCON subset is closer to being centered on zero than the full OCO-2 distribution for all three observation modes. Keep in mind that only "good" soundings were considered, and these are already filtered for extreme values of dP depending on observation mode, so these distributions are artificially truncated to begin with. Despite this difference in mean and median, the TCCON subset does a fairly good job of sampling the spread of the distribution, with the exception of high dP values for land glint mode.

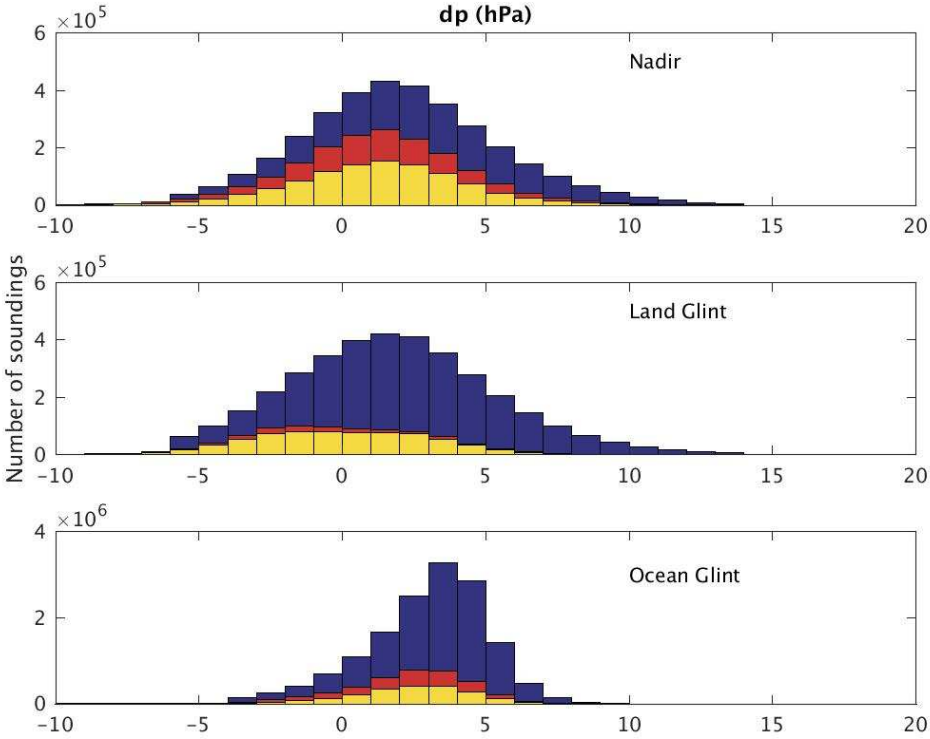


Figure 1.5. The distribution in parameter space for dP, the difference between the retrieved and prior surface pressure. All OCO-2 soundings are shown in blue, while potential TCCON collocations and actual TCCON collocations, both multiplied by 5 for visibility, are shown in red and yellow respectively. The three observation modes are plotted separately, as in Figure 1.4.

The next parameter considered is co2_grad_del, which is a unitless measure of how different (and unphysical) the vertical distribution of the retrieved CO_2 profile is from the prior. As for dP, we expect these distributions to be centered on zero, and large deviations may indicate problems with assumptions used in the retrieval. Similar to dP, we again find that the TCCON subset (both potential and actual, and both multiplied by five for visibility) is closer to being centred on zero than the full OCO-2 distribution. However in this case, the TCCON subset does a better job of sampling the parameter space fully, with the possible exception of very low values in ocean glint mode.

And finally, the retrieved parameter describing the load of coarse mode aerosols "DWS" (Dust, Water clouds and Sea salt) is presented in Figure 1.7. Here the coverage by TCCON collocations is actually quite representative, and the distributions peak in similar locations.

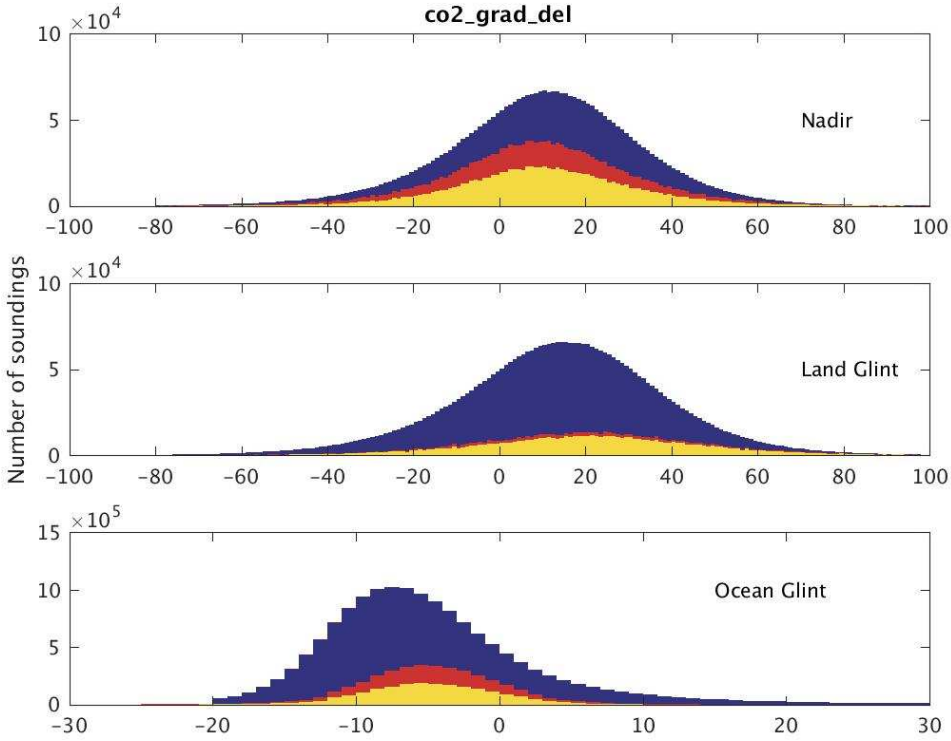


Figure 1.6. As in Figure 1.5, but for co2_grad_del, a unitless measure of deviations in the vertical structure of the retrieved CO2 profile. Again, the TCCON collocations were multiplied by five for visibility.

What may not be fully covered is the long tail, but this is cropped by the quality-filtering requirement that DWS < 0.25 for land (Osterman et al., 2017). Interestingly, the comparison of this particular aerosol measure suggests that the TCCON network represents the parameter space well, which was not the case for the aerosol parameter in the RemoTeC retrievals presented in D1.5. This is likely a question of factor of fine vs. coarse mode aerosol, with the former being more indicative of anthropogenic influence. This is highly relevant for retrieval of greenhouse gases from satellite, as multiple scattering from aerosols changes the effective path length of the light, and can lead to significant biases. Because most of the TCCON sites were chosen to be in background locations, far from local sources of fine mode aerosol, this parameter is not generally well sampled by TCCON, however this may be changing with the newly established site in Paris, and plans underway for at least one FTIR site in China to join the network over the next years.

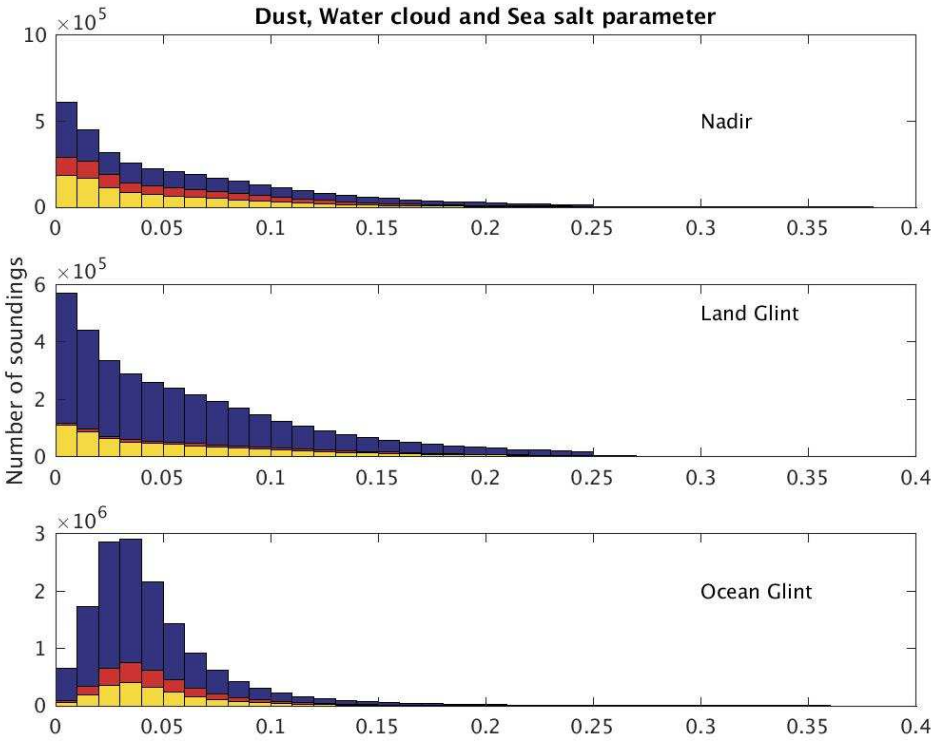


Figure 1.7. As in Figure 1.5 and 1.6, but for DWS, a unitless measure of the coarse mode aerosol, representing Dust, Water clouds, and Sea Salt.

Finally there is one more factor, and potential gap in the validation network that must be considered. The availability of the TCCON data generally lags by some months, or even more than a year. The TCCON data used in this study were retrieved from the repository on November 1, 2017. At this point, only 8 of the 25 stations had processed and made available their measurements up to the very end of 2016. On the one hand this is a question of time and resources, but there is also a one-year embargo policy in place for many of the sites in the face of lacking operational funding. In the case of funded validation requests (as recently for Sentinel-5P), these data have been made available more quickly to a subset of users. Particularly in the case of new missions and when investigating anomalous signals (such as the 2015-2016 El Niño), such an embargo can be considered a temporary temporal gap in the validation network.

2.2 Air quality, carbon monoxide (BIRA)

2.2.1 Scope of the assessment

This part of the study addresses the added value of vertical CO profiles at the Fourier-transform infrared spectroscopy (FTIR) stations of the Network for the Detection of Atmospheric Composition Change (NDACC). Suborbital CO measurements are essential to validate Earth Observation (EO) measurements of CO and to evaluate model-based analyses aimed at a better understanding of the CO budget. Although FTIR CO total column data from NDACC are known to be of great value in such activities, it is unclear (a) to what extent the currently available FTIR profile data can also contribute, (b) whether additional FTIR profile measurements would be desirable, either as more frequent measurements at existing sites, or as measurements at additional sites, and (c) whether the quality of the FTIR retrievals is sufficient, or if it needs to be improved to meet the needs of satellite characterisation.

To address those questions, we have compared modelled CO fields constrained by satellite CO data with a wide range of suborbital CO measurements, including FTIR CO profiles. Besides constraining the global sources of CO as well as its chemical sink by reaction with OH, this study also aims to assess the added value and the compatibility of CO FTIR profile measurements with respect to other types of measurements. We also conducted a model-independent evaluation of FTIR profile data against several aircraft measurement datasets.

2.2.2 Model used and simulations

Model description

The model used is the Chemistry Transport Model (CTM) IMAGES (Intermediate Model for the Annual and Global Evolution of Species) (Stavrakou et al., 2016; Bauwens et al., 2016) which calculates the distribution of 170 chemical compounds at 2°×2.5° resolution with 40 vertical levels between the surface and the lower stratosphere (44 hPa). Meteorological fields are provided by ECMWF ERA-Interim analyses. The model uses anthropogenic emissions from the Hemispheric Transport of Air Pollution inventory version 2 (HTAPv2) (Janssens-Maenhout, 2015), with the non-methane volatile organic compound (NMVOC) speciation from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) (Lamarque et al., 2010), pyrogenic emissions from the Global Fire Emissions Database (GFED4s) (Randerson et al., 2012; Giglio et al., 2013; http://www.globalfiredata.org) and biogenic Volatile Organic Carbon (VOC) emissions calculated using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2006; Müller et al., 2008; Bauwens et al., 2016). Biogenic CO emissions and CO dry deposition on soils are also taken into account (Stavrakou and Müller, 2006). The chemical degradation mechanism is described in Bauwens et al. (2016) and includes recent isoprene mechanism updates (e.g. OH recycling) as well as an explicit treatment for 16 other NMVOCs.

To minimize the impact of model errors associated with stratospheric chemistry and the boundary condition at the model top, the CO mixing ratios in the 7 uppermost layers (44, 55,

67, 80, 96, 113 and 133 hPa) are fixed and set to the reanalysis of the Aura Microwave Limb Sounder (MLS) observations (Santee et al., 2017) by a stratospheric composition assimilation system, the Belgian Assimilation System for Chemical Observations (BASCOE) (Errera et al., 2017). The IMAGES CO mixing ratios constrained by BRAM (BASCOE Reanalysis of MLS observations) were compared with ER-2 aircraft measurements of the SEAC^{4}RS campaign (Toon et al., 2016). As seen on Fig. 2.1, BRAM is found to overestimate the aircraft data by ca. 23% on average above 113 hPa, and to underestimate CO by 15% at 133 hPa.

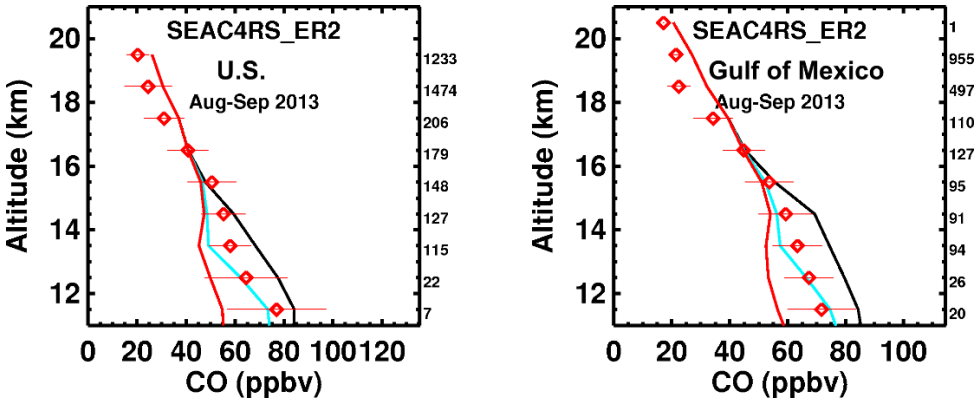


Figure 2.1. Average CO mixing ratio measured by the ER-2 aircraft of the SEAC^{4}RS campaign (red diamonds with whiskers denoting standard deviation). The curves are IMAGES-modeled CO mixing ratios. Black: the 6 uppermost levels are set to the BRAM dataset; turquoise: 7 levels; red: 8 levels.

Inversion methodology

The source inversion methodology follows Stavrakou and Müller (2006). The emission flux for a given source category is expressed as

$$G(\mathbf{x}, t, \mathbf{f}) = \sum_{j=1}^{m} \exp(f_j)\, \varphi_j(\mathbf{x}, t)$$

with f the state vector of the emission parameters, x the spatial coordinates (latitude and longitude), t the temporal coordinate (month and year), and ϕj the a priori emission distributions. The optimization seeks values of f minimizing the cost function

$$J(\mathbf{f}) = (H(\mathbf{f}) - \mathbf{y})^T \mathbf{E}^{-1} (H(\mathbf{f}) - \mathbf{y}) + \mathbf{f}^T \mathbf{B}^{-1} \mathbf{f}$$

with y the vector of CO observations; H the model operator; B and E are the covariance error matrices of the observations and of the emission parameters, respectively.

E is assumed to be diagonal, and includes instrumental uncertainties as well as representativity and model uncertainties. B is non-diagonal. The spatiotemporal correlations among the a priori uncertainties on f are defined following Stavrakou and Müller (2006), with a few adaptations.

The cost function is minimized using an iterative quasi-Newton optimization algorithm involving the calculation of the gradient of the cost function by the adjoint of the model (Müller and Stavrakou, 2005).

Observations used to constrain CO sources: IASI total columns

The observational dataset (y) consists of monthly averaged CO total columns from the Infrared Atmospheric Sounding Interferometer (IASI) onboard MetOp-A (Clerbaux et al., 2009) for 2013. IASI provides a global coverage of CO profiles twice a day. IASI CO data were previously used in global assimilation and inverse modelling studies (e.g. Klonecki et al., 2012; Fortems-Cheiney et al., 2009) and evaluated against other satellite datasets (George et al., 2009). The retrieval code of the dataset used here is the 20100815 version of FORLI-CO (George et al., 2009). The instrument sensitivity is highest in the mid- to upper troposphere. The vertical information content is coarse, usually between one total column and two partial columns, and the profiles are therefore not used in the inversion. The model operator accounts for the IASI averaging kernels (Rodgers and Connors, 2003) and for the IASI temporal sampling involved in the monthly averages.

Since the information content is especially low at high latitudes during the winter, due to the low temperatures and low signal-to-noise ratio (Pommier et al., 2010), we filter out the winter months, defined by: October-May (>75N), November-April (65-75N), November-March (55-65N), November-February (48-55N), and similarly in the Southern Hemisphere (shifted by 6 months). This definition was guided by comparisons with FTIR total column data as detailed further below.

The IASI retrieval error is typically very low (2-5%) over continents during summer, and somewhat higher (5-15%) elsewhere. The number of IASI data per model grid cell and per month is generally very high (200-2000) and only the systematic part of the retrieval error contributes non-negligibly to the measurement error of the super-observations, i.e. the monthly averaged columns at 2°×2.5°. The large number of measurements per super-observation also reduces greatly the representativity error associated with CO column variability within the same grid cell and month (Miyazaki et al., 2012). The dominant error component is the model error, which is more difficult to assess. It is here taken to be 35% of the CO column. This choice is supported by the χ^2 diagnostic which should be close to one after optimization (Klonecki et al., 2012). Its value is ca. 0.8 after optimization.

FTIR total columns used for evaluation

CO total columns and vertical profiles were retrieved from ground-based FTIR observations at 16 sites of the NDACC network in 2013 (ftp.cpc.ncep.noaa.gov/ndacc/station). Their distribution is shown in Fig. 2.2.

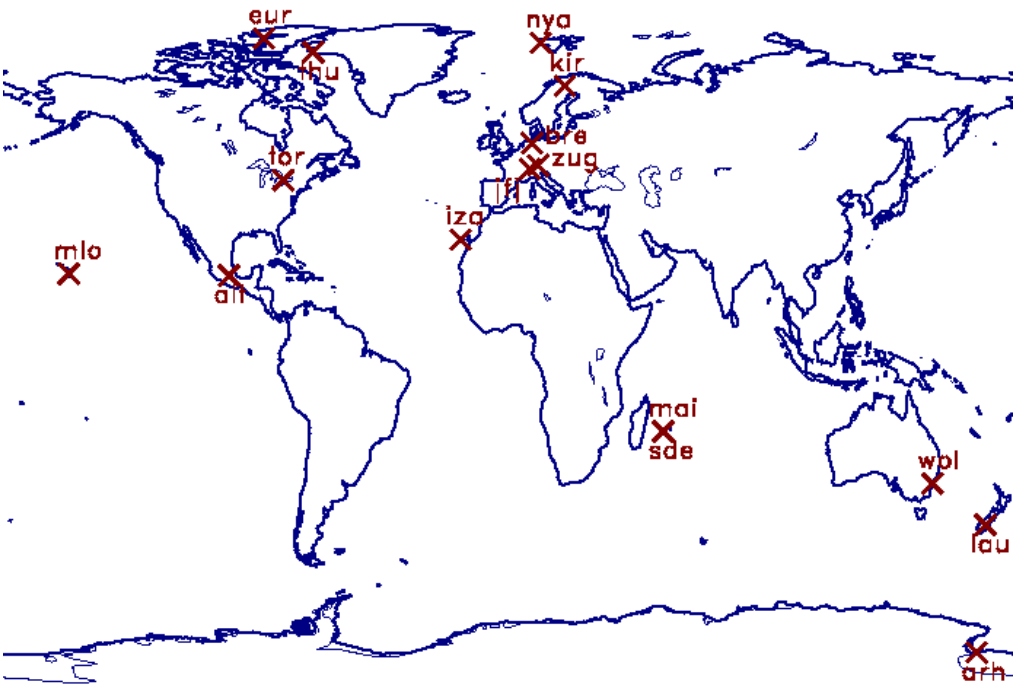


Figure 2.2. Distribution of the 16 NDACC sites with FTIR CO data used in this work. The coordinates of the sites are given in Fig. 2.3.

NDACC FTIR CO data have proven useful for validating satellite datasets including IASI (Dils et al., 2006; Senten et al., 2008; Kerzenmacher et al., 2012) or for evaluating models (Stavrakou et al., 2006; Hooghiemstra et al., 2012; Inness et al., 2015). All stations use Bruker spectrometers, but the retrieval methodologies are not harmonized, e.g. the retrieval code, the *a priori* profiles and covariances, the treatment of instrument lineshape are generally different for different sites. A harmonized FTIR CO dataset is being developed in the framework of the EU project QA4ECV (Quality Assurance for Essential Climate Variables), but it is of more limited extent than the NDACC dataset: for 2013, data for only 6 stations were delivered to date. The number of degrees of freedom (DOF) is variable among the sites but lies in the range 1.9-3.0 on annual average. When provided, the estimated systematic uncertainties of the total columns in the NDACC dataset (excluding the smoothing errors) are typically 2-4% except at a few stations (ca. 0.6% on average in 2013 at Ny-Ålesund, ca. 12% at Eureka). The average systematic uncertainties for the mixing ratios are larger, most often between 10 and 20% but ranging between 3-4% (Maïdo, St. Denis, Mauna Loa, Thule) and 30% (Bremen).

The FTIR total columns are useful to evaluate the IASI CO data used in the source inversions. To this end, we compare the monthly averaged FTIR column with the monthly averaged IASI total column in the corresponding 2°×2.5° grid cell (Fig. 2.3). This procedure is meaningful for most sites because, thanks to the monthly averaging, strong horizontal gradients are generally not expected except possibly at urban sites. The IMAGES-calculated vertical profiles are used to account for differences in surface elevation and vertical sensitivity. In particular for high altitude stations, the fraction of the total IASI column lying above the station altitude is assumed to be reasonably well determined by the model. More precisely, the IASI total columns (and their estimated retrieval uncertainties) are corrected using

$$O'_{\mathrm{IASI}} = O_{\mathrm{IASI}} \frac{M_{\mathrm{FTIR}}}{M_{\mathrm{IASI}}}$$

with OIASI the IASI column, MIASI the modelled total column using the IASI averaging kernel, and MFTIR the modelled total column using the FTIR averaging kernel. OIASI, MIASI and MFTIR are monthly averages in the 2°×2.5° cell around the station, sampled temporally as the FTIR data.

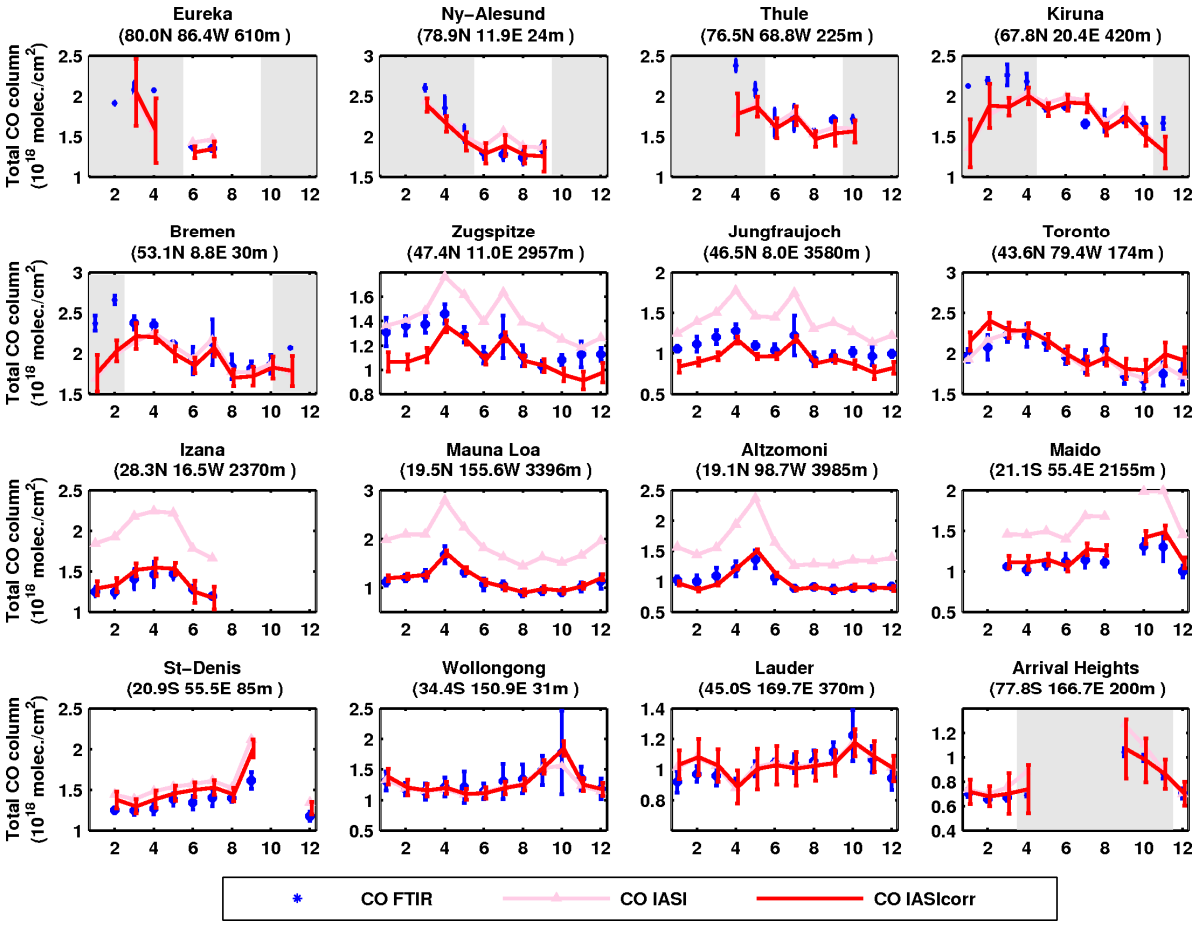


Figure 2.3. Comparison of monthly averaged FTIR CO (blue dots) with IASI (pink line) and IASI columns corrected for differences in surface elevation and vertical sensitivity (red line); see text for more details. The blue bars represent the standard deviation of FTIR columns. The red error bars are the IASI monthly averaged retrieval uncertainties. The shaded areas indicate the wintertime IASI data excluded in the source inversions.

IASI and FTIR total columns are in fairly good agreement at most locations, except at high Northern latitudes (>45°N) during winter, where IASI systematically underestimates the FTIR CO columns, by up to 30%. This issue is likely the result of the low temperatures and therefore low signal-to-noise ratio in the IASI retrievals. Biases in the FTIR columns cannot be excluded, but our comparisons of FTIR CO profiles with aircraft data from ARCTAS (Arctic Research of the Composition of the Troposphere from Aircraft and Satellites) and IAGOS (In-Service Aircraft for a Global Observing System) (see further below) do not suggest any overestimation of FTIR CO. The IASI retrieval uncertainties (red error bars on Fig. 2.3) are therefore very probably underestimated in winter at the Arctic stations, Bremen, and possibly the Alpine stations Jungfraujoch and Zugspitze. For the latter two stations, a part of

the difference could be due to uncertainties in the model profiles. As seen in the Figure, the exclusion of high-latitude wintertime IASI data appears justified in the Northern Hemisphere; although a good agreement between IASI and FTIR is found at the only high-latitude station in the Southern Hemisphere, the simple filter described above is used in both Hemispheres, for consistency.

OH fields used in CO inversions

Although the model calculates the concentrations of the hydroxyl radical (OH), we replace these values in the troposphere by OH distributions in line with observational constraints based on methylchloroform (MCF, CH_3CCl_3) observations (Prather et al., 2012; Patra et al., 2014). The OH fields are crucial to any CO modeling studies, because reaction with OH is by far the largest CO sink, and also because methane oxidation by OH is among the largest CO sources (Stavrakou and Müller, 2006; Kopacz et al., 2010). Model-calculated OH fields show very large differences among models, with e.g. the global tropospheric chemical lifetime of methane ranging between 7.1 and 13.9 years in a recent model intercomparison (Voulgarakis et al., 2013), whereas the best MCF-based estimate is 11.2 years with an uncertainty range of 9.8-12.5 years (Prather et al., 2012). Furthermore, models generally fail to reproduce the interhemispheric OH ratio derived from MCF analyses: while models predict higher OH in the Northern than in the Southern Hemisphere, constraints indicate an interhemispheric N/S ratio (R_{NS}) very close to 1 (Patra et al., 2014). Five source inversions are performed with the model, each with a different OH field. The OH concentrations of the standard run (STD) are consistent with the best estimates of Prather et al. (2012) and Patra et al. (2014), whereas the four other runs either maximize (VHNH & VHSH) or minimize (VLNH & VLSH) [OH] in either hemisphere while remaining consistent with the uncertainty estimates (Table 2.1).

Table 2.1. Description of OH fields used in the CO source inversions conducted with IMAGES. $\tau_{CH_4}^{OH}$ denotes the tropospheric chemical lifetime of methane, R_{NS} is the interhemispheric ratio.

Run label	Description	[OH] in NH (10^5 cm^{-3})	[OH] in SH (10^5 cm^{-3})	$\tau_{CH_4}^{OH}$ (years)	R_{NS}
STD	Standard OH	9.2	9.5	11.2	0.97
VHNH	Very high OH in NH	11.2	10.1	9.8	1.10
VLNH	Very low OH in NH	7.8	9.0	12.5	0.85
VHSH	Very high OH in SH	9.8	11.5	9.8	0.85
VLSH	Very low OH in SH	8.8	8.0	12.5	1.10

The [OH] fields rely on the climatological tropospheric OH distributions of Spivakovsky et al. (2000) adjusted in order to match the MCF constraints. This distribution achieves R_{NS} ≈1, with hemispherically-averaged OH concentrations $[OH]_{NH}$= 11.4×10^5 and $[OH]_{SH}$=11.5×10^5 cm^{-3}, and it implies a global tropospheric chemical lifetime of methane of 9.1 years. This field is scaled in each hemisphere to match the hemispheric mean OH concentrations (Table 2.1) calculated using

$$[OH]_{NH} = X_g \frac{2\,R_{NS}}{R_{NS}+1}\,,\ [OH]_{SH} = X_g \frac{2}{R_{NS}+1}$$

where X_g is the globally-averaged OH concentration consistent with Prather et al. (2012) (best value: 0.932×10^6 cm^{-3} , uncertainty range =(0.835-1.065) $\times10^6$ cm^{-3}), and R_{NS} the N/S concentration ratio from Patra et al. (2014) (best value: 0.97, range=0.85-1.1).

2.2.3 Analysis

Inversion of CO sources

The IASI CO columns are illustrated in Fig. 2.4 along with the IMAGES CO columns with either *a priori* or optimized emissions. In contrast with previous modelling studies (Stavrakou and Müller, 2006; Kopacz et al., 2010; Fortems-Cheiney et al., 2009), the modelled columns of the VLNH simulation overestimate the IASI columns over most of the Northern Hemisphere, due to the lower OH concentrations and higher CO lifetime compared to those studies. The source optimizations bring the model much closer to IASI data, e.g. the χ^2 diagnostic is reduced by almost 65% in both hemispheres in the VLNH optimization. Among the different optimizations, VLNH achieves the lowest χ^2 (0.70), followed by VLSH and the standard run STD (both 0.74), whereas the source optimizations with higher OH levels (VHNH and VHSH) performed comparatively worse (0.86 and 0.84). The VLNH inversion is also the simulation which achieves the lowest mean bias in both hemispheres (-0.1% in NH and 0.2% in SH), whereas e.g. VHNH and VHSH perform worse (-1.5% and -2.1% in their respective hemisphere, respectively).

