Cross comparisons report

Deliverable title	Cross comparisons report
Deliverable number	D5.3
Revision	05
Status	Final
Planned delivery date	30/11/2014
Date of issue	01/12/2014
Nature of deliverable	Report

This work has received research funding from the European Community's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n°284421.

DOCUMENT PROPERTIES

	FUNCTION	NAME	ORGANISATION	DATE	SIGNATURE
LEAD AUTHOR	Research Associate	S. Henne	EMPA	01/12/2014	
CONTRIBUTING AUTHORS	Research Associate	T. Blumenstock	KIT-IMK		
	Research Associate	M. Gil Ojeda	INTA		
	Research Associate	F. Hendrick	BIRA-IASB		
	Research Associate	E. Mahieu	ULig		
Reviewed by	Coordinator	M. De Mazière	BIRA-IASB	07/12/2014	

Table of Contents

EXECUTIVE SUMMARY 4
1. INTRODUCTION 6
2. METHODS 8
2.1. Remote sensing data 8
2.2. Remote sensing averaging kernels and lower free tropospheric average 8
2.3. Reference datasets 9
2.3.1. Simulations for MAXDOAS horizontal mean retrievals of NO_2 and O_3 10
2.4. Comparison statistics 11
2.5. Analysis of temporal variability 11
3. RESULTS 12
3.1. Diurnal cycle of surface in-situ observations 12
3.2. Carbon Monoxide observations from FTIR 14
3.2.1. Jungfraujoch 14
3.2.2. Izaña 18
3.3. Ozone observations from FTIR 21
3.3.1. Jungfraujoch 22
3.3.2. Izaña 24
3.4. Methane observations from FTIR 27
3.4.1. Jungfraujoch 28
3.4.2. Izaña 31
3.5. Nitrogen dioxide observations from MAXDOAS 34
3.5.1. Jungfraujoch 34
3.5.2. Izaña 37
3.6. Ozone observations from MAXDOAS 40
3.6.1. Izaña 40
4. CONCLUSIONS 41
REFERENCES 44
LIST OF FIGURES 45
LIST OF TABLES 49
ACRONYMS, VARIABLES AND ABBREVIATIONS 49

Executive Summary

The objective of NORS work packages (WP) 5 was the integration/inter-comparison of surface in-situ observations with NORS products at the two demonstration sites Jungfraujoch and Izana. Since surface in-situ observations of a number of traces gases can be very precise and traced back to international standards, WP5 served as a demonstration of the quality of the NORS products in the troposphere. However, in order to facilitate a meaningful comparison the representativeness of the surface in-situ and ground based remote sensing observations needs to be taken into account. Within the main activity of the WP a novel method was developed that uses backward Lagrangian Particle Dispersion Modelling (LPDM) to characterise the representativeness of different types of observations. The LPDM simulations were tailored towards each specific sampling volume. They helped to 1) characterise the history of each sampled air mass (remotely sensed and in-situ), and in turn their representativeness, and 2) to generate high-resolution model profiles specific for the remote sensing volumes. The latter were then merged with the surface in-situ observations and yielded the aspired reference profiles against which the remote sensing data were validated.

The analysis was carried out for FTIR profile retrievals of CO, CH_4, and O_3 from both demonstration sites and as well for MAXDOAS NO_2 profile retrievals from Jungfraujoch. Simulations with the LPDM FLEXPART for individual partial columns in the troposphere were performed for the years 2009-2011 for the FTIR and 2011 and 2012 for MAXDOAS. In addition, FLEXPART simulations were conducted for O_3 and NO_2 from horizontal path retrievals (modified geometric approach, MGA) at Izaña and the years 2011 and 2012. FLEXPART was driven in a nested configuration with high resolution (2 km x 2 km over the Alpine domain, 7 km x 7 km over Western Europe and 1° x 1° globally) input wind fields from the operational weather prediction model COSMO, run by MeteoSwiss, and ECMWF operational analysis. The model results were used to obtain model profiles of the mentioned species by combining recent emissions picked up by the dispersion model with global baseline concentrations from global scale models. Input from global scale models was taken from 3 different sources: 1) MACC re-analysis (CO, O_3, CH_4, NO_x), 2) TM5 (CH_4) and 3) FLEXCTM (CO, CH_4). The model profiles were fused with surface in-situ observations using the newly developed method as described in NORS deliverable **D5.1** (Henne, 2014). The method uses the information obtained from the backward dispersion simulation to identify partial columns for which the surface observation is representative and adjusts simulated concentrations according to the surface observations. The resulting profiles are referred to as in-situ profiles while pure model profiles are referred to as model-only profiles. For a final validation of the remote sensing profiles, the in-situ profiles were folded with the remote sensing averaging kernels and averaged over the tropospheric column. Different depths for the tropospheric column were analysed.

The cross comparison results can be summarised as follows (see Figure 1): All remotely sensed tropospheric columns correlated reasonably well with surface in-situ data that were observed at the same time as the remote sensing observation. Correlations were generally better for CO than for any other trace gas. For Izaña correlations were slightly improved against night-time in-situ observations. Night-time observations were used because they should be less influenced by local disturbances that only affect the surface in-situ observations but not the remote sensing techniques. When comparing against in-situ and model-only columns the correlations with the remote sensing columns largely improved for all species and both sites. At the same time the bias corrected root mean square error (BRMS) largely decreased. Improvements were largest for O_3 for which surface observations were expected to have a rather limited representativeness for the vertical column. Bias corrected RMSE for hourly data were in the order of 10 ppb for CO, O_3 and CH_4 and around 50 ppt for NO_2. Regression slopes for CO and O_3 retrieved by FTIR versus in-situ columns were not significantly different from 1, while they were around 1.4 for CH_4 columns and 1.2 for NO_2 columns. In case the of CO and O_3 this presents an improvement over the pure surface in-situ comparisons for which either slopes were

significantly different from 1 or uncertainties of the slopes were large. While the other comparison statistics improved when using in-situ columns instead of surface in-situ data, the overall absolute bias increased for CO and CH_4 (only Izaña). Biases were about 5 ppb for O_3 (at Jungfraujoch) and CH_4 (at Jungfraujoch) and 10 ppb for CO (both sites) and O_3 (at Izaña). For CH_4 a bias between 40 and 80 ppb was determined at Izaña depending on the baseline model used for the generation of the in-situ profile. Virtually no bias (<20 ppt) was determined for the MAXDOAS NO_2 observations at Jungfraujoch. In the case of the MGA retrievals at IZO the inclusion of model data did not improve the comparison statistics significantly. Strong local influence of emissions (NO_2) and deposition (O_3) on the surface in-situ observations is probably the most important reason for this behaviour. The LPDM simulations were not able to reproduce these very local influences and adequately account for them in the data integration process.

Comparisons versus model-only and in-situ columns were very similar for most species and both sites. This does not necessarily imply that the whole comparison method and surface in-situ integration deteriorates to a model validation exercise as is the main goal of the NORS project. The model profiles used here differ from those used in the Copernicus Atmospheric Monitoring Service in the sense that 1) they are dedicated transport simulations for specific remote sensing volumes, 2) we used different baseline models in the study and carefully selected those that exhibited no systematic bias compared to the surface in-situ observations at the comparison sites. The comparison of the remote sensing tropospheric columns with the baseline models themselves was usually less good than for our dedicated FLEXPART simulated profiles. Furthermore, the influence of incorporating the surface data was larger when tropospheric columns were calculated up to lower altitudes. In that case comparison statistics improved from model-only to in-situ columns, but the overall agreement was generally weaker due to the limited independent information the remote sensing retrieval could obtain from a less deep column. In one case, profile retrievals of a MAXDOAS were compared, which had a much finer vertical resolution than the FTIR retrievals. Hence, the targeted lower tropospheric column was not very deep and the influence of the surface in-situ data was comparatively large while also the model performance was relatively weak. In this case a clear improvement in the comparison statistics was visible when using the in-situ columns as compared with the model-only columns.

Another important aspect of the cross comparison was the analysis of the resolved scales of variability. When comparing with in-situ surface data correlations for hourly de-trended and de-seasonalised data were usually small and overall good correlations largely resulted from the seasonal cycle, which was common to both datasets. The integration of in-situ and model data especially improved the correlations on the hourly time scale. This leads to the conclusion that the remote sensing techniques are well able to resolve hourly variability. The same conclusion could not have been drawn from the comparison of surface in-situ data alone.

In summary, the cross comparison of the remotely sensed tropospheric columns with model assisted in-situ columns proves the high quality of the FTIR derived tropospheric columns of CO and O_3 and MAXDOAS retrievals of NO_2 at Jungfraujoch. For FTIR CH_4 tropospheric columns larger uncertainties remain that may partly be related to uncertainties in the reference profiles but mostly seem to stem from uncertainties in the CH_4 retrievals related to the quality and consistency of the CH_4 line parameters. Further improvements of the CH_4 retrieval strategy are ongoing.

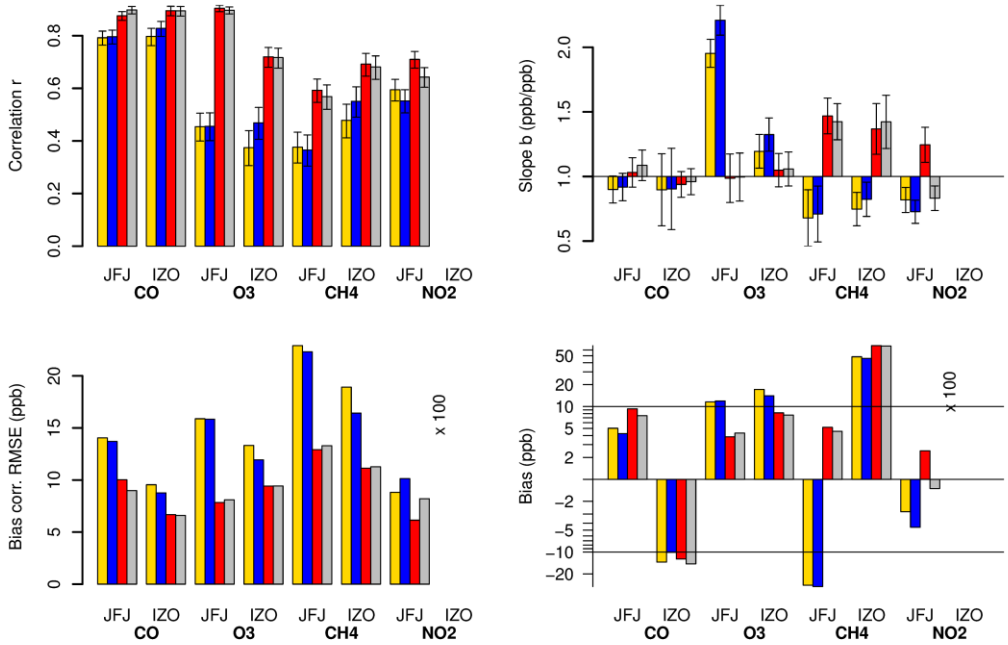

Figure 1: Comparison statistics for FTIR retrieved tropospheric columns of CO, O_3 and CH_4 at Jungfraujoch and Izaña and MAXDOAS lower tropospheric columns of NO_2 at Jungfraujoch. (top, left) correlation coefficient, r, (top right) slope of linear regression, (bottom left) bias corrected root mean square error, (bottom right) remote sensing bias. Error bars on correlation and slope represent 95% confidence estimates. Results for four different reference data sets are shown: (yellow) day-time surface in-situ observations, (blue) night-time surface in-situ observations, (red) in-situ columns, and (gray) model-only columns. In-situ columns were obtained from the model-only profiles after merging with surface in-situ observations. Both model-only and in-situ profiles were folded with the respective averaging kernel before the comparison. The analysis was performed on hourly aggregates. Note that the values for bias and bias corrected root mean square error were multiplied by 100 in the case of NO_2 observations.

1. Introduction

In-situ observations of many trace gases offer an excellent atmospheric reference since they are routinely calibrated at site and can be traced back to international standards. Depending on the gas, the combined uncertainty of in-situ measurements is usually small. Hence, in-situ measurements may offer validation data for both remote sensing (RS) and model products. Nevertheless, crucial limitations of in-situ data sets are their availability and representativeness. On the one hand, only limited in-situ data are available in the vertical dimension since these need to be obtained from cost intensive airborne platforms such as aircraft and balloon soundings. On the other hand, surface in-situ data are available from many surface locations worldwide but may not be representative of the volume of air represented by a model grid box or of the volume sampled by a RS technique. This is mainly due to inhomogeneous surface fluxes that strongly affect surface concentrations. Especially the vertical representativeness of surface in-situ data is limited, with most sites located in the planetary boundary layer, which often exhibits distinctly different trace gas concentrations than the free troposphere above.

Within work package 5 of the NORS project (Integration of tropospheric products) an effort was undertaken to extend surface in-situ observations in the vertical and, hence, produce a reference tropospheric profile which can be compared with RS profiles and tropospheric column data. The extrapolation of the surface in-situ data is tailored towards individual RS data sets, since it takes the actual sampling volume of both RS and in-situ observation into account. The approach, which was

detailed in NORS deliverable **D5.1** (Henne, 2014), uses backward simulations of a Lagrangian Particle Dispersion Model (LPDM) to 1) characterise the air mass history of each sampling volume and 2) to generate high-resolution model profiles. The latter were adjusted to match the surface in-situ observations and yield the aspired reference profiles for comparison with RS data. The LPDM FLEXPART was driven by wind fields with high horizontal resolution (2 km x 2 km in the Alpine domain) in order to capture the local flow in the vicinity of the comparison sites as good as possible and in a more realistic way than in a global scale transport model. This may not be important for sampling volumes removed from the surface, but in order to estimate the representativeness of the surface in-situ observations these need to be simulated as precisely as possible.

While NORS deliverable report **D5.1** (Henne, 2014) provides details on how reference profiles were generated, this report serves as a cross comparison of the ground based remote sensing data that were partly re-evaluated and generated within the NORS project. As such this report has two main aims: On the one hand it should demonstrate the usefulness of the surface in-situ data extension and on the other hand it should validate the ability of the remote sensing techniques to capture tropospheric variability and quantify the overall uncertainty of the remotely sensed tropospheric columns. The main target for the validation of the remote sensing products was the lower tropospheric mean column from the surface (level of the RS instrument) up to a level where the RS retrieval obtained one independent piece of information or, depending on the retrieved parameter, the influence from the upper troposphere/lower stratosphere (UTLS) was not dominating the comparison.

The comparison was carried out for tropospheric profiles of carbon monoxide (CO), ozone (O_3) and methane (CH_4) as obtained by FTIR and profiles of nitrogen dioxide (NO_2) as obtained by MAXDOAS. In addition, MAXDOAS observations of NO_2 and O_3 at horizontal level, which were obtained with the modified geometric approach (Gomez et al., 2014), were compared. For MAXDOAS observations the main source of representativeness uncertainty arises from the relatively long horizontal distance from which the MAXDOAS obtains its information. Again, FLEXPART backward simulations were used to define how representative surface in-situ observations were for the line of sight of the MAXDOAS sampling.

The analysis was restricted to two remote locations that describe the variability in the tropospheric background but are also influenced by pollution events. Both sites are high elevation sites, but in very different environments. Jungfraujoch is situated in the central Swiss Alps and as such relatively close to some of the major European emission sources, but vertically removed from them (3580 m asl). In contrast, the Izana observatory is located on the island of Tenerife off the coast of North-Africa and is seldom influenced by directed near surface pollution transport from Europe or Africa. However, local emissions on the island may strongly influence the in-situ observations when these are lifted to the mountain top by day-time up-slope winds that regularly form under the prevailing fair-weather conditions. Together these locations offer the possibility to test our surface in-situ extension method under very different conditions and distinguish where the method is more appropriate to derive reference profiles. The method was not applied to classical boundary layer sites, because for these the representativeness of surface in-situ observations is mostly limited to the boundary layer itself and may therefore not be profitable for tropospheric column comparisons.

This report is divided into 2 main parts. In section 2 we define the scope of the inter-comparison, introduce different reference datasets, and define the target statistics for the validation. In section 3 the results are presented by observed parameter and instrument. A detailed analysis is given for each parameter, instrument and site. Finally, section 4 concludes the findings.

2. Methods

2.1. Remote sensing data

Details on the remote sensing techniques inter-compared in this report can be found in the NORS Data User Guide, **D4.2** (De Mazière et al., 2013), and references therein. In addition to FTIR and MAXDOAS profile retrievals, MAXDOAS horizontal mean mole fractions as obtained by the modified geometric approach (MGA; Gomez et al., 2014) for NO_2 and O_3 observations at Izaña were used in the inter-comparison. Table 1 lists the inter-comparison periods and the NDACC data versions used in the final comparison.

Table 1: Remote sensing data used in cross comparison. The version refers to the version on the NDACC data centre.

Site		FTIR CO	FTIR O_3	FTIR CH_4	MAXDOAS NO_2	MAXDOAS O_3
Jungfraujoch	Version	004	001*	002	001	no data
	Period	2009 to 2011	2009 to 2011	2009 to 2011	2010 to 2011	no data
Izaña	Version	003	003	003	NA**	NA**
	Period	2009 to 2011	2009 to 2011	2009 to 2011	2010 to 2011	2010 to 2011

*: An updated data version (004) was available by the end of the NORS project, but it contained a new vertical sub-division in the retrieval, which would have required new FLEXPART calculations.
**: Data were not delivered through the NDACC data centre and did not follow the GEOMS format convention.

2.2. Remote sensing averaging kernels and lower free tropospheric average

Remote sensing (RS) profiles that were retrieved through a profile retrieval algorithm can be characterized by their averaging kernels (Rodgers and Connor, 2003). The averaging kernel (AVK), the n x n matrix $\boldsymbol{A}$ where n is the number of vertical sub-columns, relates the retrieved mole fractions χ_{RS} to the true profile, χ, and the a-priori profile, χ_a, which was used in the retrieval

$$\chi_{RS} = \chi_a + \boldsymbol{A}(\chi - \chi_a) + \epsilon, \quad (1)$$

where ϵ represents the retrieval uncertainties.

The rows of **A** can be interpreted as sensitivity of the retrieval to the presence of the target species in each sub-column. The trace of the matrix **A** gives the degree of freedom (DOF) of the retrieved signal. The larger the DOF the more independent information could be retrieved and the more vertical details should be revealed. Summing the elements in the trace of **A** from the lowest sub-column up to a specific altitude gives the amount of independent information that was totally retrieved in the given column.

One can apply the AVK to a reference profile to derive a smoothed profile that can be directly compared to the RS profile (Rodgers and Connor, 2003). In our case the reference profile γ is folded with the AVK, **A**, and the a-priori profile,

$$\phi = \chi_a + \boldsymbol{A}(\gamma - \chi_a) \quad (2)$$

to yield the smoothed profile ϕ.

To obtain mean mole fractions that are representative for the lower free troposphere, vertical profiles were averaged up to an altitude, z_{FT}, for which the cumulative DOF (calculated from the

surface to the given sub-column) exceeded a given threshold. Averages were then calculated as weighted means using level pressures as weights

$$X = \frac{\sum \chi_k p_k}{\sum p_k} \tag{3}$$

We refer to these averages as free tropospheric (FT) means, keeping in mind that the two analysed sites are elevated sites that usually reside above the planetary boundary layer. For FTIR retrievals of CO the threshold DOF was set to 1. For FTIR retrievals of ozone and methane, for which the tropospheric sensitivities were generally lower, the DOF threshold was set to 2/3. These values represent a compromise between using independent information from the RS observations and keeping stratospheric influences on the tropospheric mean as small as possible. Furthermore, keeping a relatively small value of z_{FT} assures that a large part of the information used in the reference profiles originates from the surface in-situ observations and is not a mere model product. The dependency of the comparison on the selection of the DOF threshold will be discussed along with the main findings.

2.3. Reference datasets

In the main part of this report the time series of FTIR and MAXDOAS retrieved lower tropospheric columns are compared with different reference time series:

a) **Day-time surface in-situ**: These are the surface in-situ observations, made at the same time (window of two hours) as the RS observation.

b) **Night-time surface in-situ**: These were derived from surface in-situ observations made during the night-time hours bracketing the RS observations. They were used for species that exhibited strong diurnal variations (see section 3.1) and may be influenced by local emissions or deposition during day-time.

c) **Folded in-situ columns**: The tropospheric mean of the “in-situ profile”, which was constructed from FLEXPART simulated profiles and the surface in-situ observations (see D5.1, Henne, 2014) and folded with the RS averaging kernels (see section 2.2).

d) **Folded model-only columns**: The tropospheric mean of the FLEXPART simulated profile, again folded with the RS averaging kernels.

e) **In-situ columns**: The tropospheric mean of the “in-situ profile” as in c), but without applying the RS averaging kernel.

f) **Folded GCTM columns**: The tropospheric mean of the pure model profiles as taken from global chemistry transport models (GCTMs). These were not adjusted to surface observations, but folded with the respective RS averaging kernels.

By using different reference time series it should be tested if including model data in the in-situ comparison enhances the comparability of the RS data to the surface in-situ data. The in-situ profiles, which rely on the in-situ data and model integration, should be the ones closest to the true tropospheric profile and, after folding with the AVKs, will serve as our main reference (reference dataset c)).

