Description of methodology for data integration

Deliverable title	Methodology for data integration
Deliverable number	D5.1
Revision	00
Status	Final
Planned delivery date	30/11/2013
Date of issue	01/12/2013
Nature of deliverable	Report
Lead partner	EMPA
Dissemination level	Public

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	DATE	SIGNATURE
LEAD AUTHOR	WP 5 Lead	Stephan Henne		
CONTRIBUTING AUTHORS				

Table of Contents

SUMMARY 4
1. INTRODUCTION 4
2. METHODS 5
2.1. Observational data 5
2.1.1. Demonstration sites 6
2.1.2. Surface in-situ observations 6
2.1.2.1. Jungfraujoch, 3580 m asl, Switzerland 6
2.1.2.2. Izaña, 2373 m asl, Tenerife (Spain) 8
2.1.3. Surface remote sensing observations 8
2.1.4. Inter-comparison period 9
2.2. Sample volume characterisation 10
2.2.1. FTIR tropospheric partial columns 10
2.2.2. MAXDOAS observations 11
2.3. Backward dispersion simulations 11
2.4. Global scale model input 14
2.4.1. MACC re-analysis 14
2.4.2. FLEXCTM 14
2.4.3. TM5 14
2.5. Representative profile regions 14
2.6. Adjustment of 'in-situ' profiles 16
2.7. Comparison with remote sensing profiles 19
3. RESULTS 19
3.1. Validation of model simulations 19
3.2. Vertical extent of profile adjustment 22
3.3. Validation of adjusted profiles 24
3.4. Uncertainties of adjusted profiles 26
4. CONCLUSIONS 27
ACKNOWLEDGEMENTS 27
ACRONYMS AND ABBREVIATIONS 28
REFERENCES 28
LIST OF TABLES 29
LIST OF FIGURES 29

Summary

The final aim of NORS work package 5 is to compare and integrate surface in-situ observations and remote sensing observations in the troposphere. In this context the surface in-situ observations are seen as a reference measurement because they can be traced back to international standards and exhibit relatively small measurement uncertainties. However, surface observations are usually not representative for extended vertical regions. Here a method is described that combines surface in-situ observations with dedicated atmospheric transport simulations in order to describe vertical representativeness and transfer the surface in-situ observation in the vertical. The product of this surface data extrapolation is a dataset of vertical profiles of the NORS target species at the times of the remote sensing observations. Since these profiles were "calibrated" against the surface in-situ observations they present reference profiles that will be used to validate the remote sensing observations.

This deliverable report describes in detail the method that is used for surface data extrapolation and the resulting reference data sets. The uncertainties of the generated reference profiles are discussed by detailing the applied in-situ observation techniques and their uncertainties. The capability of the transport simulations to reproduce the in-situ observations is analysed and serves as another measure of the uncertainty of the reference profiles.

In total 36 yearly reference datasets were generated for the CO, CH_4 and O_3 remote sensing observations at Jungfraujoch and Izaña performed by FTIR during the period 2009 to 2011. While the generated reference datasets were created from FLEXPART simulations with different initial conditions taken from up to three global scale chemistry transport models, it is recommended to use those reference profiles that resulted from the model realisation with the best performance when compared to surface observations (see Table 1). Details on the reference profiles' data format are available from NORS deliverable report D5.2. The reference profile data are available from http://lagrange.empa.ch/NORS/data/ and will be used for further validation and tuning of the FTIR retrieval algorithms in NORS work package 5 and 7.

1. Introduction

In-situ observations of many trace gases offer an excellent atmospheric reference observation 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 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 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 may strongly affect surface observations. 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 workpackage 5 of the NORS project (Integration of tropospheric products) an effort is undertaken to extend surface in-situ observations in the vertical and, hence, produce a reference tropospheric profile which can be compared with RS profile 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 presented approach uses a Lagrangian Particle Dispersion Model (LPDM) to 1) characterise the air mass history of each

sampling volume and 2) to generate a high-resolution model profile. The latter is then adjusted to match the surface in-situ observations and yield the aspired reference profiles for comparison with RS data.