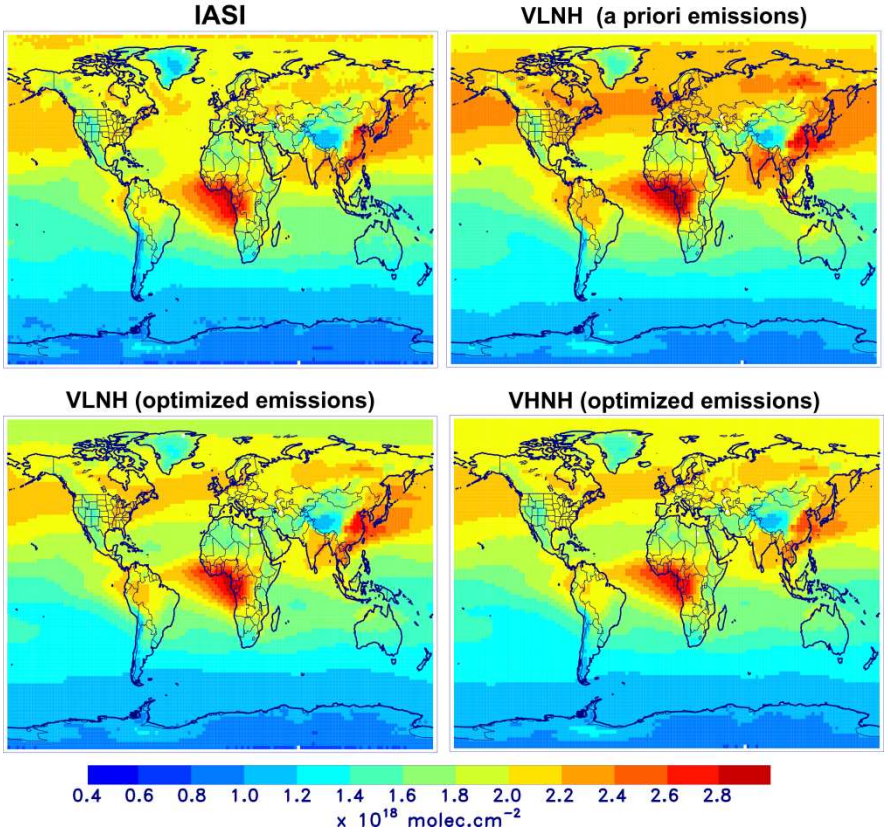


Figure 2.4. Annually averaged CO columns as observed by IASI (top left panel), and modelled with IMAGES using a priori emissions (top right) or optimized emissions (bottom panels). VLNH (low OH in Northern Hemisphere) and VHNH (high OH in NH) refer to the OH fields used in the model (see previous subsection).

Although the major features of the optimized model CO columns are similar in all five source inversions, the corresponding top-down emission estimates show large differences (Fig. 2.5 and Table 2.2), in particular for anthropogenic emissions (factor of 1.4 between global emissions in VHNH and VLNH) and for biogenic VOC emissions (factor of 1.34 between global isoprene emissions in VHSH and VLSH). When OH levels are higher (e.g. VHNH), the lifetime of CO gets shorter, which is compensated by higher emissions.

Table 2.2. Global emissions and CO and isoprene (Tg/year) in the *a priori* and in the different source inversions. The anthropogenic subtotal for the United States in also given.

	A priori	STD	VHNH	VLNH	VHSH	VLSH
Anthropogenic (TgCO/yr)	544	511	617	436	564	478
of which U.S.	*44*	*39*	*53*	*31*	*42*	*37*
Biomass burning (TgCO/yr)	291	320	352	294	354	293
Biogenic isoprene (TgC_5H_8/yr)	415	396	454	353	462	344

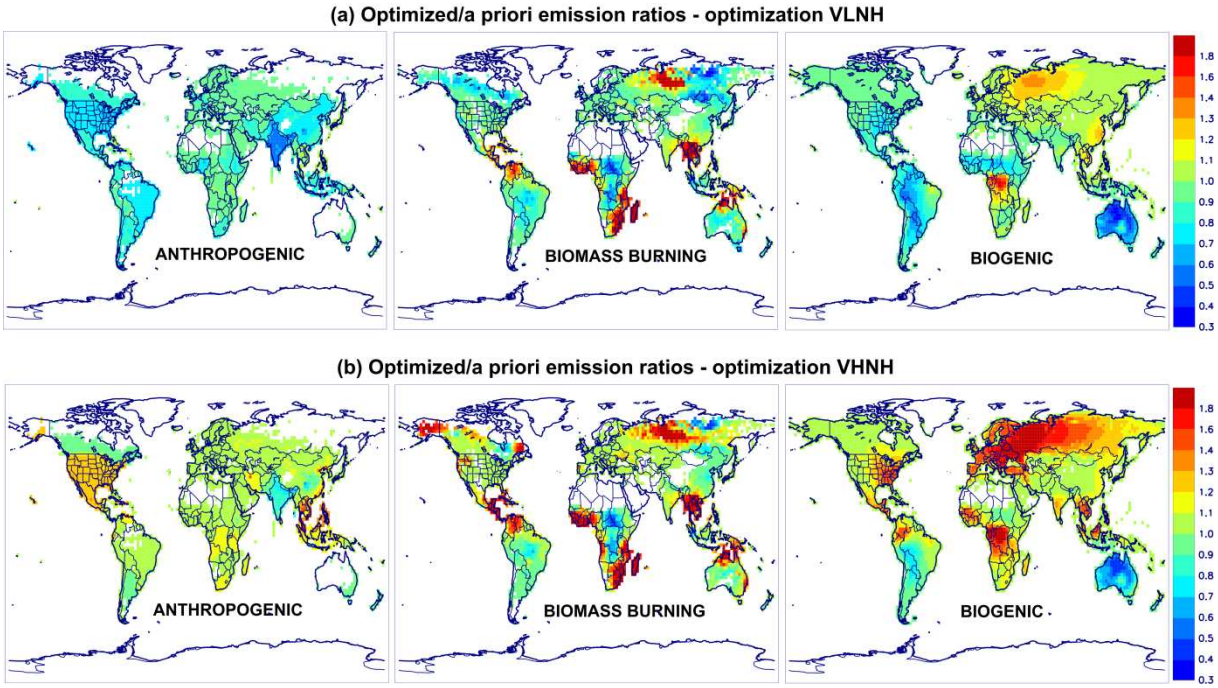


Figure 2.5. Emission updates (ratios (optimized fluxes)/(a priori fluxes)) for the three source categories (anthropogenic CO, biomass burning, biogenic VOC), annually averaged, in the optimization (a) VHNH (highest OH in NH) and (b) VLNH (lowest OH in NH).

It is difficult to validate the top-down emission estimates. However, low anthropogenic CO emissions over the U.S. are supported by independent evaluations using aircraft measurements. Hudman et al. (2008) used data from the ICARTT (International Consortium for Atmospheric Research on Transport and Transformation) campaign and inferred a 60% overestimation of U.S. EPA emissions in 2004 (see also Kopacz et al., 2010), corresponding to a total annual emission of 26 Tg/year in 2013 when adopting the emission trend published by the U.S. Environmental Protection Agency (EPA) (www.epa.gov/sites/production/files/2016-12/national_tier1_caps.xlsx). Using a geographically more limited dataset, Anderson et al. (2014) deduced a 15% overestimation

of U.S. EPA emissions, implying a total annual flux of 42 TgCO in 2013. The CO top-down estimate of the VLNH inversion (31 Tg/yr) lies between these two campaign-based estimates, whereas the U.S. total of the VHNH inversion (53 Tg/yr) seems largely overestimated.

Model evaluation against independent CO measurements

The source inversions are first evaluated against FTIR total columns in Figs. 2.6 and 2.7.

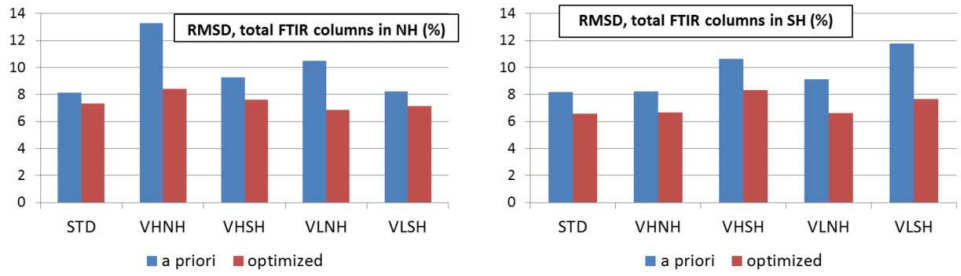


Figure 2.6. Root-mean-square relative deviations (in %) between modelled and observed monthly CO total columns in the Northern (left) and Southern Hemisphere (right).

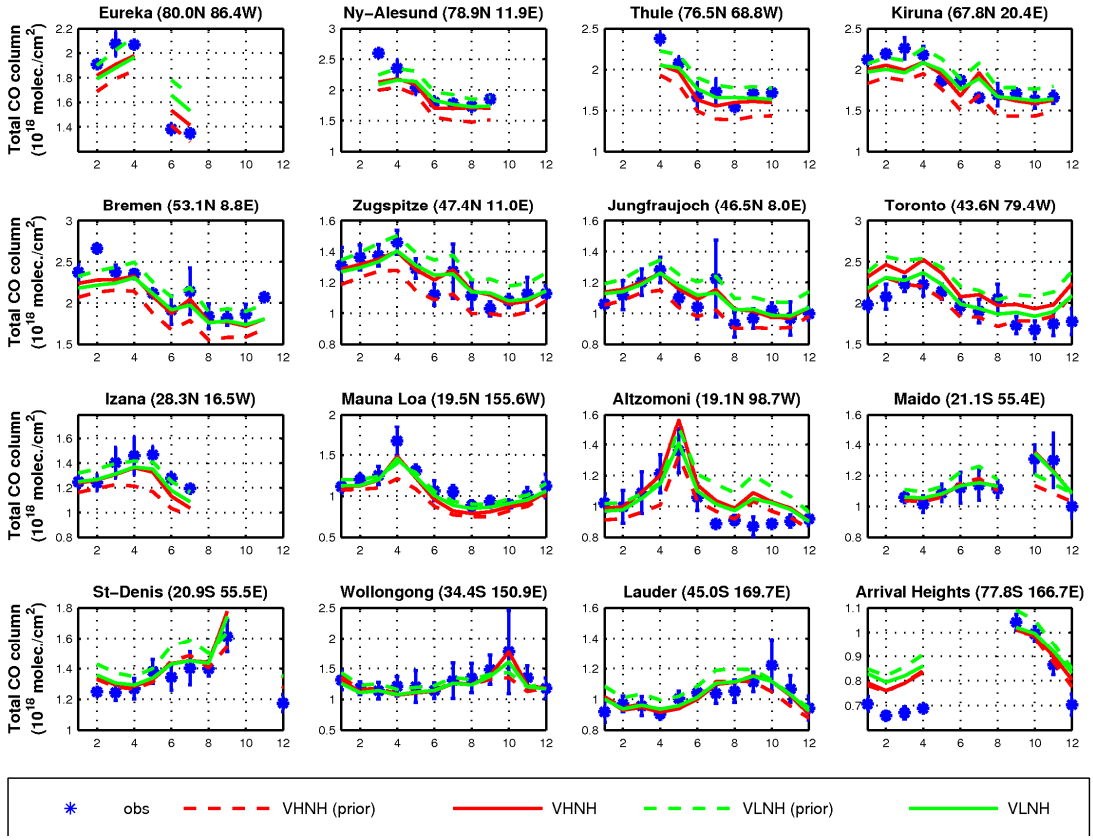


Figure 2.7. Modelled and observed CO columns at FTIR stations. In blue: FTIR monthly means with their standard deviations; dashed lines: modelled values using a priori fluxes; solid lines: modelled values using optimized fluxes. VHNH run in red, VLNH in green.

For simplicity, the discussion will focus on the optimizations VLNH and VHNH, i.e. the inversions with either the lowest or the highest [OH] in the Northern Hemisphere, which is

where most evaluation data is available. In all simulations, the root-mean-square deviation (RMSD) between modelled and observed monthly columns is decreased in both hemispheres upon IASI-based emission inversion. The results in the Northern Hemisphere indicate a clear improvement when [OH] is low (RMSD of 6.9% for VLNH, 8.4% for VHNH). A large part (about half) of the difference between the RMSD of VLNH and VHNH inversions is associated with the stations of Toronto and Mauna Loa. As seen in Fig. 2.7, the higher CO emissions of the VHNH inversion lead to a significant overestimation (+10%) at Toronto which is located in the direct vicinity of anthropogenic emission sources. At Mauna Loa, far away from the source areas and where CO is therefore more sensitive to OH levels, VHNH underestimates the FTIR columns in spite of its higher emissions. Interestingly, the RMSD between the model and FTIR columns is very similar to the RMSD between the model and IASI columns (7.3% for VLNH, 8.3% for VHNH in the Northern Hemisphere), suggesting an excellent general consistency between IASI and FTIR CO total columns, at least when high-latitude wintertime IASI data are filtered out. As for the Southern Hemisphere, both FTIR and IASI data indicate a better match (lower RMSD) when OH levels are close to the standard case (runs STD, VLNH, VHNH, with RMSD very close to 6.6% against FTIR in all 3 cases) than when OH is either very high (VHSH, 8.3%) or very low (VLSH, 7.7%).

Next, we evaluate the source optimizations against surface CO mixing ratios from the measurement networks of GMD (Global Monitoring Division of the National Oceanic and Atmospheric Administration, NOAA, ftp://aftp.cmdl.noaa.gov/data/trace_gases/co/flask) and GAW (Global Atmospheric Watch, ds.data.jma.go.jp/gmd/wdcgg/cgi-bin/wdcgg/catalogue.cgi), for a total of 123 stations providing data in 2013. The uncertainty on those measurements is typically 2-5 ppb (Novelli and Masarie, 2014). The comparison is summarized in Fig. 2.8. The seasonal variation of CO mixing ratios at selected GMD stations is illustrated in Fig. 2.9.

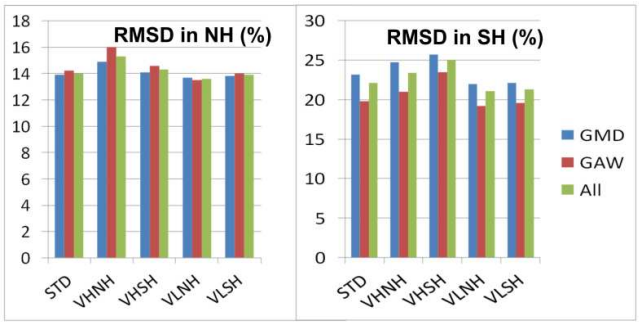


Figure 2.7. Root-mean-square relative deviations (in %) between modelled (optimized fields) and observed monthly CO mixing ratios at GMD and GAW stations in the Northern Hemisphere (left) and Southern Hemisphere (right).

Overall, the best (worst) performance in the Northern Hemisphere is again realized by the VLNH (VHNH) simulation. This VLNH simulation provides also the best match with Southern Hemisphere data. The RMSD is strongly decreased in the Southern Hemisphere upon source inversion, e.g. from 49% to 21% for VLNH. This is for a large part due to the reduction of biomass burning emissions over Indonesia in June, causing a strong reduction of the bias at Bukit Kotobatang in Sumatra (Fig. 2.9). The RMSD reduction upon optimization is

comparatively smaller in the Northern Hemisphere (from 14.5% to 13.6%). As in the case of the FTIR CO columns, the model generally underestimates the wintertime observations at high Northern latitudes (Fig. 2.9), possibly due to the exclusion of high latitude wintertime IASI data in the inversions. As for the FTIR total columns, the VHNH simulation overestimates CO over source areas (e.g. over Spain, Colorado, Utah) and it underestimates CO at remote oceanic sites (Midway, Mauna Loa, Guam), whereas the VLNH achieves a comparatively better match at these stations. The general conclusions are therefore consistent with those drawn from total CO column data. Nevertheless, the comparisons for surface mixing ratios show a larger spread (reflected by larger RMSD) than in the case of total columns. This is likely due to the larger model uncertainties for surface mixing ratios than for total columns, which are less affected by uncertainties associated with vertical mixing.

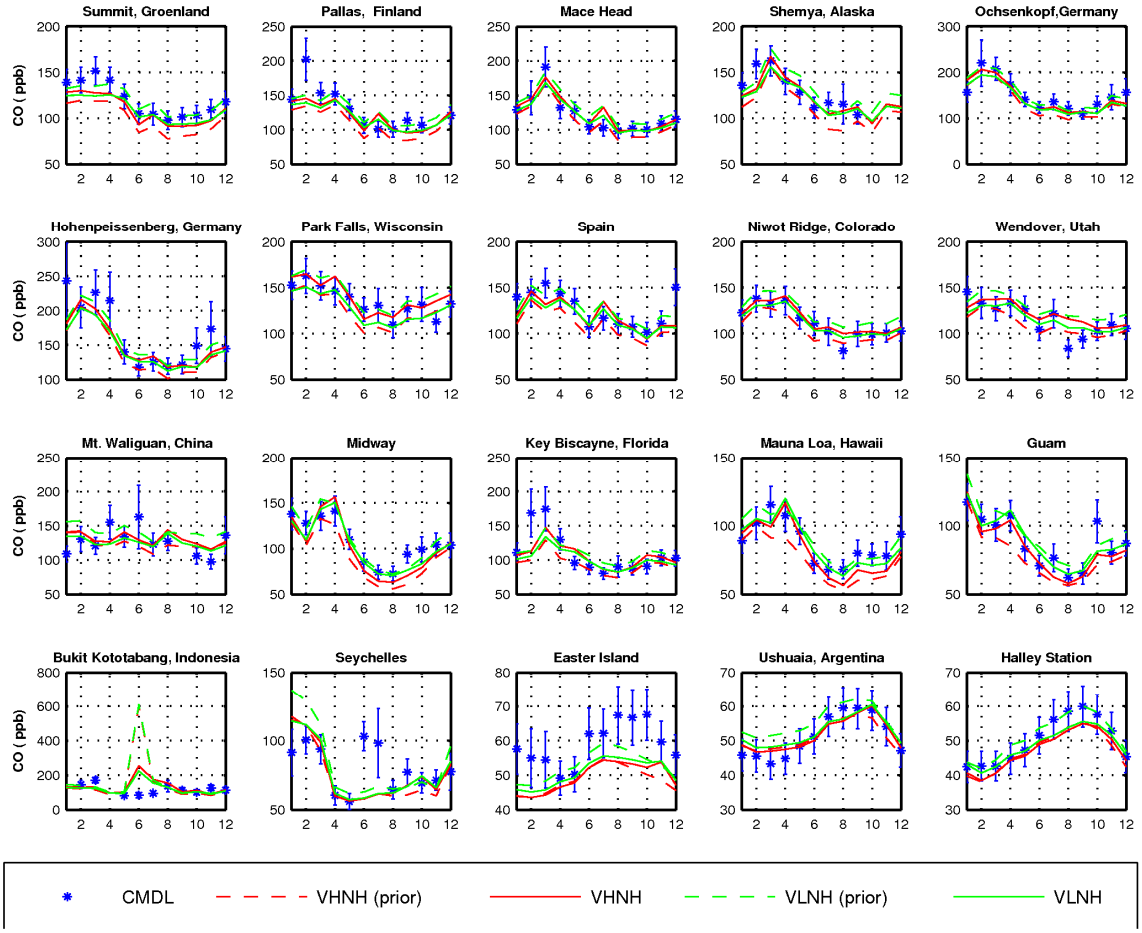


Figure 2.9. Modelled and observed CO mixing ratios at selected GMD stations. In blue: GMD monthly means with their standard deviations (denoted by the vertical bars); dashed lines: modelled values using *a priori* fluxes; solid lines: modelled values using optimized fluxes. VHNH run in red, VLNH in green.

The modelled profiles of CO mixing ratios over North America are evaluated against aircraft measurements from (a) the SEAC^{4}RS (Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys) field campaign conducted in August-September 2013 (Toon et al., 2016), (b) the DC3 (Deep Convective Clouds and Chemistry) campaign conducted in May-June 2012 (Barth et al., 2015), and (c) the the Global

Greenhouse Gas Reference Network (GGGRN) aircraft programme of the Earth System Research Laboratory (ESRL/NOAA) (https://www.esrl.noaa.gov/gmd/ccgg/aircraft , data courtesy of Colm Sweeney, ESRL) between 2012 and 2014. For both DC3 and SEAC4RS, the CO measurements used here were performed by tunable diode laser (TDL) spectroscopy onboard a DC-8 aircraft. Urban plumes ([NO_2]>4 ppb or [NOx]/[NOy]>0.4) are excluded from our analysis, as well as open fire plumes ([CH_3CN]>225 ppt) and stratospheric air ([O_3]>[CO]>1.25 mol mol^{-1}) (Hudman et al., 2007). GGGRN samples were collected by small aircraft (Cessna and similar) flown near their bases at rural locations throughout North America (see map on Fig. 2.10), and subsequently analyzed at ESRL/NOAA. The modelled mixing ratios shown on Fig. 2.10 were obtained from 2011-2014 simulations with IMAGES with emissions updated on the basis of the (optimized)/(a priori) emission ratios from the source inversions STD, VHNH and VLNH.

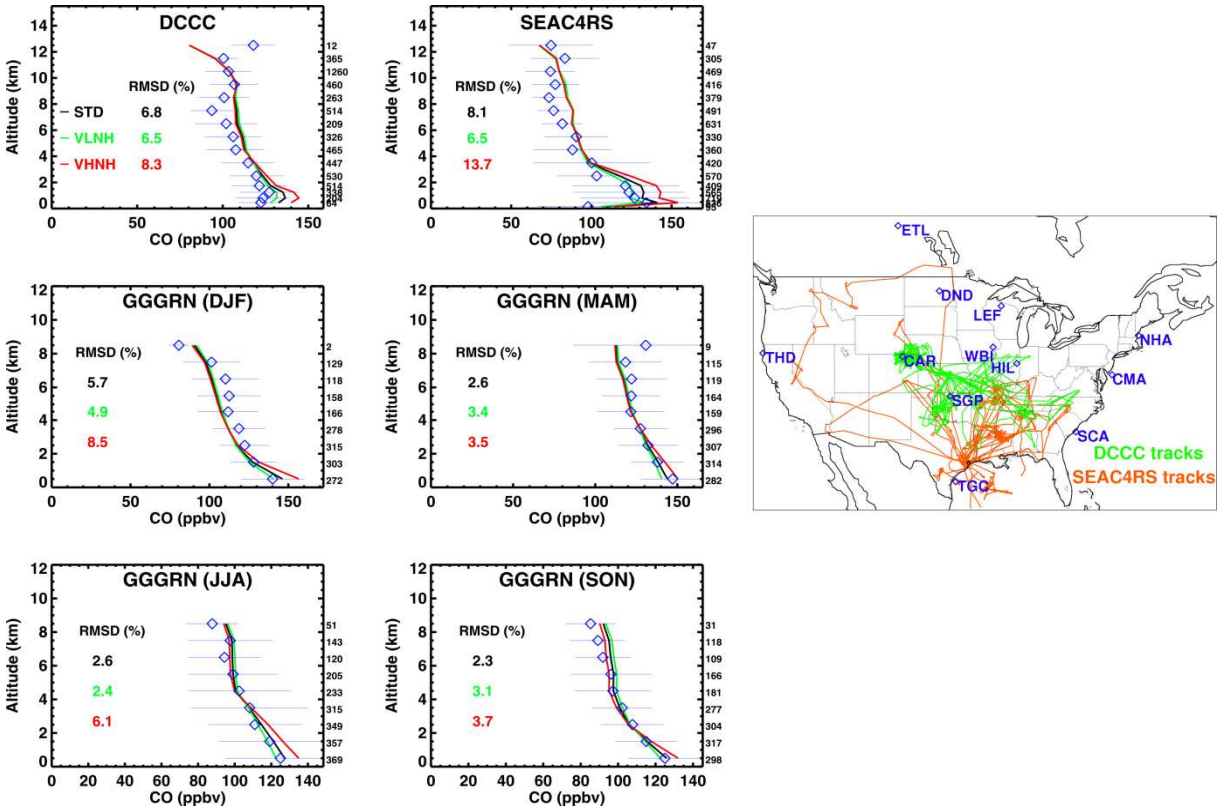


Figure 2.10. Averaged modelled CO mixing ratios using emissions based on source inversions STD (black), VLNH (turquoise) and VHNH (red) against aircraft data (blue diamonds with whiskers denoting standard deviation) from the DCCC campaign, the SEAC^{4}RS campaign, and the seasonally-averaged GGGRN data in 2012-2014, averaged over the US and Southern Canada (domain shown on the right panel). The number of measurements per altitude bin is indicated to the right of each plot. The root-mean-square relative deviations are also given.

The modelled CO mixing ratios from the three optimizations are very similar in the middle and upper troposphere, where IASI sensitivity is highest. Despite their very different emissions and OH fields, the IASI constraint brings all simulations to similar mixing ratios in this altitude range (4-12 km). The model tends to overestimate CO around 8 km altitude for both SEA^{4}RS and DCCC, but no clear pattern emerges from the comparison with GGRN profiles. Closer to the surface, however, significant differences are found between the different optimizations, reaching up to 20% between VHNH and VLNH. The higher emissions

of the VHNH simulation lead to near-surface CO overestimation in all campaigns and (for GGGRN) at all seasons, although the discrepancy is very small in MAM and SON. The VLNH achieves the best match (lowest RMSD) with observations of both DC3 and SEAC^{4}RS campaigns, as well as of GGGRN data in JJA and DJF, and it also performs better than the VHNH inversion against GGGRN data in MAM and SON.

To summarise, the source inversion using the lowest OH levels in the Northern Hemisphere achieves the best agreement against all tested CO datasets in the Northern Hemisphere, including IASI and FTIR CO columns, GMD and GAW surface concentrations, and aircraft vertical profiles from 3 datasets. The U.S. anthropogenic emissions derived in that inversion are consistent with previous top-down estimates based on aircraft campaigns. In addition, this source inversion realizes also the best agreement with measurements in the Southern Hemisphere, although other source inversions (e.g. STD) achieve very similar results in the Southern Hemisphere.

Finally, the source optimizations are evaluated against the FTIR vertical profiles. Although VLNH remains the source inversion achieving the best agreement (lowest RMSD) with FTIR mixing ratios (Fig. 2.11, upper panels), in accordance with the conclusions drawn from total column and other measurements, several features of this comparison cast doubt on its applicability for the purpose of this evaluation:

- the source inversions lead to only marginal decreases in RMSD (-6% on average in the Northern Hemisphere), in contrast with the much more substantial improvement found for total CO columns (-22%)
- in the Northern Hemisphere, the different optimizations lead to very similar *a posteriori* RMSD of mixing ratios (RMSD of VHNH only 7% higher compared to VLNH), contrasting with the larger differences seen for the total columns (23% difference between VHNH and VLNH)
- more alarmingly, the source inversion causes a systematic worsening of the model predictions for the vertical gradient of tropospheric CO within both hemispheres (Fig. 2.11, lower panels); in particular, VLNH realizes the worst agreement for the vertical profiles in the Northern Hemisphere.

The low sensitivity of the model/data discrepancy to the OH fields suggests that the FTIR profiles cannot help much in the evaluation of those inversions. More worrisome, the contradiction between the conclusions drawn from FTIR vertical profiles and those drawn from other measurements (total columns & in situ) suggest the existence of data issues in the FTIR profile retrievals, although model uncertainties (e.g. transport, intrahemispheric OH variability and seasonal variation) might also contribute.

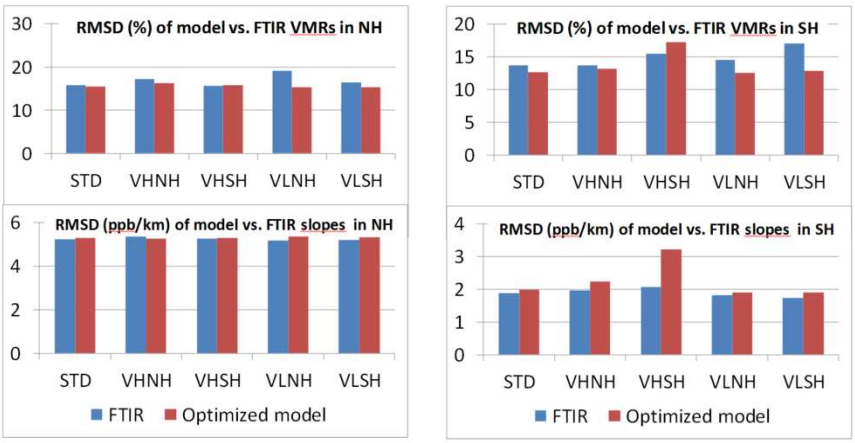


Figure 2.10. Top panels: root-mean-square deviations between monthly averaged FTIR mixing ratios and their modelled values (smoothed by FTIR averaging kernels and temporally sampled as the observations) using optimized emissions in the Northern (left) and in the Southern Hemisphere (right). Bottom panels: same for the average gradient (in ppb/km) in the troposphere, calculated as the slope of a linear regression of the mixing ratios below the tropopause.

A closer examination of the retrieved and modelled profiles (Fig. 2.12) is useful to characterize the instances of model/data mismatch requiring further investigation. Striking examples of discrepancies include e.g.

- the large model underestimation of near-surface CO and overestimation of upper tropospheric CO at the Arctic sites, especially Eureka and Ny-Ålesund,
- the strong discrepancy in the profile shape at Jungfraujoch: the retrieval indicates a strong negative gradient, not seen in the model result or in the FTIR profile at nearby Zugspitze,
- the overestimation of modelled upper tropospheric CO at Toronto,
- the large underestimation of lower tropospheric CO at Bremen, and
- the wrong sign of the CO gradient at Izaña

Note that a previous model comparison with FTIR profiles at 4 stations showed very similar discrepancies (Inness et al., 2015). In order to pinpoint the possible causes of these issues, the following subsections present an evaluation of FTIR profiles using independent measurements (aircraft, surface mixing ratios).

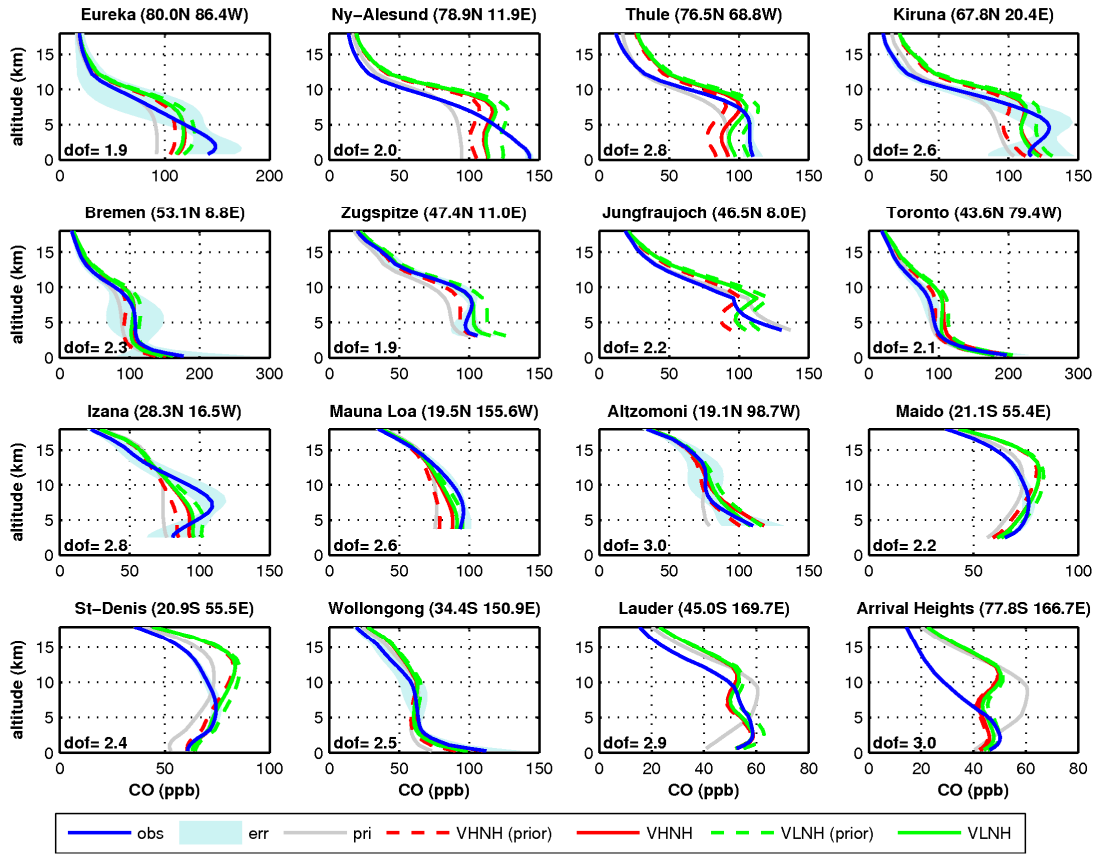


Figure 2.12. Annually-averaged vertical profiles of FTIR CO mixing ratios (blue curve) and corresponding modelled values using either a priori (dashed lines) or optimized emissions (solid lines). VHNH results in red, VLNH in green. Grey lines: FTIR a priori profiles. Shaded blue areas: FTIR confidence interval defined by the retrieval systematic error. Retrieval error information is not available at Jungfraujoch, Lauder, and Arrival Heights. The average number of degrees of freedom of the FTIR retrieval (dof) is indicated in each plot.