As outlined in detail in NORS deliverable report **D5.1** (Henne, 2014) FLEXPART backward simulations were performed for individual sampling volumes of 1) the surface in-situ observations and 2) tropospheric partial columns of the FTIR and MAXDOAS profile retrievals. The backward simulation was carried out in a nested configuration that used high resolution input in the Alpine domain around JFJ (2 km x 2 km horizontal resolution taken from COSMO analysis of MeteoSwiss) nested into a coarser resolution over Western Europe (7 km x 7 km, COSMO) and finally the global domain (1° x 1°, IFS operational analysis by ECMWF). For IZO FLEXPART was driven by ECMWF global scale

meteorology only. Each individual air mass was followed for 10 days backward in time. On the one hand, these simulations provide surface (or emission) sensitivities along the backward simulations. On the other hand, when combining the final position of the model particles (in the backward simulation) with global atmospheric composition fields one can obtain the initial state of the air mass prior to the arrival at the sampling location. The former can be combined with emission inventories to yield mole fraction enhancements during the period of the backward simulation while the latter can be seen as the baseline mole fraction for the sampled air mass. The resulting total mole fractions for each RS partial column represent the FLEXPART model-only reference dataset (see reference dataset d)).

Baseline levels were obtained from simulated atmospheric composition fields from three different global scale models: 1) MACC re-analysis for CO, O_3, CH_4, NO_x (Inness et al., 2013), 2) TM5 for CH_4 (Bergamaschi et al., 2013) and 3) FLEXCTM for CO, CH_4 (Henne et al., 2013). In NORS deliverable report **D5.1** (Henne, 2014) the differences between different baseline (or GCTM) inputs for the FLEXPART simulations at the surface in-situ observation were compared. For each comparison species and site a best suited baseline model was suggested. While focussing the analysis on the comparison of the RS columns with this "best" model realisation, for discussion purposes we also present comparison results for FLEXPART profiles derived from the other baseline models.

The FLEXPART model-only profiles were combined with surface in-situ observations using additional information available from the backward transport simulation. The information on when emissions were picked-up along the backward simulation was used to identify different vertical layers in the troposphere by applying a clustering algorithm to the results for different RS partial columns and the surface in-situ simulation. Surface in-situ observations were then ascribed to those partial columns situated in the same vertical layer as the surface in-situ observation. Instead of simply transferring the surface in-situ value to the whole layer, baseline and above baseline contributions were adjusted separately. The resulting (adjusted) FLEXPART model profiles are referred to as "in-situ profiles" (see reference datasets c) and e)). By using high-resolution FLEXPART simulations specific for the RS sampling volumes we expect to significantly reduce the representativeness uncertainty that usually arises in any kind of model comparison with point (line) observations.

In contrast to the FLEXPART derived profiles, reference data f) were obtained from a direct interpolation of the 3-dimensional output from the global scale chemistry transport models onto the location of the remote sensing sampling volume. The latter took the viewing geometry and effective sampling volume of the RS technique into account. However, a rather large representativeness uncertainty between the point (line) observations and the GCTM simulations can be expected due to the rather coarse model resolution. This is especially true for a site in complex terrain like JFJ and for the lower part of the troposphere.

All datasets were aggregated to hourly intervals for the comparison.

2.3.1. Simulations for MAXDOAS horizontal mean retrievals of NO_2 and O_3

The previous deliverable report **D5.1** (Henne, 2014) did not contain information on how model profiles were obtained for MAXDOAS observations. Modifications of the method were required due to a) different viewing geometry of the MAXDOAS instrument as compared to the FTIR, b) targeting a very reactive species (NO_2). The latter had to be described in the FLEXPART backward simulations, which was not necessary for the other trace gases that were treated as passive tracers considering the relatively short time of transport.

The sampling volume of the MAXDOAS was characterised by the horizontal viewing distance, *d*, as derived by the modified geometrical approach (Gomez et al., 2013), which is using differential absorption in the O_4 band. For MAXDOAS MGA estimates, the sampling volume was taken as a line from the instrument extending in the viewing direction up to distance *d*. For MAXDOAS profile

retrievals the sampling volume was defined as the vertical slice defined by the instrument's viewing azimuth angle and the horizontal viewing distance, vertically limited by the boundaries defined by the retrieval algorithm.

In contrast to CO, CH_4 and O_3, NO_2 has a much shorter lifetime in the troposphere. Furthermore, it is rapidly photolysed to form NO, but is also rapidly replenished by the reaction of NO with O_3. In the background troposphere NO_2, NO and O_3 remain in a close relationship described as photostationary equilibrium. As such the NO_2 to NO ratio only depends on the O_3 concentration, the reaction rate of NO with O_3 and the photolysis rate of NO_2. Due to this close link of NO and NO_2 we did not transport NO_2 separately in the model, but transported the sum of NO and NO_2: NO_x. Also emissions were given as NO_x emissions in mass units of NO_2. We parameterised NO_x loss as a first order loss process with a seasonally variable lifetime reaching from 5 hours in summer to 22 hours in winter (Schaub et al., 2007). This loss was applied to recent emission contributions. For background contributions an average lifetime of 14 days was assumed, which also takes into account the replenishing of the NO_x pool from more stable nitrogen species (organic nitrates, nitric acid). To derive NO_2 concentrations at the sampling volume the derived NO_x mole fractions were converted to NO_2 by assuming a photostationary equilibrium between O_3, NO and NO_2. In order to calculate the equilibrium ratios of NO_2 to NO, one requires O_3 number concentrations, NO_2 photolysis rates and NO+O_3 reaction rates. The first was obtained by interpolating the MACC re-analysis O_3 model simulations to the receptor volume. The second was obtained from the photolysis parameterisation given by Henne et al. (2005), which was derived for conditions above Jungfraujoch. Since this parameterisation describes clear sky conditions, we further scaled the obtained photolysis rates by the ratio of calculated (clear sky) global radiation and that observed at Jungfraujoch. The reaction rate coefficient was calculated temperature dependently.

2.4. Comparison statistics

For each comparison pair the following statistics were calculated from hourly aggregates. The Pearson correlation coefficient r for all data in the time series was calculated and gives a measure of how much of the observed temporal variability is common to the two time series. A linear regression analysis was applied that takes uncertainties in both regression variables into account (weighted total least square regression) (Krystek and Anton, 2007). The regression slope, b, is evaluated to estimate systematic scaling offsets between two time series. The mean difference between two time series is referred to as the bias

$$bias = mean(\mathrm{X}_{RS}) - mean(\mathrm{X}_r) \tag{4}$$

where X_{RS} is the remote sensing tropospheric column and X_r is the reference column.

The bias corrected root means square error was calculated as the standard deviation of the differences between reference and RS mole fractions

$$BRMS = std(\mathrm{X}_{RS} - \mathrm{X}_r) \tag{5}$$

The BRMS can serve as a measure of the overall random uncertainty (precision) of the comparison while the bias is an indication for the systematic uncertainty (accuracy). The BRMS should also give the upper limit of the remote sensing precision. Nevertheless, it may contain a significant contribution from the uncertainty of the reference profile itself. Unfortunately, it was not possible to determine the latter in any mathematical way in the current study.

2.5. Analysis of temporal variability

In order to analyse the temporal scales of the relationship between the RS data and the reference datasets the time series were separated into a part describing inter-annual variability, a part

describing the seasonal cycle and a remaining part reflecting variability down to hourly time scales. Note again that the data used in the comparison was aggregated into hourly bins.

We extract the different parts of variability by fitting a statistical model to each time series that describes the inter-annual variability by a 2nd order polynomial and the seasonal variability by first, second and third order annual harmonics. The fit residuals then give the remaining part of the variability (hourly).

3. Results

The results are presented in the following order: In section 3.1 we motivate why for some of the reference data sets night-time surface in-situ data were used in contrast to the day-time RS observation. In the following sub-sections, comparison results are presented by species and remote sensing technique.

3.1. Diurnal cycle of surface in-situ observations

Both sites analysed in this study are high elevation sites and as such are often situated in the free troposphere. However, both sites may be influenced by local up-slope flow systems that may bring air masses to the sites that are not representative of the sites altitude but more of the PBL below the site. The FLEXPART transport simulations should be able to pick up this local PBL influence for the surface in-situ and also for the RS sub-columns. This should especially hold for the simulations for JFJ, for which high-resolution (2 km x 2 km) wind fields were used. Nevertheless, it may still be argued that instead of day-time in-situ observations the bracketing night-time observations should be used to avoid any PBL influece. This was previously done in the study by Sepúlveda et al. (2012) for CH_4 FTIR vs. in-situ comparisons at Izaña.

Here the typical diurnal cycle of surface in-situ observations at both sites are analysed for different months of the year. For Jungfraujoch (Figure 2) a pronounced diurnal cycle with afternoon maximum and morning minimum was observed for ambient temperatures in all months of the year. It was especially pronounced for late spring and summer months. Along with temperature also CO showed a characteristic afternoon maximum in the summer months. The average diurnal amplitude for CO was in the order of 4 ppb. For O_3 no pronounced diurnal cycles were observed. A slight reduction of afternoon O_3 may indicate transport from the PBL in the summer months, but O_3 mole fractions in the PBL may have been very similar to those in the free troposphere, so that the PBL transport was not discernable in O_3. For CH_4 diurnal cycles with amplitudes up to 6 pbb were visible for April and May, but were less pronounced in the summer. Overall the diurnal cycles were relatively small in magnitude at JFJ. Hence, reference profiles were constructed from in-situ data taken at the same time as the remote sensing observation.

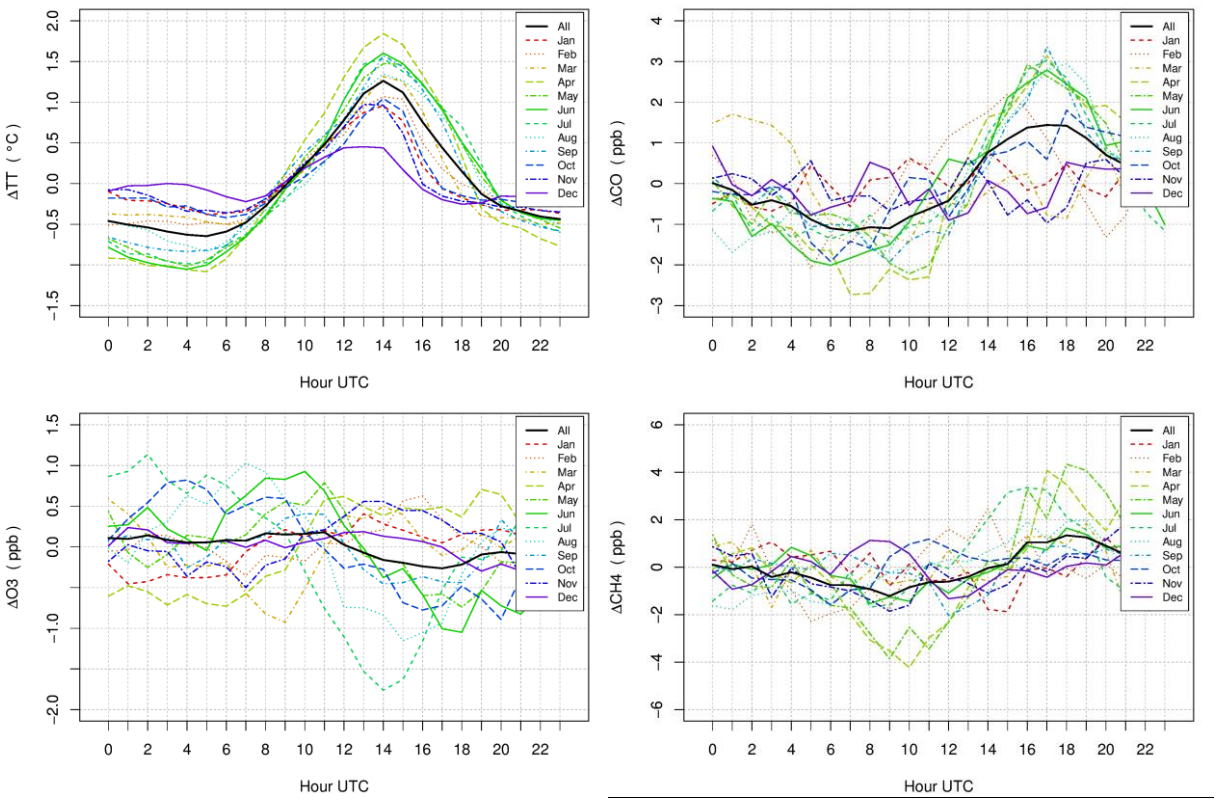

Figure 2: Diurnal cycle of surface in-situ observations of (top left) temperature, (top right) carbon monoxide, (bottom, left) ozone, and (bottom right) methane at the Jungfraujoch observatory (3580 m asl) for the period 2009 to 2011.

For Izaña the observed diurnal cycles were more robust as compared to JFJ, owing to the sub-tropical location of the site with year-round fair-weather conditions (Figure 3). Similar to JFJ, CO mole fractions were increased at the site during the late afternoon hours indicating transport of fresh emissions from the island environment. The amplitude of the CO diurnal cycle at IZO was comparable to that during summer-time at JFJ (4-6 ppb). For CH_4 the diurnal cycle was actually reversed as compared to that of CO, showing late afternoon minimum and an amplitude of about 4-6 ppb throughout the whole year. This afternoon decrease in CH_4 suggests that there are no or only minor local CH_4 sources on the island itself and that CH_4 mole fractions in the marine boundary layer, due to photochemical destruction, are smaller than in the lower free troposphere. Similarly, O_3 showed a pronounced afternoon minimum throughout the whole year. Again, smaller O_3 mole fractions in the marine boundary layer as compared to the free troposphere may be the main reason for this effect. In addition, local dry deposition of O_3 during up-slope transport of air masses from the boundary layer may cause a reduction in the O_3 levels. The FLEXPART transport simulations should simulate the exchange between marine boundary layer and free troposphere. At the location of IZO this does not happen by explicitly resolving the up-slope flow (model resolution only 0.2° x 0.2°) but rather indirectly by a mixing layer height that is increased due to sub-grid orographic variability (Stohl et al., 2005). However, the model does not describe O_3 dry deposition in any way. Therefore, the in-situ profiles for O_3 at Izaña were constructed from bracketing night-time data.

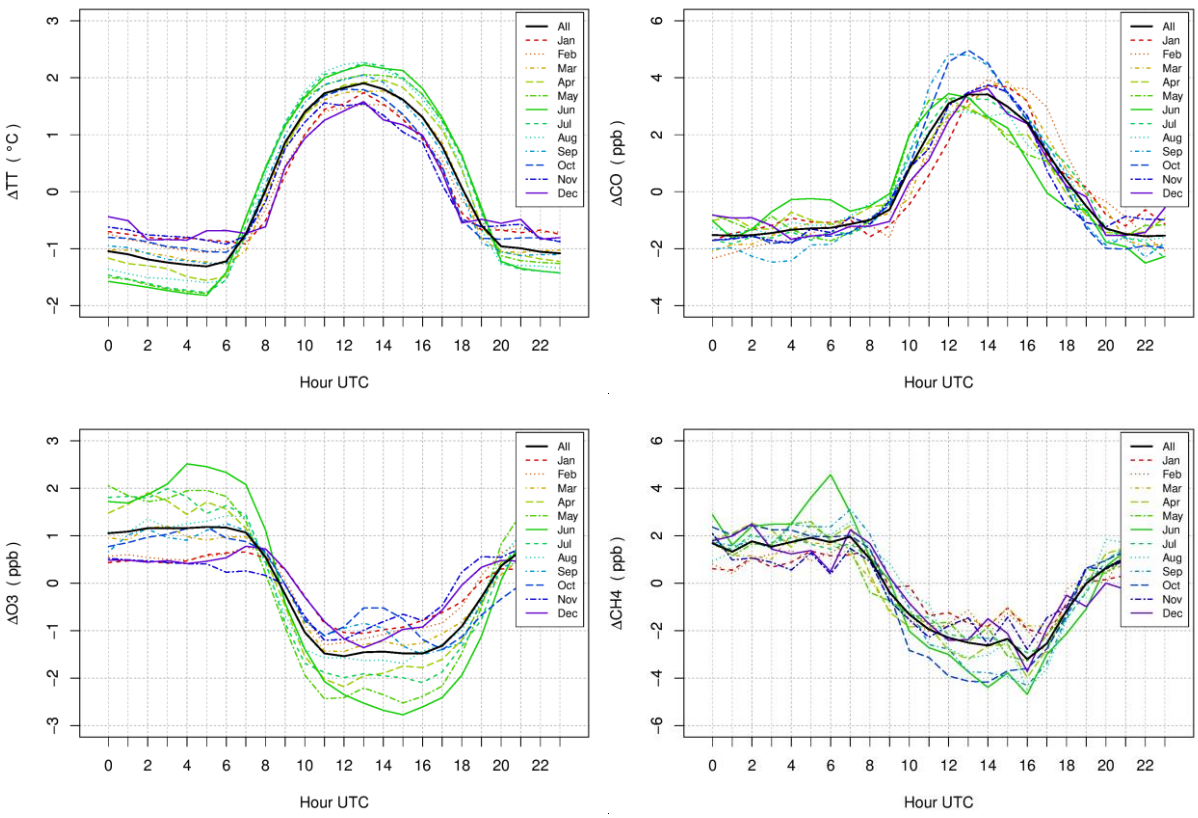

Figure 3: Diurnal cycle of surface in-situ observations of (top left) temperature, (top right) carbon monoxide, (bottom, left) ozone, and (bottom right) methane at the Izaña observatory (2373 m asl) for the period 2009 to 2011.

3.2. Carbon Monoxide observations from FTIR

The typical row averaging kernels of the FTIR CO retrieval indicated strongest sensitivities close to the surface. This was similar for both sites but much more pronounced for JFJ. On average the sum of the diagonal elements of the AVK reached 1 (indicating one independent piece of information from the retrieval) when summed over the lowest 5 and 7 FTIR sub-columns for JFJ and IZO, respectively. This corresponds to an average column top of 7.3 km asl and 7.2 km asl for JFJ and IZO. These level heights were used to derive the main comparison results for the tropospheric column. The influence of this choice will be discussed along with the main findings.

3.2.1. Jungfraujoch

Figure 4 shows the regression analysis of the CO free tropospheric column (up to 7.3 km) as obtained by the FTIR at Jungfraujoch. These values present hourly aggregates. They are compared against the different reference datasets as described in section 2.3. The FTIR columns were well correlated with the surface in-situ observations, regardless if these were taken at the same time as the FTIR observation (r=0.79, Figure 4 top left) or at night-time (r=0.80, Figure 4 bottom left). However, the scatter as measured by the bias corrected root mean square error (BRMS) was relatively large (~14 ppb) for this comparison and also the regression slopes were significantly different from 1 with the surface observations showing more variability than the FTIR observations. When the folded in-situ columns were used as reference (Figure 4, top center), the comparison improved in terms of an increased correlation coefficient (r=0.88) and a reduced BRMS (~10 ppb). Also the regression slope become close to unity but there remained a positive bias of 10 ppb: FTIR columns in general showed larger mean mole fractions. Not applying the AVK to the in-situ profiles (Figure 4, top left) and the use

of pure large scale model data (Figure 4, bottom left) decreased the comparability of the FTIR data for all comparison statistics. However, when using the folded FLEXPART mode-only columns, which were not adjusted by surface measurements, comparison statistics were as good as for the comparison using the in-situ profile (Figure 4, bottom center). This does not imply that including the surface data was altogether in vain. The FLEXPART model simulations were selected because they showed no general bias for the surface in-situ observations (Henne, 2014). Hence, the adjustment due to surface observations only minimally changed the model profiles and, because even at the high resolution simulation the model was not entirely able to describe the local surface variability, added some scatter to the profiles that was not observed by the FTIR.

The ability of the FTIR to detect atmospheric variability at different time scales (see section 2.5 for details of the method) is assessed in Figure 5 again for a column height of 7.3 km asl and three different reference datasets: day-time surface in-situ, folded in-situ columns, and in-situ columns. The regression plots for inter-annual, seasonal and hourly variability show large correlations for all time scales. The correlation for hourly variability strongly improved when the folded in-situ column was used as reference instead of the surface in-situ data (Figure 5, center). The large correlation coefficient for inter-annual variability should not be over interpreted since it is only based on three years.

The inter-comparison results presented in Figure 4 and Figure 5 for the lower tropospheric column as retrieved by the FTIR strongly depend on the height up to which the tropospheric column was calculated. In Figure 6 these dependencies are shown for the different comparison statistics and different reference datasets: day-time and night-time surface in-situ, folded in-situ columns and folded model-only columns. For the latter two, columns constructed from different baseline models are shown as well to illustrate the dependence of the results on the choice of the baseline model. In the case of CO at JFJ the best estimate reference profile, was based on FLEXPART simulations using FLEXCTM background fields. In addition, comparison results for the reference profiles constructed from FLEXPART simulations using MACC re-analysis as background fields are shown. For all column heights the correlation coefficient was larger and the BRMS was smaller for model-only and in-situ column comparisons than for surface in-situ comparisons. As already seen for a column height of 7.3 km asl in Figure 4, model-only profile means (gray lines and symbols) showed a slightly larger correlation than in-situ profile means (red lines and symbols). This was similar for both FLEXPART simulations with either FLEXCTM or MACC re-analysis as background values. In contrast, regression slopes were closer to unity for the in-situ columns than the model-only columns. This was especially pronounced for the model columns based on MACC re-analysis for which the FTIR regression slope of ~1.2 ppb/ppb for the model-only column changed to a slope of 1.07 for the in-situ column. The MACC re-analysis for CO is known to under-estimate near surface concentrations of CO in Central Europe (Inness et al., 2013). This was also visible in the comparisons when the in-situ simulations of the model were compared against surface in-situ observations (Henne, 2014). The inclusion of the surface data removed this bias and lead to an overall similar comparison of the MACC in-situ columns as for the FLEXCTM in-situ columns.