Here, the method for surface in-situ data extrapolation is detailed and documented by the application to FTIR observations of CO, O_3, and CH_4 at the two demonstration sites Jungfraujoch and Izaña. However, the method is not limited to FTIR observations and the requirements for the application of the method to other RS and model datasets are discussed.

The selected sites are elevated sites that are usually situated above the planetary boundary layer and as such are not directly influenced by strong surface fluxes (emissions or deposition). Hence, it is usually assumed that they are horizontally representative of a larger area than boundary layer sites. Nevertheless, the question of the vertical representativeness remains for the selected sites and needs to be addressed when their surface in-situ observations should be compared with RS profile retrievals. Previous inter-comparison exercises, which were done at JFJ [*Barret, et al.*, 2003; *Dils, et al.*, 2011] and IZO [*Sepúlveda, et al.*, 2012], made the assumption that the surface observation is representative of the (lower) free troposphere and that the RS observation provided one independent piece of information that could be directly compared to the in-situ observation. While in general these studies showed good correlation between FTIR and surface observations, they did not provide linear regression slopes between FTIR and in-situ observations that were close to unity and even revealed different long term behaviour [*Dils, et al.*, 2011]. Hence, these comparisons cannot serve as an absolute validation of the FTIR observation. This is not surprising since their comparison approach mostly neglects any vertical variability of the target species within the troposphere and also the integrating nature of the RS retrieval, which is expressed by its averaging kernels. A fair comparison of the RS data with a reference data set needs to take these averaging kernels into account. However, this is only possible if more than a point observation at the surface is available as reference. By producing reference profiles that are “calibrated” with the surface in-situ observation, the present study aims to overcome these shortcomings.

The details on how the reference profiles were derived from a combination of model profiles and in-situ observations are outlined in sections 2.3, 2.5, and 2.6. One main advantage of the presented method is the dedicated transport simulation for individual measurement volumes. This is in contrast to simply interpolating the output of a three-dimensional model onto the position of the observation, which may create large representativeness uncertainties due to the typically coarse resolution of the models. Furthermore, the application of a LPDM allowed for transport simulations that were driven by high-resolution meteorological input beyond the possibilities of today’s global and even regional scale chemistry transport models. Last but not least, it was possible to use the LPDM simulations also to characterise different sampling volumes in terms of their air mass history and thereby defining the profile region for which the surface observations are representative.

2. Methods

2.1. Observational data

The observational data used in this study comprise the surface in-situ observations and the surface RS data (FTIR and MAXDOAS) obtained at the NORS demonstration sites Jungfraujoch and Izaña. The RS data are not directly used here but reference data is generated specifically for the air volumes sampled by these remote sensing techniques. This requires the availability of metadata that describes the RS data in terms of sampling time, sampling volume and further retrieval details such as averaging kernels and a-priori profiles. Parts of these metadata are readily available as they are included in the GEOMS-conform (Generic Earth Observation Metadata Standard) data format that is

provided within NDACC and NORS. Additional information required to describe the sampling volume of MAXDOAS observations was generated within NORS and is used here.

2.1.1. Demonstration sites

The Jungfraujoch observatory (JFJ, 46.54749°N 7.98509°E) is located in the northern Swiss Alps on a steep mountain saddle between the two mountains Jungfrau (4158 m asl) and Mönch (4099 m asl). Strong anthropogenic emission sources are found south of the Alps in the Po Basin and north of the Alps on the Swiss Plateau and, at a larger distance, in southern and western Germany. While JFJ is generally located in the free troposphere it intermittently receives polluted boundary layer air from both southern and northern sources [*Henne, et al.*, 2010]. The vertical extent of these boundary layer pollutants can vary strongly depending on the weather condition and the transport process responsible for lifting.