Evaluation of FTIR CO profiles against aircraft measurements

Direct evaluation of FTIR CO profiles is difficult because independent measurements are generally not well co-located with FTIR data. However, due to the long lifetime of CO (≈2 months), its distribution generally shows only weak horizontal gradients when averaged over sufficiently long periods. Fig. 2.13 presents the data used here for evaluating FTIR profiles. In all cases, the aircraft profiles are temporally averaged, re-gridded onto the FTIR vertical grid and smoothed using (monthly-averaged) FTIR averaging kernels. In the case of seasonal averages over several years, the FTIR average profile accounts for the temporal sampling of aircraft data. The detailed comparisons are presented in the Appendix, the main conclusions are provided below.

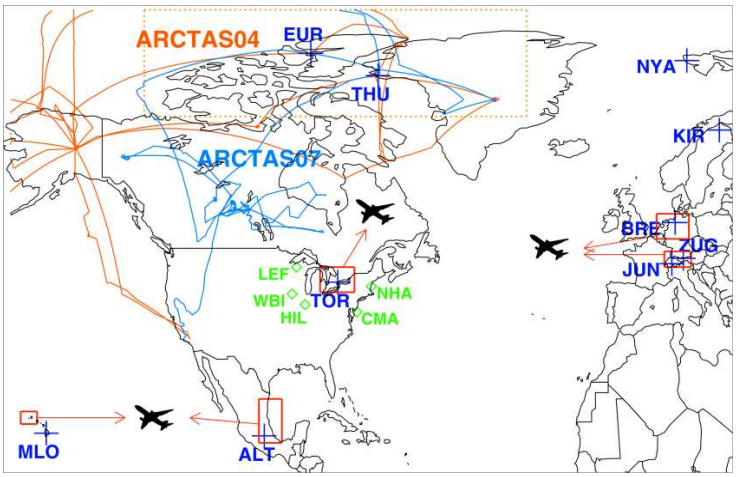


Figure 2.13. The CO profiles retrieved at FTIR stations (blue crosses) are evaluated against ARCTAS (tracks in orange and blue), IAGOS (red rectangles), and GGGRN (green diamonds) data. The dotted rectangle indicates the averaging areas for ARCTAS. EUR, Eureka; NYA, Ny-Ålesund; THU, Thule, KIR, Kiruna; BRE, Bremen; ZUG, Zugspitze; JUN, Jungfraujoch; TOR, Toronto; ALT, Altzomoni; MLO, Mauna Loa.

Although the Arctic sites of Eureka and Thule are relatively close to each other (ca. 550 km) and very distant from major emission sources, the Eureka FTIR profile show systematically lower mixing ratios in the upper troposphere compared to Thule (Fig. A1 and Fig. 2.11). This difference cannot be rationalized in terms of sources, sinks or long-range transport. The comparison with ARCTAS shows a 30% underestimation of upper tropospheric CO in April and July.

The average FTIR profiles at the two Alpine sites of Zugspitze and Jungfraujoch also show unwarranted differences given their proximity and the large number of data involved in the averages. The Zugspitze presents only a weak negative gradient in the troposphere, whereas the Jungfraujoch retrievals show a strong negative gradient, especially in winter and spring. The biases between the Zugspitze profiles and the IAGOS profiles are small and appear to be statistically non-significant. At Jungfraujoch, , the strong negative gradient of the FTIR retrieval is contradicted by the IAGOS measurements during all seasons but especially in spring, with a low-level overestimation exceeding largely the standard deviation of the IAGOS data.

The FTIR profiles at Bremen are in fairly good agreement with IAGOS data, except for a clear underestimation in the free troposphere, especially in winter (12% bias). Both FTIR data and IAGOS boundary layer data overestimate by up to 50 ppb the surface CO mixing ratio at the a nearby GAW station and therefore cannot be considered representative of a larger area around the site.
The comparisons for Toronto lead to conclusions similar to those for Bremen, with FTIR underestimations at all seasons in the free troposphere, and negligible biases in the boundary layer. In addition, comparisons with aircraft data of the GGRN programme of ESRL (green symbols in Fig. 2.13) also show FTIR underestimation in winter, as well as overestimation in the boundary layer.

At Mauna Loa, a good agreement with IAGOS data is found at all seasons, considering the statistical variability. At Altzomoni, the FTIR profile in April/May overestimates aircraft data at 7-9 km altitude, but it agrees well with both IAGOS and GMD data at the altitude of the station.

Evaluation against surface (GMD and GAW) measurements

Finally, we use the model results (simulation VLNH, which provides the best match with both column and in situ CO measurements) to evaluate near-surface FTIR mixing ratios against GMD and GAW measurements at nearby sites, for the year 2013. Surface concentration measurements should not be directly compared to near-surface FTIR mixing ratios, since the dof of the retrieval most often lies between 2 and 3, but we expect the model discrepancies with respect to near-surface FTIR data to be roughly similar to model discrepancies for surface concentration measurements at nearby stations, except in the case of important horizontal gradients or surface elevation differences. We have already mentioned the discrepancies (overestimations) of near-surface FTIR concentrations at Bremen and Toronto compared to their rural surroundings. However, most other FTIR sites are located away from urban centers and other localized sources. The locations of FTIR and in situ measurement stations used for comparison are shown on Fig. 2.14. The network stations are all relatively close to FTIR stations, except for Alert, located at about 480 km from Eureka, in a region with little horizontal gradients given the absence of local sources. Furthermore, the FTIR stations and their corresponding GMD/GAW stations lie at similar altitudes. Note that both Altzomoni and the GMD station of Mexico are located away from Mexico City.

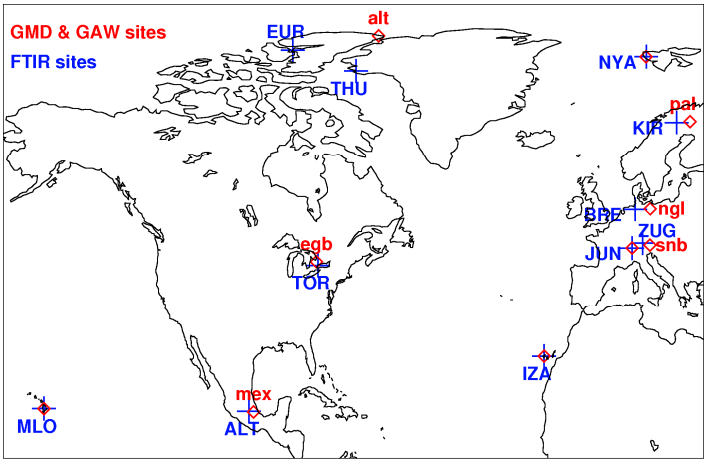


Figure 2.14. Location of FTIR, GMD and GAW stations used for comparisons. FTIR sites: EUR (Eureka), Thule (THU), NYA (Ny-Ålesund), KIR (Kiruna), BRE (Bremen), ZUG (Zugspitze), JUN (Jungfraujoch), TOR (Toronto), IZA (Izaña), MLO (Mauna Loa), ALT (Altzomoni). Other sites: alt (Alert), pal (Pallas), ngl (Neuglobsow), snb (Sonnblick), egb (Egbert), mex (Mexico).

As seen in Fig. 2.15, the model performs almost systematically worse compared to FTIR near-surface mixing ratios than compared to surface mixing ratios at nearby sites. This is especially true:

- at the Arctic sites of Eureka and Ny-Ålesund, where both the RMSD and the bias are much higher for FTIR mixing ratios than for GMD/GAW data,

- at Jungfraujoch, where the FTIR/model RMSD of 25% is ~2.5 times larger than the RMSD with respect to GAW measurements at the same site,
- at Izaña, where the FTIR/model RMSD of 19% is about 3 times higher than the RMSD with respect to GMD data at the site

Besides the special cases of the Bremen and Toronto FTIR stations, for which overestimations are expected, the comparison suggests positive biases of near-surface FTIR mixing ratios (as compared to surface measurements) at Eureka (ca. 22 ppb on annual average), Ny-Ålesund (16 ppb), Thule (11 ppb), and Jungfraujoch (18 ppb) and negative biases at Kiruna (-10 ppb), Zugspitze (-27 ppb), Izaña (-20 ppb) and Altzomoni (-17 ppb). These biases are often the same order of the estimated FTIR systematic uncertainty (24 ppbv at Eureka, 18 ppbv at Izaña, 22 ppbv at Altzomoni), except at Ny-Alesund (3 ppbv), Thule (6 ppbv) and Zugspitze (13 ppbv), where they are significantly larger, and at Kiruna (31 ppbv) where they are much smaller. Uncertainty estimates are unavailable for Jungfraujoch.

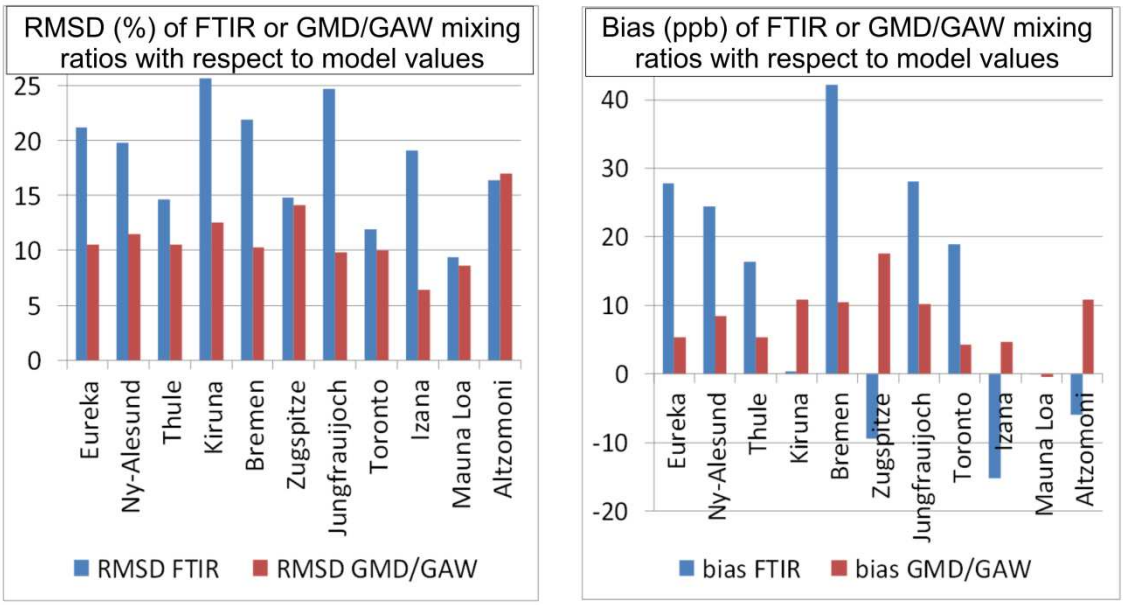


Figure 2.15. (a) Root-mean-square relative deviations (%) between either the FTIR mixing ratio at lowermost level or the surface mixing ratio at nearest GMD or GAW station with respect to modelled values of the VLNH optimization. (b) Same as (a) for the absolute bias. The nearest GMD or GAW station is as shown on Fig. 2.14. The FTIR averaging kernels are accounted for in the model comparisons with FTIR mixing ratios.

Combining these findings with the results of the comparisons of FTIR profiles with aircraft data leads to a reinterpretation of the model comparison with FTIR profiles (Fig. 2.12):

- at Arctic sites (Eureka, Ny-Ålesund, Thule), a large part (but not all) of the model underestimation of lower tropospheric CO is due to an overestimation of FTIR CO (GMD/GAW comparison, Fig. 2.15),
- at Eureka, the steep tropospheric CO gradient of the FTIR profile is unrealistic, and upper tropospheric FTIR CO is underestimated (ARCTAS comparison, Fig. A1),

- at Bremen and Toronto, the FTIR profiles are contaminated by low-level urban pollution, explaining a large part of the model underestimation near the surface (Fig. A4),
- at Toronto, the model overestimation in the free troposphere is partly due to an FTIR underestimation in winter (IAGOS and ESRL comparisons, Figs. A5 and A6), but also to a high bias of the model with respect to aircraft data in spring and summer (Fig. 2.10 (and additional comparisons, not shown)),
- at Zugspitze, the model underestimation in JJA and SON is likely due to FTIR underestimation of CO in the lower layers (IAGOS comparison, Fig. A2),
- at Jungfraujoch, the shape discrepancy is due to an overestimation of the tropospheric gradient in the FTIR profiles for all seasons (IAGOS comparison, Fig. A3),
- at Izaña, the overestimation of lower tropospheric CO by the model as well as the shape discrepancy (wrong gradient sign) with respect to FTIR is likely due to an underestimation of low-level CO in the FTIR profile (GMD/GAW comparison, Fig. 2.15),
- at Mauna Loa, the FTIR profiles appear consistent with independent measurements (IAGOS, Fig. A7 and GMD, Fig. 2.15),
- at Altzomoni, the mid-tropospheric model overestimation with respect to FTIR could be partly related to FTIR underestimations (IAGOS comparison, Fig. 2.15), but more data are needed to confirm this.

2.2.4 Results and Recommendations

The source inversion adopting the lowest average OH abundance in the Northern Hemisphere provides the best match with most CO observations. We acknowledge the existence of uncertainties due to model uncertainties associated with e.g. model transport and the seasonal evolution and intrahemispheric distribution of OH concentrations. Nevertheless, we note that the top-down CO emission estimates of the best (VLNH) inversion are corroborated by independent analyses based on CO aircraft data over the Eastern United States.

However, the model comparisons with FTIR vertical profiles at 16 NDACC stations do not fit well with the above conclusions. Not only is the model agreement with FTIR profiles only barely improved as a result of the source inversions, but it shows also very little sensitivity to the different OH setups of the inversions – in other words, *the FTIR profiles do not help much in discriminating between different CO budgets*, suggesting that model/data discrepancies for the most part cannot be traced to uncertainties in CO sources and sinks. Evaluation of FTIR profiles using independent measurements shows that the model/data discrepancies can be attributed for a substantial part to potentially erroneous FTIR profiles. For example, two pairs of FTIR stations (Eureka vs. Thule, Jungfraujoch vs. Zugspitze) display very different profiles when they are expected to be very similar by virtue of their geographical proximity. Model uncertainties surely also contribute to model mismatches with FTIR profiles, but the fact that very similar discrepancies were found in the evaluation of an assimilation of MOPITT and other data by the C-IFS system of ECMWF against FTIR profiles at 4 NDACC

stations (Inness et al., 2015) suggests that those mismatches are to a large extent model-independent.

In conclusion, extreme caution is obviously needed when using FTIR CO profiles for validating models – and by extension, for validating vertical information in satellite CO retrievals. In that regard, although the FTIR NDACC network has a great potential, our assessment is that priority should be given to strengthening the quality and robustness of the FTIR profile retrievals. Additional measurements (either more frequent measurements at existing sites or FTIR retrievals at additional sites) will most likely not contribute to narrow the vertical CO information gap. More analyses using models and in situ measurements would be welcome in order to better validate FTIR profiles. As noted in the GAIA-CLIM Gaps Analysis and Impacts Document, profiles with AIRCORE co-located with FTIR measurements may be a means to improve the FTIR profiles.

2.3 Aerosols (FMI)

2.3.1 Scope of the assessment

Atmospheric aerosol particles remain the largest source of uncertainty in the estimates of current drivers of climate change (IPCC, 2013). A better constraint of global aerosol fields both in polluted and pristine environments is therefore an urgent research topic. Given the short lifetime of atmospheric aerosols (from hours to ~1 week in the troposphere) and the relatively sparse temporal resolution of satellite-based observations (pass over a specific site every 0.5-16 days depending on the instrument and latitude), it is likely that satellites miss out on some aspects of the atmospheric aerosol variability. Further limitations of most satellite instruments are that they can observe aerosols only in cloud-free conditions and there are no observations during nighttime. On the other hand, the surface-based Aerosol Robotic Network (AERONET) (Holben et al., 1998) provides long-term, continuous measurements of AOD and other aerosol optical and radiative properties from dozens of (mostly continental) measurements sites. These continuous measurements enable a good characterization of local aerosol variability; however reasonable geographical coverage is obtained only over parts of Europe and US, and elsewhere long-term AERONET measurements are very scarce (IPCC, 2013). Given these respective limitations of both satellite and surface-based observations, it would be important to identify those geographical regions that currently suffer from poor observational constraints of aerosol properties in order to guide the development of future observational capabilities.

To our knowledge, the use of correlation lengths has not been applied before to study AOD variability in a global scale. It needs to be noted that the interpretation of the results becomes more complex than typical and more straight-forward model-to-measurement comparisons. However, our motivation here is to apply this novel approach to be possibly able to reveal such coherent spatial structures that would not be possible otherwise. This can then give insights for instance regarding the representativeness of measurement stations and thus also about possible geographical gaps in the current ground-based measurement capabilities and coverage.

Within GAIA-CLIM Task 1.5 we have investigated observational gaps from two perspectives: First, we studied whether a global aerosol model can provide information on the diurnal variability of AOD (information that AERONET measurement can provide but satellite measurements cannot) and thus indicate in which regions without AERONET coverage such cycles are important. Second, the spatial correlation structures of AOD were analyzed to study the representativeness of aerosol observations made at a single location with respect to surrounding areas.

2.3.2 Methods

The global aerosol-climate model used in both strands of the investigation was ECHAM-HAM-SALSA [Stier et al., 2005; Bergman et al., 2012; Kokkola et al., manuscript in preparation]. The large scale meteorology (vorticity, divergence, and surface pressure) was nudged towards the ECMWF reanalysis data from ERA Interim. For sea surface temperatures

and sea ice distributions we used the climatologies from the Atmospheric Model Intercomparison Project AMIP2 (Taylor et al., 2012). The emission fluxes for the mass of each species from anthropogenic sources are based on Aerocom II - ACCMIP emissions (Riahi et al., 2007; 2011). For the emission fluxes of mass of individual species from biomass burning, the first strand of the investigations (i.e. diurnal variability study) used ACCMIP data (Riahi et al., 2007; 2011), whereas the second strand (i.e. spatial correlation structure study) utilized the GFASv1 database, multiplied by a factor of 3.4 following the recommendation by Kaiser et al. (2012). Online emissions for mineral dust were calculated online according to Cheng et al. (2008). The online sea salt emissions followed the scheme detailed in Bergman et al. (2012) in the first strand of the work, and Long et al. (2011) in the second strand. The horizontal resolution of the model was T63 (roughly 1.9° × 1.9°).

The first strand of our investigation, looking at the diurnal variability of AOD, had a simulation period of 2005-2010 with a three-month spin-up. The model was run with 31 vertical pressure levels (top level 0.01 hPa) and produced three-hourly averaged output. The model was sampled at the temporal frequency and spatial location of the selected AERONET site observations used in the comparison. For these runs, interactive biogenic emissions were included via Model of Emissions of Gases and Aerosols from Nature (MEGAN), which determines emissions of terpenes and isoprenes as a function of temperature, available solar radiation, soil moisture, and carbon dioxide concentrations. To account for atmospheric SOA formation, a scheme based on the volatility basis set (VBS) (Donahue et al., 2011) was used.

In the second strand of the investigation, the representativeness of observations made at a single location was investigated by analyzing spatial correlation structures of AOD. For this work, we utilized AOD fields from the ECHAM-HAM-SALSA model simulations and MERRA-2 re-analysis and MODIS satellite observations. For both models, hourly output data corresponding to year 2005 were used and the data were co-located with MODIS observations to enable the comparison of the results. The ECHAM-HAM-SALSA model run was preceded by a one year spin-up period. The model was run with 47 vertical levels (top level 0.01 hPa).

The MERRA-2 reanalysis data (Gelaro et al., 2017) is based on GEOS atmospheric model (Rienecker et al. 2008; Molod et al. 2015) and it is produced by the Global Modeling and Assimilation Office of NASA. In MERRA-2, various satellite, airplane, and ground-based observations are assimilated into the system. MERRA-2 has an aerosol analysis (Randles et al. 2016, 2017) and the aerosol related inputs assimilated into the MERRA-2 aerosol analysis include AVHRR and MODIS (both Terra and Aqua) reflectances, MISR AOD retrievals, and AERONET AOD observations. MERRA-2 system uses GOCART model (Chin et al. 2002, Colarco et al. 2010) coupled with GEOS to simulate the evolution of five different aerosol species. Here we used the total AOD at 550 nm from MERRA-2.

The combined MODIS AOD data product based on both the Dark Target (Levy et al. 2013) and Deep Blue (Hsu et al. 2013) algorithms was used from both the Terra and Aqua satellites (AOD_550_Dark_Target_Deep_Blue_Combined_Mean). Both of the algorithms perform AOD

retrieval over land areas and the Dark Target over ocean retrievals are used for AOD retrieval over water. Level 2 satellite data was aggregated to an hourly output dataset with resolution of 1 degree by 1 degree. This aggregated dataset was used in the analysis.

2.3.3 Analysis

In the first strand of the work, the model performance was assessed by comparing the measured and modelled AOD diurnal variation at 108 AERONET locations around the world. The stations were chosen based on a requirement of a minimum of 100 days of data available, each day having at least one observation both before and after the local solar noon. The AERONET level 2.0 direct sun data of AOD at 500 nm (from 5:00 to 17:00, note that some of the satellite and model AOD values are given at 550 nm) were used as input to form the diurnal patterns during the summer season covering several years, depending on the location. All the individual summer AOD observations were used to calculate the departures from the daily mean. Each hourly departure was then averaged for each hour of the local time to form the average daily cycle of the departures at each site. The relative departure was then calculated by dividing each hourly-mean absolute departure by the overall summer AOD mean. In the process of model validation, the 3-hourly average was used to facilitate the comparison to the model output.

In the second strand of the study, the main quantity of interest was the correlation length of AOD. We defined the correlation length as the radius of an area within which the AOD is highly correlated. To find the highly correlated areas, we performed variogram analysis (Matheron 1963, Bohling 2005) of AOD using data both from the models and satellite data. In variogram analysis, a covariance function was fitted to experimental data, and we used the fitted covariance function parameters to compute the correlation length. In our analysis, we re-projected all data into 1 degree by 1 degree grid, used randomly selected data pixel pairs (250 pixel pairs per time step for each pixel), binned the data points into 35 bins based on distance between the pixels, and computed the so-called empirical semivariogram for each pixel. Then we fit a semivariogram model into this empirical semivariogram data. In this study, we used semiovariogram γ model of the form

$$\gamma(d) = n^2 + s^2\ (1 - \exp(-3(d/r)^p))$$

where d denotes the distance between the pixels, n^2 denotes the so-called nugget and represents the local component of the AOD variance (no spatial correlation). The nugget term is sometimes interpreted as the measurement or retrieval error related term. The sill s^2 describes the variance related to the spatially correlated component of AOD and p is an exponent term. In this model, r denotes a range term. This model corresponds to covariance function C of the form:

$$C(d) = \begin{cases} n^2 + s^2, & d = 0 \\ s^2 \exp\left(-3\left(\frac{d}{r}\right)^p\right), & d > 0 \end{cases}$$

With this covariance function, the covariance (or correlation) decreases the further away we go from the pixel of interest. Based on this decreasing covariance, we defined the correlation length as the length at which there was 90 % of the sill covariance left (or the covariance has decreased by 10 %). The correlation length R_{corr} was computed as:

$$R_{corr} = \left(-\frac{1}{3} r^p \log(0.9)\right)^{\frac{1}{p}}$$

We compared the correlation structures for the different datasets to: a) evaluate the correlation structure around AERONET sites; and b) evaluate the model simulations against satellite data. For statistical purposes, DOF in the different datasets were also evaluated (Henriksson et al., 2017).

2.3.4 Results and Recommendations

The results of the first strand of the investigation have been presented in earlier project reports in more detail, and are only briefly summarized here. Out of the 108 AERONET sites studied, the ECHAM-HAM-SALSA model is able to reproduce diurnal variability consistent in magnitude and pattern with the observations only at 3 sites. For the rest of the stations, the simulation shows considerable deficiencies. The daily trend analysis is illustrated in Figure 1, which shows the comparison between the observations and three model runs using different diurnal cycles for anthropogenic aerosol emissions. The model reproduces the AERONET AOD relative departure within a factor of 2 in fewer than 20% of the cases. It also indicates that all three simulations with different daily cycles of anthropogenic emissions are in general unable to reproduce the daily maxima/minima.

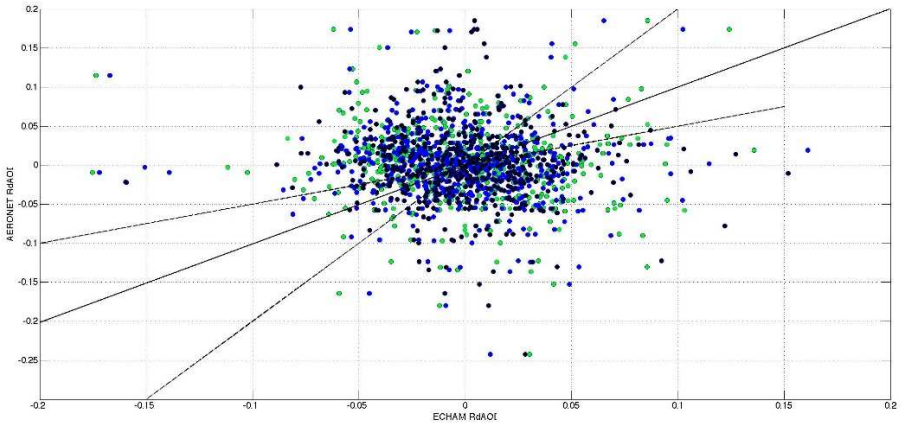


Figure 3.1. Scatter plot of the observed (AERONET) and simulated AOD relative departures. The different colored dots represent the three model simulations with different diurnal cycle of anthropogenic aerosol emissions. 1:1 and 2:1 lines are shown for clarity.

For the second strand of the study, the global correlation length maps are shown in Figure 3.2. The correlation length maps for both model based datasets (ECHAM and MERRA-2, top row) are similar. Based on the model-based data, the longest correlation lengths (over 1500 km) were observed over Northwest Sahara and over the Atlantic Ocean near the coast of North Africa. In ECHAM data, long correlation lengths were also observed over latitude 75 degrees north. In MERRA-2, there were long correlation length areas in the middle of Pacific Ocean and also over Indonesia and Philippines. In MODIS satellite data, the longest correlation lengths were observed over oceans, Southeast Asia, in East Africa over Somalia, Kenya, and Ethiopia, in the South part of Red Sea, over Spain and Portugal, and over the

coast of Venezuela. In MODIS satellite data, the typical AOD correlation length outside the long correlation areas was below 500 km.

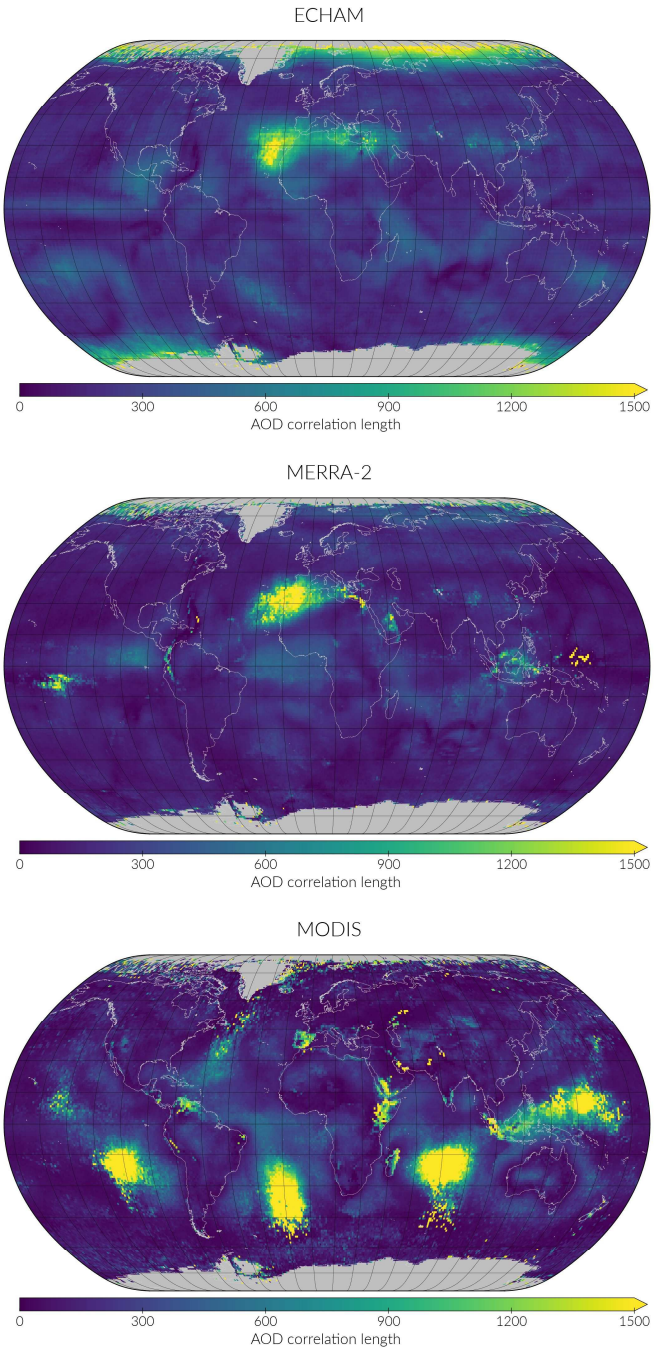


Figure 3.2. AOD correlation lengths corresponding to ECHAM-HAMMOZ, MERRA-2, and MODIS data. Blues denote shorter distances and greens and yellows longer distances.

In Figure 3.3, the AOD correlation lengths over Europe are shown. Based on model data, the typical correlation length in Europe was 200-300 km with the highest values of about 1000 km over southern Iberia. The MODIS satellite data results in slightly shorter, mostly below 150 km, correlation lengths. Similarly to model data, the longest correlation lengths over Europe were observed over the Iberian Peninsula.