Correlations between model-only and in-situ tropospheric columns on the one hand and FTIR tropospheric columns on the other hand only improved slightly with higher column tops (Figure 6, top left). Consistently, the BRMS was reduced for higher column tops (Figure 6, bottom right). The BRMS for surface in-situ comparisons remained large for all column top heights. Regression slope and bias decreased with increasing column height (Figure 6, top right and bottom left). For the in-situ profile the bias basically vanished when the tropospheric column was averaged up to 10.6 km asl and also the slope became equal to unity in this case. These changes in slope and bias with increasing tropospheric column height can be understood when comparing the vertical structure of the profiles obtained by the FTIR with those of the folded in-situ profiles. Figure 7 shows the time series of

monthly mean FTIR profiles and monthly mean differences between the FTIR and the folded in-situ profiles. The FTIR time series illustrates the typical annual cycle of CO in the northern hemispheric mid-latitudes with increased CO mole fractions in the lower troposphere during winter and decreased CO mole fractions in summer. In general CO steadily decreased from the surface to the stratosphere. The differences between FTIR and folded in-situ profile reveal an interesting relocation of CO in the FTIR profiles. While FTIR values were generally lower than in-situ profile values in the troposphere above 6 km, they were larger in the lower troposphere. In addition to this vertical shift, an intensification of the upper tropospheric deficit was detected during the winter month, while the lower tropospheric surplus remained relatively constant throughout the year with a slight amplification during the summer months. The red horizontal line in Figure 7 indicates the column height for which on average one independent piece of information was retrieved by the FTIR (DOF=1). From the vertical profile differences it becomes obvious that the lower the column height the larger the positive FTIR bias will be (compare Figure 6).

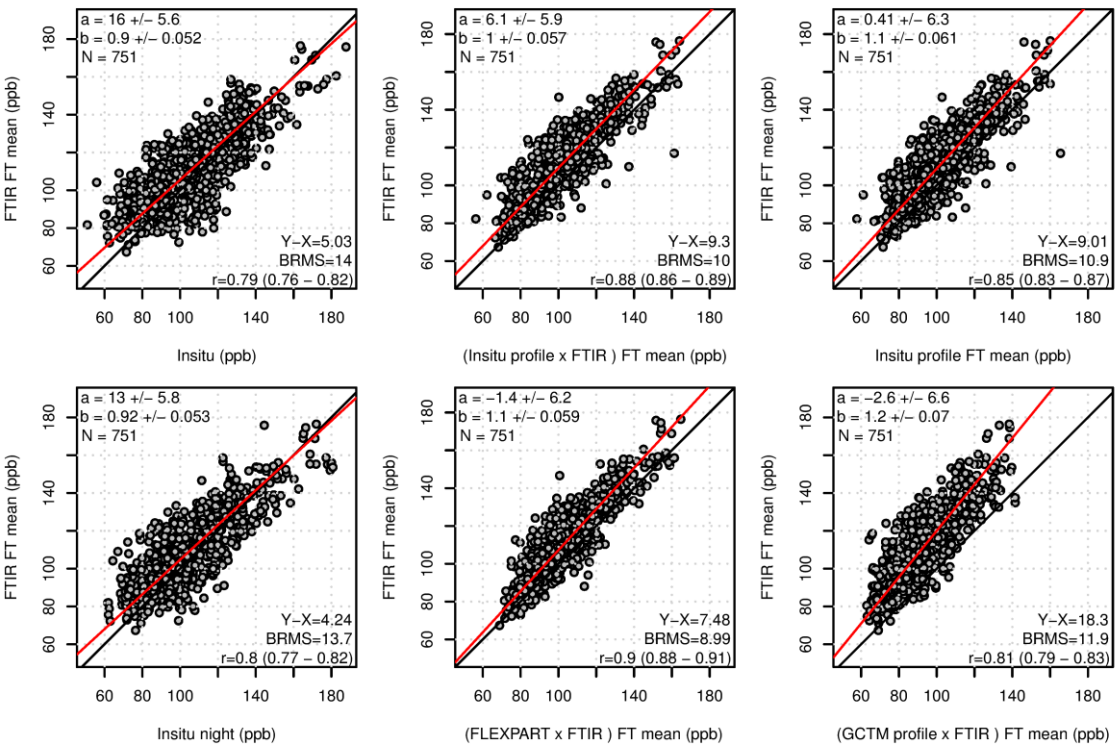

Figure 4: Linear regression analysis of mean free tropospheric (FT) CO mole fractions as obtained by FTIR at Jungfraujoch. For the lower free tropospheric mean all FTIR sub-columns up to 7.3 km asl were averaged (pressure weighted mean). The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as FTIR observations, (bottom left) night-time in-situ observations bracketing the FTIR observation, (top center) in-situ profile folded with FTIR AVK, (bottom center) model profile folded with FTIR AVK, (top right) mean in-situ profile, (bottom right) GCTM profile folded with FTIR AVK. GCTM profiles were taken from FLEXCTM. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits.

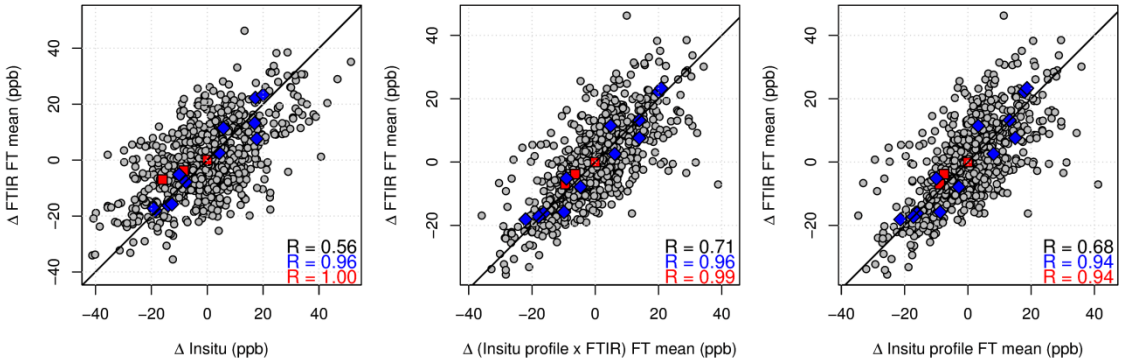

Figure 5: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The mean free tropospheric CO mole fractions as obtained by FTIR at Jungfraujoch were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.3 km asl. The black line represents the one to one line.

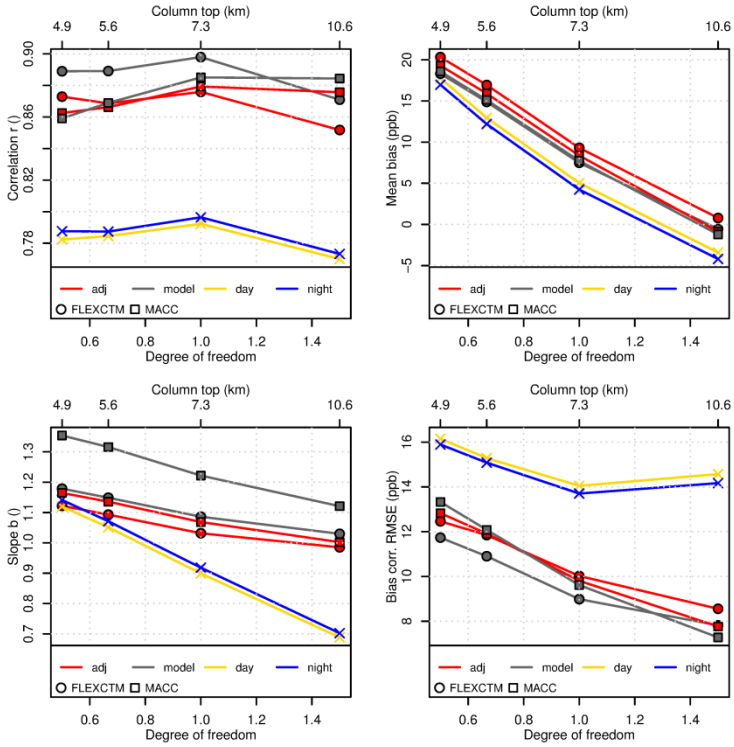

Figure 6: Dependency of comparison statistics on vertical integration depths for CO at JFJ. (top left) correlation coefficient, (top right) regression slope, (bottom left) bias (FTIR minus reference), (bottom right) bias corrected root mean square error. Comparison results are shown for a comparison versus (yellow) day-time in-situ observations, (blue) night-time in-situ observations, (red) in-situ profiles from two different baseline models, (gray) raw model profiles from two different baseline models (FLEXCTM and MACC re-analysis).

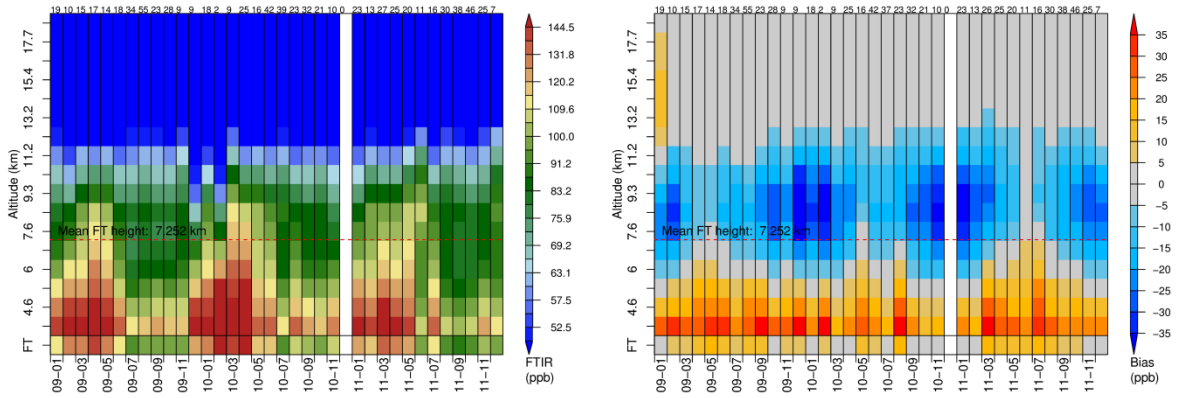

Figure 7: Time series (x-axis) of (left) monthly mean vertical profiles (y-axis) of CO mole fractions as retrieved by FTIR and (right) monthly mean differences of the FTIR and FTIR folded in-situ profiles for Jungfraujoch. The lowest level marked by FT represents the free tropospheric mean calculated from the lowest level up to the height indicated by the red dashed line. The numbers at the top of each monthly column gives the number of valid hourly FTIR aggregates in the respective month.

3.2.2. Izaña

Similar results as for Jungfraujoch were also obtained for FTIR retrieved lower tropospheric columns of carbon monoxide at Izaña. The comparison with surface in-situ observations indicates a negative bias in the FTIR and also a regression slope smaller than 1 (Figure 8, top left, bottom left). With values smaller 10 ppb the BRMS was smaller for CO at IZO than at JFJ. This may be due to the overall smaller variability in observed CO at this location. All comparison statistics except the regression slope improved when using night-time in-situ observations indicating that day-time influence from up-slope winds complicates the direct comparison. The comparison with folded in-situ columns (Figure 8, top center) showed a regression slope not significantly different from 1 and a small negative bias of the FTIR column of about -13 ppb. As for JFJ, comparisons with model-only profiles were very similar to the one using in-situ profiles. In comparison to JFJ observed correlations at IZO were generally similar and BRMS were smaller. In contrast to JFJ a negative FTIR bias as compared to the reference column was detected when the height of the tropospheric column was 7.2 km asl.

When looking at the time scales of observed variability (Figure 9) one can discern a strong improvement in the comparison for the sub-monthly variability when using the in-situ columns (gray circles in Figure 9, center) instead of the surface in-situ data (Figure 9, left). This improvement was more pronounced for IZO than for JFJ, which may relate to the existence of elevated layers of increased CO at IZO (see discussion below) that were not reflected in the surface in-situ data but were picked up by the FTIR and the model. Such elevated pollution layers are less typical for JFJ and most of the sub-monthly variability in the tropospheric column above JFJ is expected to result from variability close to the surface.

When looking at the dependence of the comparison on the tropospheric column depth (Figure 10), similar observations as for JFJ can be made. Correlations were increasing with column depth for the in-situ columns based on MACC re-analysis, while they were first increasing then decreasing for the surface in-situ observations and the in-situ columns based on FLEXCTM. With increasing tropospheric column depth the absolute mean bias and also the BRMS decreased for the in-situ columns based on MACC re-analysis. The BRMS remained large for the other reference datasets. The regression slope approached unity for the highest tropospheric column top of 10.6 km asl. But in contrast to JFJ, the bias did not vanish completely when integrating the tropospheric column up to 10.6 km asl, but remained at around -5 ppb. Differences between the model-only profile and the related in-situ profile were not significant. Again this does not mean that the incorporation of the in-situ data was not

necessary. FLEXPART-MACC was selected because of its unbiased performance for simulating the surface in-situ data. However, it is interesting to notice that the comparison statistics for the model-only FLEXPART-FLEXCTM columns improved after fusion with the in-situ data, but did not achieve similarly good results as the FLEXPART-MACC model-only and in-situ columns.

The vertical profile of CO as retrieved by the FTIR showed a pronounced seasonal cycle with largest values during the winter months and at elevated altitudes of about 9 km asl (Figure 11, left). During the summer months a lower tropospheric minimum was observed below 5 km asl. Compared with the reference profile (Figure 11, right) a consistent negative (positive) bias of the FTIR in the lower (upper) troposphere was observed. The biases were more pronounced during times when largest CO mole fractions were detected by the FTIR in the upper troposphere. The vertical profile of the bias was opposite to the one detected for CO at JFJ and was also very similar when using FLEXCTM instead of MACC re-analysis as baseline model (not shown), as it was done for JFJ. When assuming that the general vertical profiles in the model are consistent between different locations, the reversal in the bias location must be a feature of the FTIR retrievals. The MACC reanalysis was validated against aircraft data and showed a general good agreement for the middle troposphere, but is known to under-estimate CO concentrations in the planetary boundary layer close to emission sources (Inness et al., 2013). However, at levels close to the surface the additional FLEXPART simulation, which only uses the MACC-reanalysis as a background level, and the additional combination with the surface in-situ observations corrects the model profile. Hence, we don't expect such under-estimation in our reference profiles and did not see it in the model vs. in-situ comparison. In central Europe the MACC reanalysis over-estimates CO in the upper troposphere/lower stratosphere above an altitude of about 10 km (Inness et al., 2013). Due to the cross-dependency in the FTIR AVK this may also impact the comparison at lower altitudes.

However, it is interesting to note that the FTIR profiles showed the largest mole fractions at different levels for the two sites. Also, the model simulations showed similar tendencies at these altitudes, although with smaller amplitudes. At JFJ large variations close to the surface were caused by close-by emissions in the European planetary boundary layer that are vertically lifted to the site. At IZO larger variability occurred aloft and originated from long range transport from the North American east coast. One possible interpretation of the more pronounced vertical profile in the FTIR observations is that the FTIR detects CO at the correct vertical location but over-emphasizes its values. In turn the retrieval compensates the increased levels at other altitudes, but observes the total tropospheric column correctly.

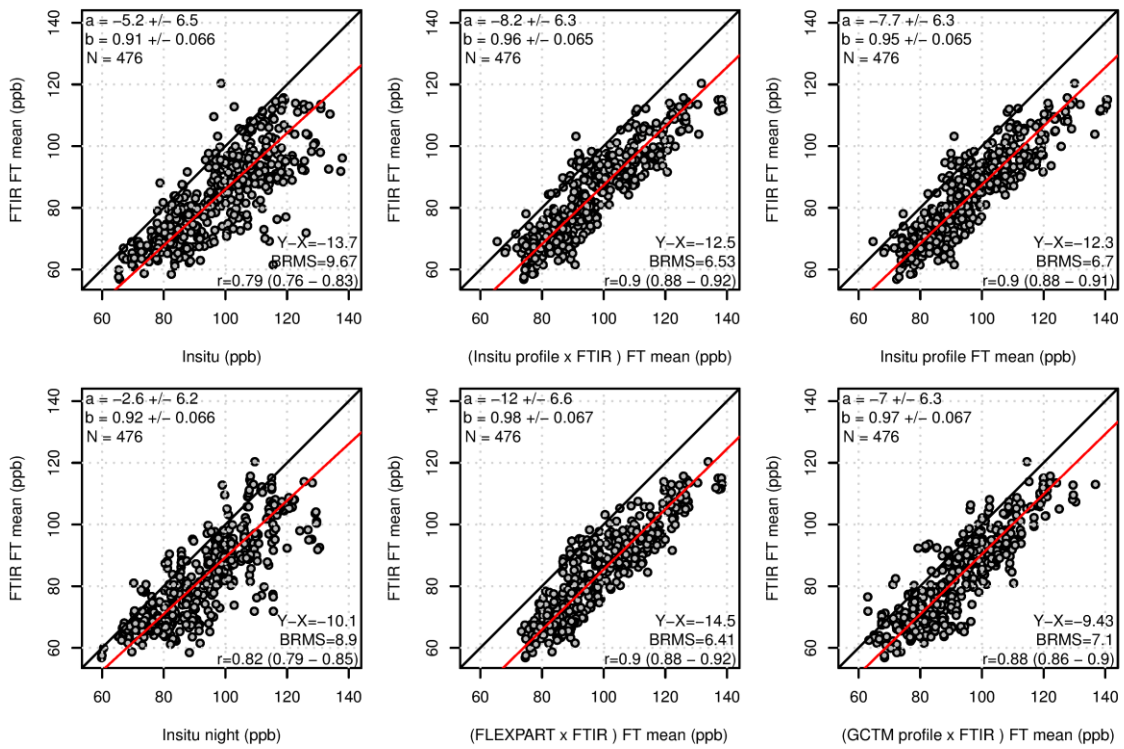

Figure 8: Same as Figure 4 but for CO as observed by FTIR at Izaña. Lower free tropospheric profiles were averaged up to 7.2 km asl. GCTM profiles were taken from MACC re-analysis.

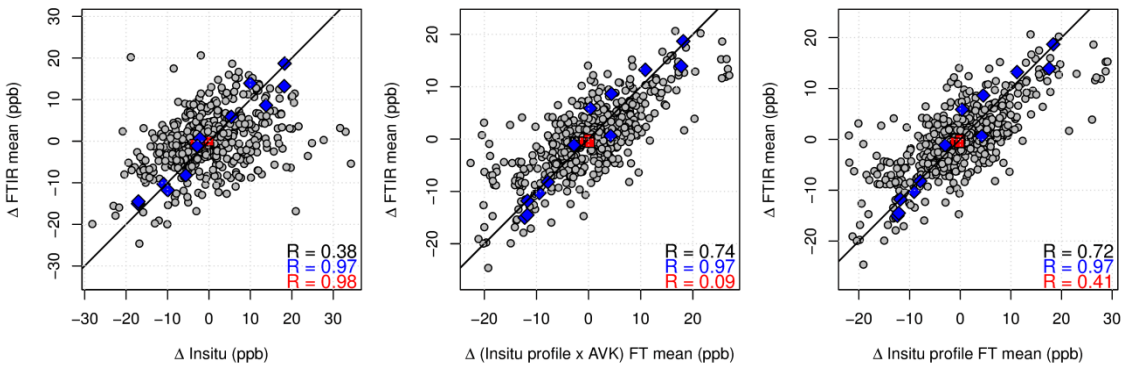

Figure 9: Same as Figure 5 but for CO as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.2 km asl.

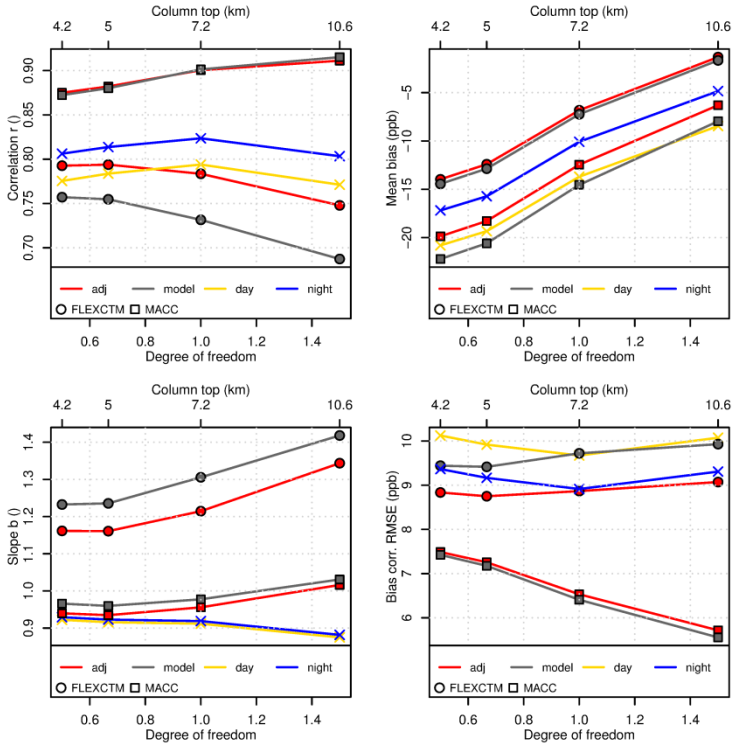

Figure 10: Same as Figure 6 but for CO as observed by FTIR at Izaña.