The Izaña observatory (IZO, 28.30900°N 16.49940°W) is located on Tenerife Island ~15 km from the coast at 28.3° North and -16.5° East at an altitude of 2373m asl on top of a mountain ridge. The close surroundings of the site are uninhabited with settlements mostly along the coast and below 1000 m asl. The distance to the European continent is ~1300 km and ~350 km to Africa.

2.1.2. Surface in-situ observations

A summary of the surface in-situ measurement techniques and their uncertainties is given here in order to document the minimal uncertainty that is associated with the generated reference profiles. Details on calibration procedures and quality control can be found in the referenced literature.

2.1.2.1. Jungfraujoch, 3580 m asl, Switzerland

Surface in-situ observations of many trace gases are preformed routinely at JFJ by Empa as part of the WMO GAW programme and the Swiss national air pollution network NABEL. Ambient air is drawn into the observatory through a stainless steel inlet, which is maintained at a temperature of 10°C (inner diameter 90 mm, flow rate of 50 m^3/hour, residence time about 1 s).

Continuous **CO** measurements have been performed with commercial instruments (Horiba APMA-360 & APMA-370, Kyoto, Japan) using Cross Flow Modulated Non-Dispersive Infrared Absorption technology [*Zellweger, et al.*, 2009a]. Sample gas and reference gas are injected alternately (1 Hz frequency) into the measurement cell using solenoid valve modulation. Sample air is taken to generate CO-free reference gas by using a catalyst to oxidize CO to CO_2. Since the same gas is used for both the sample gas and the reference gas, zero drifts and interference effects are minimized. The sample air is dried using a Nafion drier to remove potential water vapour interferences. A zero check of the instrument is performed every 49 h using externally generated zero air by means of a CO/CO_2 converter, molecular sieve 3 °A and metal catalysts to convert CO (SOFNOCAT). Calibration of the instrument is performed twice a month using CO standards in the low-ppm range that are calibrated against NIST (National Institute of Standards and Technology, USA) and NMI (Nederlands Meetinstituut, The Netherlands) standards. The detection limit was about 30 ppb. Data are routinely recorded as 10 min averages. The overall measurement uncertainty for these values is estimated to be <10% below 100 ppb and <5% above 100 ppb.

Since 2005, **CH_4** has been measured by gas chromatography and subsequent flame ionization detection (GC-FID). A sample of 10 ml volume is injected every 12 to 15 minutes into the measurement system. Each ambient air sample is bracketed with calibration runs, using real-air standards (working standards) with concentrations representative for Northern Hemisphere tropospheric concentrations. The real-air standards are analysed for CH_4 (WMO NOAA2004 scale) by the GAW World Calibration Centre (WCC-Empa) at Empa, based on NOAA/ESRL certified gas

cylinders. Since January 2010, Empa uses the wavelength-scanned cavity ring down technique (Picarro G1301 (January 2010 till July 2011) and Picarro G2401 (September 2011 till present)) coupled to a custom-built calibration/drying unit as primary CH_4 analyser. The sample air was initially dried prior to analysis by means of a Nafion dryer. Along with CH_4, the instrument is also capable to measure CO_2 and H_2O. Thus, the CH_4 data can be corrected for interferences of remaining water traces. From August 2010 till July 2011, no water vapour removal was used anymore and CH_4 dry air mixing ratios are determined by application of an empirical humidity correction to the fully unaltered humid gas stream accounting for dilution and pressure broadening effects. In September 2011 the G1301 analyser was replaced by the newer G2401 model that is also capable to monitor CO mole fractions. Since then, the drying of the sample gas was again applied mainly due to beneficial effects with respect to the precision of the CO analysis. Calibrations are performed every 46 hours with two calibration gases. In addition, a target gas is analysed every 15 hours to detect potential short-term instrument sensitivity changes. The concentrations of the calibration gases were determined by WCC-Empa. The CH_4 data are reported on the NOAA2004 scale.