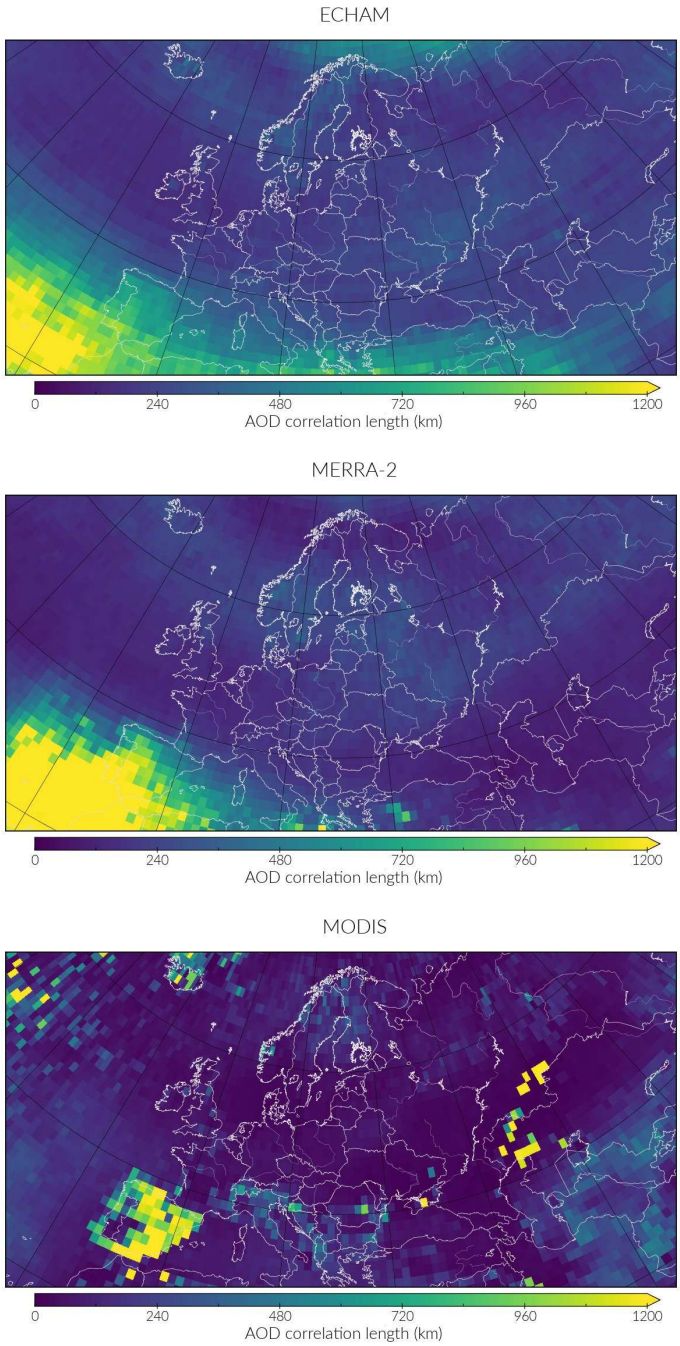


Figure 3.3. AOD correlation lengths over Europe corresponding to ECHAM-HAMMOZ, MERRA-2, and MODIS data.

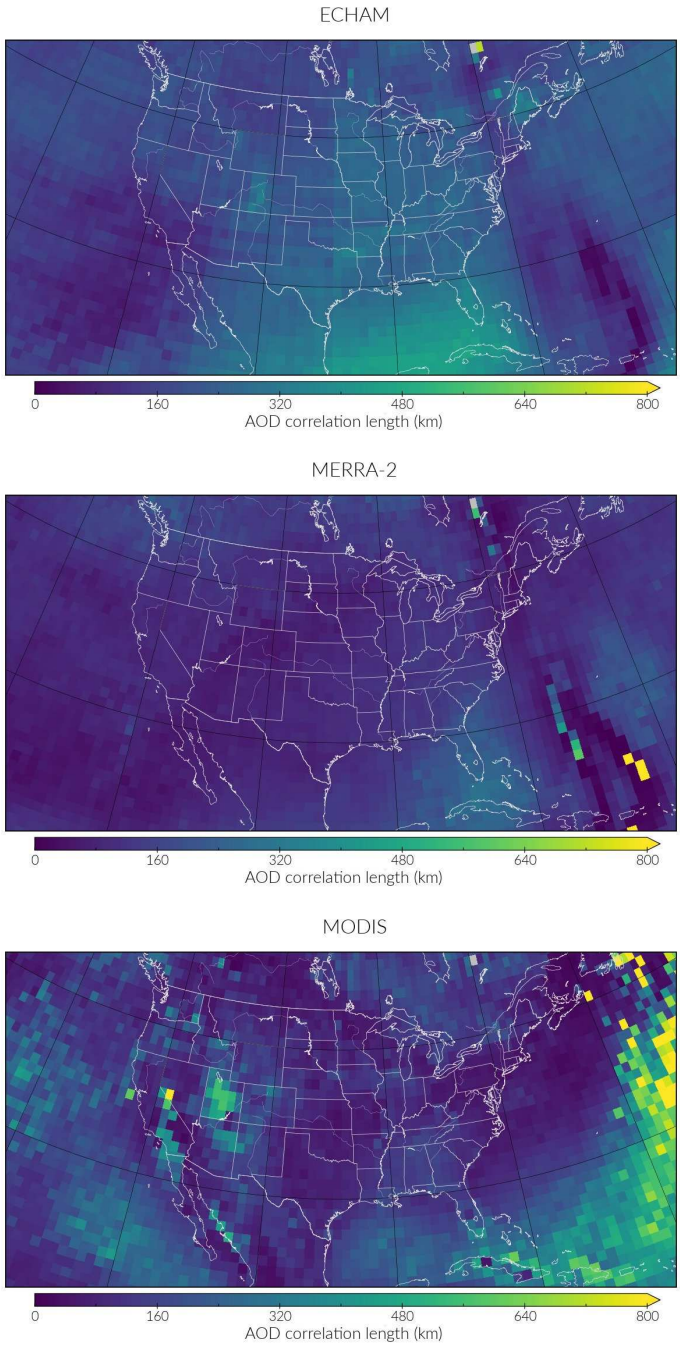


Figure 3.4. AOD correlation lengths over USA corresponding to ECHAM-HAMMOZ, MERRA-2, and MODIS data.

In Figures 3.4 and 3.5, the AOD correlation lengths over USA and Asia are shown, respectively. ECHAM data results in slightly longer average correlation length over the US than MERRA-2 or MODIS data. MODIS data shows longer correlation lengths over Utah and in a single pixel in western Nevada. The correlation lengths over Asia were similar to other areas. In Asia, the longest correlation lengths were observed over the Indonesia and Philippines, and over the Pacific Ocean North of Papua-New-Guinea based on MODIS data.

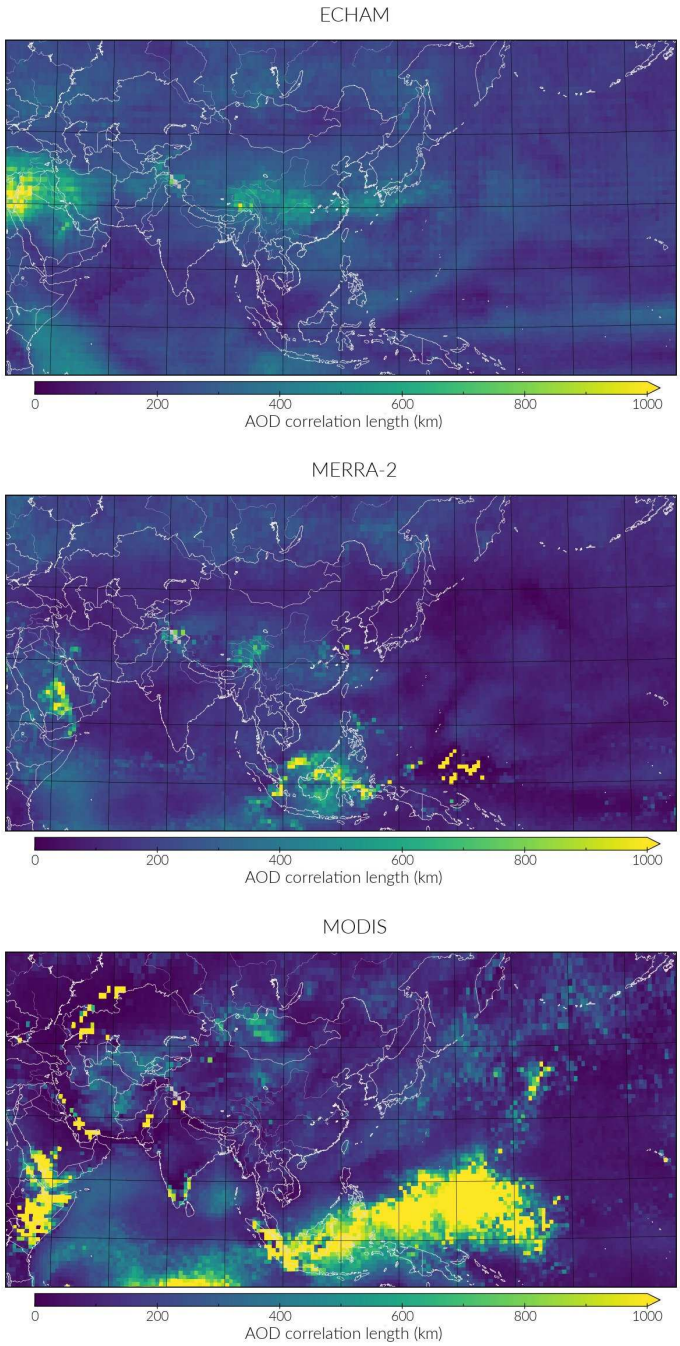


Figure 3.5. AOD correlation lengths over Asia corresponding to ECHAM-HAMMOZ, MERRA-2, and MODIS data.

While regional AOD fields of the satellite retrievals and ECHAM-HAMMOZ simulations share qualitative agreements, correlation length comparisons were not as successful. The reasons behind this have been investigated, but to date no clear, single reason has been found. For example, the regional standard deviation or variance, whether normalized by average AOD or not, was not a good predictor of whether the model and satellite products agreed. It is likely that there exist several distinct reasons for the discrepancy which interact together in a complex manner. While the weather was nudged towards reanalysis data, it should be noted that the nudging is only performed for vorticity, divergence, temperature and surface pressure. Thus variables that are important for aerosol transport, chemistry and removal,

such as convection, turbulence, cloud cover and rainfall can feature significant deviations from the reanalysis. Further, the reanalysis data can deviate from measured data (Pfeifroth et al., 2013). For reasons described above, an article in preparation is more focused on investigating how well measurements at AERONET sites represent the AOD field around them.

We further studied the correlation lengths at the locations of AERONET stations. The correlation length of AOD describes the representativeness of AOD in the surrounding area. AERONET stations with long correlation lengths mean that the AERONET AOD values will faithfully represent the AOD values for relatively large surrounding areas. Short correlation lengths mean that aerosol conditions may significantly vary in the surrounding areas and AOD values based on AERONET measurements are only representative of the relatively small area near the station. The AOD correlation length estimates can therefore be used for evaluating the regional representativeness of AERONET stations and the results can be used, for example, to recommend new locations for new AERONET stations to make the whole network globally more representative. In Figure 3.6, the MODIS correlation lengths for selected AERONET stations are shown and also compared with the Max-Planck-Institute Aerosol Climatology version 1 (MAC-v1)▪ aerosol climatology related AERONET station range score (Kinne et al. 2013). In large proportion of AERONET stations, the MODIS correlation length is in relatively good agreement with the MAC-v1 range score. The largest difference in correlation lengths is observed over Spain where MODIS based estimates are significantly longer than the MAC-v1 range scores. The MODIS correlation lengths corresponding to AERONET stations over the Pacific Ocean are significantly shorten than the range scores. The MODIS correlation lengths were systematically shorter than the range scores also in West Africa and in central North America. The global average MODIS correlation length corresponding to AERONET site locations was about 160 km. The average MODIS correlation length corresponding to AERONET stations and different regions were also computed. The following correlation lengths were observed: Europe 190 km, North America 130 km, South America 180 km, Asia 140 km, Africa 140 km, and Australia 100 km.

For reasons described above, an article in preparation (Lipponen et al., in preparation) is more focused on investigating how well measurements at AERONET sites represent the AOD field around them.

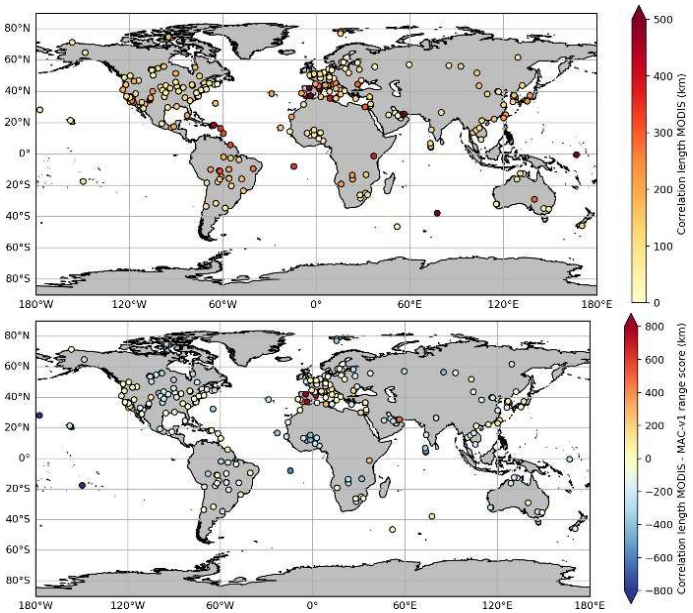


Figure 3.6. Left) MODIS AOD correlation lengths corresponding to selected AERONET stations. Right) The difference between the MODIS AOD correlation lengths and MAC-v1 aerosol climatology AERONET station range score.

Based on the two strands of investigation, it is evident that using a global aerosol-climate model to identify gaps in observational networks is very challenging. In our analysis, the simulated AOD diurnal variability and spatial correlation lengths differ notably from observations from AERONET sites and satellites, respectively. This suggests that aerosol processes in global models need to be further developed and constrained to make sure that they can realistically capture short-term AOD variability on regional scales. It is noteworthy, however, that process description improvements alone are likely not enough to remedy all the model deviations from observations. This is because according to our correlation length analysis, results based on meteorological reanalysis, which is often used to nudge global models, can also show significant deviations from satellite observations. Since the reanalysis data has a strong impact on the transport, removal, etc. in the global aerosol models, further research is required to uncover the reasons behind the differences between reanalysis and satellite correlation lengths.

2.4 Ozone and related trace gases (KIT)

2.4.1 Scope of the assessment

Climate monitoring of ozone (profiles and total columns) requires the reliable detection of long-term changes at the 1% level or even better, as this is the scale of past and projected future ozone changes. Long-term satellite records of ozone need to be validated by ground-based reference networks, to ensure long-term stability and characterize potential drifts in the satellite data. NDACC (www.ndacc.org) constitutes a well characterized reference network for ozone observations with multiple instruments and techniques at individual stations. One important question then is, how representative long-term records at the individual NDACC stations are for validation of satellite observations over a latitude band. We address this question by comparing the long-term ozone evolution from a chemistry-climate model sampled at the location of the NDACC stations with the model results over the wider latitude band. Fig. 4.1 shows a map of the active NDACC stations considered in this study. The colour code used indicates zonal bands. The majority of stations are located in mid-latitude Europe (green), fewer than ten stations are based in the tropics (red) and only one station lies in the northern hemisphere subtropics (golden, Izaña, Tenerife, Spain). Among the tropical stations, only two are in the deep tropics close to the equator (Paramaribo, Suriname and Natal, Brasil).

This modelling study of the representativeness of long-term ozone changes at the NDACC stations is complemented by a statistical analysis of the modelled ozone fields to derive structure functions for the root mean squared (rms) difference of ozone as a function of distance. We place here specific emphasis on the tropical regions for two reasons:

(a) the tropics are a region where past and expected future total ozone changes are influenced by a number of different underlying causes that result in very different long-term trends for the stratospheric and tropospheric partial columns, as we will discuss below.

(b) The tropics are still a rather data sparse region with only a few long-term reference-quality stations. Given this sparseness of stations in the deep tropics, we are going to investigate how representative those available data are in terms of zonal and global ozone trends.

The evolution of the stratospheric ozone layer over the 21st century will be controlled by the decreasing concentration of ozone depleting substances (ODS) following the Montreal Protocol and its amendments, as well as by climate change due to the increase of greenhouse gases. Tropospheric ozone, itself a greenhouse gas, will also be influenced by changes in precursor emissions and indirectly by climate change due to the increase of greenhouse gas concentrations. One sensitive region is the tropics, where chemistry-climate models predict that stratospheric ozone will not recover following the reduction of ODS, but will further decrease due to changes in the large-scale circulation, namely the enhancement

of the overturning Brewer-Dobson Circulation (BDC), causing increased concentrations of greenhouse gases. However, a clear attribution of past changes in tropical column ozone is complicated by uncertainties in concurrent increases in tropospheric ozone.

Here we will assess gaps in the geographical coverage of ozone column measurements (– as well as gaps in relevant parameters), and in particular we will assess whether the coverage of existing networks is sufficiently representative to detect expected climate change signals.

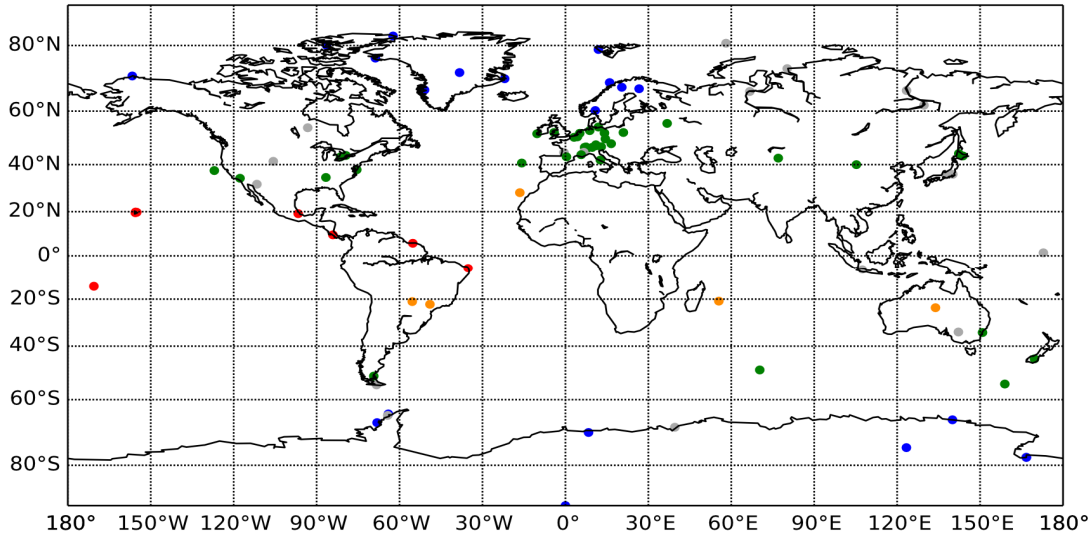


Figure 4.1. Map of current NDACC reference measurement stations considered in our analysis. The colour coding indicates the zonal band: Red: tropics (20° N – 20° S), golden: subtropics (20°-30° N/S), green: mid-latitude (30°-60° N/S), blue: high-latitude (60°-90° N/S); grey: presently inactive stations. See http://www.ndacc.org

2.4.2 Model system and simulations

The data used for our analysis are part of a Chemistry–Climate Model Initiative (CCMI) recommended set of simulations by the Earth System Chemistry Integrated Modelling (ESCiMo) consortium. The ESCiMo simulations have been performed at DKRZ in Hamburg using version 2.51 of the ECHAM/MESSy Atmospheric Chemistry (EMAC) model. The long-term simulation comprises a time span of 150 years starting in 1950 and combines hindcast and projections. The future greenhouse-gas emissions are based on the RCP6.0 scenario, while the sea-surface temperatures and sea-ice cover are taken from HadGEM2 simulations. The ozone data are available as monthly averages in a spatial resolution of T42L47 (approx. 2.8°x2.8° and with a top level at 0.01 hPa). For further detailed information, we refer to the corresponding publication by Jöckel et al. 2016.

In addition, we have analysed higher resolution model results over a limited time period. These additional model simulations are run at the higher resolution of T106L90,

corresponding to about 1.25°x1.25° at 90 vertical levels. (with the top level again at 0.01 hPa) These simulations have been initialized from the long-term simulation in mid 2015, and results are analysed for January 2016.

2.4.3 Analysis

From the 150 years of ESCiMo simulation, the total column ozone has been retrieved. For eight tropical stations, we have evaluated data at the nearest model grid point with respect to the real location in both the horizontal and vertical position. A moving average with a time window of seven years has been used to smooth periodical ozone variation (e.g. seasons). The eleven years solar activity cycle , which is well known to affect total ozone columns, remains as an underlying periodical signal.

In various climate simulation studies, the tropical upwelling has been found to be increasing by the end of the 21st century. As a consequence of the accelerated BDC, there is very likely to be a decrease of tropical lower-stratosphere ozone. Beside this dominant dynamical affect, a recent study (Meul et al. 2014) identifies subsiduary contributions from changes in chemical ozone production and destruction, which may be expected to introduce altitude-dependent changes. While the contribution of chlorine compounds to the ozone depletion declines, the importance of HO_x and NO_x rises. Regarding the ozone production, methane photolysis may slightly gain in importance, which is noteworthy in the case of the real-world potentially following future scenarios seeing a significant rise in methane emission (RCP8.5). The tropospheric ozone is mainly affected by this increase in methane.

In Fig. 4.2 to 4.5, the resulting ozone-column trends are depicted for zonal mean (bold blue line) and existing NDACC stations located within these latitude bands (differently coloured dashed lines). The ozone changes are shown with respect to the annual mean of 1980. The shaded area represents the standard deviation of the moving average of the tropical zonal mean. In these figures, the total column, tropospheric column, and stratospheric column ozone are displayed individually.

For the total column ozone (Figs 4.2-4.5), three major periods can be identified:

- Ozone depletion (~1979-2010)
- Recovery due to ODS decline (~2010-2060)
- Accelerated circulation due to climate change (~2060-…)

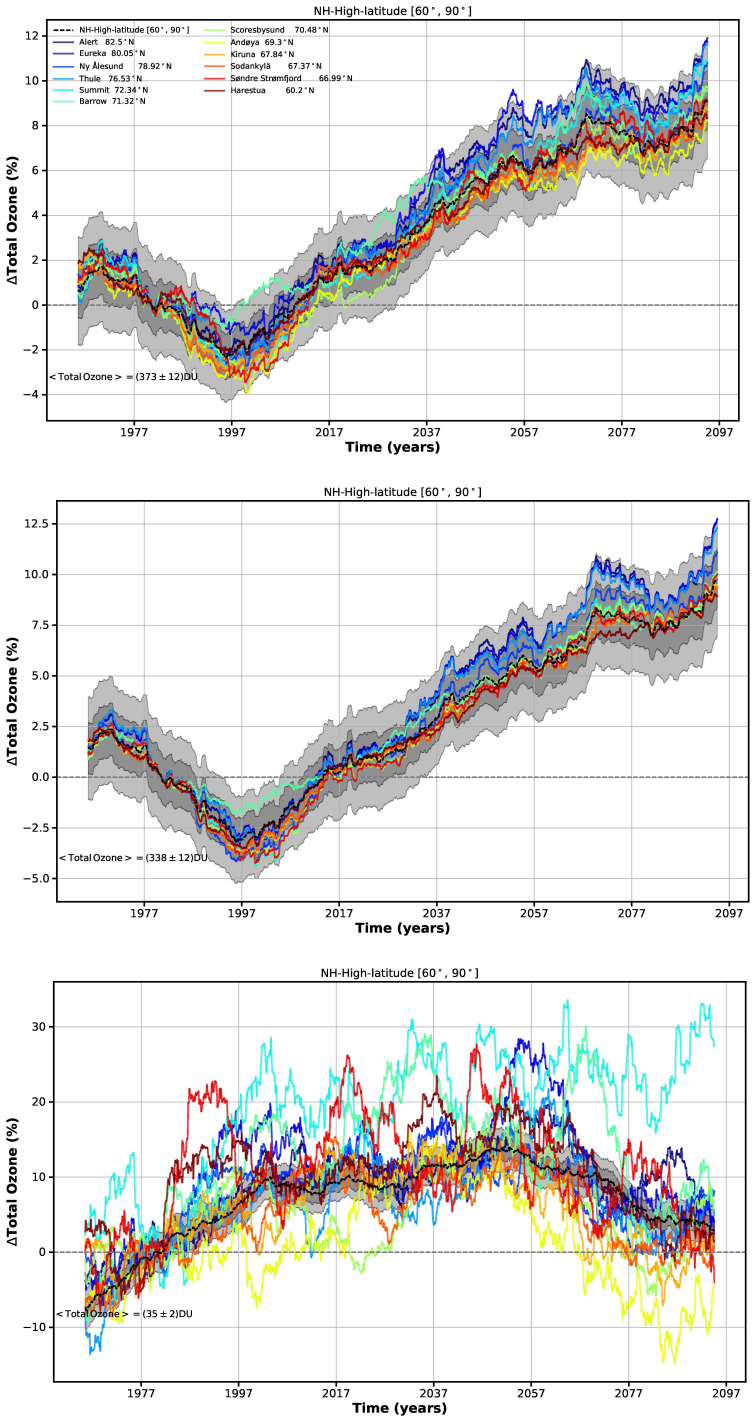


Figure 4.2. EMAC simulations of ozone-column changes between 1970 and 2095 at Northern Hemisphere high latitudes (latitude band 60°N – 90°N), expressed in percent change relative to 1980. Dashed line and gray shading represents the model averaged over 60°N – 90°N, the coloured lines results from the model simulation sampled at individual NDACC stations to assess the representativeness of the existing ground-based network for the wider latitude band. Top panel shows total column changes, middle panel stratospheric column and bottom panel tropospheric column changes.

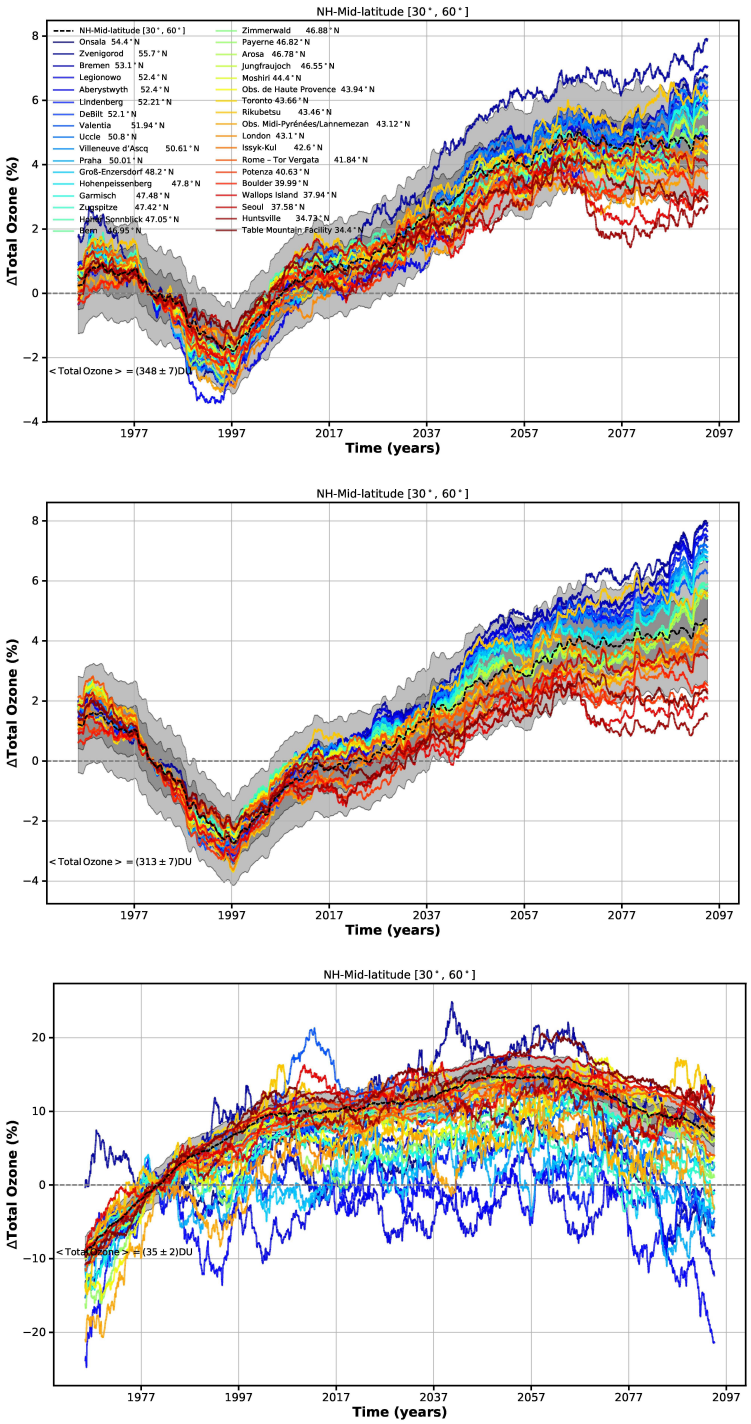


Figure 4.3. As Fig. 4.2 but for Northern Hemisphere mid-latitudes (30°N – 60°N).

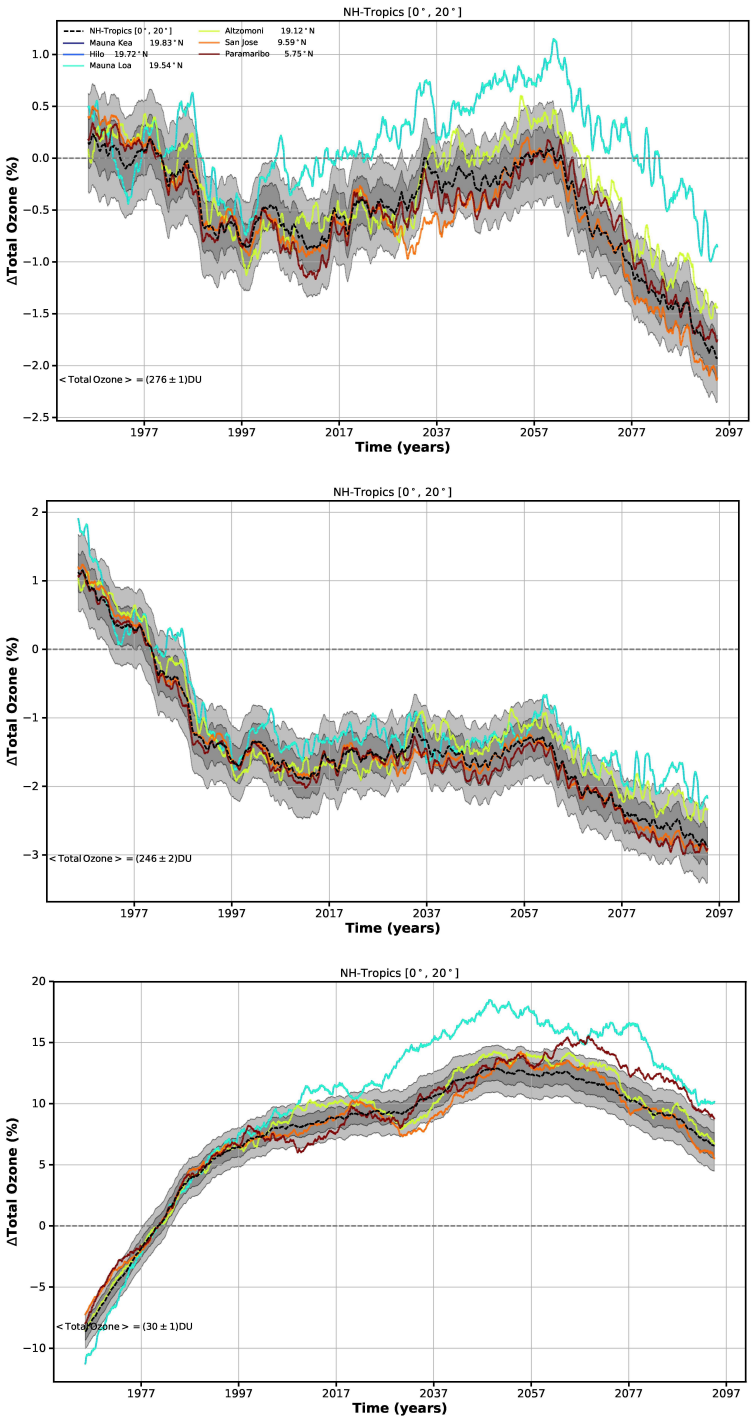


Figure 4.4. As Fig. 4.2 but for Northern Hemisphere tropics (Equator – 30°N).