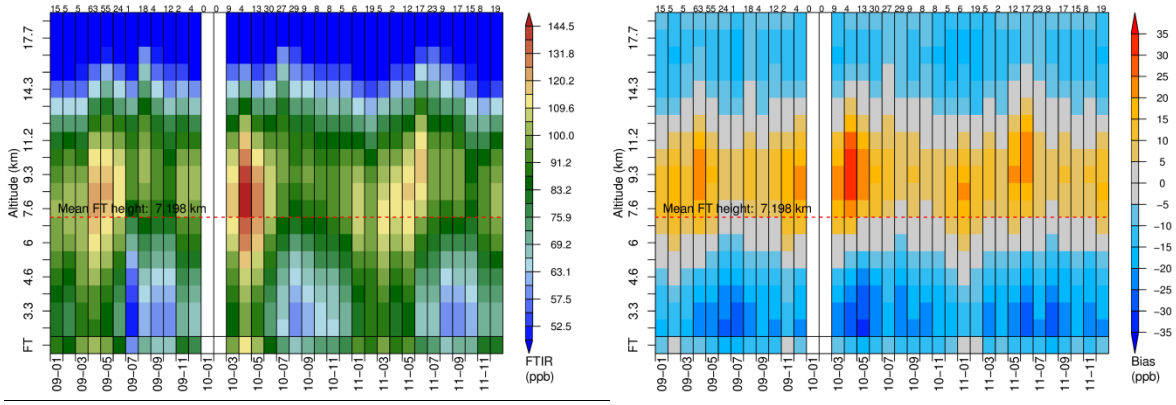

Figure 11: Same as Figure 7 but for CO as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile.

3.3. Ozone observations from FTIR

Typical row averaging kernels of the FTIR O_3 retrieval exhibited strongest sensitivities in the lower to middle troposphere. Row averaging kernels for different tropospheric sub-columns were relatively similar. The same general structure of the AVKs was inherent at both sites. On average the sum of the diagonal elements of the AVK reached 1 when summed over the lowest 7.4 and 11.6 FTIR sub-columns for JFJ and IZO, respectively. This corresponds to an average column top of 11.1 km asl and 11.3 km asl for JFJ and IZO. To restrict the influence from the UTLS region the main comparison results for the tropospheric column were derived for a tropospheric column height of 7.8 and 7.2 for JFJ and IZO, respectively, which represents an average degree of freedom of 2/3.

3.3.1. Jungfraujoch

The tropospheric O_3 mean column retrieved by FTIR at JFJ was averaged up to an altitude of ~7.8 km asl, which is the altitude for which on average the profile retrieval indicated a degree of freedom larger than 2/3. In Figure 9 this tropospheric column is compared against different reference data sets. The direct comparison with day-time surface in-situ data shows that FTIR mole fractions were generally larger as indicated by a slope of b=1.2. However, no overall bias was detected between the two datasets. That the FTIR retrieved tropospheric column would be larger than the surface in-situ is not surprising considering the general increase of ozone with altitude in the troposphere. Nevertheless, a reasonable correlation between both data sets was established with r=0.53. When comparing the day-time FTIR observations with the bracketing night-time surface in-situ observations the comparison statistics only improved marginally. This is in agreement with the observation of only minor systematic diurnal variations of surface O_3 at JFJ (Figure 2). The agreement between the FTIR tropospheric column and model-only columns (Figure 12, bottom center and bottom right) is generally much better than for surface in-situ data. Even better comparison statistics were obtained against the in-situ columns (our main reference; Figure 12, top center). In this case, the regression slope was close to unity (b=1.1) and the BRMS was reduced to ~7 ppb. A small overall bias of 3.5 ppb remained (FTIR – in-situ column).

To highlight the benefit of the FLEXPART simulations it should be pointed out that the model-only columns taken directly from the MACC re-analysis at the location of the FTIR observation did not agree as well with the FTIR as the FLEXPART simulations did (Figure 12, bottom left versus bottom center). This discrepancy between FTIR and MACC re-analysis seems to be mostly due to representativeness uncertainties that were reduced with the high-resolution and dedicated (specific for the sampling volume of the FTIR) FLEXPART simulations.

The benefit of referencing the FTIR column against the folded in-situ column is even more pronounced when looking at the sub-monthly variability (Figure 13). For the surface in-situ comparison most of the common variability originated from the corresponding seasonal cycle while the correlation for sub-monthly variability was relatively small (r=0.32, Figure 13 left). In contrast, the FTIR column correlated much better with the in-situ column for sub-monthly variability (r=0.61 before applying the AVKs and r=0.79 after applying the AVKs, Figure 13 center and right). Again, the negative correlations for inter-annual variability should not be interpreted, since they were derived from three years of data only and average values for these years were very similar.

As for CO the comparison statistics depend on the height of the tropospheric column. In Figure 14 it can be seen that correlation and regression slope for a comparison against in-situ columns and model-only columns further improve with a deeper tropospheric column. At the same time the comparisons with surface in-situ data deteriorate. However, mean bias and BRMS also slightly increased for the model-only and in-situ columns. While the comparison with in-situ columns was generally slightly improved as compared to model-only columns, this difference decreased with increasing column depth. The latter is not surprising since the influence of the surface in-situ data is limited to the lower troposphere and loses importance with increasing column height. It should be noted that for a tropospheric column reaching up to ~11 km, the altitude where the retrieval usually achieved to obtain one independent piece of information (DOF=1), the comparison with model-only and in-situ column was almost perfect with r=0.9, b=1, bias=4 ppb, and a BRMS of 8 ppb.

The monthly mean vertical O_3 profiles as retrieved by the FTIR are shown in Figure 15. The vertical and temporal variability displayed generally agrees with our expectations for a site in the northern hemisphere mid-latitudes, showing a higher tropopause during the summer (low values of O_3 reaching up to > 11 km) and a lower tropopause during winter. In addition, tropospheric O_3 decreased during fall and winter and was largest during spring and summer. When compared to the folded in-

situ profiles the FTIR profile showed a generally small, but positive bias in the troposphere (already seen above), which was increased in the upper troposphere during the summers of 2009 and 2011. In 2011 the overall agreement in the troposphere was improved. The positive bias may relate to a negative model bias that was observed for the MACC re-analysis when compared to ozonesonde and aircraft data (Inness et al., 2013). An average model bias of about -5 % was found for European sites. The bias however improved after 2008 when a problem with the variational bias correction was resolved (Inness et al., 2013) and the period covered in this study (2009 – 2011) should be unaffected. This is also supported by the very good agreement of the model data with the surface in-situ observations as shown in NORS deliverable D5.1 (Henne, 2014). Above the tropopause the FTIR was negatively biased as compared to the in-situ profile. While this bias appears relatively large in absolute terms (< -20 ppb), it was comparatively small in a relative sense (~5 %).

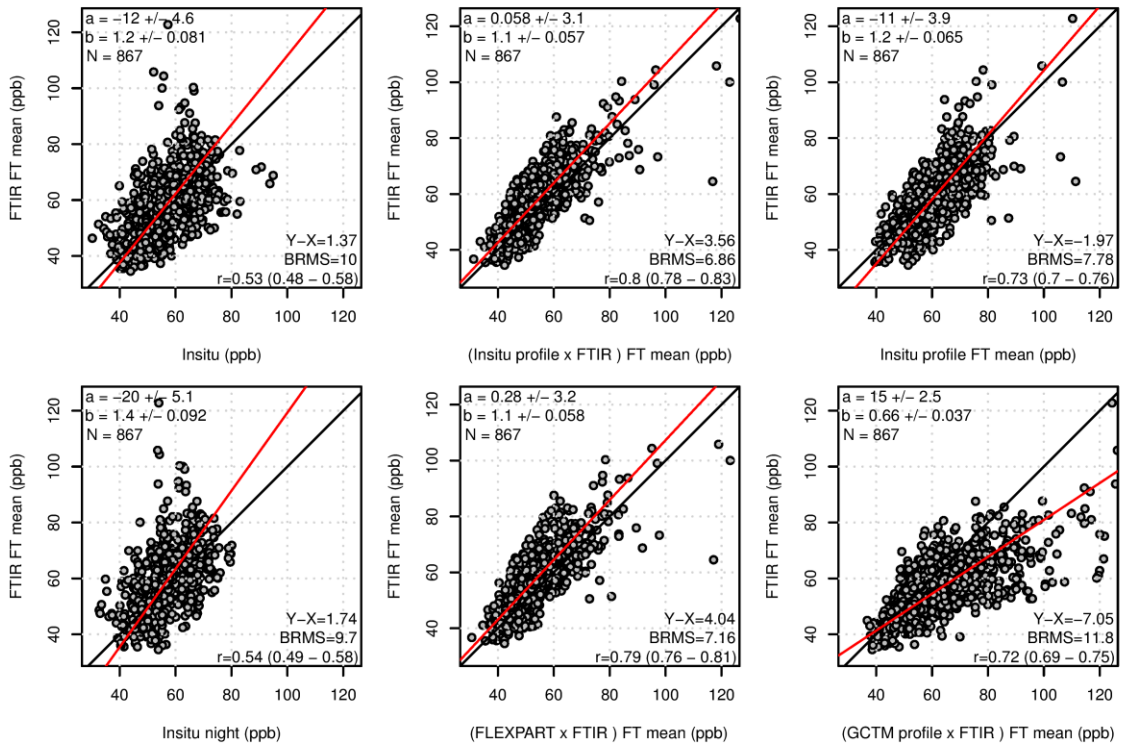

Figure 12: Same as Figure 4 but for O_3 observations from FTIR at Jungfraujoch. Lower free tropospheric profiles were averaged up to 7.8 km asl. GCTM background and profiles were taken from MACC re-analysis.

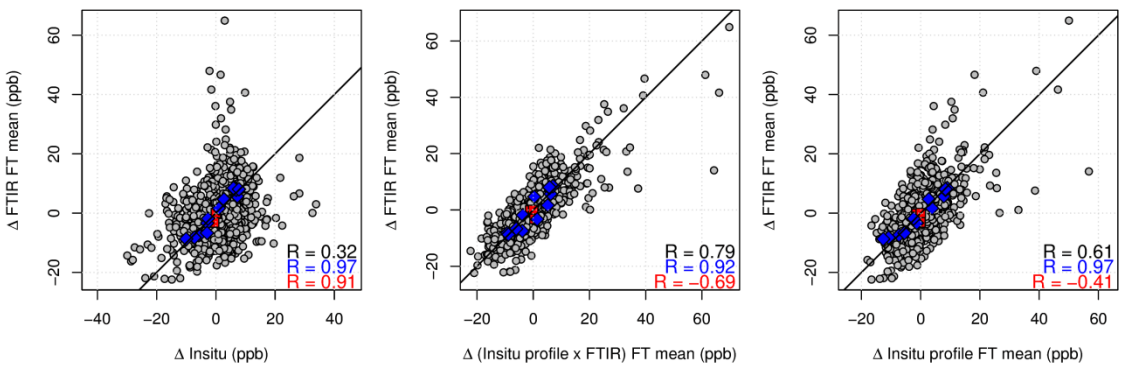

Figure 13: Same as Figure 5 but for O_3 as observed by FTIR at Jungfraujoch. The mean free tropospheric FTIR columns were analysed versus (left) surface in-situ observations at the same time as FTIR observations,

(center) folded in-situ columns, (right) in-situ columns. The FT averaging ranged from the surface to 7.8 km asl.

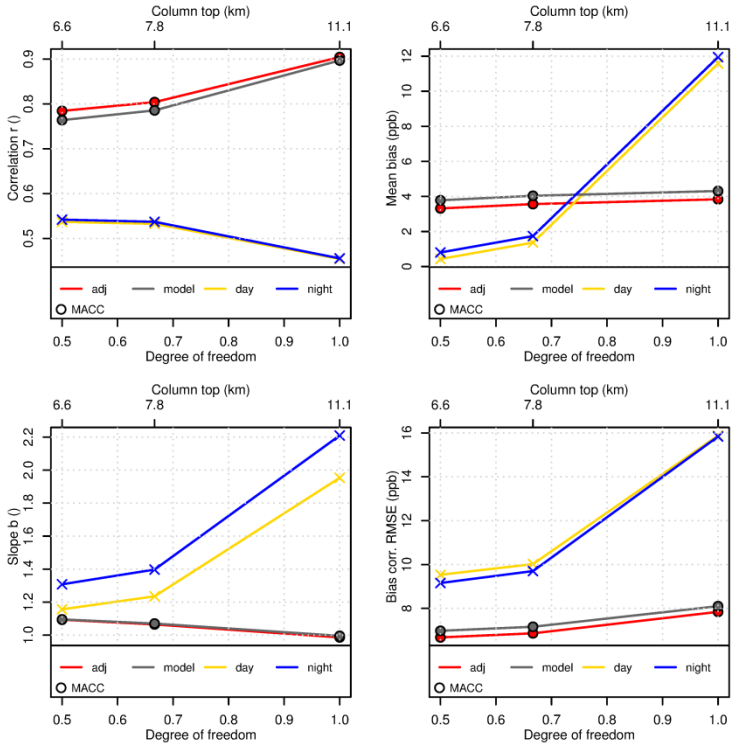

Figure 14: Same as Figure 6 but for O_3 as observed by FTIR at Jungfraujoch.

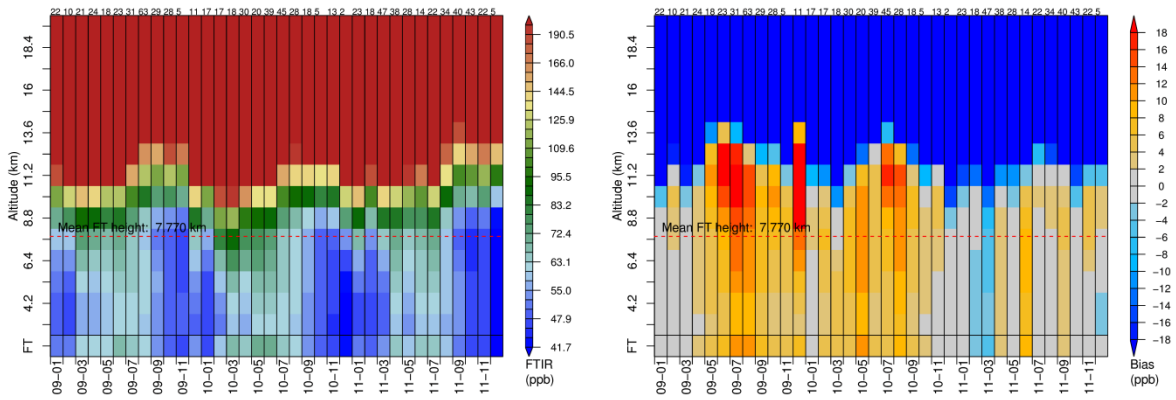

Figure 15: Same as Figure 7 but for O_3 as observed by FTIR at Jungfraujoch. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile.

3.3.2. Izaña

The tropospheric O_3 column as retrieved by FTIR at Izaña and averaged up to an altitude of 7.2 km asl did not compare well with surface in-situ observations taken at the same time as the FTIR observations (Figure 16, top left). As discussed in section 3.1 day-time transport of marine boundary layer air and additional surface deposition of O_3 along the mountain slope systematically decrease ozone surface concentrations at the site. Consequently, the direct comparison with FTIR observations, which were only minimally influenced by this up-ward transport, indicated a positive FTIR bias and regression slope larger than 1. The comparison of the FTIR column with night-time surface in-situ observations showed better agreement in terms of correlation but still showed a positive FTIR bias

and regression slope larger than 1 (Figure 16, bottom left). This is similar to the findings at JFJ and due to the generally increasing O_3 mole fractions with altitude. All comparisons against model-only and in-situ columns showed largely improved comparison statistics: correlations were larger, bias and BRMS were reduced (Figure 16). It is interesting to note that the comparison of the FTIR against the in-situ column exhibited better comparison statistics (increased correlation coefficient, reduced bias and BRMS) than the comparison against the in-situ column that was folded with the AVK before the calculation of the tropospheric mean. However, the regression slope for the first was relatively large (b=1.4) while it was closer to 1 (b=1.2) for the latter. The reason for this is currently not understood but may relate to uncertainties in the O_3 retrieval.

The results of the comparison of different temporal scales of variability are similar to those obtained for O_3 at JFJ. The separation of the signals reveals that especially on the sub-monthly time scale correlations between FTIR and reference data improved when in-situ columns were used instead of surface in-situ data (Figure 17, center vs. left). With respect to the previous observation of reduced correlation for folded in-situ profiles, it is interesting to note that when the variability was separated onto the different time scales, correlations for hourly variability were as good for the folded in-situ columns as for the non-folded in-situ columns (Figure 17, center and right). At the same time the correlation for inter-annual variability was very poor for the folded in-situ columns. Although, the inter-annual correlation should be viewed cautiously, this may suggest that the problems with the AVK could be related to temporal changes in the retrieval algorithm.

As for O_3 at JFJ comparison statistics improved for a deeper tropospheric column at IZO as well. For a tropospheric column reaching up to 11.4 km, the height for which the AVK indicated the retrieval of one independent piece of information, the regression slope between FTIR column and folded in-situ column was close to unity (b=1.05) and the BRMS was ~9 ppb (Figure 18). However, a positive FTIR bias of ~8 ppb remained. Again it can be seen that including the surface in-situ observations has only limited influence on the comparison statistics, but was strongest for a lower tropospheric column top.

The temporal evolution of the monthly mean FTIR profile of O_3 in Izaña was similar to that at JFJ in the sense that it also showed largest tropospheric ozone mole fractions in spring and summer and lowest mole fractions in fall and winter. Compared to JFJ the tropopause, where O_3 mole fractions sharply increase with altitude, was at higher altitudes at IZO. This is consistent with the sub-tropical location of the site. Consistent with the FTIR observations at JFJ a positive bias as compared to the folded in-situ profiles was also established for IZO. The magnitude of the bias was larger than at JFJ and it was visible for all years. Furthermore, the bias was more emphasized during summer when relatively high O_3 mole fractions occurred. As discussed for JFJ, the MACC O_3 re-analysis was shown to exhibit only small tropospheric biases after 2008. Our model comparison with night-time surface in-situ observations (not shown here) only revealed minor seasonality in the bias with simulated values being larger than in-situ values in summer and fall. While this model bias was corrected close to the surface it cannot be ruled out that the same positive model bias also occurred at higher altitude. Nevertheless, a positive model bias as compared to reality (surface in-situ) and a positive FTIR bias when compared to the in-situ profiles constructed from this model would result in an even more pronounced FTIR bias as compared to reality. In summary, the observed positive FTIR bias as compared to the in-situ profiles does not seem to originate from the model but seems to be driven by the FTIR observations.

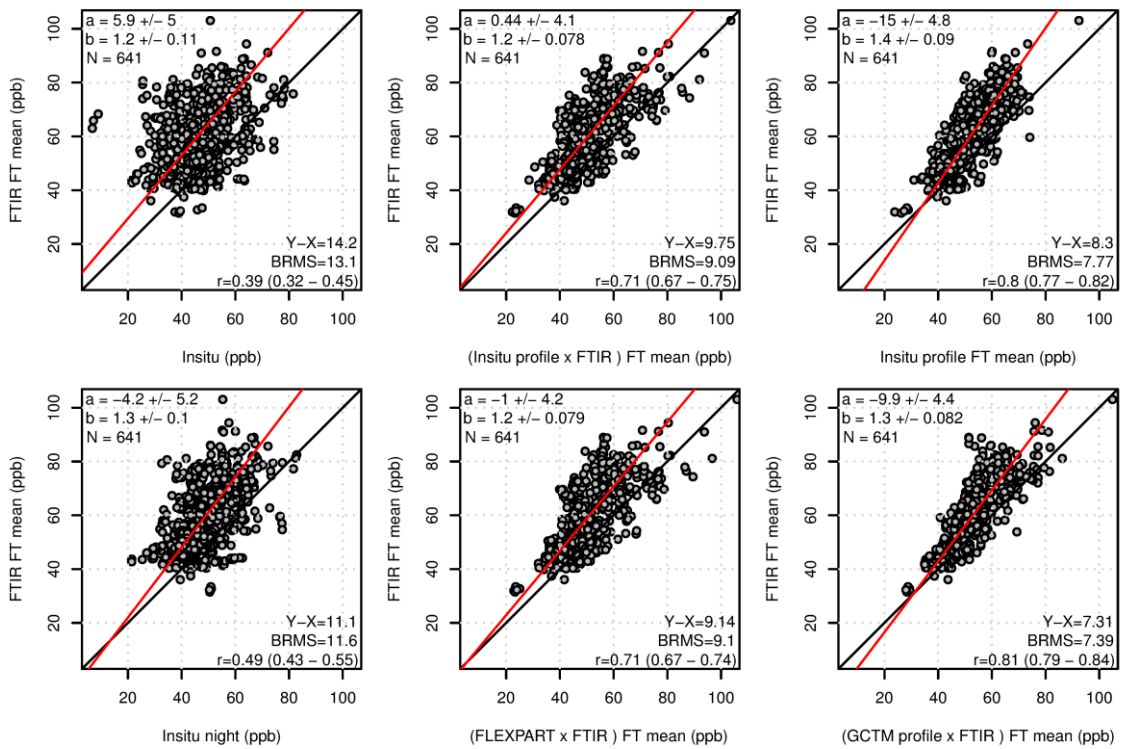

Figure 16: Same as Figure 4 but for O_3 observations by FTIR at Izaña. Lower tropospheric profiles were averaged up to 7.2 km asl. GCTM profiles were taken from MACC re-analysis.