Ozone is measured by UV absorption (TEI 49C, Thermo Electron Corporation, Environmental Instruments). The O_3 analyser is calibrated on site every three months with a TEI 49-PS reference instrument (traceable to NIST SRP#15). The procedure follows a multipoint calibration at approximately 30, 60, 90, 120 ppb O_3 and a zero point and an O_3 scrubber test at 400 ppb. The analyser is corrected accordingly if the offset varies >0.5 ppb and/or the slope deviates >0.5 %. The combined measurement uncertainty for hourly averages is 1.2 ppb and 2% for values below and above 60 ppb, respectively.

NO and **NO_x** measurements are performed by commercial analysers (CLD 89p, Eco Physics, Switzerland) based on chemiluminescence detection with temperature-controlled reaction chambers (an individual chamber for NO and NO_x). NO_x is measured as NO after photolytic conversion of NO_2 (Eco Physics, PLC 762). Automated calibrations of the chemiluminescence detectors (CLDs) are performed every 35 hours. The CLDs are calibrated with NO standards (about 5.0 ppm; referenced against NIST Standard Reference Material (SRM 2629a) and NMI Primary Reference Material (PRM BD11)) diluted with synthetic air to a concentration of about 45 ppb. The conversion efficiencies of the photolytic converter are measured every 70 hours by generating a known amount of NO_2 by gas phase titration of NO with ozone. The efficiency of the photolytic converter ranged from 45 to 62%. The detection limit (zero + 3sigma of the zero signal) of the CLD 89p (30 min averages) for NO measurements was 15 ppt whereas it was 25 ppt for the NO_2 channel due to the incomplete conversion and the determination of NO_2 by the difference of two measures (NO and NO + converted NO_2) signals. The overall measurement uncertainty for NO_x is 5%.

The in-situ measurements of CO and CH_4 were obtained under dry air conditions either by using an upstream Nafion dryer or by application of a humidity correction. Hence, reported mole fractions are dry air mole fractions. On the contrary, FTIR observations represent moist air mole fractions. The difference between dry and moist mole fractions is in the order of 1 % close to the surface and diminishes with altitude due to strongly decreasing specific humidity. For CH_4 observations differences in the order of 1 % would lead to significant biases in an inter-comparison. Hence, in-situ surface data were converted to moist air mole fractions applying

$$\chi_{moist} = \chi_{dry} \frac{(p-e)\mu_{air}}{(p-e)\mu_{air} + e\mu_{H2O}}, \quad (1)$$

where χ is the mole fraction of the observed trace species, p the total ambient pressure, e the water vapour pressure and μ_{air}=28.96 g mol^{-1} and μ_{H2O}=18.015 g mol^{-1} the molar masses of air and water vapour, respectively. For JFJ, surface in-situ observations were available with a 10 minute temporal resolution, but were aggregated to 1-hourly values.

For the adjustment of model profiles a separation of the simulated mole fractions into a background and a pollution signal was undertaken (see section 2.3). These parts are adjusted differently, but require the knowledge of the same separation of the observed mole fractions. We used the statistical method developed by Ruckstuhl et al. [2012] to separate the observed mole fractions. The method was run with a window width of 60 days. It yields a background signal smoothly varying with time and a constant uncertainty estimate of this background. For values where the observed surface mole fraction, χ_s, is larger than the estimated background mole fraction, $\chi_{s,b}$, we separate the observed signal into background and pollution, $\chi_{s,p}$,

$$\chi_s = \chi_{s,b} + \chi_{s,p} \quad (2)$$

At times when the estimated background is larger than or equal to the actual observation, we set $\chi_{s,b} = \chi_s$ and $\chi_{s,p} = 0$.