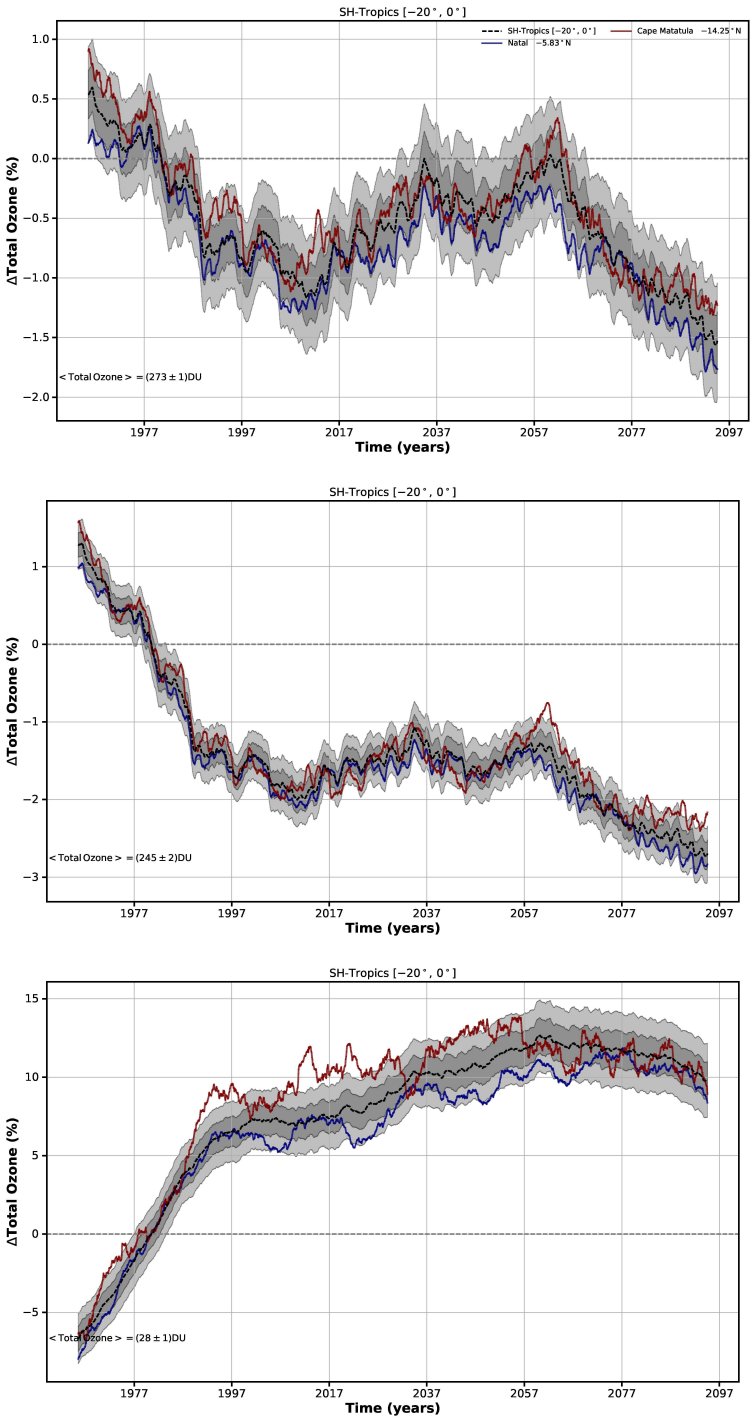


Figure 4.5. As Fig. 4.2 but for Southern Hemisphere tropics (Equator – 30°N).

In the tropics, the divergence between individual stations and the true area average varies within 2% over 150 years. Within the ozone depletion period, all local ozone columns agree rather well with the zonal mean. In the recovery period, the northern-most stations at Hawaii (Mauna Kea, Hilo, Mauna Loa) diverge most strongly from the zonal mean. Splitting the total column into troposphere (Fig 4.4b) and stratosphere (Fig 4.4c) fraction, an underlying rise in troposphere ozone in the course of the 21st century can be identified,

which is strongest for the northern-most stations. In agreement with other studies, the stratospheric ozone column in the tropics does not see a return to pre-1980 levels.

Our results indicate that observations from stations such as Paramaribo (Suriname), San Jose (Costa Rica), and Cape Matatula (American Samoa) cover the zonal ozone column trend well. While these few individual stations are generally representative for stratospheric ozone observations in the tropics as a whole, for tropospheric ozone there are larger discrepancies. Here, we also show that the expected changes in total ozone in the tropics are caused almost equally by changes in stratospheric ozone and by changes in tropospheric ozone. This has implications for the validation of long-term ozone observations in the tropics.

Evaluation of structure functions for ozone variability

For a direct comparison with the statistical analysis of small scale ozone variability based on observational data (GAIA-CLIM Deliverable Report D1.10), we have computed structure functions of ozone from our EMAC chemistry climate model simulations, as well. As our long-term simulations have a rather coarse horizontal resolution of about 2.8°x2.8° (T42 truncation), for this purpose, we have included an additional EMAC model simulation run at the higher resolution of 1.25°x1.25° (T106 truncation) for a shorter time period from mid 2015 to mid 2016 with wind and temperature fields nudged towards the ERA-Interim re-analysis. The structure function or variogram is defined as

$$D(\rho) = D(|r_1 - r_2|) = \left\langle (O_3(r_1) - O_3(r_2))^2 \right\rangle ,$$

were $O_3(r_1)$ and $O_3(r_2)$ are pairs of modeled ozone at two locations, with ρ the distance between r_1 and r_2. The mean is then calculated over all squared differences. The following plots present the square root of the structure function D, i.e. rms differences as a function of distance or separation length, expressed in percent of the mean modelled ozone at that altitude and latitude band. While the structure functions based on observed ozone as performed by FMI (GAIA-CLIM Report D.10) are calculated from collocated (or nearby) observations of the MIPAS and GOMOS instruments, which are due to orbit characteristics restricted to high latitudes, we calculate the structure function at all latitudes as a function of longitudinal separation. Stratospheric ozone typically shows much stronger gradients in latitudinal than longitudinal directions. Tropospheric ozone exhibits strong small-scale structure, often associated with low ozone values in deep-convection areas and high ozone values affected by stratosphere-troposphere exchange. Total ozone in contrast shows less small-scale variability due to integration (in some sense smoothing) over all altitude levels.

Ozone structure functions calculated from the EMAC model for January shows different levels of rms variability for the different altitude regions. Results are shown for the stratosphere at 1 hPa (approximately 45km altitude), 10 hPa (about 30km altitude) and 50 hPa (about 20km), as well as for the mid-troposphere (approximately 500 hPa) and total

column ozone. The results for Arctic winter at 75°N-85°N may be compared to the corresponding analysis of measured ozone by FMI as part of Task 1.4 (see Figure 4.7).

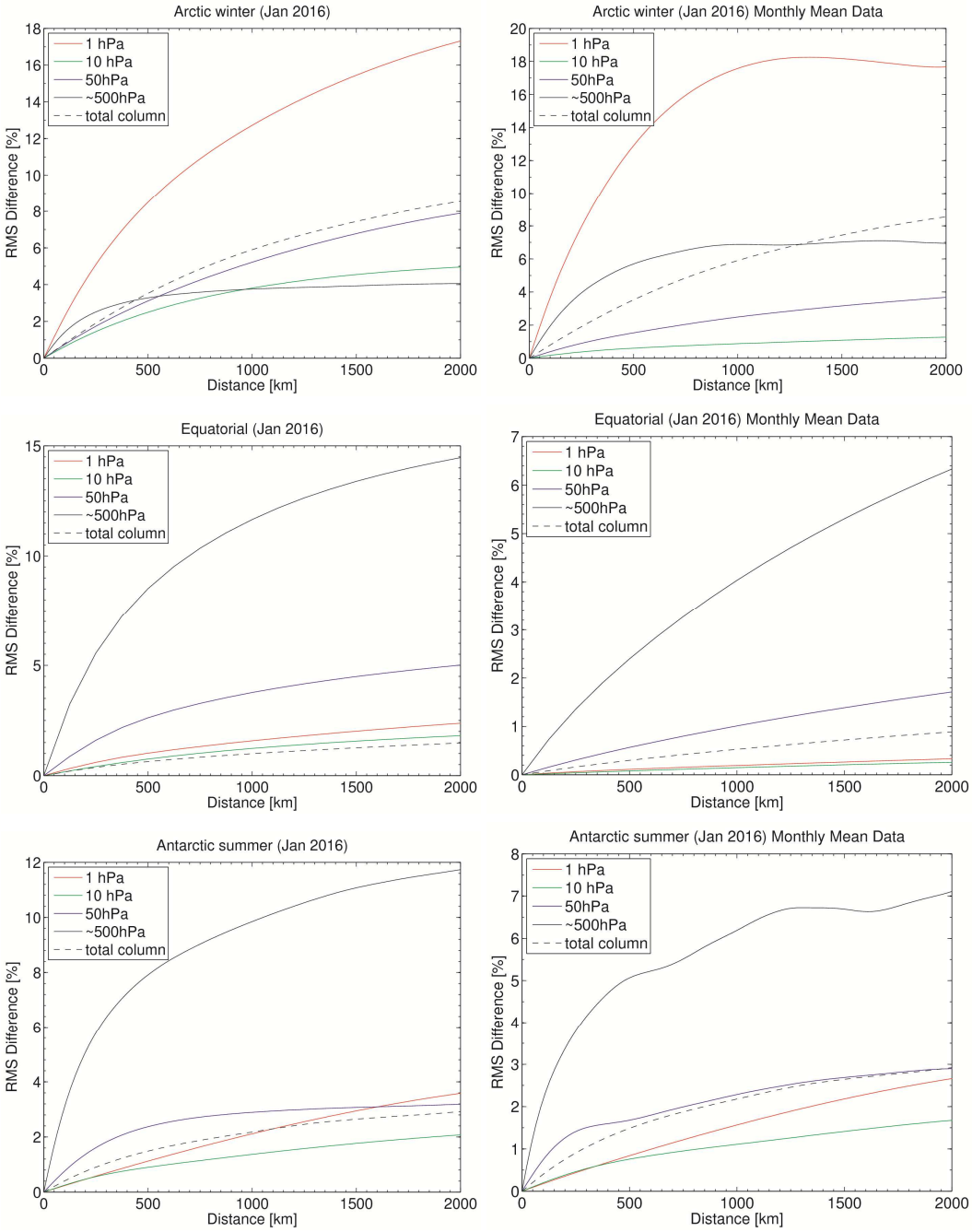


Figure 4.6. Ozone structure functions evaluated from the EMAC model for different pressure levels in the stratosphere (1 hPa, 10 hPa, 50 hPa), the mid-troposphere (aprox. 500 hPa) and for total column ozone. Top: Arctic winter (75°N – 85°N), middle: tropics (15°N – 15°S), bottom: Antarctic summer (75°S – 85°S). Note different vertical scales.

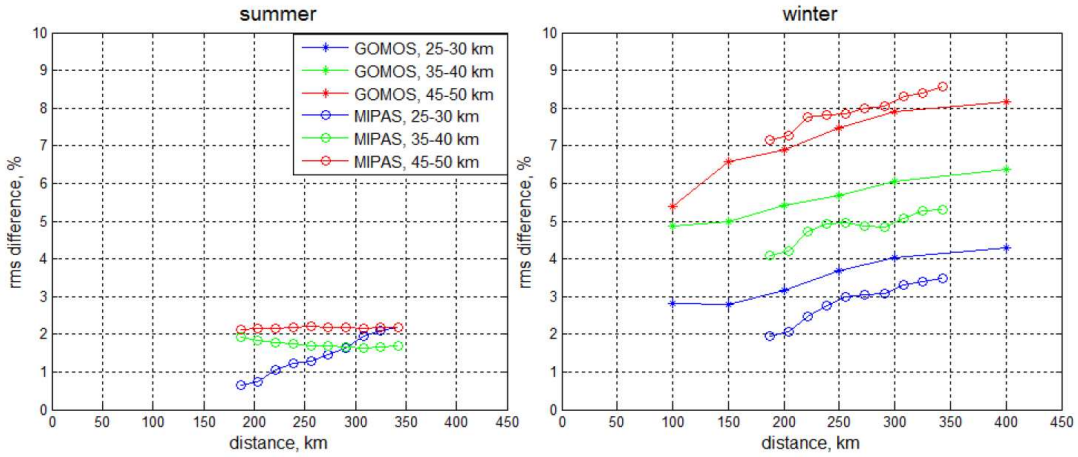


Figure 4.7. Ozone structure functions near the North Pole, calculated from MIPAS and GOMOS satellite observations. Taken from the Task 1.4 Technical Note (GAIA-CLIM Deliverable Report D.10) for direct comparison, provided by FMI.

By definition, the ozone structure functions go to zero for zero distance. For our simulations, the shortest real correlations can be evaluated at the grid scale, i.e. at distances of the order of 100 km for the T106 simulation. At long distances, the structure functions approach the variance (or standard deviation, if the square root is taken) of ozone in this latitude band and altitude. The ozone variance is rather different at different altitude levels and so is the long-range limit of the structure function. The correlation length, i.e. over what length the structure function stays below a certain threshold when normalized by the long-range variance, is more similar for different altitude and latitude regions, but shows also marked differences: tropospheric ozone shows generally shorter correlation lengths than stratospheric ozone. Structure functions for total ozone are typically similar to those for the (lower) stratosphere, which is not surprising as these regions contribute most to the ozone columns.

A closer comparison of the model derived ozone structure functions with those obtained from satellite observations shows quantitative agreement for the Arctic winter stratosphere. The satellite observation based structure function at 45 – 50 km altitude shows RMS differences between about 6% at 100 km distance, increasing to about 8% at 400 km distance. This agrees well with the modeled RMS difference at 400km of about 7%, while the modeled value at 100 km distance at about 2% is smaller than the observed value. Modeled RMS differences at lower altitudes are smaller, in qualitative agreement with the observed values. Observed RMS differences at the 200 to 300 km scale at 25-30 km altitude from MIPAS are around 2-3%, while the modeled RMS difference at 10 hPa (~30km altitude) is only about 1% and at 50 hPa (~20 km altitude) is 2-3%. As noted above, the structure function for total ozone is similar to the one for ozone at 50 hPa.

For Antarctic summer (75°S-85°S), the stratospheric ozone structure functions (and thus variance) are substantially smaller than those for Arctic winter. This is qualitatively in line with the observationally derived structure functions for Arctic summer: Modelled values for Antarctic summer are in the range of 1-2% at the 200-350 km length scale, as are the MIPAS-based structure functions for Arctic summer.

In the tropical stratosphere, the variance is significantly smaller than at high latitudes. Structure functions for the mid- and upper stratosphere, as well as for total ozone, are below 1% for distances below 500 km. Only in the tropical lowermost stratosphere at 50 hPa, variance is larger with values approaching 3% at 500 km distance. In contrast, variance for tropical tropospheric ozone is much larger, with values of 7-8% at 500km distance. This indicates that reference observations at a given location in the tropics are more representative for a larger region in the stratosphere, but much less so in the troposphere. This is likely due to the smaller dynamical scales of tropospheric weather systems, compared to stratospheric dynamics, the effect of convection in the troposphere and due to the more localized sources of tropospheric ozone.
In Fig. 4.6, we compare the instantaneous structure functions with structure functions calculated on monthly-mean data. This provides us some information on how representative monthly-mean values at a given location are for the monthly means of wider latitude regions, a question less important for calibration and validation of the satellite data, but highly important for characterizing the long-term stability of satellite-ozone observations.

2.4.4 Results and Recommendations

The chemistry-climate model simulations show the needs and current shortcomings of the ground-based network to validate the stability of long-term satellite ozone observations. We have provided analyses for various latitude bands but emphasize the importance for tropical ozone for two reasons: (a) this is a critical region were model projections show a complex interplay of changes due to climate change, due to changes in ozone depleting substances, and due to changes in tropospheric ozone, and (b) this is a data-sparse region. Projected ozone changes are in the range of a few per cent for total ozone, with changes in the stratospheric column and the tropopspheric column of similar absolute size. Changes in tropospheric ozone thus partly mask changes in stratospheric ozone, which is an important climate variable. This demonstrates the challenges for a ground-based network for the purpose of validating long-term satellite ozone records, in particular important for nadir viewing satellites that provide either only total column ozone or have broad weighting functions / averaging kernels were signals are impacted by both the troposphere and the stratosphere.

From the model simulations, we find that the long-term changes in tropical stratospheric ozone are sufficiently well represented by the existing stations of the reference network.

This is more critical for tropical tropospheric ozone, where single stations are less representative for the tropics as a whole.

These findings are further corroborated by the statistical analyses of the model data using structure functions or variograms. These structure functions show that the rms difference between two ozone observations is expected to be in the order of a few per cent for more than 1000 km, while the variance in tropospheric ozone is much larger and rms differences of more than 10% are expected to be reached for distances of about 500km.

2.5 Assessing geographical gaps for reference temperature and humidity in-situ data (ECMWF)

2.5.1 Scope of the assessment

Satellite instruments sensitive to atmospheric temperature and humidity are regularly assimilated into Numerical Weather Prediction (NWP) and Re-analysis systems and contribute to the atmospheric (re)analysis products of temperature, humidity, and vector winds. Before assimilating a new instrument, the radiances are compared to the background fields (short-range forecasts from the previous cycle) in radiance space, in order to assess the quality of the data. The background fields (B) are transformed into radiance space using a radiative transfer forward model and then compared to the observed radiances (O) for the different instruments. Observation minus background (O - B) values provide an important diagnostic tool for the calibration/validation (cal/val) of new satellite instruments since the NWP background fields are highly accurate global fields, allowing for cal/val over the full dynamic range of the instrument. What remains are a combination of uncertainties in the observations, the forecasts and the radiative transfer forward models. Global statistics of O - B tend to show non-zero biases, as well as mean values that vary geographically or as a function of orbital parameters. Systematic uncertainties in the observations (e.g. due to calibration) or in the radiative transfer forward model must be corrected either prior to or during the assimilation, in order to prevent these uncertainties from propagating into the reanalysis products and/or the forecasts of NWP systems. In some cases, the biases can be easily diagnosed as due to the instrument, such as the orbital-dependent biases found for the SSMI/S instrument (Bell et al 2008; Geer et al; 2010), which were diagnosed as due to solar intrusions into the warm calibration load and emission from the main reflector of the instrument. In other cases, however, it is more difficult to diagnose the source of the biases. Furthermore, while the presence of instrument-related uncertainties can often be detected, it is difficult to quantify them since the uncertainties in the forecasts are not known exactly.

Reference in-situ radiosonde measurements for temperature and humidity with calculated uncertainty values, such as the GCOS Reference Upper Air Network (GRUAN), have the potential to quantify how much of the bias observed in O - B statistics is due to systematic uncertainties in NWP forecast models, and give confidence intervals for these values. In work package 4 of the GAIA-CLIM project, a GRUAN processor has been developed which transforms the GRUAN observations and uncertainties into radiance space for different satellite instruments. The GRUAN processor allows a comparison between the reference in-situ observations and the NWP model background fields in radiance space. The use of this processor for assessing systematic forecast uncertainties in radiance space would greatly support the use of NWP fields in the cal/val of new satellite instruments.

However, while the GRUAN data have a high vertical resolution and traceable estimates of uncertainties (Dirksen et al; 2014), they are only available for a limited number of locations and it is not clear whether the network is geographically capable of assessing the types of bias observed in O - B statistics, in particular biases that vary geographically. In this study, we

analyse the mean O - B values for key satellite microwave temperature and humidity sounders, and sub-sample these values at the locations of the GRUAN sites. Comparing the sub-sampled O - B values to the full dataset will allow us to better understand which types of bias the GRUAN data can diagnose, as well as areas where we may have geographical gaps. Note that we consider here the types of bias where it is difficult to identify the source, i.e. global and geographical biases. There are other types of bias which can be diagnosed unambiguously as being due to the instrument using NWP short-range forecasts, such as biases due to solar intrusions, or asymmetric scan-angle biases. These latter biases are not considered in this study.

This report is structured as follows. In section 3.5.1 we present firstly the radiosonde site locations, and then in section 3.5.2 we describe the measurements made by the microwave satellite instruments considered in this study. In section 3.5.3 we outline how the O - B statistics are calculated for these instruments, using the ECMWF model background fields. Finally, we present an analysis of O - B statistics, in sections 3.5.4 and 3.5.5.

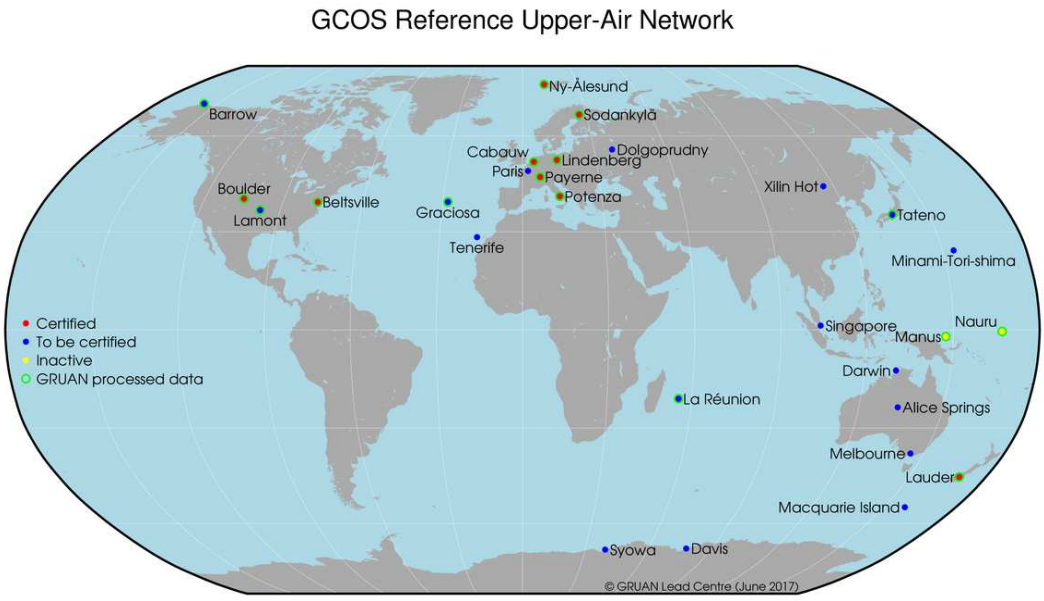


Figure 5.1. Map of the GRUAN site locations, including certified GRUAN sites, non-certified GRUAN sites, GRUAN sites currently reporting (both certified and non-certified), and two GRUAN sites that are now inactive.

2.5.1 The radiosonde sites

The GRUAN Network consists of a number of sites where Vaisala RS92 radiosondes are launched and the data processing follows the criteria set out by Immler et al (2010) for reference measurements. This includes estimates of measurement uncertainty as described by Dirksen et al (2014). The locations of the sites are shown in Fig. 5.1 and data are made available online by NOAA[1]. There are currently nine certified sites, and a further 17 sites yet to be certified. Of these un-certified sites, five already provide data processed with the

[1] ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/RS92-GDP/version-002

GRUAN data processing software, and with estimated uncertainties. In addition, there are two inactive sites for which data are available in 2013.

Reference in-situ radiosonde measurements for temperature and humidity with calculated uncertainty values, such as the GCOS Reference Upper Air Network (GRUAN), have the potential to provide a quantative estimate of how much of the bias observed in O - B statistics is due to systematic uncertainties in NWP forecast models, and give confidence intervals for these values. For the current network, we consider both the 9 certified sites and sites whose data are available on NOAA's ftp in 2016. The list of these latter sites is given in Table 5.1, along with the mean number of reports per month for 2016, the number of reports in July 2016, and the regular launch times of the sites. Note that there were no reports available for Lauder in 2016 and so this network is restricted to sites in the Northern Hemisphere. This list of sites represents those included in the GRUAN processor of work package 4 for 2016. For the full GRUAN network, we consider all sites excluding the inactive sites of Manus and Nauru.

In addition to GRUAN, we calculate the biases in O - B sub-sampled at sites in the full RAOB network. This is because these sites provide an indication of where the GRUAN network could most easily be expanded to in the future. It also gives us an upper bound of the sampling we can expect from the current full radiosonde network (excluding ships and buoys). Furthermore, these are the sites considered in the statistical study of task 1.4. The locations of these sites are shown in Fig. 5.2.

Table 5.1. GRUAN sites whose data are available in 2016 on the ftp site, along with the average number of reports per month (rounded to the nearest integer), the number of reports in July 2016, and the regular launch times. Note that launches at other times are sometimes also performed for some sites.

GRUAN site	**Average report number per month**	**Number of reports in July 2016**	**Regular launch times (UTC)**
Barrow	36	45	5:30, 17:30
Boulder	6	2	Evening (18:00 – 20:00)
Cabauw	20	21	0:00, 12:00
Lindenberg	102	68	0:00, 6:00, 12:00, 18:00
Ny Alesund	31	26	12:00
Payerne	3	2	0:00, 12:00
Potenza	1	0	Evening (17:30 – 19:00)
Lamont	67	82	5:30, 11:30, 17:30, 23:30
Sodankyla	33	50	0:00, 12:00
Taterno	3	2	0:00, 12:00
Tenerife	40	50	0:00, 12:00

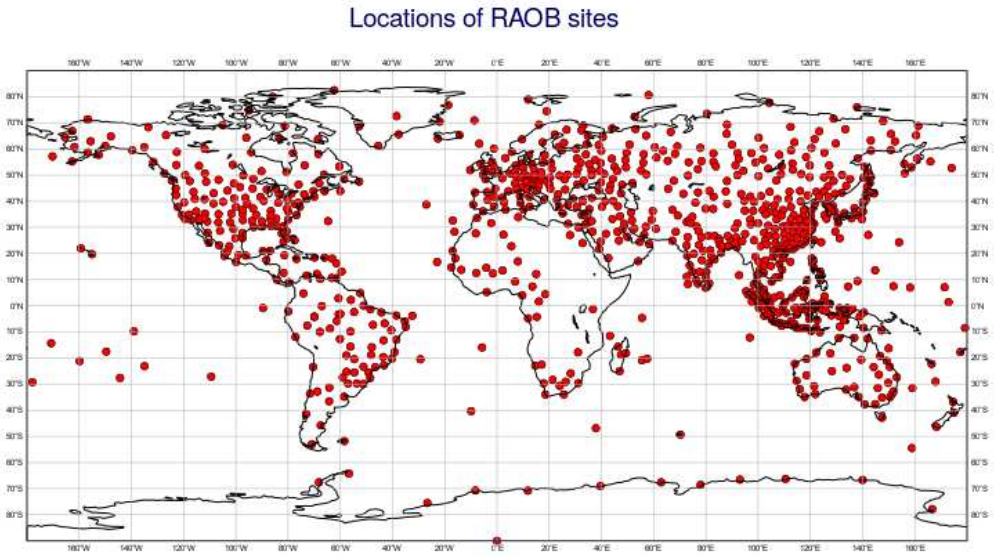


Figure 5.2. Map of the RAOB site locations

2.5.2 The AMSU-A, ATMS and MHS satellite instruments

The main microwave temperature sounding instruments used in NWP and reanalysis are the Advanced Microwave Sounding Unit-A (AMSU-A) and the Advanced Technology Microwave Sounder (ATMS). These instruments have temperature sounding channels at frequencies between the lines in the 50 - 60 GHz oxygen band which are sensitive to temperature at different heights of the atmosphere. The channel frequencies are given in Table 5.2, and the sensitivities of each channel to different atmospheric heights is shown in the Jacobians plotted in Fig. 5.3 (for ATMS channel numbers). To date, there have been seven AMSU-A's flown on both American and European satellite platforms, six of which are still currently operational, one ATMS flown on the Suomi-NPP platform and an additional ATMS launched on the NOAA-20 satellite in November 2017 for which data are not yet available. Conical scanners such as the Special Sensor Microwave Imager/Sounder (SSMI/S) also have 50 - 60 GHz channels, but in this study, we will focus on the cross-track scanning instruments of AMSU-A and ATMS.

Microwave humidity sounding channels measure around the water vapour absorption line at 183 GHz. Humidity sounding channels at 183 GHz are found on the Microwave Humidity Sounder (MHS) instruments, ATMS, the MicroWave Humidity Sounder (MWHS) on FY-3B, and the MicroWave Humidity Sounder -2 (MWHS-2) instrument flown on the FY-3C satellite, all of which are cross-track scanning instruments. There are currently four MHS instruments flown on two NOAA satellites and two MetOp satellites, and one ATMS. The MHS instruments have 3 sounding channels around 183 GHz. The ATMS instrument has the same channels and an additional two channels. The channel numbers and frequencies of MHS and ATMS are given in Table 5.3. Each sounding channel on these instruments is sensitive to relative humidity at different heights of the atmosphere. The sensitivity to different heights can be illustrated by the humidity Jacobians, shown in Fig. 5.3 for the ATMS channels. It is worth noting that the Jacobians themselves vary with atmospheric humidity so that in a dry

atmosphere the channels peak at lower heights and in a wetter atmosphere they peak higher up. This is different to the temperature sounding channels, where the Jacobian heights do not greatly vary (since oxygen is a well-mixed gas).

Table 5.2. Temperature-sounding channel numbers and frequencies for the AMSU-A and ATMS instruments

Channel frequency (GHz)	AMSU-A channel number	ATMS channel number
53.596 ± 0.115	5	6
54.4	6	7
54.94	7	8
55.5	8	9
57.29	9	10
57.29 ± 0.3222 ± 0.217	10	11
57.29 ± 0.3222 ± 0.048	11	12
57.29 ± 0.3222 ± 0.022	12	13
57.29 ± 0.3222 ± 0.010	13	14
57.29 ± 0.3222 ± 0.0045	14	15

Table 5.3. Humidity Sounding channel numbers and frequencies for the MHS and ATMS instruments

Channel frequency (GHz)	MHS channel number	ATMS channel number
183.31 ± 1.0	3	22
183.31 ± 1.8	-	21
183.31 ± 3.0	4	20
183.31 ± 4.5	-	19
183.31 ± 7.0 (ATMS)/ 190.311(MHS)	5	18

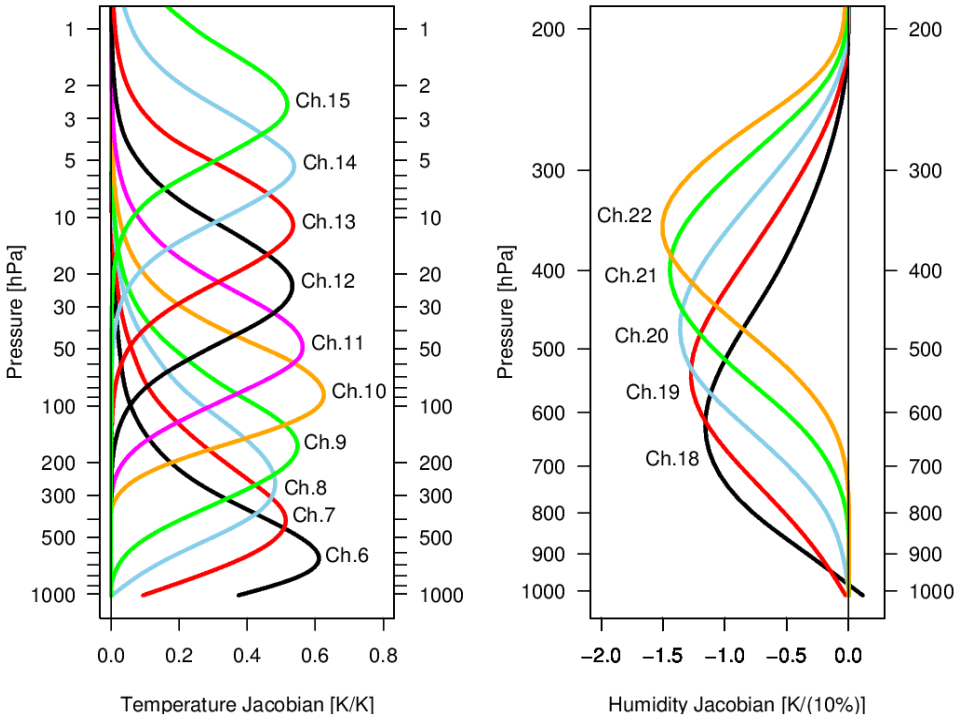


Figure 5.3. Temperature and humidity Jacobians for ATMS sounding channels (normalised by the change in ln pressure of each model level)

Some of the AMSU-A instruments assimilated at ECMWF have a so-called 'γ-correction' applied for channels 5 - 8 in the radiative transfer calculations. For these channels, the optical depth is scaled by a factor (γ), which varies by satellite and channel. This factor acts to remove air mass dependent biases observed in O – B, but it is empirically calculated from O - B values themselves and so ideally we would want to assess the O - B statistics without this scaling. We have therefore concentrated our analysis on the AMSU-A instruments, which do not have a scaling applied: MetOp-A, MetOp-B, and NOAA-19. However, previous studies have been carried out without the γ-correction for AMSU-A (e.g. Lupu et al; 2016) and a brief analysis of O - B values from the experiments in this study indicates that the bias patterns are broadly similar for all AMSU-A instruments once the γ-correction has been removed (results not shown). We therefore think that the conclusions of this study can reasonably be extrapolated to all AMSU-A instruments.