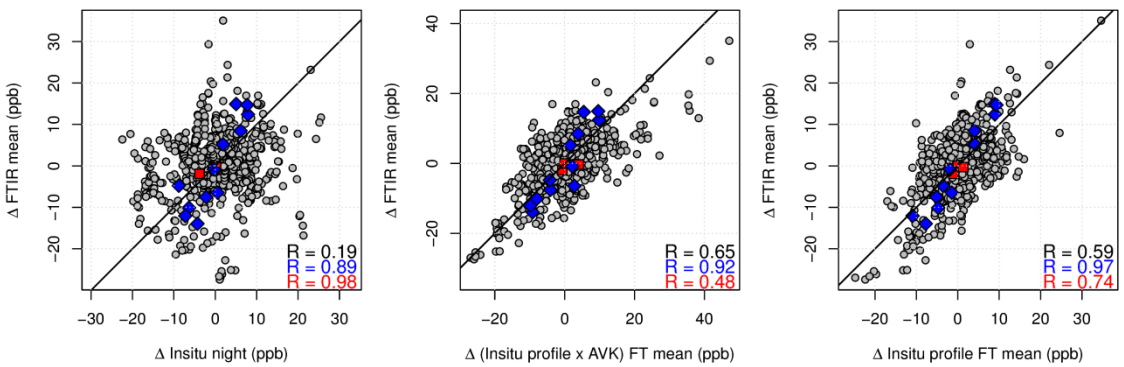

Figure 17: Same as Figure 5 but for O_3 as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) night-time surface in-situ observations, (center) folded in-situ columns, (right) in-situ columns. The FT averaging ranged from the surface to 7.2 km asl.

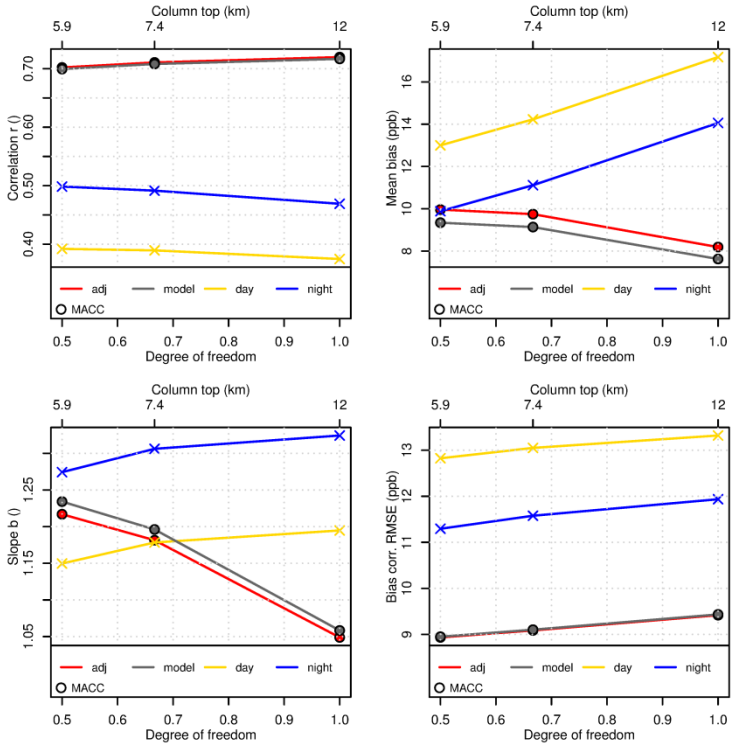

Figure 18: Same as Figure 6 but for O_3 as observed by FTIR at Izaña.

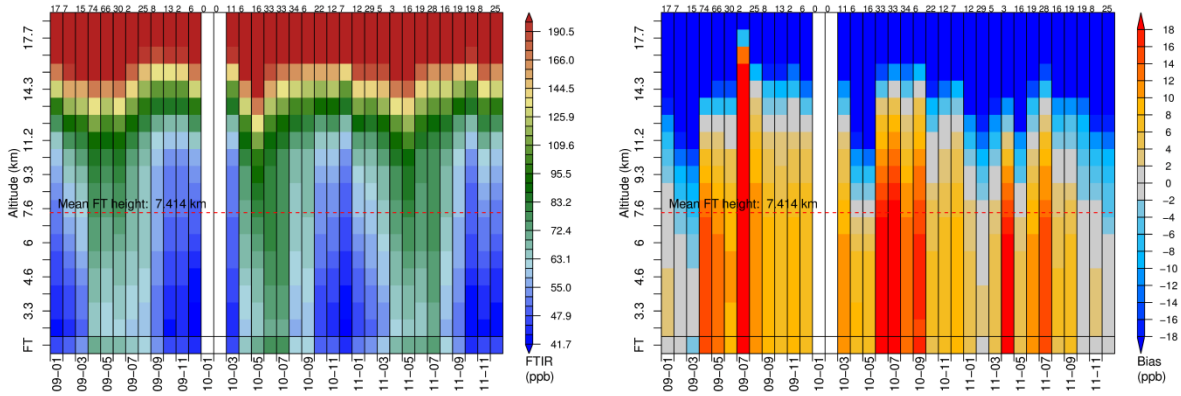

Figure 19: Same as Figure 7 but for O_3 as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile.

3.4. Methane observations from FTIR

The typical row averaging kernels for the FTIR CH_4 retrieval indicated strongest sensitivities to mid tropospheric CH_4, while sensitivities close to the surface were relatively small. This was similar for both sites, but with generally smaller sensitivities close to the surface at JFJ than at IZO. In contrast, sensitivities were more pronounced at mid-tropospheric levels at JFJ than at IZO. It should be pointed out that the methane line parameters remain unsatisfactory, still presenting some inconsistencies. To avoid producing unrealistic oscillations in the FTIR profiles, the retrieval is under-constrained to get a reduced but decent DOF of 2 on average for JFJ. On average the sum of the diagonal elements of the AVK reached 1 (indicating one independent piece of information from the retrieval) when summed over the lowest 10 and 12 FTIR sub-columns for JFJ and IZO, respectively. This corresponds to an average column top of 12.2 km asl and 11.6 km asl for JFJ and IZO, respectively. However, considerable cross influences from altitudes up to 20 km were also contained in the tropospheric

profile. To restrict the stratospheric influence on the main comparison results the tropospheric column height was restricted to 9.3 km asl and 7.6 km asl for JFJ and IZO, respectively. The influence of this selection is discussed along with the main findings.

3.4.1. Jungfraujoch

Consistent with the analysis of FTIR observed CO at JFJ also FTIR observed tropospheric CH_4 columns at JFJ showed less variability as compared to the surface in-situ observations (Figure 20, top left). The correlation coefficient for hourly data was r=0.53, lower than for the CO comparison. Comparison with night-time data did not significantly change the comparison statistics (Figure 20, bottom left). Better agreement was established between FTIR columns and the in-situ columns (Figure 20, top right), although the improvement was mainly in terms of a slope closer to unity and reduced bias and BRMS but not in correlation. Applying the FTIR AVK to the in-situ profile resulted in a surprising reduction of the data range of the folded in-situ columns (Figure 20, top center). While correlations slightly improved in this comparison and the BRMS was further reduced, the regression slope increased to b=1.5, indicating much larger variability in the FTIR observations than in columns obtained from the folded in-situ profiles. The same was observed when folding the model-only profiles (FLEXPART and GCTM, Figure 20, bottom center and right).

The separation of different scales of variability revealed that the observed seasonal cycle largely agreed between the FTIR column and the in-situ column (Figure 21, right). In contrast to the surface in-situ comparison, the FTIR sub-monthly variability was in slightly better agreement for the in-situ columns (r=0.47 vs. r=0.40). As seen above, when using the folded in-situ column a regression slope significantly larger than 1 was obtained both for the seasonal cycle and the sub-monthly variability (Figure 21, center).

Comparison statistics depended on the depth of the analysed tropospheric column (Figure 22). For FLEXCTM derived in-situ columns correlations and bias decreased with increasing column depth (Figure 22, red lines), while the BRMS slightly increased. The regression slope remained similarly large (b~1.5) for all column depths. As expected, comparison statistics got worse with increasing column height for surface in-situ comparisons. Two other baseline models were used for the construction of in-situ profiles. While comparisons using in-situ columns obtained with TM5 model data were similar to those with FLEXCTM model data, the comparisons with in-situ columns using MACC re-analysis data were much worse in terms of correlation, regression slope and BRMS. This was already seen when comparing the FLEXPART in-situ simulations based on MACC re-analysis with the surface in-situ data (Henne, 2014).

The monthly time series of the FTIR obtained CH4 profiles at JFJ are shown in Figure 23. It is interesting to note that the time series shows almost no CH_4 variability in the lowest layers. This corresponds to the low sensitivities of the retrieval at these levels as it is reflected in the AVKs. Consequently, the values in this layer remained close to the a-priori values used in the retrieval. Otherwise, largest CH_4 mole fractions were observed at an altitude around 6 km and during winter. The monthly bias plot showed a consistent positive bias in the troposphere and strong negative bias in the lower stratosphere. The latter was less pronounced when using TM5 in the construction of the in-situ profiles. Tropospheric biases were largest at times of largest FTIR observed mole fractions.

The reason for the systematic change in regression slopes against the FTIR observation when folding the in-situ and model-only profiles with the FTIR AVK needs to be discussed in more detail. As discussed above, the AVKs indicated very small sensitivity in the lowermost troposphere. Folding the reference profiles with the AVK thus resulted in values that were very close to the a-priori values at these levels. The same is true for the FTIR profiles as well, and overall no bias was seen between the two in the lowest FTIR sub-column (Figure 23, right). In contrast to the reference profiles, the FTIR

showed larger CH_4 values again at a layer around 6 km asl, where the folded profiles did not show large CH_4 values. Similarly as discussed above for CO, it seems that some of the near surface variability in CH_4 was detected by the retrieval in the mid troposphere although the AVK does not reveal any surface influence for these levels and consequently folding reference profiles with the AVK results in generally lower column values as in the FTIR column.

Another possibility for the increased slope when comparing the FTIR columns versus the folded in-situ columns is that of cross influence from the stratosphere. However, when comparing the datasets for different tropospheric column depths (Figure 21) similar values were observed for the regression slope between FTIR columns and folded in-situ columns. Furthermore, comparison results were very similar for two different baseline models (FLEXCTM and TM5) that were used for the FLEXPART simulations. Both models showed relatively large differences in the UTLS region. So if there was an important influence from the UTLS, one would expect comparison results to be very different, which they were not. For FLEXCTM no independent model validation against profile data is available. TM5 was validated against a number of different aircraft data sets (Bergamaschi et al., 2013). A small positive model bias of about 5-10 ppb was observed for the UTLS region (200-300 hPa) on average. Average vertical gradients in the UTLS were sharper and average mole fractions in the UTLS were smaller in TM5 as compared to FLEXCTM. Despite these differences, the FTIR comparison against the surface in-situ profiles constructed from these two models was very similar. A third in-situ column was constructed using CH_4 from the MACC re-analysis. While in general these MACC in-situ columns did not compare as well with the FTIR observations as those of the two other models (Figure 21), the same behaviour (increased regression slope) was observed after folding with FTIR averaging kernel. In addition to the use of different baseline models, we also tried to restrict the stratospheric influence in the tropospheric part of the FTIR profile by applying the profile correction suggested by Sepúlveda et al. (2014), which focuses the AVK onto the tropospheric column. While the correction removed part of the observed bias it did not change the regression slope. From this we conclude that the decrease in variability in the in-situ columns after folding with the AVK does not originate from possible model biases in the UTLS, but rather from the limited sensitivity of the JFJ FTIR retrievals to the lowermost levels. Efforts are ongoing to improve the methane line parameters. Their implementation should help pushing the JFJ FTIR retrievals further, allowing increasing the DOF, in particular the sensitivity to the first layers, while producing more realistic vertical distributions of methane.

Considering the problems that occurred when applying the AVK to the in-situ and model-only profiles, it seems to be reasonable to report the comparison statistics of the FTIR column against the in-situ column for which the AVK was not applied. For the tropospheric column reaching on average up to 12.2 km asl at least a DOF of 1 was obtained by the retrieval. The correlation for this tropospheric FTIR column with the in-situ column was around 0.6, while a slope of b=1.0 and a BRMS of ~13 ppb was determined. The overall bias for the tropospheric column was -8 ppb (FLEXCTM), but depended on the choice of baseline model (-20 ppb for TM5).

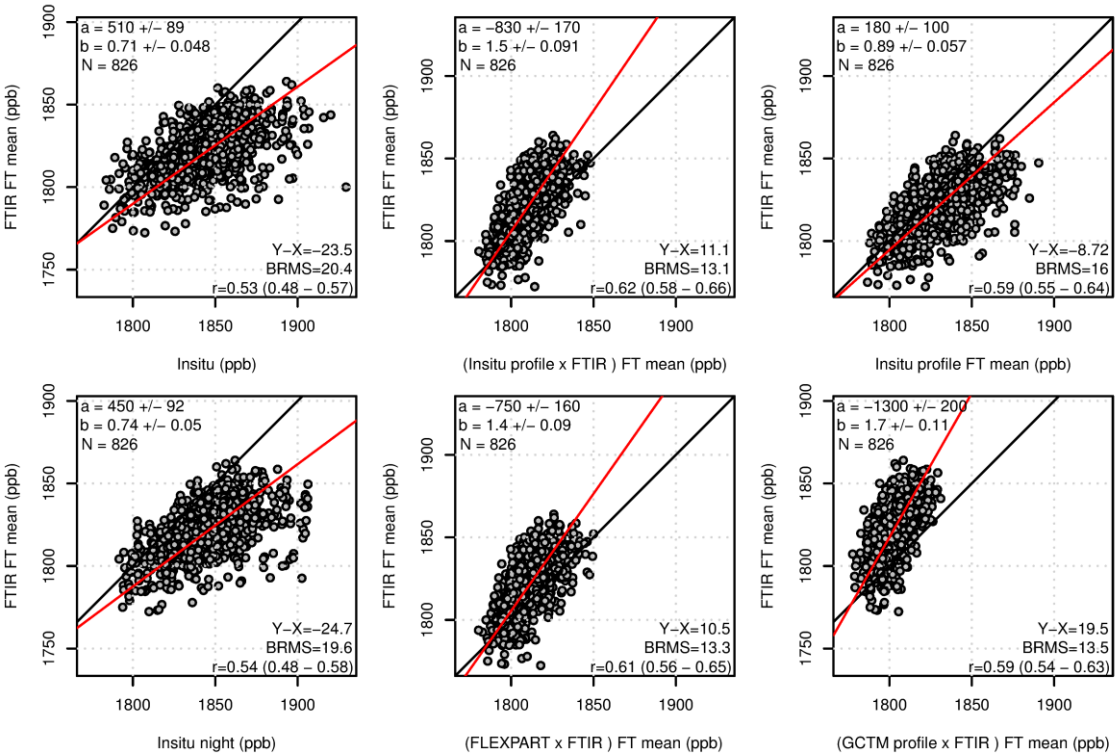

Figure 20: Same as Figure 4 but for CH_4 as observed by FTIR at Jungfraujoch. Tropospheric columns were averaged up to 9.3 km asl. GCTM profiles were taken from FLEXCTM simulations.

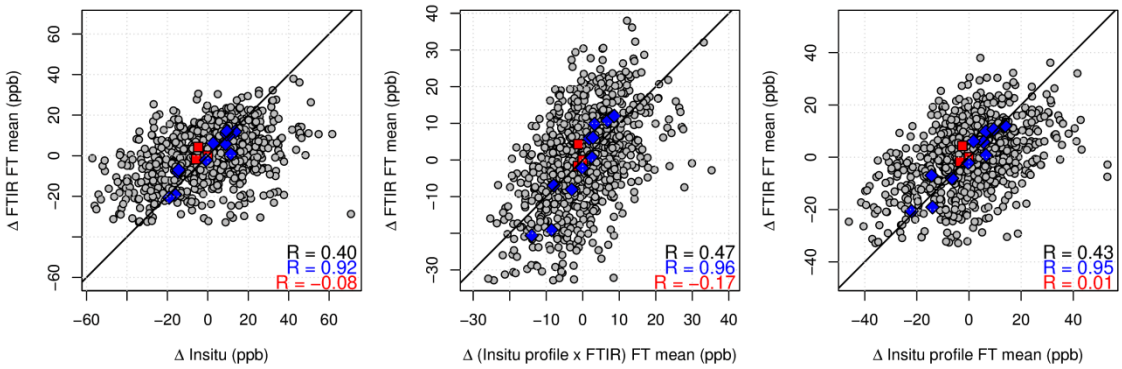

Figure 21: Same as Figure 5 but for CH_4 as observed by FTIR at Jungfraujoch. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 9.3 km asl.

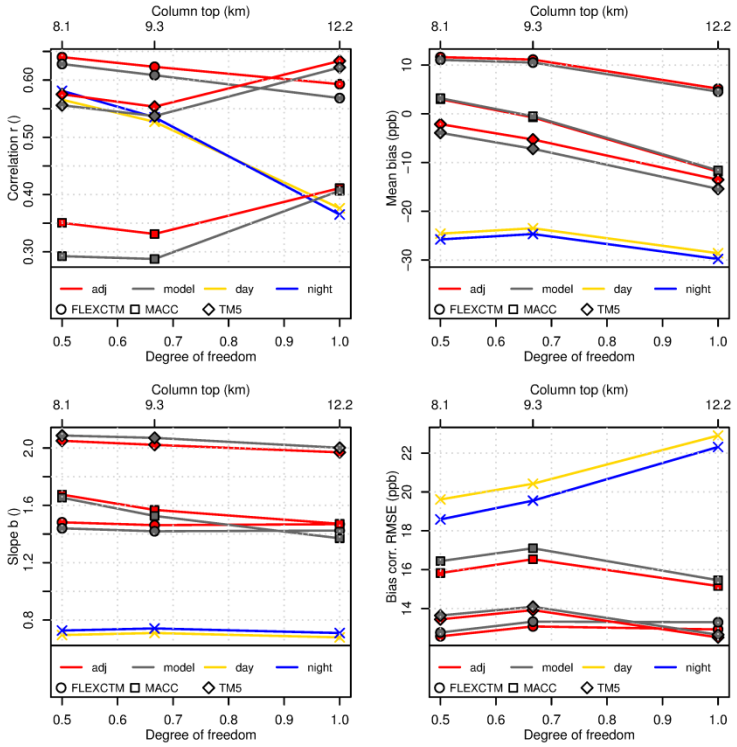

Figure 22: Same as Figure 6 but for CH_4 as observed by FTIR at Jungfraujoch.

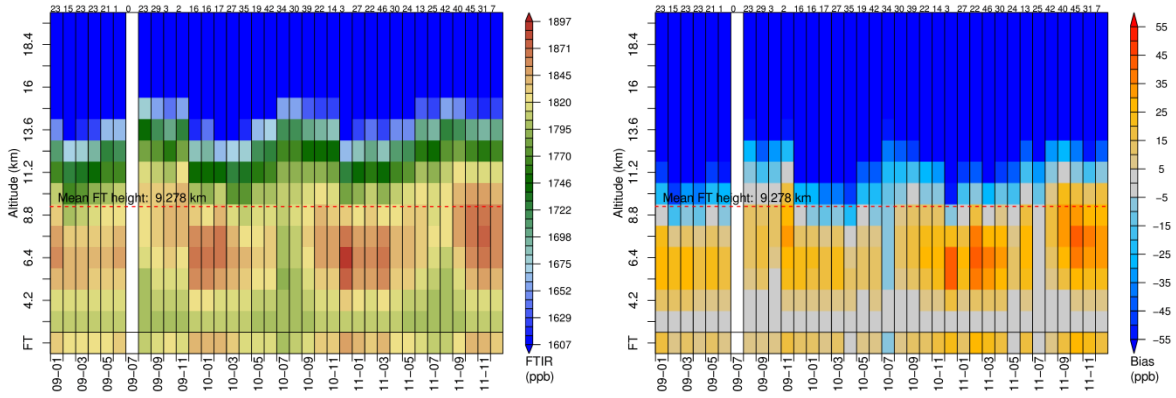

Figure 23: Same as Figure 7 but for CH_4 as observed by FTIR at Jungfraujoch. (left) FTIR mole fractions, (right) FTIR bias as compared to reference profile.

3.4.2. Izaña

At Izaña FTIR observed tropospheric CH_4 columns were calculated up to an altitude of 7.6 asl. When compared to day-time and night-time surface in-situ observations, FTIR columns showed less overall variability and regression slopes smaller than 1 were obtained. In general, there was large scatter in the comparisons (BRMS of ~18 ppb) and a positive FTIR bias of ~60 ppb was detected. The comparability improved when using night-time in-situ observations instead of day-time data. Our results are similar to those obtained in Sepúlveda et al. (2012) in terms of correlation and regression slope. However, they did not report a relevant bias between the two data sets. Part of our large bias may be explained by our transformation of in-situ mixing ratios from dry air mole fractions to moist air mole fractions, thereby decreasing them by about 1 % (~19 ppb) on average. No such correction was reported by Sepúlveda et al. (2012).

The comparability was increased in terms of correlation and BRMS when using model-only and in-situ columns instead of surface in-situ data, but the bias further increased (77 ppb for the folded in-situ profile, Figure 24 top center). In addition, a regression slope significantly larger than unity was obtained for all model derived profiles. For the folded profiles this is similar to the results for CH_4 at JFJ, where it was argued that the AVKs introduced a problem in the comparison. Since for IZO also the in-situ columns that were not folded with the AVK showed a slope much larger than unity a general overestimation of tropospheric CH_4 by the FTIR could be suspected. Surface in-situ values (dry air mole fractions) were in very good agreement with simulated dry air mole fractions (no bias for FLEXCTM). Therefore, we don't think that the observed FTIR bias and large regression slope originate from unrealistic reference profiles.