2.1.2.2. Izaña, 2373 m asl, Tenerife (Spain)

Izaña is equipped with a Trace Analytical RGA-3 GC-system for simultaneous measurements of **CO** and H_2. The general ambient air inlet is an 8 cm inner diameter stainless steel pipe, has a high flow rate and is mounted 30 m above ground (since June 2005). Upstream of the GC sample loop (1 ml size), a cold-trap is used to reduce the dew point of ambient air to -49° C. The instrument is calibrated every two weeks against five CO primary standards (prepared by WMO GAW CO CCL), that are certified by NOAA/ESRL based on the WMO-2004 calibration scale. Good agreement between the GC and independent standards was observed during the 2009 WMO GAW system and performance audit [*Zellweger, et al.*, 2009b].

CH_4 mole fractions are measured using a DANI 3800 GC with a FID detector. Ambient air is cooled to −70° C to remove water vapour before the sample loop. The high quality of the measurements is supported by repeated comparison with NOAA flask samples taken at the same site and was confirmed during the 2009 WMO GAW system and performance audit [*Zellweger, et al.*, 2009b], which showed a combined uncertainty of the GC system of ~5 ppb. Data are reported on the NOAA04 scale.

O_3 is measured using commercial TEI 49C ozone analysers that are calibrated against a TEI 49-PS reference instrument traceable to NIST SRP, [*Cuevas, et al.*, 2013]. The combined measurement uncertainty for hourly averages is 0.8 ppb and 1 % for values below and above 60 ppb, respectively [*Zellweger, et al.*, 2009b].

NO and **NO_2** are measured using a chemiluminescence analyser, model 42C-TL (Thermo Electron Corporation), with a photolytic converter.

Adjustment of observed dry air mole fractions to moist air mole fractions, separation into background and pollution contribution and aggregation to hourly intervals were done in the same way as for the observations at JFJ.

2.1.3. Surface remote sensing observations

No details on the remote sensing observations are given here. The reader is referred to the NORS Data User Guide (D4.2) for a detailed description of the methods and the produced types of data.

For the comparison/integration of tropospheric RS with the method described here, RS data has to meet the following requirements

- RS data has to be available for several tropospheric partial columns resulting from a profile retrieval and these partial columns need to contain at least one piece of independent information in the troposphere and not exhibit averaging kernels that are sensitive to the stratosphere. If only total tropospheric column observations are available this column needs to be sub-divided into sub-columns for the analysis of representativeness.
- The a-priori profiles and averaging kernels as used by and obtained from the retrieval algorithm have to be available in order to assure a fair comparison with reference data.
- The sampling volume of the RS observation needs to be described so as to enable the analysis of representativeness of different sampling volumes

The first two requirements are fulfilled by the GEOMS-conform data of FTIR and MAXDOAS observations delivered within NDACC and NORS. The third requirement is described in more detail in section 2.2.

2.1.4. Inter-comparison period

The period for which inter-comparison data were produced extended from 2009 to 2011, when well calibrated and quality assured in-situ and FTIR data were available. MAXDOAS observations at both sites commenced in 2010 and the inter-comparison period was set to 2011.

Figure 1 shows the time series of surface in-situ observations of the target species CH_4, CO and O_3 from both sites for which also FTIR observations (not shown) are available. Differences in trace gas levels at the two sites are apparent and can be understood from their respective location. Since JFJ is located relatively close to the central European emission sources, levels of CH_4 and CO are usually larger than at IZO even if the altitude of the observatory is considerably higher. Both sites experience typical northern hemispheric annual cycles of CH_4 and CO, although the amplitude in case of CH_4 is less pronounced at JFJ. Ozone levels are generally larger at JFJ with a summer-time maximum. The first can be explained by the higher altitude of the observatory and, hence, a larger contribution of stratospheric ozone. The latter is a result of summer-time photochemical production of ozone in polluted European air masses.