2.5.3 Calculating O – B

The observation minus background statistics used in this study were calculated from an offline run of the ECMWF Integrated Forecasting System (IFS), run at a horizontal resolution of TCo399 (~25 km, a lower resolution than operations), with 137 levels in the vertical (model top of 0.01 hPa) and for a recent version of the IFS (version 43r3). Statistics were calculated for December 2015 and July 2016. The background radiances for AMSU-A and ATMS were calculated for clear-sky conditions using RTTOV version 11.2 (Saunders et al; 2006, Hocking et al; 2015, Lupu et al; 2015a, 2015b). The background radiances for MHS were calculated in all-sky conditions, using RTTOV-SCATT (Bauer et al; 2006, Geer et al; 2014a) since this instrument is assimilated operationally in all-sky conditions. Before

calculating O - B statistics (both global and sub-sampled), a cloud screening was applied for the lower-peaking channels which includes all humidity sounding channels, channels 5 - 8 of AMSU-A, and channels 6 - 9 of ATMS. It is important to do this since cloud and precipitation effects can otherwise dominate the statistics for these channels. Additional quality control screens were also applied, removing data where there are known model problems or known satellite instrument problems. The cloud and precipitation screening and the quality control are described in the following and also summarised in Table 5.4 for the different instruments and channels.

For AMSU-A and ATMS the operational cloud-screening (as described by Lawrence et al (2015)) was applied along with an additional check of the liquid water path, which is calculated from 24 GHz and 89 GHz window channels over ocean. Any data with a liquid-water path greater than 0.10 kg/m^3 were removed over oceans. There is no operational cloud and precipitation screening applied for MHS in the assimilation of these data, since it is assimilated in all-sky conditions. However, for the purposes of this study, we used the scattering index calculated from channels 1 (90 GHz) and 2 (150 GHz) which is normally used as a cloud predictor in the calculation of the observation error covariance matrix (see Geer at al (2014b) for more information). Since cloud effects are included in the background for MHS, we screened for the background cloud using the background scattering index (SI_{fg}) as well as the observed cloud using the observation scattering index (SI_{ob}). (The calculation of SI_{fg} and SI_{ob} is given by Geer et al (2014b).) Any data with the observation or the background scattering index greater than 10 over ocean or greater than 5 over land were screened.

Table 5.4. Cloud screening and quality control applied to the data before calculating O – B statistics

Instrument	Channels	Cloud screening	Additional Quality Control
AMSU-A	All (5 – 12)	-	Remove the first and last 3 scan positions
AMSU-A	5 – 8	Operational & remove data with liquid water path > 0.1 kgm^{-3} over ocean	Operational orography screening. Remove data over seaice.
ATMS	6 – 9	Operational & remove data with liquid water path > 0.1 kgm^{-3} over ocean	Operational orography screening. Remove data over seaice.
ATMS	18 – 22	As for channels 6 - 9	As channels 6 – 9 & remove data with T_{skin}<278 K over land
MHS	3 - 5	Remove data with: Ocean: SI_{ob}>10 or SI_{fg}>10 Land: SI_{ob}>5 or SI_{fg}>5	Operational orography screening. Remove data with T_{skin}<278 K over land. Operational cold-air outbreak screening (ch5). No data over seaice.

Surface emissivity is an important parameter for surface-sensitive channels and must be estimated in order to calculate the surface contribution to the background in radiance space. Over ocean, the surface emissivity is calculated from skin temperature (T_{skin}) and 10m wind speed using the FASTEM model (version 6) and over land and sea-ice, it is estimated from window channel observations following the method of Karbou et al (2005). The skin temperature used for the surface emission is based on the OSTIA product (Donlon et al 2012).

In addition to the cloud screening, further screenings were applied in order to remove data in areas where the forecast model and surface emission are known to be less accurate. Additional screenings were applied as follows. The first and last three scan positions of AMSU-A were removed, as is done in the operational assimilation. The 183 GHz channels were screened over land for surface skin temperature values lower than 278 K, in order to remove data over snow where the surface emissivity calculation may be less reliable. Channel-5 MHS data were screened for cold-air outbreaks in the extra-tropics using the operational screening of Lonitz et al (2015) since this is a known model problem. All humidity sounding channels and the surface-sensitive temperature sounding channels (AMSU-A 5 - 7 and ATMS 6 - 8) were also screened over sea-ice. This was done since the biases over sea-ice can be quite different, which is thought to be due to a less accurate estimated surface emission (rather than potential satellite instrument biases). The lower peaking channels were screened over high orography following the operational screening (see Lawrence et al (2015) for more details).

2.5.4 O - B statistics for AMSU-A and ATMS temperature-sounding channels

Here we evaluate the O - B statistics of the temperature sounding channels of AMSU-A and ATMS and sub-sample these statistics at radiosonde locations. Note that we do not assess any radiosonde data in this study - we assess the satellite instrument O - B values at radiosonde locations. For both, AMSU-A and MHS, we consider only the global and geographically varying biases of O - B. These are biases where the GRUAN data could be useful in estimating the contribution of systematic uncertainties in the short-range forecasts. There are other types of instrument bias, which are not evaluated here, (e.g. scan-angle, orbital biases etc.), since they can generally be un-ambiguously diagnosed as due to the instrument.

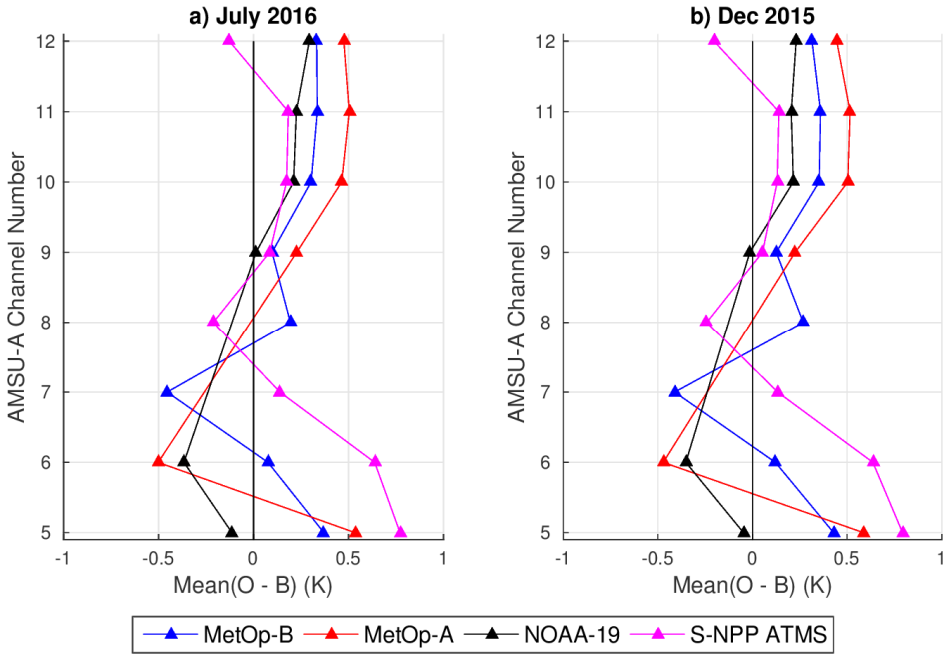


Figure 5.4. Mean O - B for AMSU-A and ATMS temperature sounding channels, plotted as a function of AMSU-A channel number and for a) July 2016 and b) December 2015.

The mean O - B (bias) values were calculated for the AMSU-A and ATMS instruments in July 2016 and December 2016. The globally averaged values for all AMSU-A and ATMS temperature sounders in July 2016 and December 2015 are shown in Fig. 5.4 as a function of channel, and maps of O - B are shown in Fig. 5.5 for two channels of MetOp-B AMSU-A and ATMS in July 2016. As these Figures show, the AMSU-A instrument has global biases of magnitude 0 - 0.5 K, and complex geographical biases, varying by up to 0.8 K (compared to instrument noise of ~0.2 K and uncertainties in the random background uncertainties of ~0.1 K). The geographical biases vary with channel and season and appear to depend on the air mass. The air-mass biases for all AMSU-A instruments are very similar, while those of ATMS are somewhat different, as can be seen in Fig. 5.5. ATMS has a different bias pattern in comparison to AMSU-A for the tropospheric channels (ATMS 6 - 8), with biases that are less 'smooth' geographically (see e.g. Fig. 5.5c compared to Fig. 5.5a). On the other hand, the higher peaking channels of ATMS (9 - 11) have a similar pattern of biases to AMSU-A but are less geographically variable, as can be seen in Fig.s 5.5b and 5.5d for AMSU-A channel 9/ATMS channel 10.

Both global and air mass dependent biases could be due either to systematic instrument uncertainties, systematic radiative transfer uncertainties or systematic forecast model uncertainties, or (perhaps most likely) a mixture of all three. The air-mass biases for AMSU-A channels 5 - 8 of some satellites have been attributed in the past either to radiative transfer uncertainties or to pass-band shifts in the instrument. However, a recent study by Lupu et al (2016) testing the suggested corrections for both of these sources of uncertainty led to unexpected degradations in forecast accuracy. Further work is therefore needed to understand the sources of these biases, and it could be particularly useful to diagnose how much of this bias is due to systematic uncertainties in the forecast, using the GRUAN network.

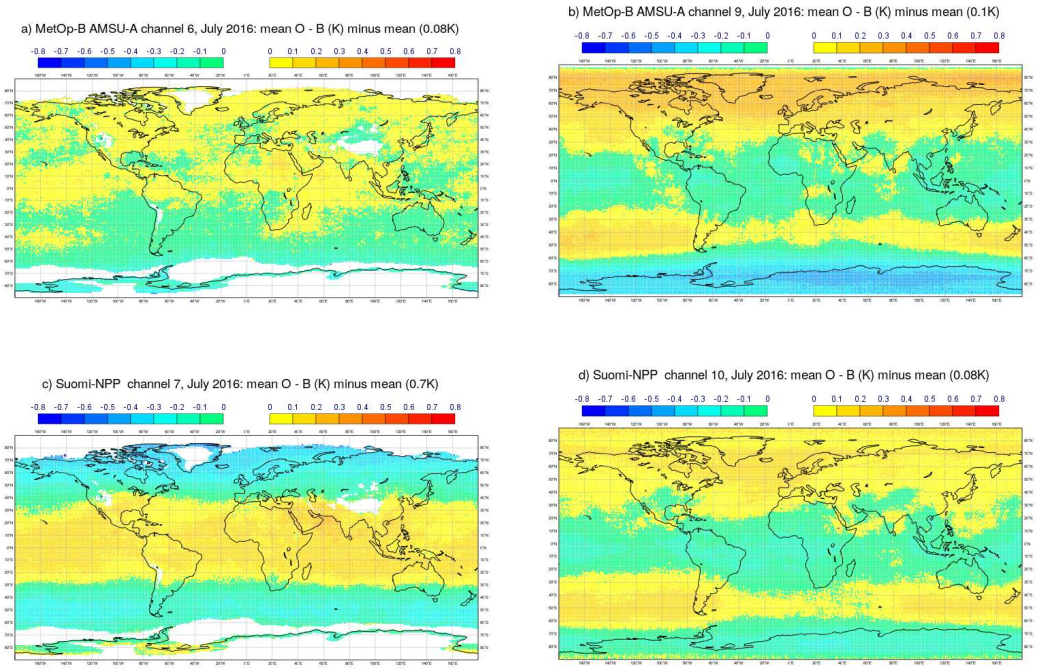


Figure 5.5. Maps of mean (O - B) minus the global average, for AMSU-A and ATMS data averaged over one month for July 2016, shown for a) MetOp-B AMSU-A channel 6, b) MetOp-B AMSU-A channel 9, c) Suomi-NPP ATMS channel 7 and d) Suomi-NPP ATMS channel 10. Note that the lower channels were screened for cloud, sea-ice and orography. The AMSU-A and ATMS channels shown here are at the same frequency.

In order to analyse the capabilities of the GRUAN network for assessing geographical biases, the mean O - B values for each channel of AMSU-A and ATMS were calculated for AMSU-A and ATMS observations sub-sampled at different radiosonde sites. The data were sub-sampled at the currently certified GRUAN sites, sites whose data are available for 2016, and all GRUAN sites, including both present and future (excluding the two inactive sites in the tropics). In addition, we calculated biases sub-sampled at sites in the full RAOB network.

For most sub-samples, we included up to two satellite O - B calculations per 24 hours, selecting those closest to the sites within a 100-km radius. We therefore did not consider the temporal sampling of the sites. However, for the 2016 GRUAN sites, we aimed to estimate the impact of the temporal sampling in addition to the geographical sampling. To do this, we calculated the average monthly reporting rate for these sites from the data available on NOAA's ftp site, and then selected the same number of O - B data. We first selected two datasets per day, then further sub-sampled these (for sites with less than 62 reports per month) by making a random selection. A random selection was chosen as a method of estimating the effects of the radiosonde temporal sampling. Perhaps a more representative method would be to select data from the days in July 2016 and December 2015 when there were launches at the different sites - this could be considered in future work. The 2016 GRUAN-reporting sites are shown in Table 5.1, along with their mean number of launches per month for 2016. Data from the ftp are included in the GRUAN processor and so this sub-

sample gives us an idea of what is currently possible with the output of the GRUAN processor.

Results of the sub-sampled biases are illustrated in Figs 5.6, and 5.7, which show global and hemispheric biases for different seasons for MetOp-B AMSU-A, and Suomi-NPP ATMS respectively. The results for MetOp-A and NOAA-19 AMSU-A are consistent with MetOp-B AMSU-A and so are not shown here. The error bars given in these plots ($\sigma_{sub\text{-}sample}$) are the 95% confidence intervals in the mean, calculated as follows:

$$\sigma_{sub-sample} = 1.96\frac{\sigma(O-B)}{\sqrt{n}} \tag{5.1}$$

where *n* is the number of data in the sub-sample, and $\sigma(O-B)$ is the standard deviation of O - B for the full sample (either global or hemispheric). These error bars give us an indication of the margin of uncertainty that we might expect from the sample size in a random Gaussian distribution - any deviations from this are an indication that the sample we have chosen is not what would be expected of a random sample of size *n* from a Gaussian dataset with the same variance.

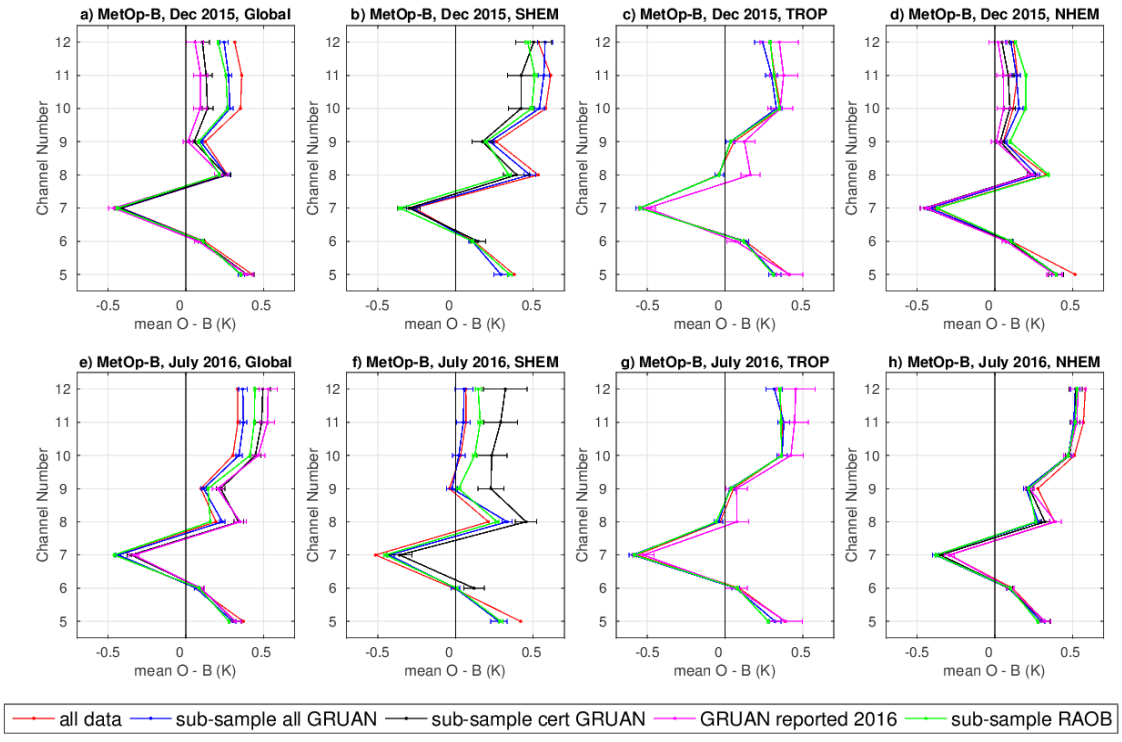


Figure 5.6. MetOp-B AMSU-A global and hemispheric mean O - B values shown as a function of channel number and for two seasons - a one-month average in December 2015 and a one-month average in July 2016. Each plot shows the average over all data (after quality control) and data sub-sampled at different sites including the full GRUAN network, the certified GRUAN network, the GRUAN network with data reported in 2016, and the full RAOB network.

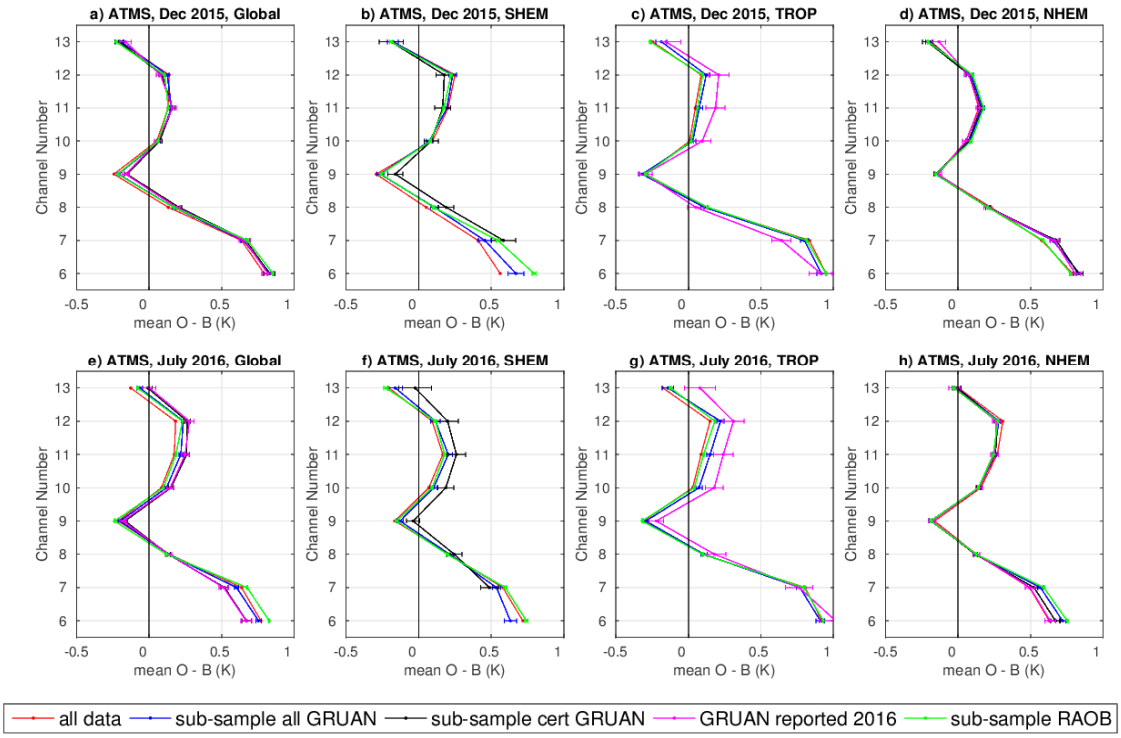


Figure 5.7. As Fig. 5.6 but data are shown for the equivalent channels of Suomi-NPP ATMS.

The results of Fig. 5.6 show that the sampling of GRUAN sites is not able to capture the global biases of AMSU-A for many of the upper channels, and this is due to the sampling in the Southern Hemisphere extra-tropics. Increasing the sampling to the full radiosonde network brings the biases closer to the global value but there is still approximately 0.1 K difference to the full dataset for some channels, particularly channels 10 - 12. These channels have the largest variation in bias across the hemispheres. However, the full GRUAN sampling (certified and future stations) is good enough to capture the hemispheric averages in the tropics and Northern Hemisphere for the AMSU-A instruments.

The global and hemispheric mean O - B, sub-sampled at different radiosonde sites, is shown for ATMS in Fig. 5.7. As these results show, the sampling of the GRUAN network, even the depleted 2016 sample, appears good enough to capture the ATMS global and hemispheric biases. The reason for this is that the ATMS biases have less geographical variation in comparison to AMSU-A (as can be seen in Fig. 5.5).

In order to evaluate the geographical capabilities of GRUAN for assessing these biases, we have plotted the locations of the GRUAN sites on maps of O - B (averaged in a 1 x 1 degree grid), shown in Fig.s 5.8 – 13 for selected channels of MetOp-B AMSU-A and Suomi-NPP ATMS (left plots). We have also plotted the mean (O - B) for each one-degree grid point as a function of latitude, and indicated the points belonging to the sub-samples for the full GRUAN locations and the GRUAN 2016 sites. These are shown alongside the maps in Figs. 5.8 – 13 (right plots). Note that all points are satellite instrument O - B values including the

values sub-sampled at radiosonde sites, shown with the blue and green crosses in the right-hand plots.

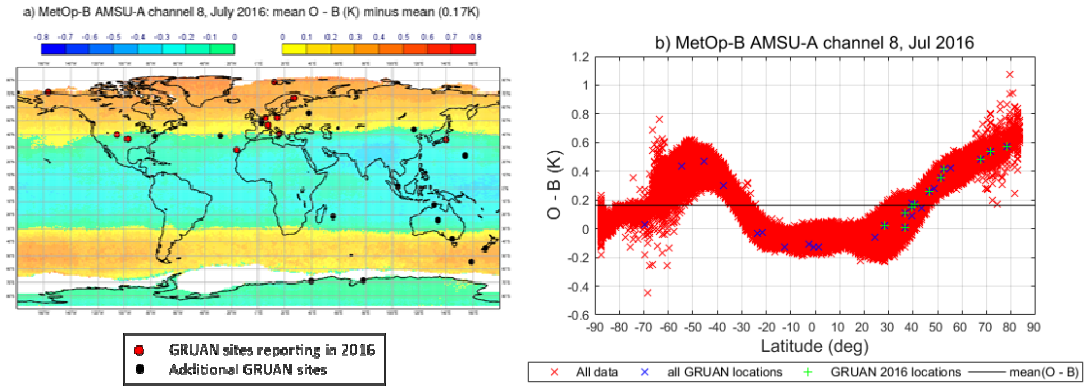


Figure 5.8. a) Map of the 1-month mean (O - B) minus global bias for MetOp-B AMSU-A channel 8 in July 2016 with GRUAN sites superposed. All GRUAN sites are shown, with those reporting in 2016 shown in red and additional sites in black. b) The mean O - B values from a) shown as a function of latitude, with data sub-sampled at all GRUAN and GRUAN in 2016 also shown.

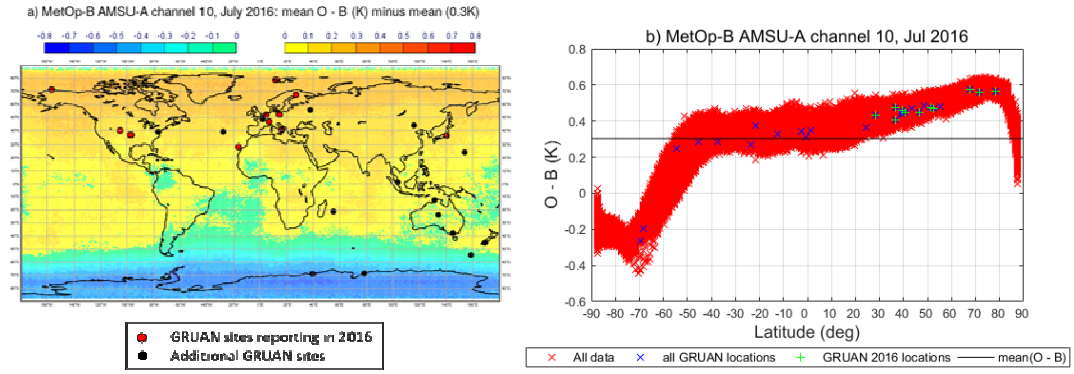


Figure 5.9. As Fig. 5.8 but for MetOp-B AMSU-A channel 10.

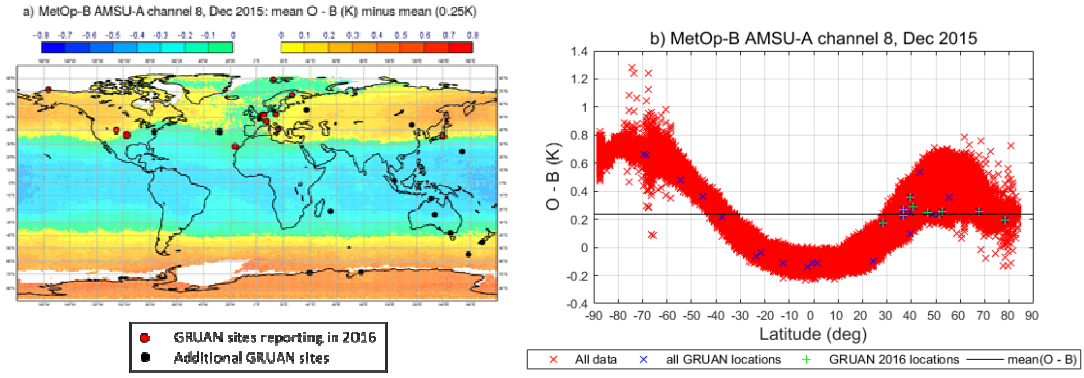


Figure 5.10. As Fig. 5.8 but for December 2015.

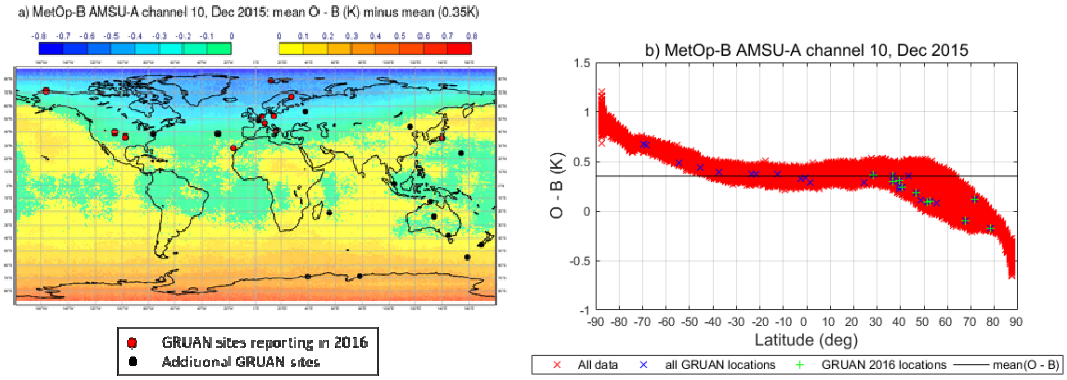


Figure 5.11. As Fig. 5.8 but for MetOp-B AMSU-A channel 10 in December 2015.

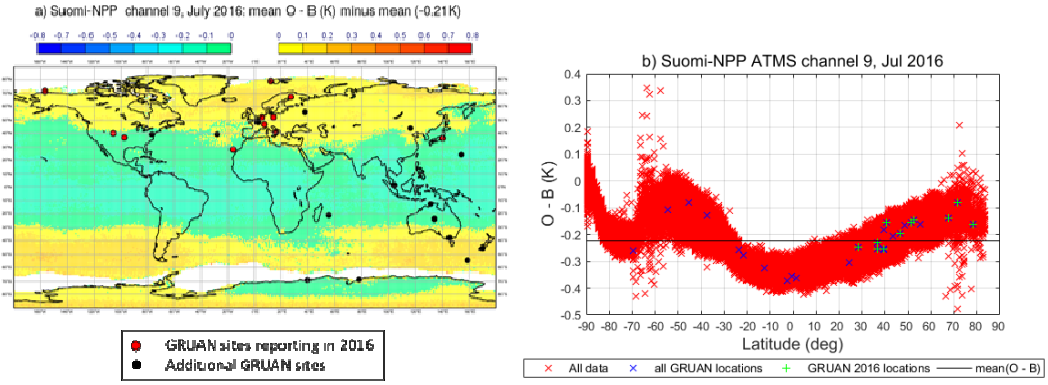


Figure 5.12. As Fig. 5.8 but for Suomi-NPP ATMS channel 9, shown for July 2016.

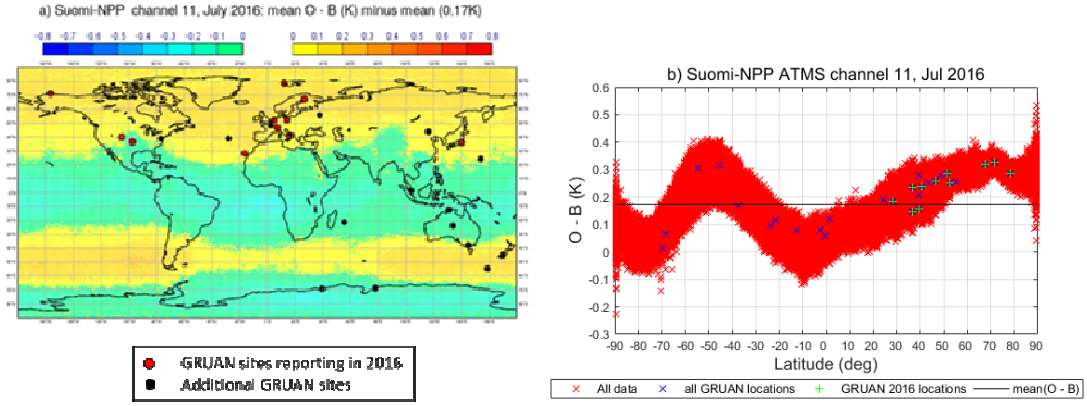


Figure 5.13. As Fig. 5.8 but for Suomi-NPP ATMS channel 11, shown for July 2016.

As Fig.s 5.8 – 13 show, the full GRUAN network samples the main zonal geographical variation in biases well for most channels, particularly the variations in the tropics and Northern Hemisphere. For some channels, there are sharp changes in bias with latitude for latitudes south of about 50°S, which are not very well captured by the GRUAN stations, however. These may well come from systematic uncertainties in the background or radiative transfer models, but instrument uncertainties are also possible (e.g. non-linearity uncertainties). This occurs in the Southern Hemisphere winter, and is particularly true for AMSU-A channels 10 - 11, as can be seen in Fig. 5.9 for channel 10 for example. It is also interesting to notice that these biases over the South Pole are different between the equivalent channels of AMSU-A and ATMS. This indicates an error in one of the instruments,

but we cannot know which one without additional information such as from GRUAN. Extending the GRUAN network in the Southern Hemisphere below 50°S could help in diagnosing this. This might be difficult, however, as there are very few operational radiosondes in the Southern Hemisphere due to the sparsity of land (see e.g. Fig. 5.2). The only RAOB stations below 50°S which are not planned to be included in the GRUAN network are some of those on Antarctica and one site on the Falkland Islands. It would be useful to add any of these into the GRUAN network (preferably with a spread of latitudes). There are also areas of high bias for AMSU-A channels 7 and 8 (peaking around the tropopause) in the Northern Hemisphere, during the Northern Hemisphere winter (December 2015). These areas can be seen in Fig. 5.10a, just south of Alaska and South of North-Eastern Russia. As shown in Fig. 5.2, these areas do have RAOB stations, however, and so it could be useful to add these into the GRUAN network.