As seen for all previous comparisons, the improved correlation in the FTIR comparison against in-situ columns mainly resulted from an increased correlation on the sub-monthly time scale (r=0.3 to r=0.66; Figure 25). However, in the case of CH_4 at IZO correlation coefficients also slightly increased for the seasonal time-scale, when using in-situ columns instead of surface data.

Correlations and BRMS were not very sensitive to the tropospheric column height for the in-situ comparisons and TM5 derived in-situ columns (Figure 26). The bias decreased with increasing column height. It remained especially large for in-situ columns derived from FLEXCTM (~70 ppb) and was considerably smaller for TM5 derived in-situ columns (~40 ppb). In contrast to in-situ simulations based on FLEXCTM, TM5 based simulations showed an average surface bias of ~18 ppb, explaining part of the difference between the two bias estimates. Both FLEXCTM and TM5 derived in-situ simulations did not show a large bias at JFJ. Also the column comparison at JFJ did not reveal a large difference between in-situ column and FTIR column. Therefore, it seems likely that either tropospheric CH_4 is systematically over-estimated by the FTIR at IZO or that the models did not reproduce the tropospheric profile shape at IZO. The latter seems unlikely since IZO is situated away from major CH_4 emission sources and agreement with surface in-situ data was very good. Overall the influence of surface in-situ data on the in-situ profiles was small, which was to be expected due to small surface sensitivity of the retrieval. In the case of FLEXCTM, the simulated in-situ concentrations were in close agreement with the surface in-situ measurements, which only led to minor adjustments of the model-only profiles.

The temporal evolution of the FTIR observed tropospheric profile can be seen in Figure 27. As seen before, the variability of the tropopause height can be clearly discerned from the sharp CH_4 gradient in the UTLS. Elevated CH_4 concentrations reached higher during summer when mean tropospheric concentrations were smallest. In contrast to the FTIR observations at JFJ, largest CH_4 mole fractions were retrieved close to the surface, while at JFJ these prevailed in an elevated layer. The strong FTIR bias as compared with the folded in-situ profiles did not show a clear seasonal dependence. It was positive throughout the whole troposphere up to 14 km asl. Above a negative FTIR bias prevailed throughout the whole comparison period. The negative bias in the lower stratosphere is consistent with the one observed at JFJ, while the positive bias in the troposphere was much less pronounced at JFJ. The positive tropospheric bias at IZO showed an increasing trend over the three year period (Figure 27, right). The trend mainly originated from the trend in the FTIR observations, while no such trend could be seen in the FLEXCTM derived profiles. In contrast, TM5 derived profiles also showed a small positive trend over the investigation period.

In summary, one can say that the CH_4 tropospheric columns as observed by FTIR over IZO agreed fairly well with the reference profiles in terms of correlation. Integrating model profiles and in-situ observations and model profiles, helped to improve the comparability of the FTIR observations. However, a large FTIR bias of 40 ppb to 70 ppb for the tropospheric column remained. The magnitude

of the bias depends on the baseline model used in the FLEXPART simulations and was largest for the baseline model that showed best performance for the surface in-situ observations.

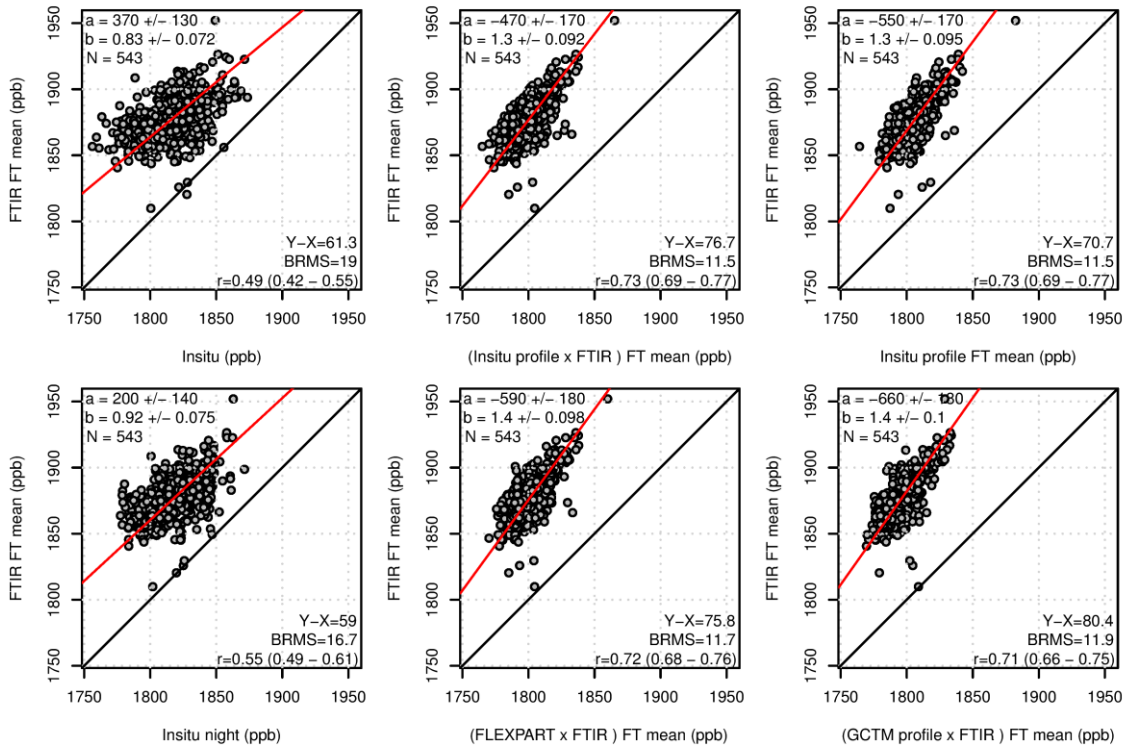

Figure 24: Same as Figure 4 but for CH_4 as observed by FTIR at Izaña. Free tropospheric profiles were averaged up to 7.6 km asl. GCTM profiles were taken from FLEXCTM.

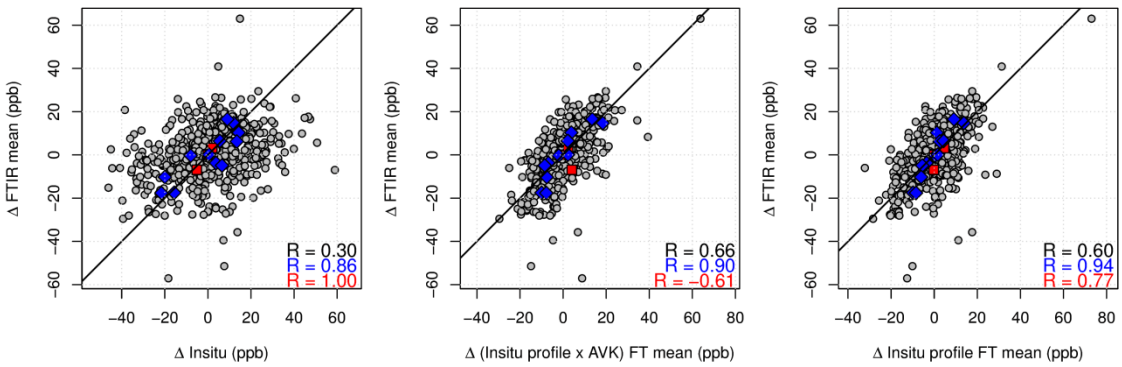

Figure 25: Same as Figure 5 but for CH_4 as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.6 km asl.

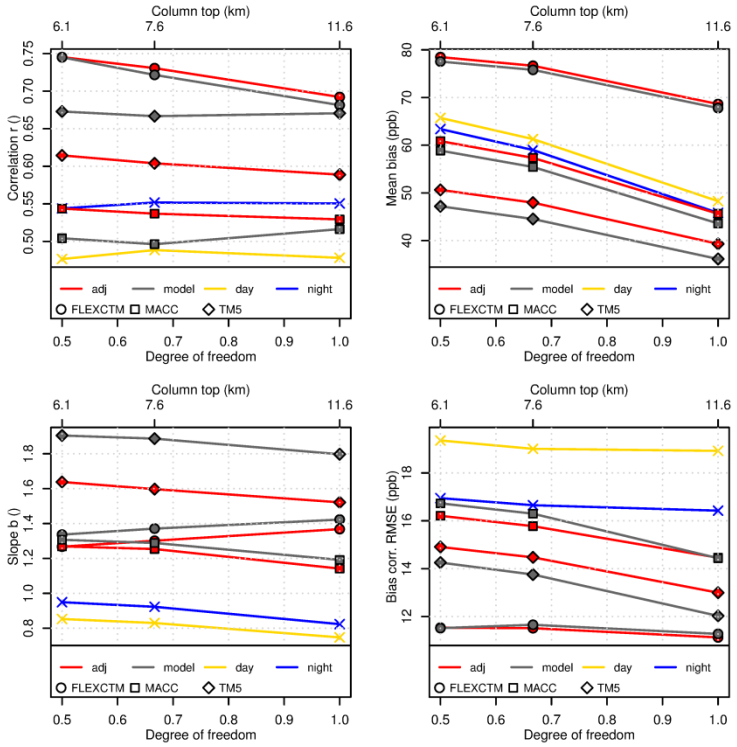

Figure 26: Same as Figure 6 but for CH_4 as observed by FTIR at Izaña.

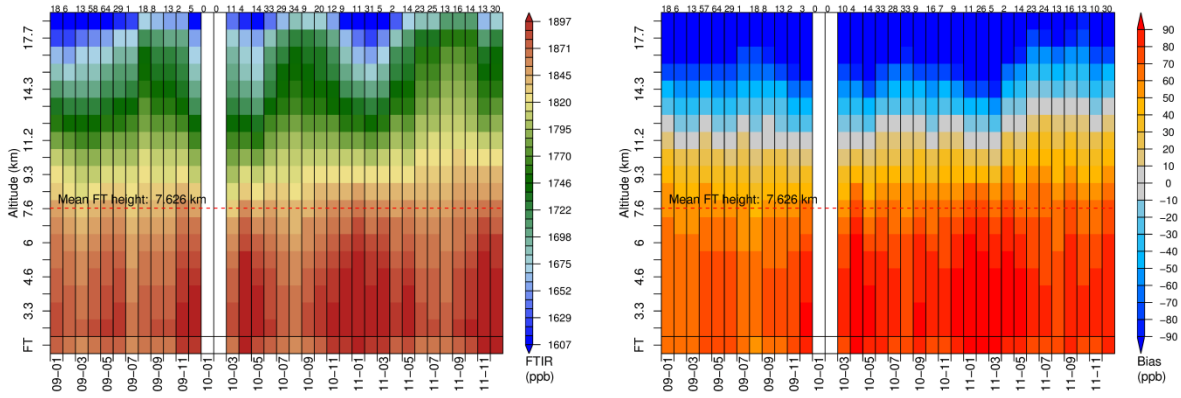

Figure 27: Same as Figure 7 but for CH_4 as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile.at IZO.

3.5. Nitrogen dioxide observations from MAXDOAS

3.5.1. Jungfraujoch

The MAXDOAS profile retrieval provided NO_2 observations up to an altitude of 7 km asl. The retrieval was very sensitive to NO_2 close to the surface. The AVKs indicated that already at an altitude of 4.1 km asl (or 500 m above the instrument) at least one independent piece of information could be retrieved (DOF>=1). While this relatively shallow vertical column should in general improve the comparability with surface in-situ data, a representativeness problem arises from the long horizontal viewing path of the instrument. In the case of the MAXDOAS at JFJ, the viewing direction was North-West towards pre-Alpine terrain. Often the viewing distances were long enough to reach over the flat terrain of the Swiss plateau. Since the mountains may considerably alter the vertical trace gas stratification, usually lifting trace species to higher altitudes during day-time over the mountainous

terrain, it is likely that the NO_2 concentrations along the horizontal viewing path of the MAXDOAS decrease with distance. FLEXPART backward simulations considered the viewing geometry and sampling volume of the MAXDOAS (see section 2.3.1). Hence, the representativeness problem should be resolved by the transport in the model and the further in-situ data integration. Overall FLEXPART showed problems in reproducing the surface in-situ observations of NO_2, which can be related to the simplifications in the model approach (see Section 2.3.1). However, the model adjustment by the surface in-situ observations compensated for these problems. Unfortunately, the adjustment was often not successful due to near zero background or pollution contributions (see method details in NORS D5.1, Henne, 2014). Therefore, it was necessary to additionally filter the comparison for cases when the adjustment produced unrealistic corrections in the model profile. Another possible source of problems with the model adjustment was the presence of large in-situ NO_2 observations that were most likely due to local activities around the site (e.g. construction work, snow cats, and helicopters). While such episodes are usually manually removed from the time series, some outliers were still present before the filter was applied. For the two year comparison period a total of 915 hourly comparison pairs remained after filtering.

The direct comparison of the MAXDOAS lower tropospheric columns (up to 4.1 km asl) with the surface in-situ observations indicated a moderate correlation (r=0.59) but large scatter also at large mole fractions. The BRMS of 90 ppt was large compared to the average surface in-situ NO_2 levels of 100 ppt. The estimated bias of -30 ppt was in the range of the detection limit of the in-situ analyser. The comparison of the MAXDOAS column with in-situ and model-only columns showed an improved agreement in terms of larger correlation coefficients (~0.7) and reduced BRMS and bias. The model-only columns (especially the one directly taken from the MACC re-analysis) showed large scatter when compared with the MAXDOAS. The adjustment of the model profiles towards the in-situ observations largely improved the comparability. In case of the folded in-situ column the correlation coefficient increased to 0.7 and the BRMS decreased to 60 ppt. Only the regression slope increased to a value significantly larger than unity (b=1.2), which could indicate a slight over-estimation of the lower tropospheric NO_2 mole fractions by the MAXDOAS. However, due to the relatively large uncertainty of the in-situ analyser and the generally large model uncertainty in the case of NO_2 as compared to the previously discussed species, the problem may also lie on the side of the reference profile. In contrast to the previously discussed FTIR retrievals the MAXDOAS AVKs are relatively focused and, especially for the lowest sub-column, did not show much cross talk from other levels. Therefore, it is not surprising that applying the AVKs to the reference profiles did not significantly change the comparison (Figure 28, top left).

Different scales of variability were resolved by the MAXDOAS and the reference datasets in a similar way. Like for the FTIR, the comparison for hourly samples benefitted most from the inclusion of model profiles, while the reproduction of the seasonal cycle was similar also for the in-situ surface observations (Figure 29).

While it was said before, that the AVKs for the MAXDOAS were relatively focused, we still analysed the dependence of the comparison statistics on the column depth. Correlation coefficients generally increased and BRMS decreased with increasing column depth (up to 4.5 km asl) for the in-situ column comparison Figure 30. For all column depths, including the in-situ data in the simulated profiles strongly improved the comparability of the MAXDOAS data in terms of correlation, BRMS, and bias. As expected, direct comparisons with surface in-situ data showed less favourable comparison statistics with increasing column height.

The monthly mean vertical profiles as obtained by the MAXDOAS showed relatively small NO_2 values below 4 km asl, a maximum around 4.5 km asl and exponentially decreasing values above. This exponential decrease was mainly owed to the exponentially decreasing a-priori profile used in the

retrieval. No distinct seasonal cycle was visible in the MAXDOAS profiles. While in 2011 a summer time maximum was visible, winter time maxima seemed to dominate in 2012. The overall MAXDOAS bias as compared to the folded in-situ profiles was positive above 4.3 km asl (possibly driven by the a-priori profile), while it was negative below 4 km asl for the year 2011 and near zero for 2012.

Overall and considering the relatively low NO_2 levels the agreement between the MAXDOAS observations and the in-situ profiles was remarkably good.

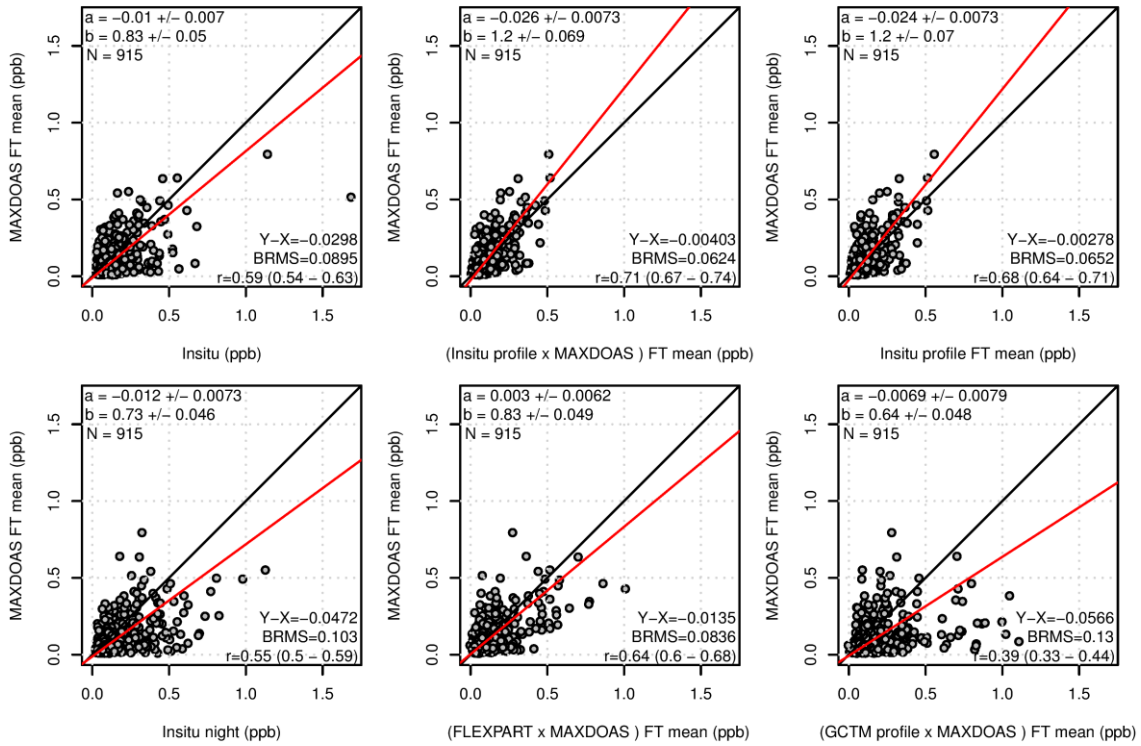

Figure 28: Same as Figure 4but for NO_2 as observed by MAXDOAS at Jungfraujoch. Tropospheric columns were averaged up to 4.06 km asl. GCTM profiles were taken form MACC re-analysis.

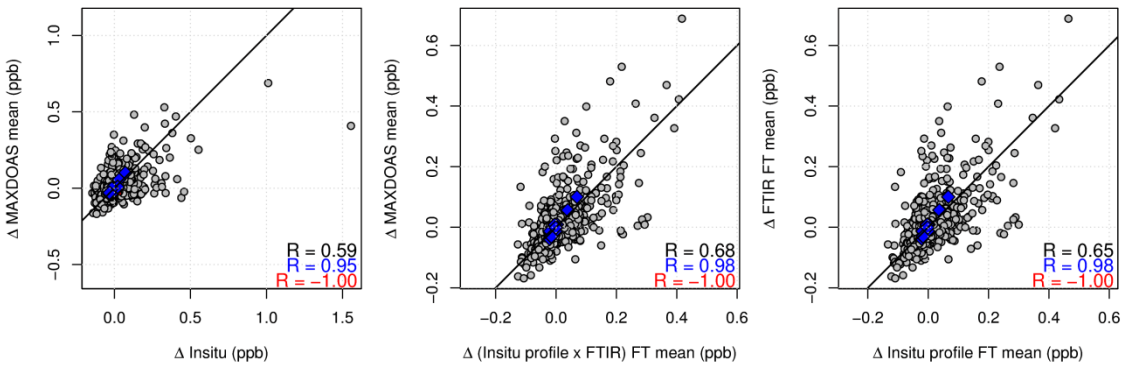

Figure 29: Same as Figure 5 but for NO_2 as observed by MAXDOAS at Jungfraujoch. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as MAXDOAS observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 4.06 km asl.