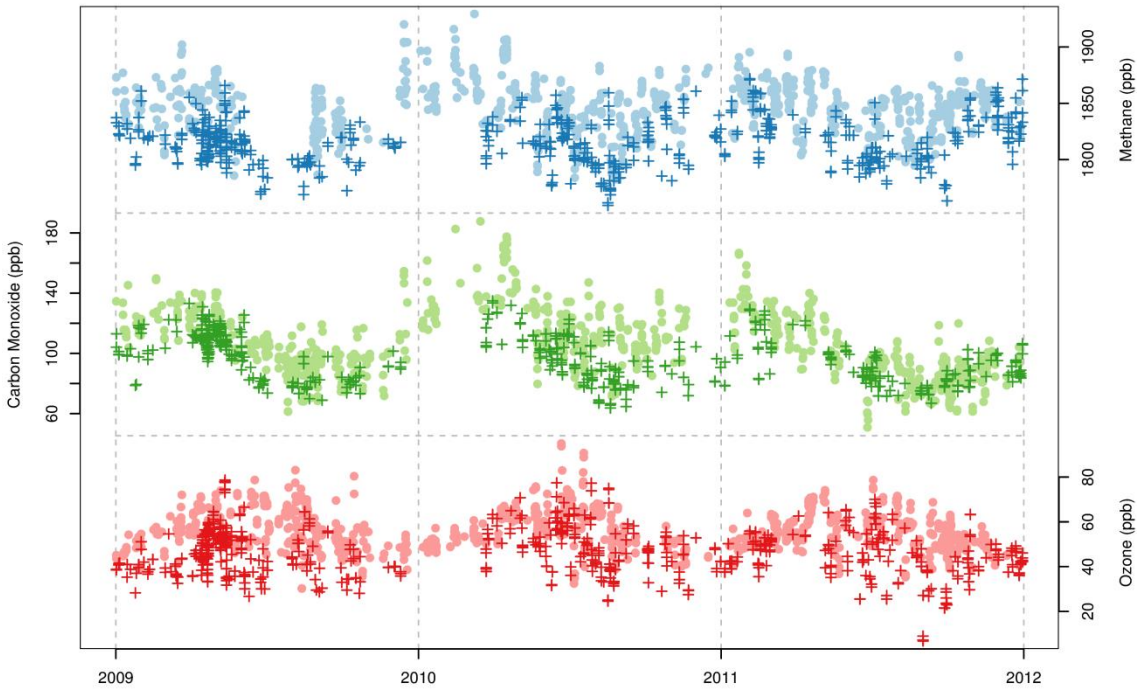

Figure 1: Time series of surface in-situ observations that are collocated with FTIR observations (not shown) for (pale dots) JFJ and (crosses) IZO.

2.2. Sample volume characterisation

2.2.1. FTIR tropospheric partial columns

The sampling volume of individual FTIR retrievals can be described relatively easily, since the FTIR observes direct solar radiation. Hence, the sampling volume lies along the line of sight of the FTIR towards the sun and can be described by the solar azimuth and solar zenith angle at the time of observation. The FTIR profile retrieval gives mole fractions of the target species for partial columns defined by the altitudes at the top and bottom of each partial column. Using the viewing geometry we calculate the longitude and latitude of each of the interfaces between partial columns.

For the backward transport simulation particles were released from each partial column of the FTIR retrieval below 16 km. This focuses on the tropospheric part of the profile, for which some representativeness of the in-situ surface observation can be expected. Above the tropopause, a connection with surface observations is unlikely. Since releases in FLEXPART simulations can only be defined as volume releases and not as line releases, the sampling volume of each FTIR partial column was approximated by the volume defined by the location of the partial column interfaces. An example of the horizontal positions of the FTIR partial columns is shown in Figure 2, where each rectangle represents the horizontal extent of the release volume and its central height is indicated by the fill colour of the rectangle. The plot indicates how important it is to take the horizontal displacement of the FTIR sampling volumes into account. In the very inhomogeneous Alpine terrain, trace gas gradients may exist as a result of large emission gradients and orographically forced transport processes.