While the full GRUAN network appears to provide adequate sampling of most geographical areas for the present purpose, the stations reporting in 2016 are only in the Northern Hemisphere and so their sampling is more limited. They do appear to sample most of the geographical bias changes in the Northern Hemisphere (with the exception of AMSU-A channels 7 - 8 in December 2015). However, without sites in the tropics nor Southern Hemisphere, the larger, inter-hemisphere bias differences cannot be sampled.

2.5.5 O - B statistics for MHS and ATMS humidity-sounding channels

Like AMSU-A, the MHS instruments have global and geographically varying biases. The global biases are of a similar order of magnitude to AMSU-A. These are shown in Fig. 5.14 for data over ocean and after cloud screening and quality control. As this Figure shows, the instruments all have a similar inter-channel bias pattern and the global biases do not vary greatly by season. The similar inter-channel biases between all 183 GHz channels has been the subject of some scientific interest recently. The possible sources of systematic error that could lead to such a bias have been reviewed by Brogniez et al (2016), but the main source of this bias has not yet been identified.

The geographical biases for MHS are quite different to those of AMSU-A. Fig. 5.17 shows maps of mean O - B for the MetOp-B MHS instrument for all channels and two seasons. These bias patterns are very similar between instruments. Broadly speaking the biases show differences between the different hemispheres, but the magnitude of variation is generally smaller than for AMSU-A. For channel 3, the biases appear to be warmer in the tropics than in the extra-tropics, and the winter hemisphere appears to be colder than the summer hemisphere for both channels 4 and 5. The latter bias appears over ocean only and is strongest for the lowest peaking channel. It may therefore be due to systematic uncertainties in the skin temperature used for the surface emission. It would be useful to verify that there are no systematic humidity uncertainties in the background, however, using the GRUAN sondes. For this, sites in the different hemispheres are needed. 'This could be done with the future expanded GRUAN network, using island sites such as La Reunion, Macquarie Island, Tenerife, and Graciosa (future sites) along with the site of Ny Alesund (current GRUAN site).

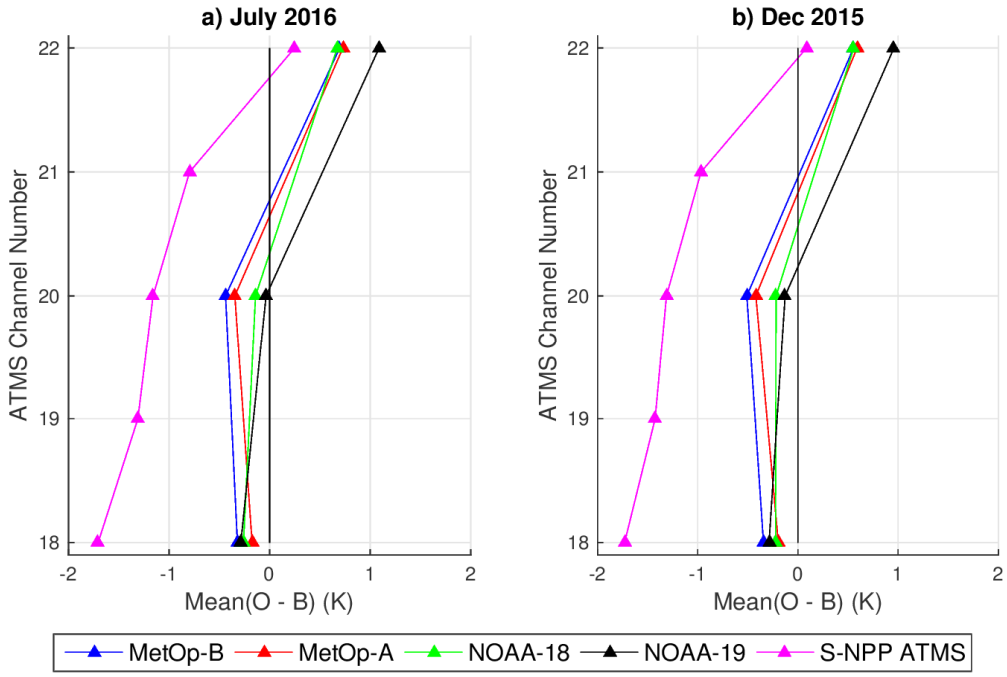


Figure 5.14. Global mean O - B for MHS and ATMS humidity sounding channels, shown as a function of ATMS channel number, and for a) July 2016 and b) December 2015 (1-month averaged values).

The MHS and ATMS O - B values were sub-sampled at GRUAN and RAOB sites and the means calculated for these sub-sampled datasets. The results are given in Fig. 5.15 for MetOp-B MHS and Fig. 5.16 for ATMS, but results are similar for all instruments. Note that we have excluded the RAOB sites over South America and Africa since there are localised biases over this area, which are likely due to localised systematic forecast uncertainties in humidity (as described in the following). As this plot shows, the current GRUAN network is enough to sample these global inter-channel biases and also some inter-satellite biases, even when the smallest number of sites is included.

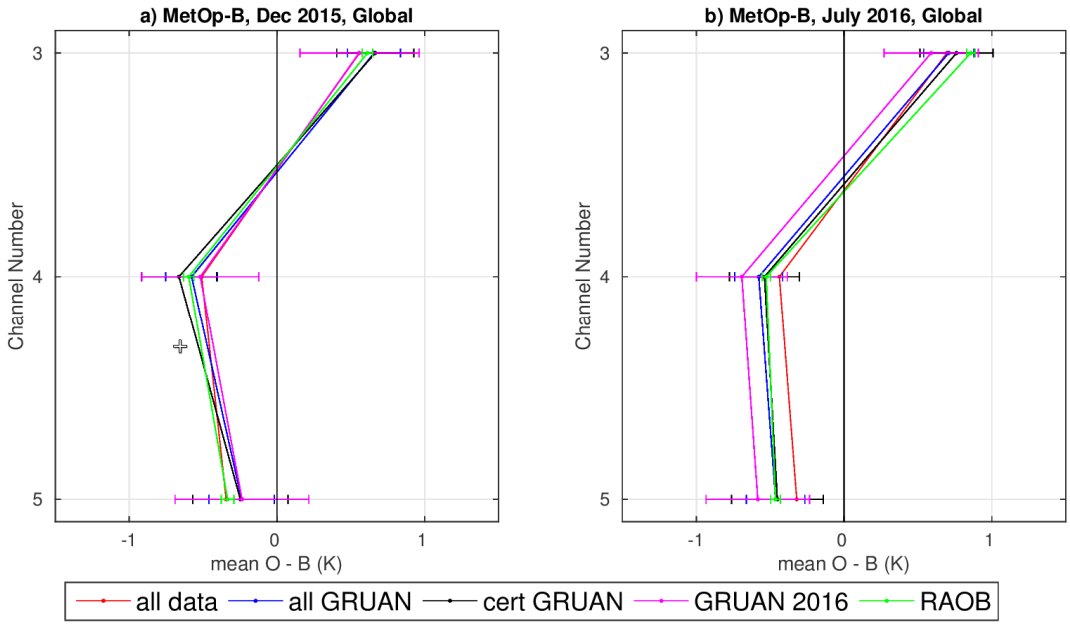


Figure 5.15. Global mean O - B for MetOp-B humidity sounding channels, shown as a function of ATMS channel number, and for a) July 2016 and b) December 2015 (1-month averaged values). The means are shown for all data after quality control, and sub-sampled at different radiosonde sites, including all GRUAN sites, certified GRUAN sites, GRUAN sites reporting in 2016 and all RAOB sites.

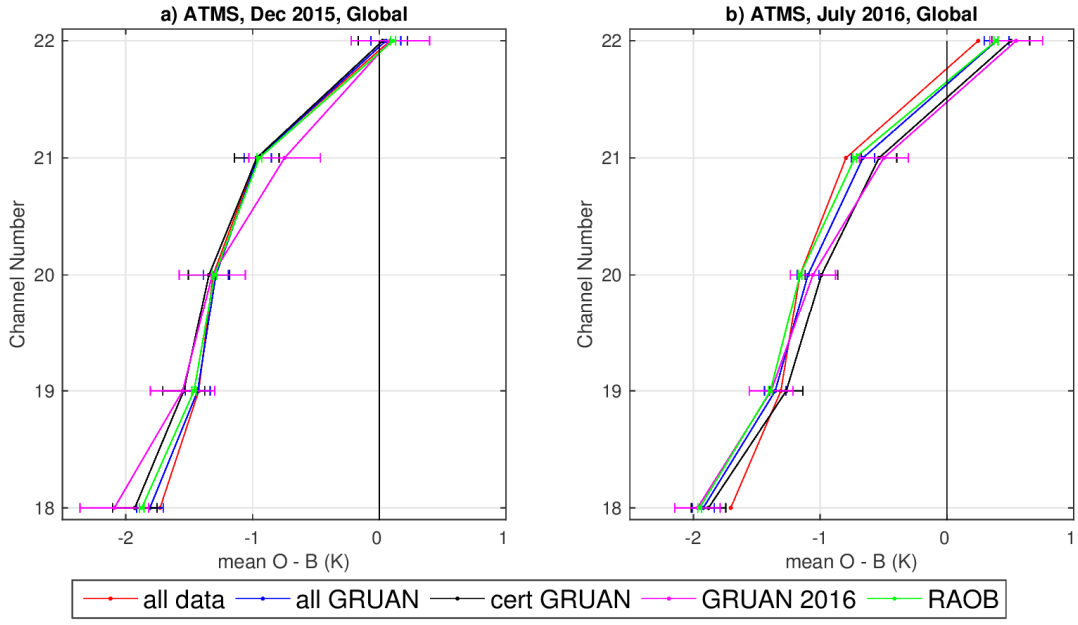


Figure 5.16. As Fig. 5.15 but for ATMS humidity sounding channels.

As Fig. 5.17 shows, there are also smaller scale geographical bias features such as the bands of blue and red around the tropics for channels 4 and 5, and over South America and Africa for channels 3 and 4. The bands of strong bias over the ocean in the tropics are most likely to be errors in the cloud and precipitation forecasts that have not been removed by the cloud screening. This can occur because the latter screen removes scattering features only, missing

areas of liquid cloud. The features over South America and Africa could also be cloud and precipitation features missed by the scattering index. However, another possibility is that there are systematic uncertainties in the humidity forecast occurring in areas of deep convection. This latter explanation is more likely since the effect is less visible for channel 5 which, as the lowest peaking channel, should be the most sensitive to cloud and precipitation. It would be interesting to analyse this bias with GRUAN radiosondes, to confirm that it is indeed a systematic error in the humidity forecast. To do this, RAOB sondes over South America and/or Africa would need to be added to the network. However, this is not necessary for the cal/val of these satellite instruments since this cannot be an instrument uncertainty.

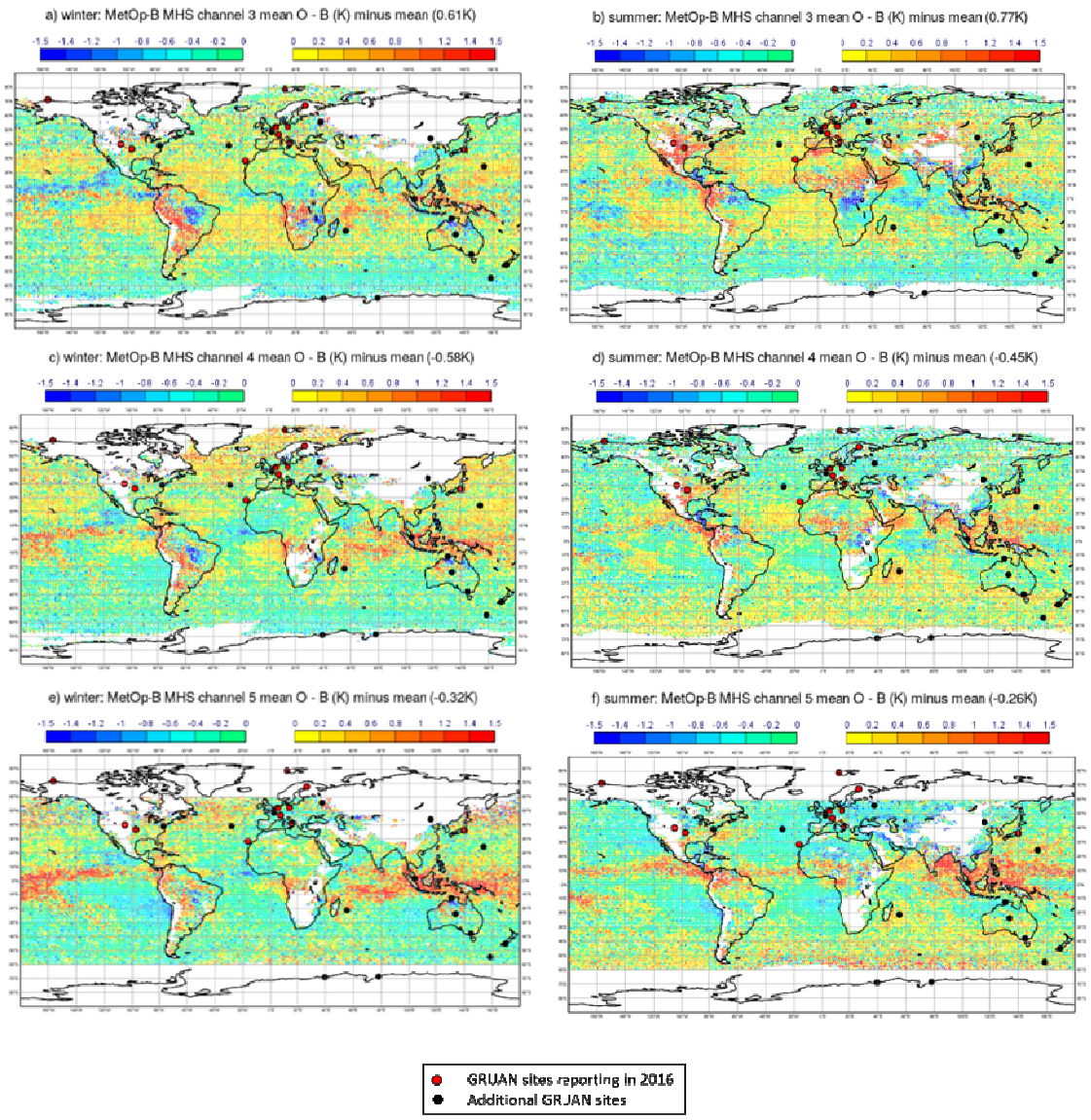


Figure 5.17. Maps of mean O - B minus the global averages for MetOp-B MHS, averaged for one month of data in July 2016 (left plots) and December 2015 (right plots). Maps are shown for the MHS sounding channels (3 - 5). The locations of the GRUAN and RAOB stations are also shown on the maps.

2.5.6 Results and Recommendations

In this study, we have analysed the global and geographical biases of AMSU-A, ATMS, and MHS temperature and humidity sounding observations minus the ECMWF short-range forecasts, in radiance space. These show a strong air-mass structure for AMSU-A temperature-sounding channels and a smoother geographical structure for MHS and ATMS humidity-sounding channels. Global biases are observed for all channels of magnitude up to 0.5 K.

We examined the locations of the GRUAN sites with respect to these biases and also sub-sampled the global O - B statistics at GRUAN sites. This gives an idea of the geographical capabilities of the GRUAN network for analysing the residual O - B biases in microwave temperature and humidity sounding satellite instruments. The biases considered were those where it can be difficult to know whether there is a contribution due to systematic errors in the forecast model. There are other types of bias not considered here where the use of NWP background fields is clearly able to identify instrument-related problems, such as solar intrusions, assymetric scan-angle biases, geolocation errors, etc.

The results indicate that the current GRUAN network, with data available on NOAA's ftp in 2016, should be able to capture the global mean O - B of MHS and ATMS and the Northern Hemisphere mean O - B of the temperature-sounding channels of AMSU-A. However, for channels 9 - 12 of AMSU-A, the sub-sample does not accurately represent the global biases - the mean O - B values for the GRUAN sub-sample are approximately 0.3 - 0.4 K colder than the full global biases. In order to analyse the global biases of these channels, and also the broad geographical structure of the AMSU-A and MHS biases, more stations are needed in the tropics and the Southern Hemisphere.

In the future, the GRUAN network is supposed to be expanded to other latitudes and these locations were also examined. Given the larger spread in latitudes, this increased coverage should help in the analysis of the geographical biases for AMSU-A and MHS. A greater variation in latitude is needed for AMSU-A when compared to MHS, since the bias structure is more geographically variable for AMSU-A. This study supports the future expansion of the GRUAN sites.

For some channels of AMSU-A, particularly 10 - 11, the biases change rapidly with latitude south of around 50°S. Radiosonde sites south of 50°S are highly desirable for analysing the source of these biases. Such sites would need to have radiosonde launches which reach a height of approximately 10 – 20 hPa, since these channels are sensitive to atmospheric heights of around 500 – 20 hPa. Additionally, strong positive biases in the Northern Hemisphere for AMSU-A channels 7 and 8 are not sampled by GRUAN in areas just south of Alaska and North-Eastern Russia. Therefore, besides the planned expansion of GRUAN sites, additional sites south of 50°S and just south of Alaska and North-Eastern Russia in the Northern Hemisphere would be desirable.

In this study, we have concentrated on O - B values for satellite data, and not considered those for radiosondes. A possible extension to the work could be a comparison of the

GRUAN temperature and humidity O - B values to those of the full global radiosonde network. Ingleby (2017) has recently analysed O - B statistics for the full radiosonde network assimilated at ECMWF and has identified commonalities in the mean O - B values, particularly for temperature. It would be interesting to examine the O - B values of the GRUAN data to see if they also show the same biases captured by the full radiosonde network in radiance space.

Following the study presented in this report, we aim to use the GRUAN processor in future work to assess the geographical and global biases of AMSU-A and MHS. For the assessment of AMSU-A in particular, it would be useful to include additional stations, as well as the current GRUAN sites. One option is to use RS92 operational radiosonde data and estimate the observation uncertainty as an average of the values given for GRUAN data. Such an approach is being used by the recently instigated Copernicus Climate Change Service C3S 311a Lot 3 contract, which is being undertaken by GAIA-CLIM partners.

3. Overall Conclusions and Recommendations

In this report we have addressed gaps in the current ground-based observing system for satellite validation, based on a range of modelling approaches. The individual modelling and analyses approaches differ because:

- Different scientific questions need to be addressed per ECV (e.g. trend versus diurnal, column versus profile).
- Available observational constraints differ in both coverage and characteristics per ECV.
- Variations exist in modelling capabilities and available diagnostics tools.

The ECVs and gaps covered by the model-based gap analysis are summarized in Table 2. Clearly, most of the analysis focuses on the spatiotemporal aspects, as these are the questions that can be directly assessed with models. None of the analyses address the technical, governance, or uncertainty-related gaps which instead are addressed in remaining work packages and the GAID[2] process.

Table 2. A summary of the gap analysis presented in this deliverable, with respect to the gap types and the ECVs from Section 3. Here the spatiotemporal gap is separated into the spatial and temporal components.

	Spatial coverage	Temporal coverage	Vertical resolution	Uncertainty (per observation)	Uncertainty (wrt. comparator measures, representativeness)	Technical (formats, etc.)	Governance (harmonisation procedures, etc.)	Parameter (need for auxiliary information/parameters)
CO_2	■				■			
CH_4					■			
CO	■	■	■					
AOD	■	■						
O_3	■		■					■
Temperature	■							
Humidity	■							

In the following we summarize the main results obtained and formulate specific recommendations:

[2] http://www.gaia-clim.eu/page/gaid

Greenhouse gases:

- The coverage of the TCCON network in geographic and parameter space was assessed using OCO-2 measurements for the years 2015-2016 (in this deliverable), and GOSAT and SCIAMACHY measurements for 2010 (in D1.5).
- Despite the 2017 addition of a station in Burgos, Philippines (offset by the removal of the Amazon station at Manaus), a gap in the tropics remains.
- Ocean glint measurements are poorly constrained polewards of 50° south and in the northern hemisphere in winter. Here regular profiles with AirCore or aircraft would be helpful, particularly in the winter.
- The processing time and/or embargo time of the TCCON data can create a temporary temporal gap in the validation network, of particular importance for emerging missions or when investigating anomalous satellite results.
- In terms of parameters space, gaps were identified at high-albedo and high-aerosol-load sites (for GOSAT), and specifically for OCO-2 for low values of co2_grad_del (i.e., very large and unphysical variation in the retrieved vertical profile of CO_2 from that assumed in the prior.), and for high values of dP (the difference between the retrieved and the *a priori* surface pressure). Conservative quality filtering likely underestimates these gaps.

Recommendation: Potential expansion of the TCCON network should prioritize (in terms of spatial coverage) the tropics and high southern latitudes.

Recommendation: Potential expansion of the TCCON network should prioritize (in terms of parameter coverage) high albedo sites, locations with high fine-mode aerosol load, and regions with large deviations in dP and co2_grad_del (the latter two specific to OCO-2).

Recommendation: A sustainable solution to long-term operation of TCCON sites would alleviate the temporary temporal gaps in the validation network as a result of data embargoes, and would ensure the continuity of the existing network going forward.

Air quality, carbon monoxide:

- The added value of profile information of carbon monoxide (CO) from NDACC sites for providing additional constraints on CO sources and OH concentrations has been explored in an inverse modeling study.
- The source inversion adopting the lowest average OH abundance in the Northern Hemisphere provides the best overall match with available CO observations.
- The FTIR observations however, do not provide a clear picture: the FTIR profiles currently do not provide sufficient information to discriminate between different CO budgets, suggesting that model/data discrepancies for the most part cannot be

traced to uncertainties in CO sources and sinks. Further improvements to models and/or observations are required to close the CO budget.

- Evaluation of FTIR profiles using independent high-quality vertically resolved CO measurements shows that the model/data discrepancies can be attributed for a substantial part to potentially erroneous FTIR profiles.

Recommendation: priority should be given to strengthening the quality and robustness of the FTIR profile retrievals. Additional measurements (either more frequent measurements at existing sites or FTIR retrievals at additional sites) without improvements in the FTIR measurement understanding and processing will most likely not contribute to narrowing the vertical CO information gap.

Recommendation: In line with the GAIA-CLIM Gaps Analysis and Impacts Document, we suggest a sustained program of profiles with AIRCORE co-located with FTIR measurements would help to better characterize FTIR profiles.

Aerosols:

- MODIS AOD correlation lengths vary from 100-200 km over populated continental areas to 1500 km in remote ocean or desert areas
- Global climate-aerosol models are not yet good enough to reproduce measured diurnal cycles or spatial correlations

Recommendation: Take local correlation structure into account when filling gaps between AERONET stations with satellite data.

Recommendation: Carry out long-term, in-depth analysis and development of aerosol transport, chemistry, and removal in models beyond simply validating weather parameters.

Ozone:

- Projected ozone changes are in the range of a few per cent for total ozone, with changes in the stratospheric column and the tropopspheric column of similar absolute size. Changes in tropospheric ozone thus can partly mask changes in stratospheric ozone, which is an important climate variable. This demonstrates the challenges for a ground-based network for the purpose of validating long-term satellite ozone records.
- From the model simulations, we find that the long-term changes in tropical stratospheric ozone are sufficiently well represented by the existing stations of the reference network. This is more critical for tropical tropospheric ozone, where single stations are less representative for the tropics as a whole owing to much smaller correlation scales of tropospheric ozone. In general a denser global network is required if we wish to diagnose tropospheric ozone from in-situ observational techniques.

- A statistical analysis from the model simulations supports the observational driven statistical analysis of small scale ozone variations (GAIA-CLIM report D1.10) and offers the potential to extend this analysis to regions were there are not enough observational data available for a statistical analysis.

Recommendation: Maintaining a ground-based network in the tropics to observe vertically resolved ozone is essential to characterize long-term ozone changes, in particular for the validation of nadir viewing satellite data that either provide only total columns or have only broad weighting functions/averaging kernels.

Temperature and humidity:

- The current GRUAN site locations are able to help diagnose the sources of some types of O - B bias for some instruments including: global MHS biases, global ATMS biases and the Northern Hemisphere biases of AMSU-A
- The planned expansion of the GRUAN network is needed for diagnosing the source of geographical bias in AMSU-A and MHS O - B values and this study supports that expansion. Sites in Antarctica are of particular importance for diagnosing O - B biases in AMSU-A, since there are strong biases over the South Pole for AMSU-A channels 10 - 11 (500 - 20 hPa) in the Southern Hemisphere winter.

Recommentation: Further expansion of the GRUAN network to sites in the Southern Hemisphere, at latitudes below 50S (capturing temperature at heights of 500 - 20 hPa), and sites just south of North-Eastern Russia and Alaska in the Northern Hemisphere (capturing temperature at heights of 1000 - 50 hPa), would be desirable. These constitute areas of high apparent bias in satellite measurements but no GRUAN sites are foreseen that could help to characterise and confirm such biases.

4. References

Agustí-Panareda, A., Massart, S., Chevallier, F., Boussetta, S., Balsamo, G., Beljaars, A., Ciais, P., Deutscher, N. M., Engelen, R., Jones, L., Kivi, R., Paris, J.-D., Peuch, V.-H., Sherlock, V., Vermeulen, A. T., Wennberg, P. O., and Wunch, D.: Forecasting global atmospheric CO_2, Atmos. Chem. Phys., 14, 11959-11983, doi:10.5194/acp-14-11959-2014, 2014.

Agustí-Panareda, A., Massart, S., Chevallier, F., Balsamo, G., Boussetta, S., Dutra, E., Beljaars, A.: A biogenic CO_2 flux adjustment scheme for the mitigation of large-scale biases in global atmospheric CO_2 analyses and forecasts, ECMWF Technical Memorandum 773, https://www.ecmwf.int/en/elibrary/15374-biogenic-c02-flux-adjustment-scheme-mitigation-large-scale-biases-global, 2015.

Agustí-Panareda, A., Diamantakis, M., Bayona, V., Klappenbach, F., Butz, A.: Improving the inter-hemispheric gradient of total column atmospheric CO2 and CH4 in simulations with the ECMWF semi-Lagrangian atmospheric global model, Geosci. Model Dev., 10, 1-18, doi:10.5194/gmd-10-1-2017.

Anderson, D. C., Loughner, C. P., Diskin, G., Weinheimer, A., Canty, T. P., Salawitch, R. J., Worden, H. M., Fried, A., Mikoviny, T., Wisthaler, A., and Disckerson, R. R.: Measured and modeled CO and NO_y in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US, Atmos. Environ., 96, 78-87, 2014.

Arola, A., Eck, T.F., Huttunen, J., Lehtinen, K.E.J., Lindfors, A.V., Myhre, G., Smirnov, A., Tripathi, S.N., Yu, H.: Influence of observed diurnal cycles of aerosol optical depth on aerosol direct radiative effect, Atmos. Chem. Phys., 13, 7895-7901, 2013.

Barth, M., Cantrell, C. A., Brune, W. H., et al.: The Deep Convective Clouds and Chemistry (DC3) field campaign, Bull. Am. Met. Soc., 96, 1281-1309, 2015.

Bauwens, M., Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf,, G. R., Wiedinmyer, C., Kaiser, J., Sindelarova, K., and Guenther, A.: Nine years of global hydrocarbon emissions based on source inversion of OMI formaldehyde observations, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-221, 2016.

Bauer, P., Moreau, E., Chevallier, F., and O'Keeffe, U.: Multiple-scattering microwave radiative transfer for data assimilation applications. Quart. J. Roy. Meteorol. Soc., 132, 1259 – 1281, 2006.

Bell W, Candy B, Atkinson N, Hilton F, Baker N, Bormann N, Kelly G, Kazumori M, Campbell W, Swadley S. 2008. The assimilation of SSMIS radiances in numerical weather prediction models. IEEE Trans. Geoscience. Rem. Sensing. 46: 884 – 900.

Bergman, T., Kerminen, V.-M., Korhonen, H., Lehtinen, K. J., Makkonen, R., Arola, A., Mielonen, T., Romakkaniemi, S., Kulmala, M., and Kokkola, H.: Evaluation of the sectional aerosol microphysics module SALSA implementation in ECHAM5-HAM aerosol-climate model, Geosci. Model Dev., 5, 845–868, 2012.

Bousquet, P., D. A. Hauglustaine, P. Peylin, C. Carouge, and P. Ciais, Two decades of OH variability as inferred by an inversion of atmospheric transport and chemistry of methyl chloroform, Atmos. Chem. Phys., 5, 2635-2656, 2005.

Brogniez, H., English, S., Mahfouf, J.-F., Behrendt, A., Berg, W., Boukabara, S., Buehler, S. A., Chambon, P., Gambacorta, A., Geer, A., Ingram, W., Kursinski, E. R., Matricard, M., Odintsova, T., Payne, V.,

Thorne, P., Tretyakov, M., Wang, J.: A review of sources of systematic errors and uncertainties in observations and simulations at 183 GHz, Atmos. Meas. Tech., 9, 2207 – 2221, 2016.

Clerbaux, C., Boynard, A., Clarisse, L., et al.: Monitoring of atmospheric composition using the thermal infrared IASI:MetOp sounder, Atmos. Chem. Phys., 9, 6041-6054, 2009.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Sweeney, C., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: The MOPITT Version 6 product: algorithm enhancements and validation, Atmos. Meas. Tech., 7, 3623-3632, doi:10.5194/amt-7-3623-2014, 2014.

Dils, B., De Mazière, M., Müller, J.-F., et al.: Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH4, CO2 and N2O, Atmos. Chem. Phys., 6, 1953-1976, 2006.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., Vomel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463 – 4490, 2014.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics, Atmos. Chem. Phys., 11, 3303-3318, 2011.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., Wimmer, W.: The operational sea surface temperature and sea ice analysis (OSTIA) system, Rem. Sens. Env., 116, 140 – 158, 2012.

Errera, Q., Braathen, G., Chabrillat, S., Christophe, Y., Santee, M., and Skachko, S.: BRAM: A reanalysis of Aura MLS chemical observations by the Belgian Assimilation System for Chemical Observations (BASCOE), presented at European Geophysical Union (EGU) General Assembly, Vienna, 2017. see also strato.aeronomie.be/index.php/2-uncategorised/6-bram

Fortems-Cheiney, A., Chevallier, F., Pison, et al.: On the capability of IASI measurements to inform about CO surface emissions, Atmos. Chem. Phys., 9, 8735-8743, 2009.

Geer A, Bauer P, Bormann N. 2010. Solar biases in microwave imager observations assimilated at ECMWF. IEEE Trans. Geoscience. Rem. Sensing. 48: 2660–2669.

Geer, A. J., Baordo, F., Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies, Atmos. Tech. Meas., 7, 1839 – 1860, 2014a.

Geer, A., Baordo, F., Bormann, N., English, S.: All-sky assimilation of microwave humidity sounders, ECMWF tech. memo., 741, 2014b.

George, M., Clerbaux, C., Hurtmans, D., et al.: Carbon monoxide distributions from the IASI/METOP mission: evaluation with other space-borne remote sensors, Atmos. Chem. Phys., 9, 8317-8330, 2009.

George, M., Clerbaux, C., Bouarar, I., Coheur, P.-F., Deeter, M. N., Edwards, D. P., Francis, G., Gille, J. C., Hadji-Lazaro, J., Hurtmans, D., Inness, A., Mao, D., and Worden, H. M.: An examination of the long-term CO records from MOPITT and IASI: comparison of retrieval methodology, Atmos. Meas. Tech., 8, 4313-4328, doi:10.5194/amt-8-4313-2015, 2015.

Giglio, L., Randerson, J. T., and Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res., 118(1), 317–328, doi:10.1002/jgrg.20042, 2013.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittay a, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended

and updated framework for modelling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, doi:10.5194/gmd-5-1471-2012, 2012.

Guenther, A., Karl, T., Harley, T., Wiedinmyer, C., Palmer, P., and Geron, C., Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, 2006.