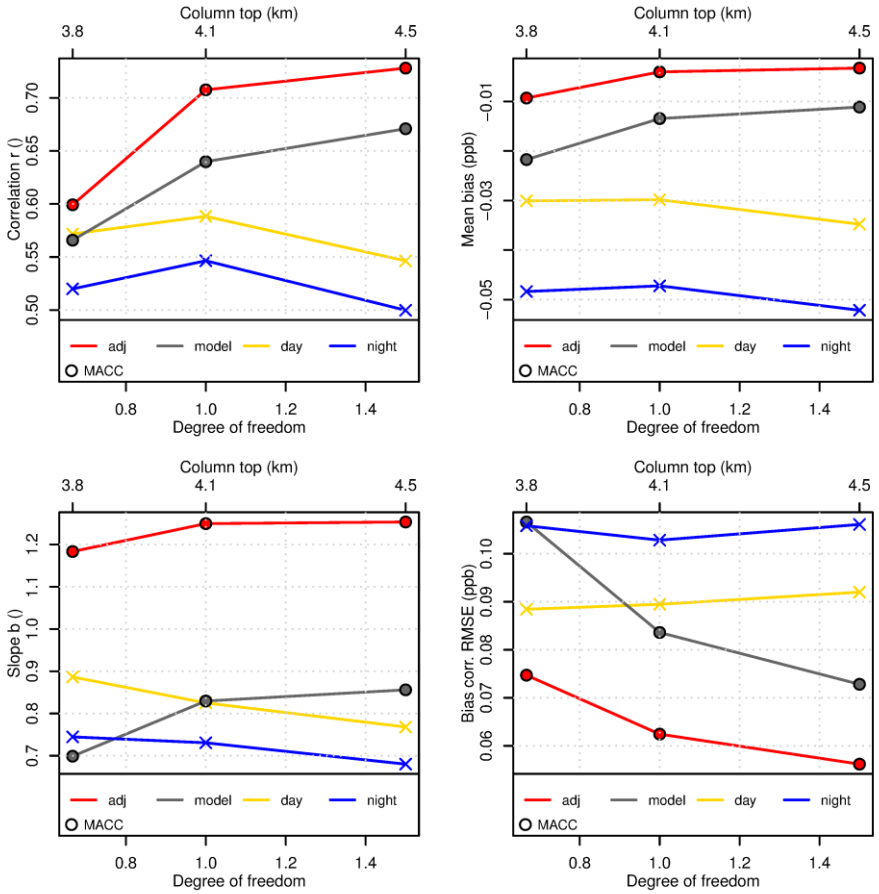

Figure 30: Same as Figure 6 but for NO_2 as observed by MAXDOAS at Jungfraujoch.

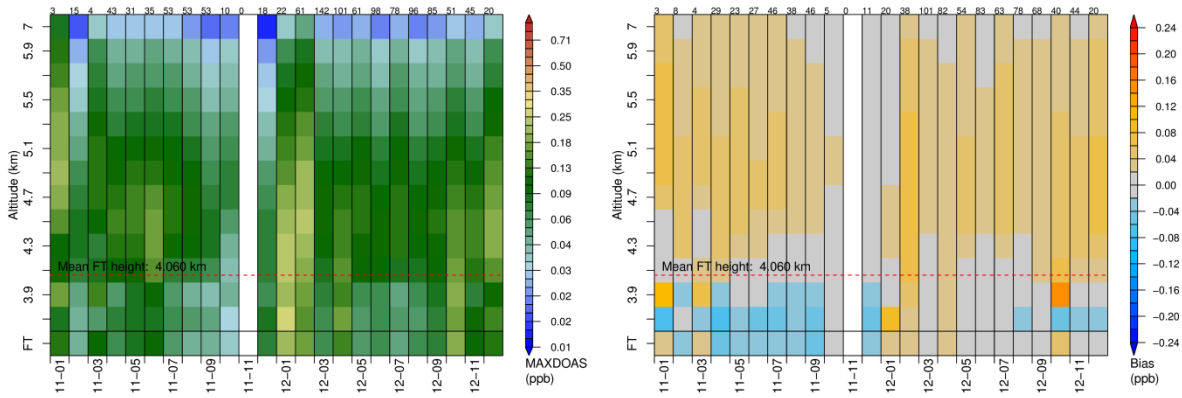

Figure 31: Same as Figure 7 but for NO_2 as observed by MAXDOAS at Jungfraujoch. (left) MAXDOAS mole fractions, (right) MAXDOAS bias as compared to folded in-situ reference profile.

3.5.2. Izaña

MAXDOAS NO_2 time series at Izaña were retrieved with the modified geometric approach (MGA) (Gomez et al., 2014). The retrieval focuses on the mean mole fraction in a horizontal line of sight from the instrument. Hence, no profile information was used in the inter-comparison. However, with the help of the FLEXPART simulations the in-situ observations were transferred onto the volume sampled by the MAXDOAS, which was assumed to be along the viewing direction of the instrument up to a distance, *d*, as determined by the retrieval (see section 2.3.1 for details). To be consistent with the

discussion of the profile retrievals, we will refer to the adjusted model concentration along the MAXDOAS sampling volume as in-situ columns as before. In addition to FLEXPART derived in-situ columns also GCTM (MACC re-analysis) values were interpolated to the MAXDOAS sampling volume and were adjusted by the surface in-situ observation following the same approach as in the case of FLEXPART.

The direct comparison of the MAXDOAS horizontal mean NO_2 mole fractions with surface in-situ observations at the observatory were hindered by the strong influence of local NO_2 sources in the island environment. For NO_2 the increase of day-time in-situ mole fractions at the observatory is especially pronounced as compared to other trace species, because emissions from a small power plant situated at the shoreline south-west of the station, was often directly transported up the slopes of the mountain. Since NO_2 levels in the marine environment and also the overlying free troposphere are very low the contrast to the advected power plant plume was especially large. The MAXDOAS usually was not influenced by this NO_2 plume because it was integrating over a rather long path pointing away from the island. Only during special situations when the mean wind at the height of the observatory transported the NO_2 plume in the viewing direction of the MAXDOAS, increased MAXDOAS NO_2 mole fractions were observed. In addition, there were situations when no NO_2 plume reaches the observatory and a comparison between MAXDOAS and surface in-situ observations should have been possible. In contrast to O_3 observations, it was also not possible to compare day-time MAXDOAS observations to night-time surface in-situ observations that should not be influenced by up-slope transport. However, night-time NO_2 mole fractions differ significantly from day-time values due to a shift between NO_2 and NO due to changed photochemical conditions at night. Hence, it was necessary to develop a filter for day-time observations that were not influenced by the advection of a NO_2 plume. Constructing such a filter from local wind observations failed and did not realistically filter large NO_2 mole fractions. Therefore, a simple filter criterion based directly on the surface in-situ observation of NO_2 was used. The filter selected only cases for which the surface in-situ NO_2 was smaller than 150 ppt and larger than -25 ppt (which is the negative value of the detection limit of the in-situ analyser). Furthermore, the comparison period had to be limited to Aug 2011 to Aug 2012 due to obvious breaks in the in-situ NO_2 time series that could not be resolved.

With these restrictions in mind, it is not surprising that the comparison of MAXDOAS versus surface in-situ data was not very favourable (see Figure 32, top left), showing only a correlation coefficient of r=0.19 and scatter (30 ppt) in the same order of magnitude as the mean MAXDOAS observed mole fractions (~45 ppt). Although there was no significant absolute bias, surface in-situ observations still showed more frequent NO_2 mole fractions larger 100 ppt than the MAXDOAS. The comparison against night-time observations (Figure 32, bottom left) indicates the anticipated problems due to a change in the photochemical regime at night. Both FLEXPART and global scale NO_2 simulations were rather uncorrelated with MAXDOAS observations (Figure 32, bottom center and right). Correcting the model simulations by the surface in-situ observations slightly improved the comparability with MAXDOAS observations (Figure 32, top center and right), but did not yield more favourable comparison statistics than the surface in-situ comparison.

When looking at the temporal scales of variability it turns out that especially the hourly variability was not common to the MAXDOAS and reference datasets (Figure 35). However, the seasonal variability was well correlated with all reference datasets. Largest NO_2 values were commonly observed/simulated during the summer months. Compared to FLEXPART derived in-situ columns, MAXDOAS observations showed a less pronounced seasonal amplitude. While the amplitude was very similar for in-situ columns that were directly derived from GCTM simulations integrating the surface in-situ observations. This good agreement of the MAXDOAS observed seasonal cycle was also shown against another chemistry transport model (CAM-Chem, Gil Ojeda et al., in preparation). Considering the uncertainties in the surface in-situ observations and the NO_2 model simulations, the disagreement

of the sub-monthly variability between MAXDOAS and reference datasets cannot be ascribed to the MAXDOAS but may rather result from the locally disturbed surface observation and mole fractions around the detection limit of the in-situ analyser. A better way to filter the in-situ dataset needs to be developed for further validation of the MAXDOAS observations.

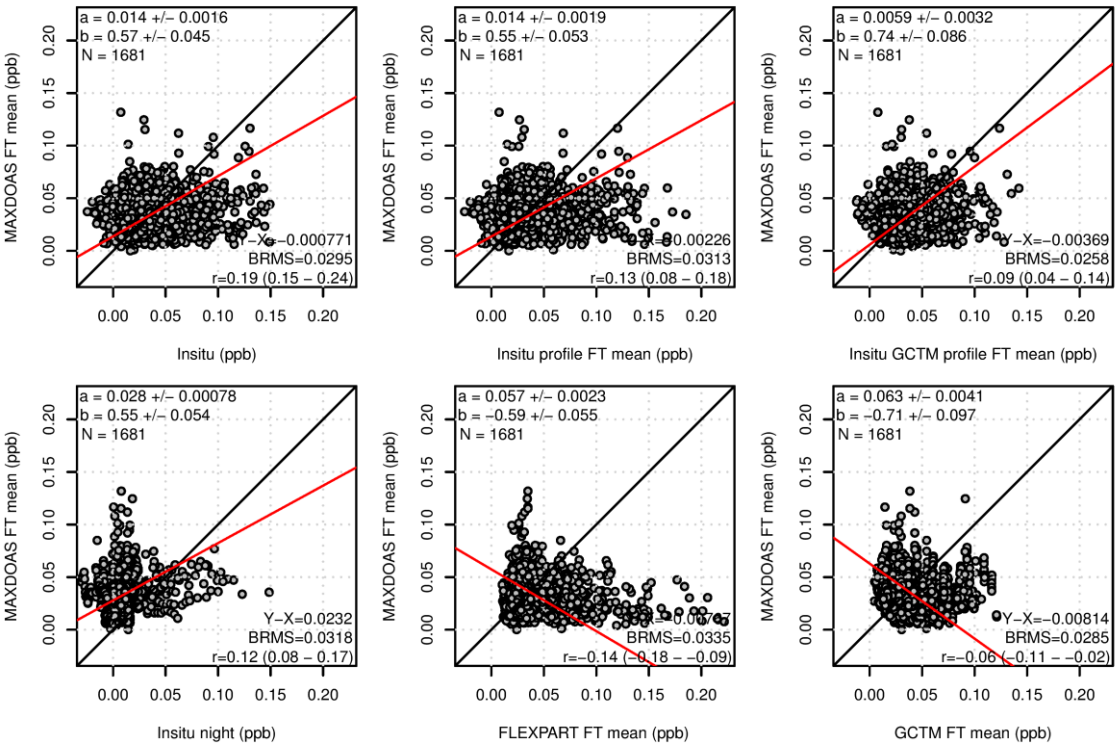

Figure 32: Linear regression analysis of horizontal mean NO_2 mole fractions as obtained by MAXDOAS MGA at Izaña. The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as MAXDOAS observations, (bottom left) night-time in-situ observations bracketing the MAXDOAS observation, (top center) in-situ columns derived from FLEXPART, (bottom center) FLEXPART derived columns, (top right) in-situ columns derived directly from MACC, (bottom right) MACC derived columns. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits.

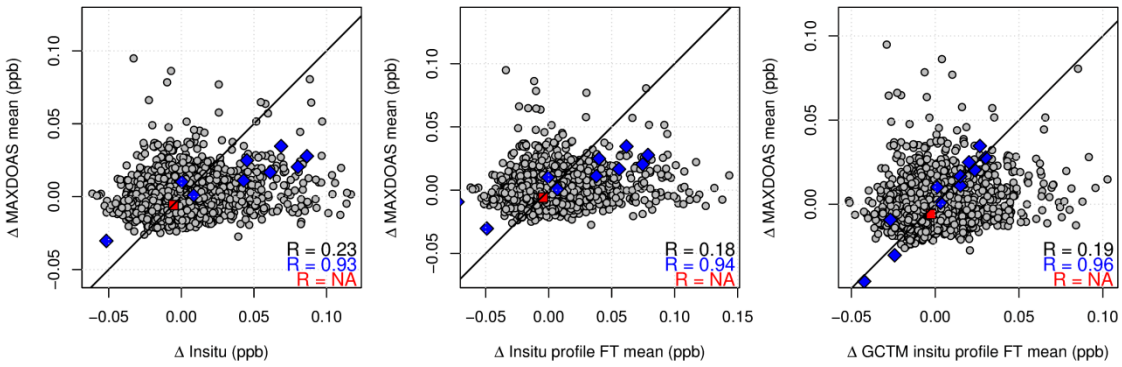

Figure 33: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The horizontal mean NO_2 mole fractions as obtained by MAXDOAS MGA observations at Izaña were analysed versus (left) night-time surface in-situ observations, (center) in-situ columns obtained from FLEXPART simulations, (right) in-situ columns directly taken from MACC re-analysis. The black line represents the one to one line.

3.6. Ozone observations from MAXDOAS

3.6.1. Izaña

As for NO_2 MAXDOAS observations at IZO, MAXDOAS O_3 time series at IZO were retrieved with the modified geometric approach (MGA) (Gomez et al., 2014). Again no profile information could be used in the inter-comparison.

A direct comparison of the MAXDOAS MGA O_3 mole fractions with the in-situ observations at the time of the MAXDOAS observation only showed a relatively weak correlation (r=0.39), large BRMS (~12 ppb) but no systematic bias and also a regression slope equal to 1 (Figure 34, bottom left). The comparison further improved when bracketing night-time surface in-situ data were used instead of day-time surface data (Figure 34, bottom left). As described in detail for the comparison of FTIR retrieved O_3, day-time in-situ measurements at IZO are strongly influenced by up-slope transport of marine boundary layer air masses and O_3 dry deposition along the slopes of the mountain. The use of night-time data overcomes this very local effect on O_3 as measured by the in-situ observation. FLEXPART simulations of O_3 did not agree well with the surface in-situ observations (not shown). Also the agreement with the MAXDOAS observations was only weak (Figure 34, bottom center). Adjusting the model simulations with the help of the in-situ data improved the comparison again, but it still featured a large BRMS (Figure 34, top center). Instead of using FLEXPART derived model values, O_3 mole fractions were also taken directly from MACC re-analysis. The model-only data agreed fairly well with the MAXDOAS observations. When adjusted by the in-situ data the comparison slightly improved but did not reach better results than for the night-time surface in-situ comparison.

The analysis of different scales of temporal variability did not reveal large differences between the reference data sets. Comparison statistics were similar for the night-time surface in-situ, the FLEXPART in-situ column and the MACC re-analysis in-situ column (Figure 35). In all cases it becomes apparent that next to the hourly variability also large parts of the seasonal variability were not aligned between MAXDOAS and reference datasets. The monthly variability agreed better for the in-situ columns, which may indicate seasonal variability in the in-situ observations that originate from varying degrees of local influence. The explained hourly variability remained small also for in-situ columns.

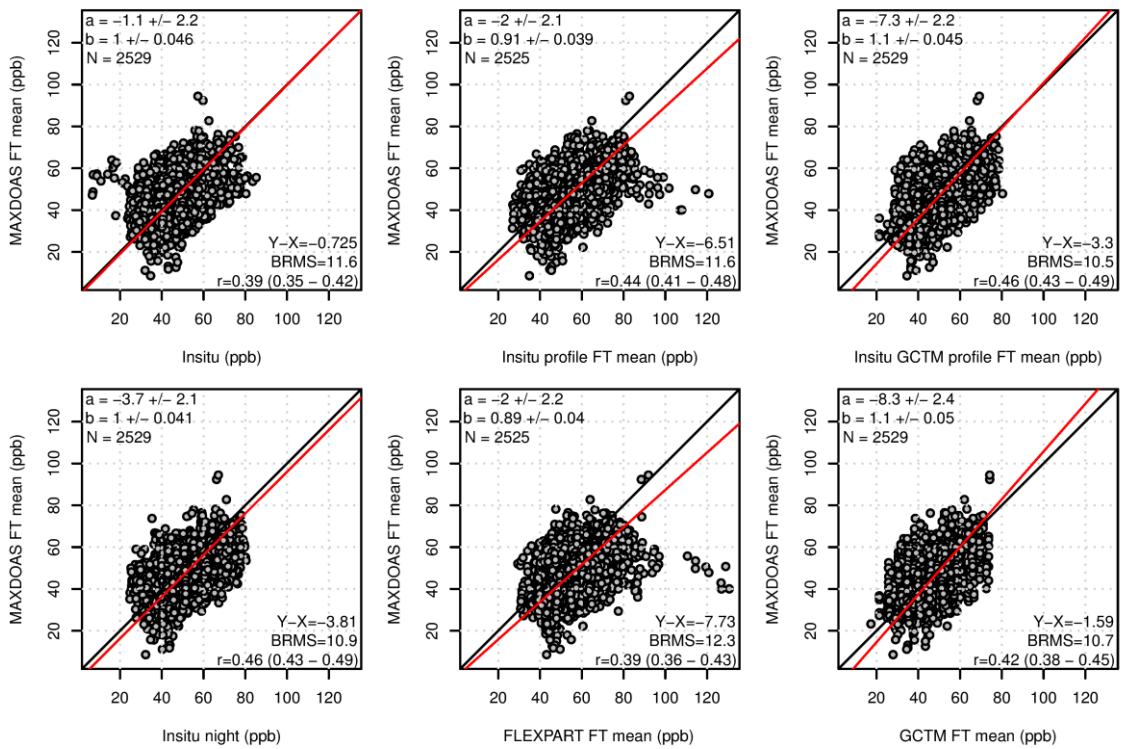

Figure 34: Linear regression analysis of horizontal mean O_3 mole fractions as obtained by MAXDOAS MGA at Izaña. The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as MAXDOAS observations, (bottom left) night-time in-situ observations bracketing the MAXDOAS observation, (top center) in-situ columns derived from FLEXPART, (bottom center) FLEXPART derived columns, (top right) in-situ columns derived directly from MACC, (bottom right) MACC derived columns. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits.

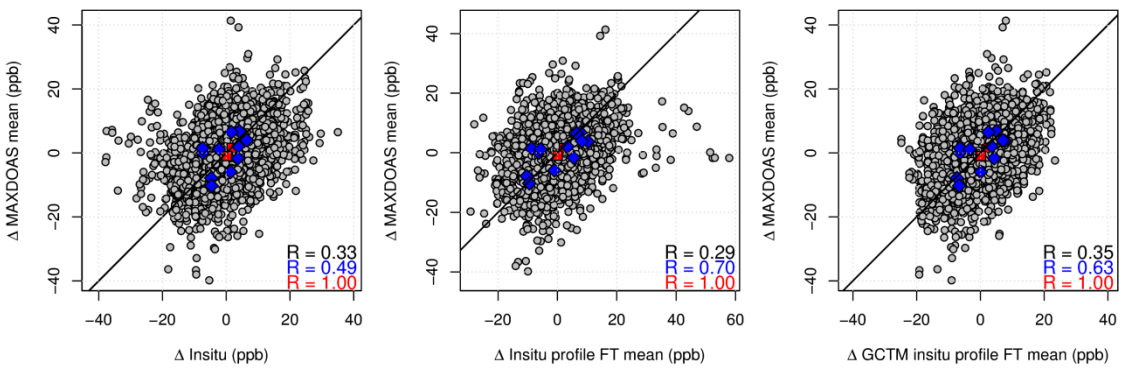

Figure 35: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The horizontal mean O3 mole fractions as obtained by MAXDOAS MGA observations at Izaña were analysed versus (left) night-time surface in-situ observations, (center) in-situ columns obtained from FLEXPART simulations, (right) in-situ columns directly taken from MACC re-analysis. The black line represents the one to one line.

4. Conclusions

The cross comparison results are summarised in Table 2 and Figure 36 for all compared profile retrievals. A direct comparison of remotely sensed tropospheric columns with surface in-situ data already revealed a reasonable correlation between the two techniques. Slight improvements in the comparison were seen when using night-time surface in-situ data instead of day-time observations

taken at the same time as the remote sensing observations. This effect was more pronounced for Izaña where recurring day-time upslope winds have a systematic impact on the surface in-situ observations. Correlations were generally best for CO as compared with all other trace gases. When comparing against in-situ and model-only columns the correlations with the remote sensing columns largely improved for all species and both sites. At the same time the BRMS largely decreased. Improvements were largest for O_3 for which surface observations were expected to have a rather limited representativeness for the vertical column. BRMSE for hourly data were in the order of 10 ppb for CO, O_3 and CH_4 and around 50 ppt for NO_2. Regression slopes for CO and O_3 retrieved by FTIR versus in-situ columns were not significantly different from 1, while they were around 1.4 for CH_4 columns. In the case of CO and O_3 this presents an improvement over the pure surface in-situ comparison for which either slopes were significantly different from 1 or uncertainties of the slopes were large. The slope of 1.25 of MAXDOAS obtained NO_2 columns against in-situ columns at JFJ was connected with a relatively large uncertainty and changed from a value of 0.82 when comparing against surface in-situ observations. While the other comparison statistics improved when using in-situ columns instead of surface in-situ data, the overall absolute bias increased for CO and CH_4 (only Izaña), but was reduced for O_3. Biases as compared to the in-situ columns were about 5 ppb for O_3 (at Jungfraujoch) and CH_4 (at Jungfraujoch) and 10 ppb for CO (both sites) and O_3 (at Izaña). For CH_4 a bias between 40 and 80 ppb was determined at Izaña depending on the baseline model used for the generation of the in-situ profile. Virtually no bias (<20 ppt) was determined for the MAXDOAS NO_2 observations at Jungfraujoch. In the case of the MGA retrievals at IZO the inclusion of model data did not improve the comparison statistics significantly. Strong local influence of emissions (NO_2) and deposition (O_3) on the surface in-situ observations, as described above, is probably the most important reason for this behaviour. The LPDM simulations were not able to reproduce these very local influences and adequately account for them in the data integration process.