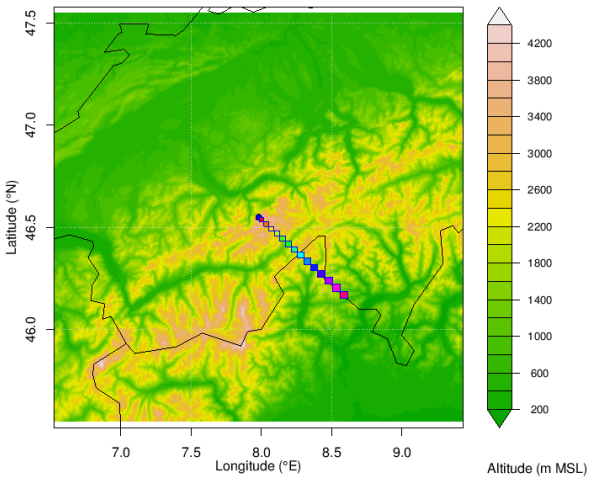

Figure 2: Example of the horizontal position (rectangles) of the tropospheric partial columns of the FTIR observation at JFJ (blue dot in centre) on 2013-01-27 09:00 UTC. The colour scale on the right refers to the height of the topography, while the colour of the rectangles gives the altitude of the partial column centre between the surface and 16 km altitude (scale not shown).

2.2.2. MAXDOAS observations

Since the MAXDOAS uses differential absorption of scattered solar radiation, the sampling volume of a MAXDOAS observation is much less well defined as the one of the FTIR. In the troposphere, the scattering of solar radiation strongly depends on the aerosol load and, hence, does the sampling volume of the MAXDOAS. In the boundary layer, the size of the sampling volume decreases with increasing aerosol load and towards shorter retrieval wavelengths. In order to obtain a good definition of the sampling volume, a 3D radiative transfer simulation would be required, which is usually not available.

As a first approximation of the MAXDOAS sampling volume the horizontal viewing distance as obtained from the modified geometrical approach (MGA) [*Gomez, et al.*, 2013] can be used together with the instrument's viewing geometry. The horizontal viewing distance is a parameter that is not part of the GEOMS-conform data delivery, but needs to be provided separately.

2.3. Backward dispersion simulations

For each sampling volume (surface or RS partial column) backward Lagrangian particle dispersion simulations were performed with the LPDM FLEXPART. For JFJ the regional scale version of the model based on output from the regional NWP COSMO [*Henne, et al.*, 2013b] was applied, nested into the global scale version of FLEXPART [*Stohl, et al.*, 2005]. For IZO only the global scale FLEXPART version was used due to the lack or regional scale input data.

The COSMO input meteorology was obtained from the operational analysis by MeteoSwiss and covers the Alpine area with a 2 km x 2 km horizontal resolution and Western Europe with a horizontal resolution of 7 km x 7 km (Figure 3). Analysis fields were available hourly. The global scale version of FLEXPART was driven by operational analysis fields obtained from the ECMWF IFS system with a horizontal resolution of 0.2° x 0.2° around the site of interest and 1° x 1° globally. ECMWF fields were available 3-hourly, combining analysis and 3-hour forecast fields.