Guerlet, S., Butz, A., Schepers, D., Basu, S., Hasekamp, O. P., Kuze, A., Yokota, T., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., Galli, A., and Aben, I.: Impact of aerosol and thin cirrus on retrieving and validating XCO2 from GOSAT shortwave infrared measurements, J. Geophys. Res. Atmos., 118, 4887–4905, doi:10.1002/jgrd.50332, 2013.

Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricard, M., Geer, A., Brunel, P., Vidot, J.: RTTOV v11 users guide, available online at: https://www.nwpsaf.eu/site/software/rttov/rttov-v11/, 2015.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – a federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Hooghiemstra, P. B., Krol, M. C., Bergamaschi, P., et al.: Comparing CO emission estimates using MOPITT or NOAA surface network observations, J. Geophys. Res., 117, D06309, 2012.

Hudman, R. C., Jacob, D. J., Turquety, S., et al.: Surface and lightning sources of nitrogen oxides over the United States: Magnitudes, chemical evolution, and outflow, J. Geophys. Res., 112, D12S05, 2007.

Hudman, R. C., Murray, L. T., Jacob, D. J., et al.: Biogenic versus anthropogenic sources of CO in the United States, Geophys. Res. Lett., 35, L04801, 2008.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., Vomel, H.: Reference quality upper air measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217 – 1231, 2010.

Ingleby, B.: An assessment of different radiosonde types 2015/2016, ECMWF tech. memo., 807, 2017.

Inness, A., Blechschmidt, A.-M., Bouarar, I., et al.: Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS, Atmos. Chem. Phys., 15, 5275-5303, 2015.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jacob, D. J., Crawford, J. H., Maring, H., et al.: The Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) mission: design, execution, and first results, Atmos. Chem. Phys., 10, 5191-5212, 2010.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., et al.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411-11432, 2015.

Jiang, Z., Jones, D. B. A., Worden, J., Worden, H. M., Henze, D. K., and Wang, Y. X.: Regional data assimilation of multi-spectral MOPITT observations of CO over North America, Atmos. Chem. Phys., 15, 6801-6814, doi:10.5194/acp-15-6801-2015, 2015.

Jöckel, P.; Tost, H.; Pozzer, A.; Kunze, M.; et al., Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, Geoscientific Model Development, 9(3), 1153-1200, doi:10.5194/gmd-9-1153-2016, 2016.

Karbou, F., Prigent, C., Eymard, L., Pardo, J.: Microwave land emissivity calculations using AMSU-A and AMSU-B measurements, IEEE Trans. Geosc. Rem. Sens., 43, 948 – 959, 2005.

Kerzenmacher, T., Dils, B., Kumps, N., et al.: Validation of IASI FORLI carbon monoxide retrievals using FTIR data from NDACC, Atmos. Meas. Tech., 5, 2751-2761, 2012.

Klonecki, A., Pommier, M., Clerbaux, C., et al.: Assimilation of IASI satellite CO fields into a global chemistry transport model for validation against aircraft measurements, Atmos. Chem. Phys., 12, 4493-4512, 2012.

Kopacz, M., Jacob, D. J., Fisher, J. A., et al.: Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmos. Chem. Phys., 10, 855-876, 2010.

Lamarque, J.-F., Bond, T. C., Eyring, V., et al.: Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017-7039, 2010.

Lawrence, H., Bormann, N.: Scene-dependent observation errors for the assimilation of AMSU-A, EUMETSAT/ECMWF fellowship programme research report, 35, 2015.

Lonitz, K., Geer, A.: New screening of cold-air outbreak regions used in 4D-Var All-sky assimilation, EUMETSAT/ECMWF fellowship programme research report, 35, 2015.

Lu Q, Bell W. 2014. Characterizing Channel Center Frequencies in AMSU-A and MSU Microwave Sounding Instruments. J. Atmos. Ocean. Tech. 31: 1713-1732.

Lu Q, Lawrence H, Bormann N, English S, Lean K, Atkinson N, Bell W, Carminati F. 2015. An evaluation of FY-3C satellite data quality at ECMWF and the Met Office. ECMWF Tech. Memo. 767.

Lupu, C., Geer, A. J.: Operational implementation of RTTOV-11 in the IFS, ECMWF Tech. Memo., 748, 2015a.

Lupu, C., Geer, A. J., Bormann, N.: Revision of the microwave coefficient files in the IFS, ECMWF Tech. Memo., 749, 2015b.

Lupu, C., Geer, A., Bormann, N., English, S.: An evaluation of radiative transfer modelling errors in AMSU-A data, ECMWF Tech. Memo., 770, 2016.

Massart, S., Agusti-Panareda, A., Aben, I., Butz, A., Chevallier, F., Crevoisier, C., Engelen, R., Frankenberg, C., and Hasekamp, O.: Assimilation of atmospheric methane products into the MACC-II system: from SCIAMACHY to TANSO and IASI, Atmos. Chem. Phys., 14, 6139-6158, doi:10.5194/acp-14-6139-2014, 2014.

Massart, S., Agustí-Panareda, A., Heymann, J., Buchwitz, M., Chevallier, F., Reuter, M., Hilker, M., Burrows, J. P., Deutscher, N. M., Feist, D. G., Hase, F., Sussmann, R., Desmet, F., Dubey, M. I., Griffith, D. W. T., Kivi, R., Petri, C., Schneider, M., Velazco, V. A.: Ability of the 4-D-Var analysis of

the GOSAT BESD XCO_2 retrievals to characterize atmospheric CO_2 at large and synoptic scales, Atmos. Chem. Phys., 16, 1653-1671, doi:10.5194/acp-161653-2016, 2016.

Meul, S.; Langematz, U.; Oberländer, S.; et al., Chemical contribution to future tropical ozone change in the lower stratosphere, Atmos. Chem. Phys.,14(6) 2959-2971, doi:10.5194/acp-14-2959-2014, 2014.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NOx emission estimates derived from an assimilation of OMI tropospheric NO2 columns, Atmos. Chem. Phys., 12, 2263-2288, 2012.

Müller, J.-F., and T. Stavrakou, Inversion of CO and NOx emissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5, 1157-1186, 2005.

Müller, J.-F., Stavrakou, T., Wallens, S., et al.: Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environmental model, Atmos. Chem. Phys., 8, 1329-1341, 2008.

Naik, V., A. Voulgarakis, A. M. Fiore, L. W. Howitz, J.-F. Lamarque, M. Lin, et al. Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 5277-5298, 2013.

Nedelec, P., Cammas, J.-P., Thouret, V., et al.: An improved infrared carbon monoxide analyzer for routine measurements aboard commercial Airbus aircraft: technical validation and first scientific results of the MOZAIC III programme, Atmos. Chem. Phys., 3, 1551-1564, 2003.

Novelli, P. C., and Masarie, K. A.: Atmospheric carbon monoxide dry air mole fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1988-2013, version 2014-07-02, available at ftp://aftp.cmdl.noaa.gov/data/trace_gases/co/flask/surface/ (last access: 18/08/2017).

Osterman, G. B., Eldering, A., Mandrake, L., O'Dell, C. W., Wunch, D., Wennberg, P. O., Fisher, B. and Marchetti, Y: OCO-2 XCO2 B8Ar Lite Products, Warn Levels, and Bias Correction, version 2.0, 22/09/2017, available at https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_XCO2_Lite_Files_and_Bias_Correction.pdf (last access: 01/12/2017).

Patra, P. K., M. C. Krol, S. A. Montzka, T. Arnold, E. L. Atlas, B. R. Lintner, B. B. Stephens, et al. Observational evidence for interhemispheric hydroxyl-radical parity, Nature 513, 219-223, 2014.

Pommier, M., Law, K. S., Clerbaux, C., et al.: IASI carbon monoxide validation over the Arctic during POLARCAT spring and summer campaigns, Atmos. Chem. Phys., 10, 10655-10678, 2010.

Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, Geophys. Res. Lett., 39, L09803, 2012.

Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning emissions from small fires, J. Geophys. Res., 117, G04012, 2012.

Riahi K, Gruebler A, Nakicenovic N: Scenarios of long-term socio-economic and environmental development under climate stabilization. Technol Forecast Soc Chang 74(7):887–935, 2007.

Riahi K, Krey V, Rao S, Chirkov V, Fischer G, Kolp P, Kindermann G, Nakicenovic N, Rafai P.: RCP-8.5: exploring the consequence of high emission trajectories. Climatic Change. doi: 10.1007/s10584-011-0149-y , 2011.

Rodgers, C. D., and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 4116, 2003.

Saunders, R., Matricard, M., Brunel, P.: An improved fast radiative transfer model for assimilation of satellite radiance observations, Q. J. Roy. Meteorol. Soc., 125, 1407 – 1425, 1999.

Santee, M. L., Manney, G. L., Livesey, N. J., et al.: A comprehensive overview of the climatological composition of the Asian summer monsoon anticyclone based on 10 years of Aura Microwave Limb Sounder measurements, J. Geophys. Res., 122, 5491-5514, 2017.

Senten, C., De Mazière, M., Dils, B., et al.: Technical note: New ground-based FTIR measurements at Ile de La Réunion: observations, error analysis, and comparisons with independent data, Atmos. Chem. Phys., 8, 3483-3508, 2008.

Spivakovsky, C. M., Logan, J. A., Montzka, S. A., et al.: Three-dimensional climatological distribution of tropospheric OH: Update and evaluation, J. Geophys. Res., 105, 8931-8980, 2000.

Stavrakou, T., and J.-F. Müller, Grid-based vs. big-region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, J. Geophys. Res., 111, D15304, doi:10.1029/2005JD006896, 2006.

Stavrakou, T., Guenther, A., Razavi, A., et al.: First space-based derivation of the global atmospheric methanol emission fluxes, Atmos. Chem. Phys., 11, 4873-4898, 2011.

Stavrakou, T., Müller, J.-F., Bauwens, M., et al.: Substantial underestimation of post-harvest burning emissions in the North China Plain revealed by multi-species space observations, Sci. Rep., 6, 32307, 2016.

Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., Ganzeveld, L., Tegen, I., Werner, M., Balkanski, Y., Schulz, M., Boucher, O., Minikin, A., and Petzold, A.: The aerosol-climate model ECHAM5-HAM, Atmos. Chem. Phys., 5, 1125–1156, 2005.

Toon, O. B., Maring, H., Dibb, J., et al.: Planning, implementation, and scientific goals of the Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS) field missions, J. Geophys. Res., 121, 4967-5009, 2016.

Voulgarakis, A., Naik, V., Lamarque, J.-F., et al.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, Atmos. Chem. Phys., 13, 2563-2587, 2013.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.: observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals, J. Geophys. Res., 115, D18314, doi:10.1029/2010JD014242, 2010.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Phil. Trans. R. Soc. A, 369, doi:10.1098/rsta.2010.0240, 2011.

Zhang, K., O'Donnell, D., Kazil, J., Stier, P., Kinne, S., Lohmann, U., Ferrachat, S., Croft, B., Quaas, J., Wan, H., Rast, S., and Feichter,J.: The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations, Atmos. Chem. Phys., 12, 2012.

List of Acronyms

ACCMIP	The Atmospheric Chemistry and Climate Model Intercomparison Project
AERONET	Aerosol Robotic Network (aeronet.gsfc.nasa.gov)
AIRS	Atmospheric Infrared Sounder
ARCTAS	Arctic Research of the Composition of the Tropopshere from Aircraft and
SatellitesAOD	Aerosol optical depth
BASCOE	Belgian Assimilation Systel fir Chemical Observations
BIRA	Belgian Institute for Space Aeronomy
BRAM	BASCOE Reanalysis of MLS observations
CCMI	Chemistry Climate Modelling Initiative (www.met.reading.ac.uk/ccmi)CTM Chemistry transport model
DC3 or DCCC	Deep Convective Clouds and Chemistry
dof	degrees of freedom of signal
ECHAM-HAMMOZ	Aerosol climate model based on the climate model ECHAM
ECHAM/MESSy	Chemistry climate model based on the climate model ECHAM and the Modular Earth Submodel System
ECMWF	European Centre for Medium Range Weather Forecasts (www.ecmwf.int)
EO	Earth observation
EPA	Environmental Protection Agency
ESRL	Earth System Research Laboratory (NOAA)
FMI	Finnish Meteorological Institute (http://en.ilmatieteenlaitos.fi/)
FORLI	Fast Optimal Retrievals on Layers for IASI
FTIR	Fourier Transform Infra-Red (ground-based) observations
GFED4	Global Fire Emissions Database (version 4)
GAW	Global Atmopsheric Watch
GGGRN	Global Greenhouse Gas Reference Network programme of NOAA
GMD	Geoscientific Model Development (http://www.geoscientific-model-development.net/)
GMD	Global Monitoring Division (part of NOAA)
GOSAT	Greenhouse Gases Observing Satellite
HTAP	Hemispheric Transport of Air Pollution
IAGOS	In-service Aircraft for a Global Observing SystemIASI The Infrared Atmospheric Sounding Interferometer
ICARTT	International Consortim for Atmospheric Research on Transport and Transformation
IGAC	International Global Atmospheric Chemistry project (www.igacproject.org)
IMAGES	Intermediate Model for the Annual and Global Evolution of Species
KIT	Karlsruhe Institute of Technology (www.kit.edu)
KNMI	Royal Netherlands Meteorological Institute (http://www.knmi.nl/over-het-knmi/about)

MACC/CAMS	Monitoring Atmospheric Composition & Climate / Coperrnicus Atmospheric Monitoring System
MCF	Methylchloroform
MEGAN	Model of Emissions of Gases and Aerosols from Nature
MOHYCAN	Biogenic Hydrocarbon Emission Modeling
MOPITT	Measurement of Pollution in the Troposphere
MPI BGC	Max Planck Institute for Biogeochemistry (www.bgc-jena.mpg.de)
NDACC	Network for the Detection of Atmospheric Composition Change (www.ndacc.org)
NMVOC	Non-Methane Volatile Organic Compound
NOAA	National Oceanic and Atmospheric Administration
QA4ECV	Quality Assurance for Essential Climate Variable
RMSD	Root-Mean-Square Deviation
SEAC4RS	Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys
SOA	Secondary Organic Aerosol
TCCON	Total Carbon Column Observing Network
VLNH	Very Low OH in Northern Hemisphere
VLSH	Very Low OH in Southern Hemisphere
VHNH	Very High OH in Northern Hemisphere
VHSH	Very High OH in Southern Hemisphere
VOC	Volatile Organic Compounds
WCRP	World Climate Research Programme (wcrp-climate.org)

Appendix 1. Evaluation of FTIR CO profiles against aircraft campaign data

Arctic sites

The DC-8 measurements of the ARCTAS campaigns conducted in April and July 2008 (Jacob et al., 2010) are compared with the FTIR profiles at Eureka and Thule (Fig. A1). Except for episodic CO plumes, the vertical CO profiles at these two sites should be relatively similar. The comparison with ARCTAS data shows an excellent agreement at Thule in April and to a lesser extent in July, whereas the Eureka profiles underestimate upper tropospheric CO by ca. 30% (40 ppb) in April and July. The statistical variability of CO measured by ARCTAS (red error bars) is larger in the upper troposphere than in the lower layers, but this cannot fully account for the analysed discrepancy.

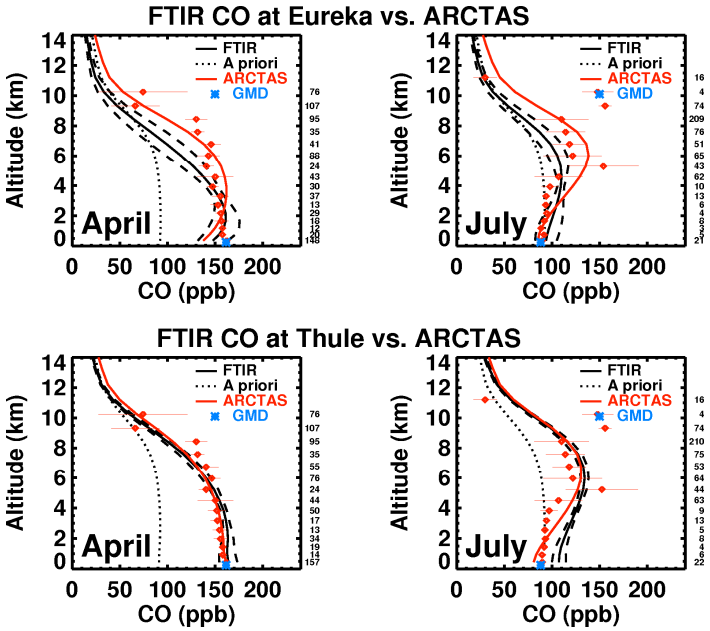


Figure A1. Monthly FTIR profiles at Eureka and Thule (solid lines) and ARCTAS average profile (red symbols and standard deviations) in April (left) and July (right). The red lines are the ARCTAS profiles smoothed by FTIR averaging kernels. Dotted line: FTIR *a priori*. Dashed lines: FTIR confidence interval. The CO concentration measured at the GMD station of Alert is shown as blue symbols. The number of aircraft data per altitude bin is indicated to the right of each plot.

Alpine sites

The FTIR profiles at Jungfraujoch and Zugspitze are evaluated using the CO measurements conducted between 2000 and 2015 from commercial aircraft of the IAGOS programme (http://www.iagos.org; data access : http://iagos.sedoo.fr) in the vicinity of the two stations (6-13 E, 46-48.5 N, see Fig. 2.13), to and from the airports of Zürich and München (see Figs. A2 and A3). The uncertainty of IAGOS measurements is ca. 5 ppb (Nedelec et al., 2003). The two stations are 250 km apart and both located at high altitude (3.0-3.6 km) and are therefore minimally impacted by boundary layer pollution. Given the long averaging period and the large number of data involved in the averages (~500 profiles per season at Jungfraujoch, ~2000 per season at Zugspitze, several thousand IAGOS measurements per season), the large differences found between the FTIR profile shapes at the two sites are

unexpected. Whereas the profile at Zugspitze generally presents only a weak negative gradient in the troposphere (and even a weak positive gradient in JJA), the Jungfraujoch retrievals show a strong negative gradient, especially in DJF and MAM, with mixing ratios about 50 ppb lower at 9 km altitude than at the station altitude. The Zugspitze profiles are perfectly consistent with IAGOS measurements (after application of the averaging kernels) in DJF and MAM, and show some low-level underestimation (8%) in the lower levels in JJA and SON. This bias appears to be statistically non-significant (the IAGOS standard deviation ≈13% at these levels), but the good agreement of IAGOS data with surface measurements at the nearby site of Sonnblick lends credence to the representativity of IAGOS data for this region.

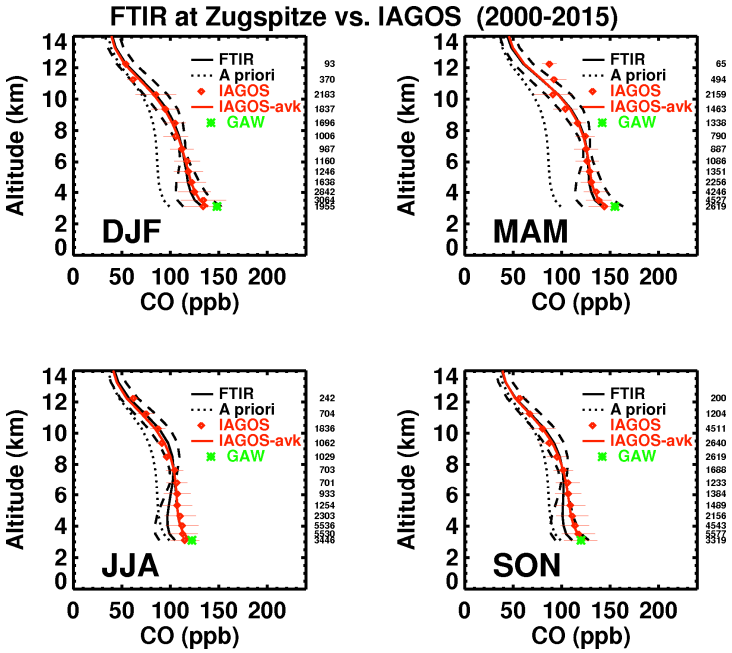


Figure A2. Seasonal FTIR profiles at Zugspitze (solid lines) and IAGOS corresponding profiles (red symbols). The red lines are the IAGOS profiles smoothed by FTIR averaging kernels. Dotted line: FTIR *a priori*. Dashed lines: FTIR confidence interval. The CO concentrations at the GAW station of Sonnblick are shown as green symbols.

At Jungfraujoch, the strong negative gradient of the FTIR retrieval is contradicted by the IAGOS measurements during all seasons but especially in spring, with a low-level overestimation exceeding largely the standard deviation of the IAGOS data. The local GAW mixing ratios at Jungfraujoch are even lower than in the IAGOS data, although still within their statistical variability.

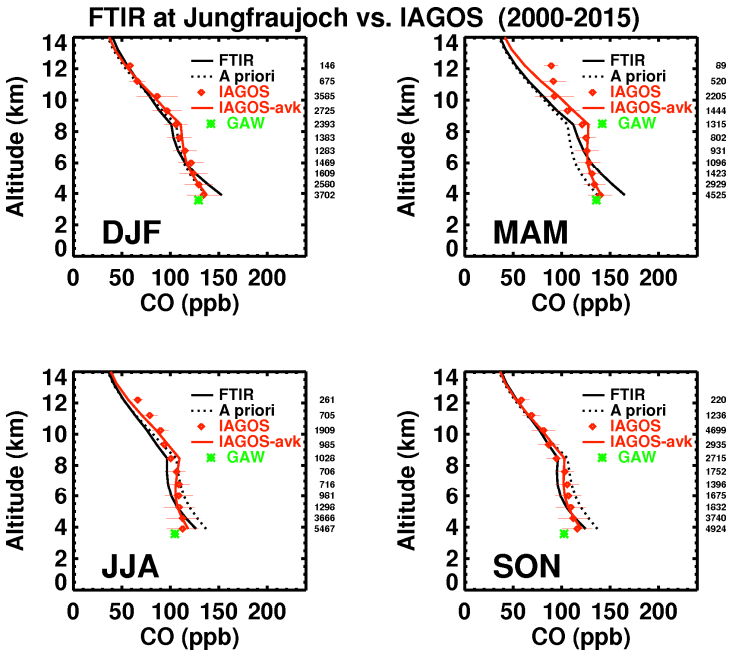


Figure A3. As Fig. A2, for Jungfraujoch. The in situ concentrations at Jungfraujoch are shown in green.

Bremen

The FTIR profiles at Bremen (Fig. A4) are in fairly good agreement with the IAGOS data (averaged between 2008 and 2016 over 50.5-54.5 N, 4-12.5 E, to and from the airports of Amsterdam, Bremen, Hamburg, Brussels, Köln, Düsseldorf and Parchim). The total number of IAGOS measurements is very large (>200,000), but most are from 2015, and the number of FTIR profiles per season for which IAGOS data is available ranges between 45 (DJF) and 94 (MAM). The comparison shows a clear underestimation of FTIR CO in the free troposphere, especially in winter (12% bias) but also in all other seasons (11% in JJA, 9% in SON). There is very little bias near the surface. As could be expected, both the FTIR measurements (in a city) and the IAGOS boundary layer data (near airports) overestimate by up to 50 ppb the surface mixing ratio measured at the nearby GAW station of Neuglobsow, a rural site 300 km away, and therefore cannot be considered typical of a larger area around the site. The systematic error of the FTIR retrieval (dashed lines on Fig. A4) is largely overestimated for this site.

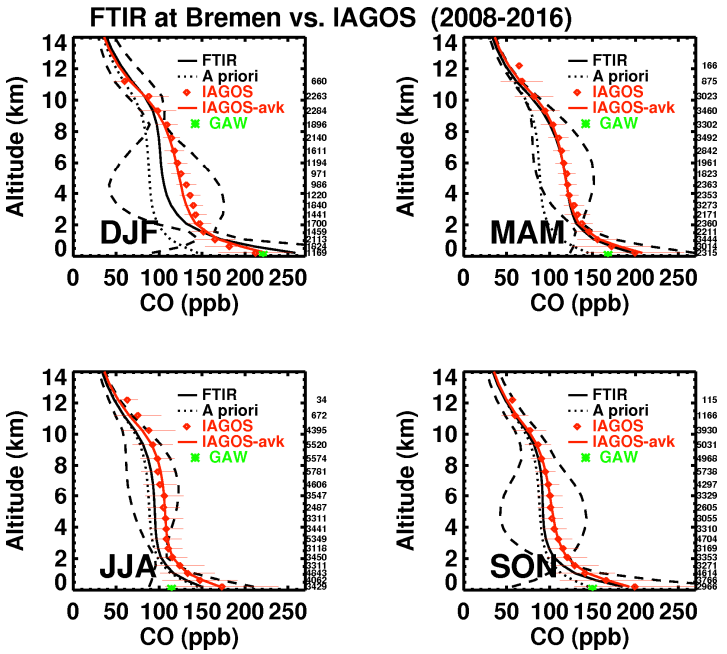


Figure A4. As Fig. A2, for Bremen. The in situ concentrations at Neuglobsow (GAW) are shown in green.

Toronto

The comparisons of Toronto FTIR profiles against IAGOS data (averaged over 2009-2014 in an area 42-46 N, 75-84 W, to and from the airports of Toronto and Chicago) lead to conclusions very similar to those for Bremen (Fig. A5). The number of FTIR profiles for which IAGOS data are available ranges between 93 (DJF) and 176 (MAM). The FTIR profiles underestimate IAGOS data at all seasons in the free troposphere (2-10 km), by 15% in DJF, 6.5% in MAM, 8% in JJA and 11% in SON. The bias is negligible in the boundary layer, and both FTIR and IAGOS overestimate the surface concentration by ca. 15% at the nearby GAW site of Egbert, 60 km away from the FTIR site.

The Toronto profiles are also compared with the aircraft data of the GGGRN program of ESRL/NOAA using measurements from the 5 nearest sites (see map on Fig. 2.13). The number of ESRL measurements used is ca. 1800 per season (considerably fewer than the number of IAGOS measurements), but the number of FTIR profiles for which ESRL flights are available is important, between 212 (DJF) and 597 (JJA). The FTIR data underestimate ESRL data in winter (-15%) and this bias is larger than the statistical variability of the ESRL averages (Fig. A5). The bias is small for remaining seasons. In the boundary layer, FTIR overestimate ESRL data by ca. 50 ppb. The near-surface ESRL data agree very well with the surface concentrations at Egbert, despite the large distance (>500 km) between Egbert and each of the ESRL aircraft bases, suggesting that the ESRL average profiles used here are valid proxies for the CO profile in the rural surroundings near Toronto. The conclusions from these comparisons are consistent with the comparisons with IAGOS when accounting for the rural character of ESRL profiles compared to both FTIR and IAGOS data.

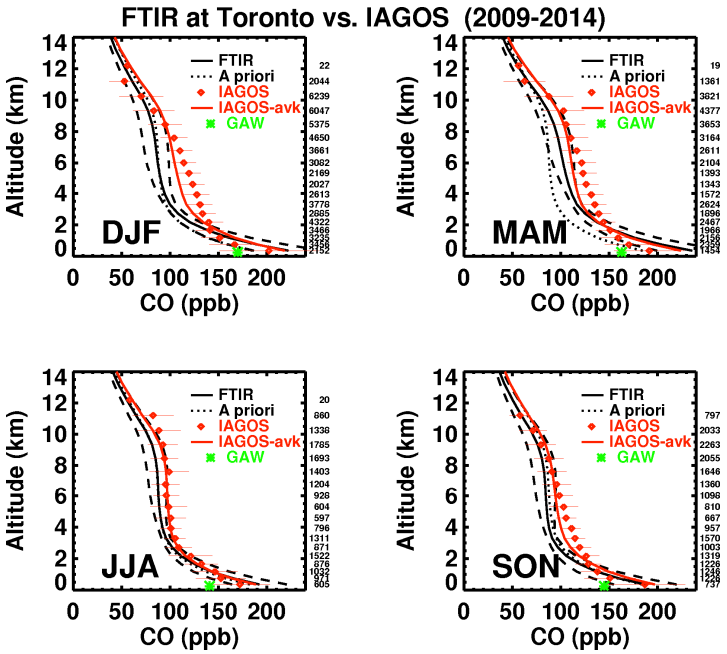


Figure A5. As Fig. A2, for Toronto. The in situ concentrations at Egbert (GAW) are shown in green.

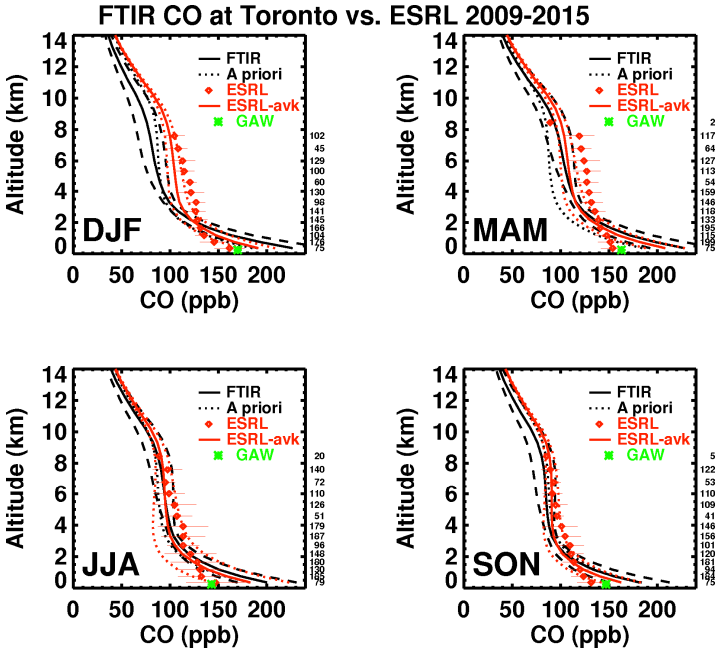


Figure A6. Same as previous Figure, except for aircraft data of the GGGRN programme of ESRL (red symbols).

Mauna Loa

FTIR profiles at Mauna Loa are compared with IAGOS data averaged over 2015-2016 (20.4-23 N, 157.5-162.5 W, to and from Honolulu) since data were unavailable prior to 2015 (Fig. A7). The number of FTIR profiles for which IAGOS data are available is 185 in DJF and 86 and SON. A good agreement is found at both seasons, considering the statistical variability of the IAGOS data, but the vertical gradient is more negative in IAGOS compared to FTIR.

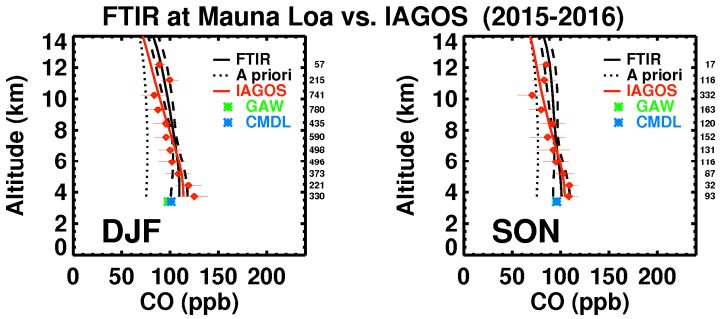


Figure A7. As Fig. A2, for Mauna Loa. The GMD and GAW surface mixing ratios are shown in blue and green, respectively.

Altzomoni

The FTIR average profile at Altzomoni in April/May is compared with IAGOS data (between 2001 and 2015 over 15-24 N, 90-100 W, flights to and from Mexico) in Fig. A8. April/May is the only time of the year for which the comparison is possible. The FTIR profile overestimates aircraft data by up to 25 ppb at 7-9 km altitude, and is in good agreement with both IAGOS and GMD data at the altitude of the station. The statistical variability of IAGOS data is very high below 10 km altitude, and of the same order as (or higher than) the FTIR bias.

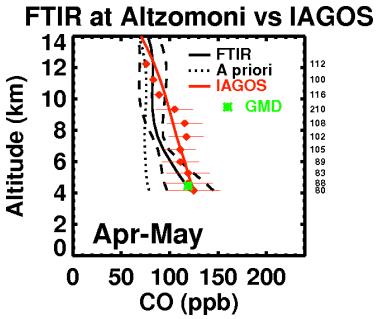


Figure A8. As Fig. 2.14, for Altzomoni. The mixing ratio at the GMD site of Mexico is shown in green.