Another important aspect of the cross comparison was the analysis of the resolved scales of variability. When comparing with in-situ surface data correlations for hourly de-trended and de-seasonalised data were usually small and overall good correlations mainly resulted from the seasonal cycle, which was common to both datasets. The integration of in-situ and model data especially improved the correlations on the hourly time scale. This leads to the conclusion that the remote sensing techniques are well able to resolve hourly variability. The same conclusion could not have been drawn from the comparison of surface in-situ data alone.

Comparisons versus model-only and in-situ columns were very similar for most species and both sites. This does not necessarily imply that the whole comparison method and surface in-situ integration deteriorates to a model validation exercise as is the main goal of the NORS project. The model profiles used here differ from those used in the Copernicus Atmospheric Monitoring Service in the sense that 1) they are dedicated transport simulations for specific remote sensing volumes, 2) we used different baseline models in the study and carefully selected those that exhibited no systematic bias compared to the surface in-situ observations at the comparison sites. The comparison of the remote sensing tropospheric columns with the baseline models themselves was usually less good than for our dedicated FLEXPART simulated profiles. Furthermore, the influence of incorporating the surface data was larger when tropospheric columns were calculated up to lower altitudes. In that case comparison statistics improved from model-only to in-situ columns, but the overall agreement was generally weaker due to the limited independent information the remote sensing could obtain from a less deep column. In one case, profile retrievals of a MAXDOAS were compared, which had a much finer vertical resolution than the FTIR retrievals. Hence, the targeted lower tropospheric column was not very deep and the influence of the surface in-situ data was comparatively large while also the model performance was relatively weak. In this case a clear improvement in the comparison statistics was visible when using the in-situ columns as compared with the model-only columns.

In conclusion, the cross comparison of the remotely sensed tropospheric columns, which contained sufficiently large independent information, with model assisted in-situ columns proves the high quality of the FTIR derived tropospheric columns of CO and O_3 and MAXDOAS retrievals of NO_2 at Jungfraujoch. For FTIR CH_4 tropospheric columns larger uncertainties remain that may partly be related to uncertainties in the reference profiles but mostly seem to stem from uncertainties in the CH_4 retrievals related to the quality and consistency of the CH_4 line parameters. However, it was also shown that column comparisons strongly depend on the depth of the chosen tropospheric column and that retrieved profile shapes can significantly differ from the reference profiles. Large parts of these differences in profile shape are inevitable since the remote sensing techniques can only obtain a limited amount of vertically independent information. This needs to be kept in mind when vertical profile shapes are compared instead of integrated columns.

Next to proving the high quality of the remote sensing data, it could be shown that the presented data integration method allows for a more quantitative validation of the remote sensing products than previously achieved by direct in-situ comparisons. Currently the method does not treat uncertainties in the obtained reference profiles. These uncertainties arise from the model simulation and the propagation of the uncertainty of the surface in-situ observation in the vertical. Since the latter process was not formulated in an optimal estimate framework we were not able to treat uncertainties explicitly in this study. Future improvements of the method should take such uncertainty propagation into account and also use it in the validation of the remote sensing data.

Table 2: Summary of comparison statistics between all profile retrievals and reference data sets: surface in-situ observations and in-situ columns. For O_3 comparisons at IZO night-time surface in-situ data were used otherwise day-time observations at the same time as the RS observation.

	FTIR CO		FTIR O_3		FTIR CH_4		MAXDOAS NO_2 *
Site	**JFJ**	**IZO**	**JFJ**	**IZO**	**JFJ**	**IZO**	**JFJ**
Bias (ppb)							
In-situ	5	-14	12	14	-29	48	-3
In-situ column	9	-12	4	8	5	69	0
BRMS (ppb)							
In-situ	14	10	16	12	23	19	9
In-situ column	10	7	8	9	13	11	6
Correlation							
In-situ	0.79	0.80	0.45	0.47	0.38	0.48	0.59
In-situ column	0.88	0.90	0.90	0.72	0.59	0.69	0.71
Regression Slope							
In-situ	0.90	0.90	1.95	1.32	0.68	0.75	0.82
In-situ column	1.03	0.94	0.99	1.05	1.47	1.37	1.25

*: Values given in units mole fractions present ppt for NO_2 comparisons.

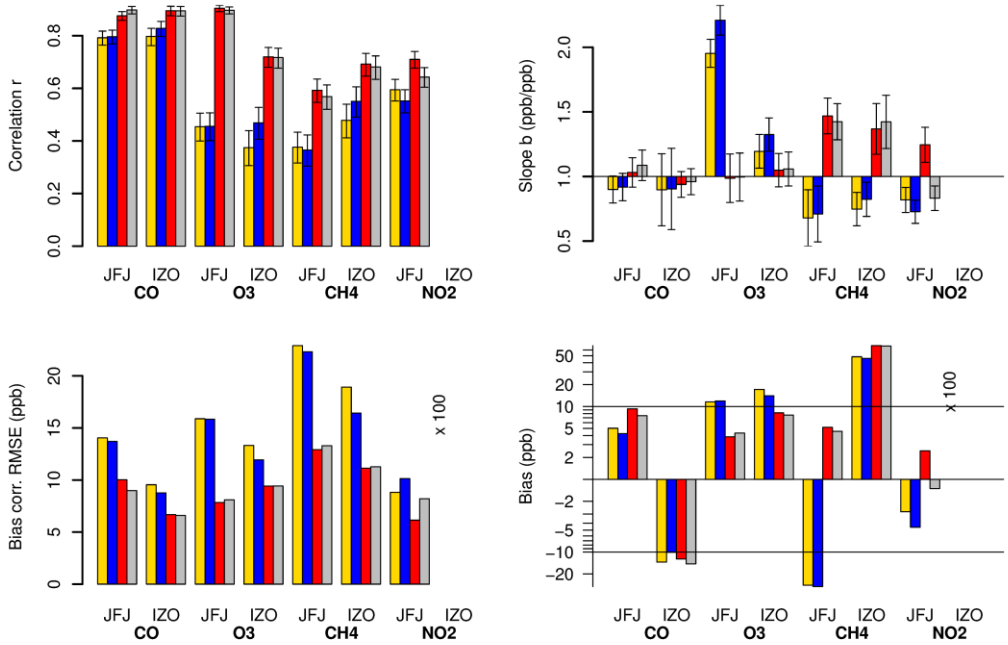

Figure 36: Comparison statistics for FTIR retrieved tropospheric columns of CO, O_3 and CH_4 at Jungfraujoch and Izaña and MAXDOAS lower tropospheric columns of NO_2 at Jungfraujoch. (top, left) correlation coefficient, r, (top right) slope of linear regression, (bottom left) bias corrected root mean square error, (bottom right) remote sensing bias. Error bars on correlation and slope represent 95% confidence estimates. Results for four different reference data sets are shown: (yellow) day-time surface in-situ observations, (blue) night-time surface in-situ observations, (red) in-situ columns, and (gray) model-only columns. In-situ columns were obtained from the model-only profiles after merging with surface in-situ observations. Both model-only and in-situ profiles were folded with the respective averaging kernel before the comparison. The analysis was performed on hourly aggregates. Note that the values for bias and BRMS were multiplied by 100 in the case of NO_2 observations.

References

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., and Gerbig, C.: Atmospheric CH4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, J. Geophys. Res., 118, 7350-7369, doi: 10.1002/jgrd.50480, 2013.

De Mazière, M., Richter, A., Pastel, M., Hendrick, F., Longerock, B., and Hocke, C.: Data User Guide, NORS Deliverable, D4.2, 45, 2013.

Gomez, L., Navarro-Comas, M., Puentedura, O., Gonzalez, Y., Cuevas, E., and Gil-Ojeda, M.: Long-path averaged mixing ratios of O3 and NO2 in the free troposphere from mountain MAX-DOAS, Atmos. Meas. Tech. Discuss., 6, 8235-8267, doi: 10.5194/amtd-6-8235-2013, 2013.

Gomez, L., Navarro-Comas, M., Puentedura, O., Gonzalez, Y., Cuevas, E., and Gil-Ojeda, M.: Long-path averaged mixing ratios of O3 and NO2 in the free troposphere from mountain MAX-DOAS, Atmos. Meas. Tech., 7, 3373-3386, doi: 10.5194/amt-7-3373-2014, 2014.

Henne, S., Dommen, J., Neininger, B., Reimann, S., Staehelin, J., and Prevot, A. S. H.: Influence of mountain venting in the Alps on the ozone chemistry of the lower free troposphere and the European pollution export, J. Geophys. Res., 110, D22307, doi: 10.1029/2005JD005936, 2005.

Henne, S., Schnadt Poberaj, C., Reimann, S., and Brunner, D.: Global-scale tropospheric Lagrangian particle models with linear chemistry, in: Lagrangian Modeling of the Atmosphere, edited by: Lin, J. C., Gerbig, C., Brunner, D., Stohl, A., Luhar, A., and Webley, P., Geophysical Monographs, AGU, Washington, DC, 235-250, 2013.

Henne, S.: Methodology for data integration, NORS Deliverable, D5.1, 33, 2014.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J. N., Thouret, V., Vrekoussis, M., Zerefos, C., and the, M. t.: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, doi: 10.5194/acp-13-4073-2013, 2013.

Krystek, M., and Anton, M.: A weighted total least-squares algorithm for fitting a straight line, Meas. Sci. Technol., 18, 3438-3442, 2007.

Rodgers, C. D., and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 4116, doi: 10.1029/2002JD002299, 2003.

Schaub, D., Brunner, D., Boersma, K. F., Keller, J., Folini, D., Buchmann, B., Berresheim, H., and Staehelin, J.: SCIAMACHY tropospheric NO2 over Switzerland: estimates of NOx lifetimes and impact of the complex Alpine topography on the retrieval, Atmos. Chem. Phys., 7, 5971-5987, doi: 10.5194/acp-7-5971-2007, 2007.

Sepúlveda, E., Schneider, M., Hase, F., García, O. E., Gomez-Pelaez, A., Dohe, S., Blumenstock, T., and Guerra, J. C.: Long-term validation of tropospheric column-averaged CH4 mole fractions obtained by mid-infrared ground-based FTIR spectrometry, Atmos. Meas. Tech., 5, 1425-1441, doi: 10.5194/amt-5-1425-2012, 2012.

Sepúlveda, E., Schneider, M., Hase, F., Barthlott, S., Dubravica, D., García, O. E., Gomez-Pelaez, A., González, Y., Guerra, J. C., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Strong, K., Weaver, D., Palm, M., Sadeghi, A., Deutscher, N. M., Warneke, T., Notholt, J., Jones, N., Griffith, D. W. T., Smale, D., Brailsford, G. W., Robinson, J., Meinhardt, F., Steinbacher, M., Aalto, T., and Worthy, D.: Tropospheric CH4 signals as observed by NDACC FTIR at globally distributed sites and comparison to GAW surface in situ measurements, Atmos. Meas. Tech., 7, 2337-2360, doi: 10.5194/amt-7-2337-2014, 2014.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461-2474, doi: 10.5194/acp-5-2461-2005, 2005.

List of Figures

Figure 1: Comparison statistics for FTIR retrieved tropospheric columns of CO, O_3 and CH_4 at Jungfraujoch and Izaña and MAXDOAS lower tropospheric columns of NO_2 at Jungfraujoch. (top, left) correlation coefficient, r, (top right) slope of linear regression, (bottom left) bias corrected root mean square error, (bottom right) remote sensing bias. Error bars on correlation and slope represent 95% confidence estimates. Results for four different reference data sets are shown: (yellow) day-time surface in-situ observations, (blue) night-time surface in-situ observations, (red) in-situ columns, and (gray) model-only columns. In-situ columns were obtained from the model-only profiles after merging with surface in-situ observations. Both model-only and in-situ profiles were folded with the respective averaging kernel before the comparison. The analysis was performed on hourly aggregates. Note that the values for bias and bias corrected root mean square error were multiplied by 100 in the case of NO_2 observations. 6

Figure 2: Diurnal cycle of surface in-situ observations of (top left) temperature, (top right) carbon monoxide, (bottom, left) ozone, and (bottom right) methane at the Jungfraujoch observatory (3580 m asl) for the period 2009 to 2011..... 13

Figure 3: Diurnal cycle of surface in-situ observations of (top left) temperature, (top right) carbon monoxide, (bottom, left) ozone, and (bottom right) methane at the Izaña observatory (2373 m asl) for the period 2009 to 2011..... 14

Figure 4: Linear regression analysis of mean free tropospheric (FT) CO mole fractions as obtained by FTIR at Jungfraujoch. For the lower free tropospheric mean all FTIR sub-columns up to 7.3 km asl were averaged (pressure weighted mean). The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as FTIR observations, (bottom left) night-time in-situ observations bracketing the FTIR observation, (top center) in-situ profile folded with FTIR AVK, (bottom center) model profile folded with FTIR AVK, (top right) mean in-situ profile, (bottom right) GCTM profile folded with FTIR AVK. GCTM profiles were taken from FLEXCTM. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits. 16

Figure 5: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The mean free tropospheric CO mole fractions as obtained by FTIR at Jungfraujoch were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.3 km asl. The black line represents the one to one line..... 17

Figure 6: Dependency of comparison statistics on vertical integration depths for CO at JFJ. (top left) correlation coefficient, (top right) regression slope, (bottom left) bias (FTIR minus reference), (bottom right) bias corrected root mean square error. Comparison results are shown for a comparison versus (yellow) day-time in-situ observations, (blue) night-time in-situ observations, (red) in-situ profiles from two different baseline models, (gray) raw model profiles from two different baseline models (FLEXCTM and MACC re-analysis)..... 17

Figure 7: Time series (x-axis) of (left) monthly mean vertical profiles (y-axis) of CO mole fractions as retrieved by FTIR and (right) monthly mean differences of the FTIR and FTIR folded in-situ profiles for Jungfraujoch. The lowest level marked by FT represents the free tropospheric mean calculated from the lowest level up to the height indicated by the red dashed line. The numbers at the top of each monthly column gives the number of valid hourly FTIR aggregates in the respective month. 18

Figure 8: Same as Figure 4 but for CO as observed by FTIR at Izaña. Lower free tropospheric profiles were averaged up to 7.2 km asl. GCTM profiles were taken from MACC re-analysis. 20

Figure 9: Same as Figure 5 but for CO as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.2 km asl. 20

Figure 10: Same as Figure 6 but for CO as observed by FTIR at Izaña. 21

Figure 11: Same as Figure 7 but for CO as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile..... 21

Figure 12: Same as Figure 4 but for O_3 observations from FTIR at Jungfraujoch. Lower free tropospheric profiles were averaged up to 7.8 km asl. GCTM background and profiles were taken from MACC re-analysis..... 23

Figure 13: Same as Figure 5 but for O_3 as observed by FTIR at Jungfraujoch. The mean free tropospheric FTIR columns were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) folded in-situ columns, (right) in-situ columns. The FT averaging ranged from the surface to 7.8 km asl. 23

Figure 14: Same as Figure 6 but for O_3 as observed by FTIR at Jungfraujoch. 24

Figure 15: Same as Figure 7 but for O_3 as observed by FTIR at Jungfraujoch. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile. 24

Figure 16: Same as Figure 4 but for O_3 observations by FTIR at Izaña. Lower tropospheric profiles were averaged up to 7.2 km asl. GCTM profiles were taken from MACC re-analysis. 26

Figure 17: Same as Figure 5 but for O_3 as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) night-time surface in-situ observations, (center) folded in-situ columns, (right) in-situ columns. The FT averaging ranged from the surface to 7.2 km asl. 26

Figure 18: Same as Figure 6 but for O_3 as observed by FTIR at Izaña. 27

Figure 19: Same as Figure 7 but for O_3 as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile. 27

Figure 20: Same as Figure 4 but for CH_4 as observed by FTIR at Jungfraujoch. Tropospheric columns were averaged up to 9.3 km asl. GCTM profiles were taken from FLEXCTM simulations. 30

Figure 21: Same as Figure 5 but for CH_4 as observed by FTIR at Jungfraujoch. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 9.3 km asl. 30

Figure 22: Same as Figure 6 but for CH_4 as observed by FTIR at Jungfraujoch. 31

Figure 23: Same as Figure 7 but for CH_4 as observed by FTIR at Jungfraujoch. (left) FTIR mole fractions, (right) FTIR bias as compared to reference profile. 31

Figure 24: Same as Figure 4 but for CH_4 as observed by FTIR at Izaña. Free tropospheric profiles were averaged up to 7.6 km asl. GCTM profiles were taken from FLEXCTM. 33

Figure 25: Same as Figure 5 but for CH_4 as observed by FTIR at Izaña. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as FTIR observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 7.6 km asl. 33

Figure 26: Same as Figure 6 but for CH_4 as observed by FTIR at Izaña. 34

Figure 27: Same as Figure 7 but for CH_4 as observed by FTIR at Izaña. (left) FTIR mole fractions, (right) FTIR bias as compared to folded in-situ profile.at IZO. 34

Figure 28: Same as Figure 4but for NO_2 as observed by MAXDOAS at Jungfraujoch. Tropospheric columns were averaged up to 4.06 km asl. GCTM profiles were taken form MACC re-analysis. 36

Figure 29: Same as Figure 5 but for NO_2 as observed by MAXDOAS at Jungfraujoch. The mean free tropospheric FTIR values were analysed versus (left) surface in-situ observations at the same time as MAXDOAS observations, (center) FT mean of folded in-situ profiles, (right) FT mean of in-situ profiles. The FT averaging ranged from the surface to 4.06 km asl. 36

Figure 30: Same as Figure 6 but for NO_2 as observed by MAXDOAS at Jungfraujoch. 37

Figure 31: Same as Figure 7 but for NO_2 as observed by MAXDOAS at Jungfraujoch. (left) MAXDOAS mole fractions, (right) MAXDOAS bias as compared to folded in-situ reference profile. 37

Figure 32: Linear regression analysis of horizontal mean NO_2 mole fractions as obtained by MAXDOAS MGA at Izaña. The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as MAXDOAS observations, (bottom left) night-time in-situ observations bracketing the MAXDOAS observation, (top center) in-situ columns derived from FLEXPART, (bottom center) FLEXPART derived columns, (top right) in-situ columns derived directly from MACC, (bottom right) MACC derived columns. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits. 39

Figure 33: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The horizontal mean NO_2 mole fractions as obtained by MAXDOAS MGA observations at Izaña were analysed versus (left) night-time surface in-situ observations, (center) in-situ columns obtained from FLEXPART simulations, (right) in-situ columns directly taken from MACC re-analysis. The black line represents the one to one line. 39

Figure 34: Linear regression analysis of horizontal mean O_3 mole fractions as obtained by MAXDOAS MGA at Izaña. The six different reference mole fractions were obtained from (top left) in-situ observations at the same time as MAXDOAS observations, (bottom left) night-time in-situ observations bracketing the MAXDOAS observation, (top center) in-situ columns derived from FLEXPART, (bottom center) FLEXPART derived columns, (top right) in-situ columns derived directly from MACC, (bottom right) MACC derived columns. The uncertainties given for regression slope, b, and regression intercept, a, are 95% confidence limits. 41

Figure 35: Scatter plots for different temporal scales of variability: (red squares) inter-annual, (blue diamonds) seasonal, (gray circles) hourly residuals. The horizontal mean O3 mole fractions as obtained by MAXDOAS MGA observations at Izaña were analysed versus (left) night-time surface in-situ observations, (center) in-situ columns obtained from FLEXPART simulations, (right) in-situ columns directly taken from MACC re-analysis. The black line represents the one to one line. 41

Figure 36: Comparison statistics for FTIR retrieved tropospheric columns of CO, O_3 and CH_4 at Jungfraujoch and Izaña and MAXDOAS lower tropospheric columns of NO_2 at Jungfraujoch. (top, left) correlation coefficient, r, (top right) slope of linear regression, (bottom left) bias corrected root mean square error, (bottom right) remote sensing bias. Error bars on correlation and slope represent 95% confidence estimates. Results for four different reference data sets are shown: (yellow) day-time surface in-situ observations, (blue) night-time surface in-situ observations, (red) in-situ columns, and (gray) model-only columns. In-situ columns were obtained from the model-only profiles after merging with surface in-situ observations. Both model-only and in-situ profiles were folded with the respective averaging kernel before the comparison. The analysis was performed on hourly aggregates. Note that the values for bias and BRMS were multiplied by 100 in the case of NO_2 observations. 44

List of Tables

Table 1: Remote sensing data used in cross comparison. The version refers to the version on the NDACC data centre. 8

Table 2: Summary of comparison statistics between all profile retrievals and reference data sets: surface in-situ observations and in-situ columns. For O_3 comparisons at IZO night-time surface in-situ data were used otherwise day-time observations at the same time as the RS observation..... 43

Acronyms, variables and abbreviations

a	intercept of linear regression
asl	above mean sea level
AVK	averaging kernel
b	slope of linear regression
Bias	Mean difference between target and reference
BRMS	Bias corrected root mean square error
CO	carbon monoxide
CH_4	methane
FTIR	Fourier transform infrared spectroscopy
GCTM	global chemistry and transport model
GEOMS	generic earth observation standard
hdf	hierarchical data format
IZO	Izaña
JFJ	Jungfraujoch
LPDM	Lagrangian Particle Dispersion Model
MAXDOAS	multi axis differential optical absorption spectroscopy
NO_2	nitrogen dioxide
NDACC	Network for the Detection of Atmospheric Composition Change
O_3	ozone
PBL	Planetary boundary layer
r	Pearson's regression coefficient
RS	remote sensing
UTLS	upper troposphere/lower stratosphere