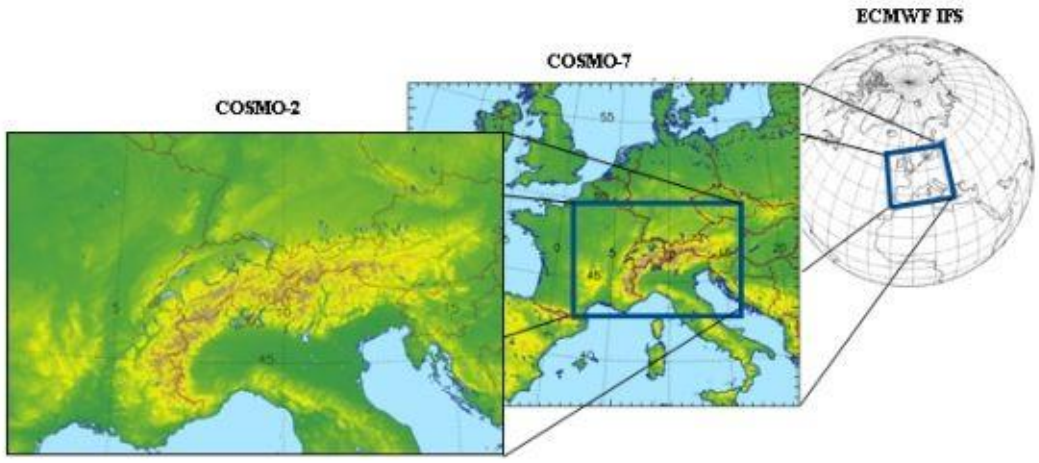

Figure 3: COSMO domains used as input for the LPDM simulations. Courtesy of MeteoSwiss.

Particles were followed 10 days backward in time. The nesting between FLEXPART-COSMO and FLEXPART-ECMWF was done in an offline mode. Particles were initially transported in FLEXPART-COSMO until they either left the COSMO input domain or had been followed for two days. Particles were then restarted in FLEXPART-ECMWF with their associated position and time when they left the COSMO domain. This treatment implies that particles that left the COSMO domain and would enter this region at a later time would still be treated by the FLEXPART-ECMWF version and not the COSMO version.

The backward dispersion simulations result in source sensitivities (also termed source-receptor relationships, srr or footprints) for the period of integration given in units s kg^{-1} m^3. These can directly be multiplied with surface emissions (E, units kg s^{-1}) and summed over all surface grid boxes to yield simulated mole fractions at the respective sampling volume

$$\psi_p = \sum_{i,j} srr_{i,j,1} E_{i,j} \tag{3}$$

These simulated mole fractions solely resulted from emissions picked up during the period covered by the LPDM simulation. We refer to this part of the total mole fractions as pollution. The remaining part of the total mole fractions can be associated with the large scale background signal (since it does not contain any fresh emissions younger than 10 days).

$$\psi = \psi_b + \psi_p \tag{4}$$

The background part of the mole fractions is calculated as the average of the mole fractions at all N particle positions at the end of the simulation, t_0, (10 days before the arrival at the receptor)

$$\psi_b = \frac{1}{N} \sum_i c(x_i, y_i, z_i, t_0) \tag{5}$$

The mole fractions, c, at each particle's position can be extracted from global scale time resolved models or climatologies. For the current study input from 3 different global scale models was used as is detailed in the next section.

Source sensitivities were evaluated on regular longitude/latitude grids with a global resolution of 0.5° x 0.5° and a nested output domain over Europe with 0.2° x 0.2. In the case of JFJ a second nest with 0.1° x 0.1° resolution centred over the Alps was generated for the FLEXPART-COSMO simulations as well.

Emissions of CO, CH_4 and NO_x were taken from the EDGAR inventory (EC-JRC/PBL.EDGAR version 4.2 http://edgar.jrc.ec.europa.eu/) [*Olivier and Berdowski*, 2001] for the reference year 2008. Total (over all sectors) emissions were available on a 0.1° x 0.1° global grid and further aggregated to the FLEXPART output grids. No annual cycle was applied to these emissions. Biomass burning and wetland emissions were not taken into account since their contribution to the sites of interest was deemed small.

Examples of FLEXPART simulated surface source sensitivities for the lowest (Figure 4) and the partial column centred around 6.8 km asl (Figure 5) for the FTIR CO retrieval at JFJ at 2013-01-29 13:00 UTC show the differences in recent transport history of the respective air masses. While the air in the lowest partial column had recent surface contact over Europe (large source sensitivities), and, hence, potentially picked up emissions of the target gases, the air within the partial column at 6.8 km was only weakly sensitive to European emissions. From a purely optical and subjective inspection of these source sensitivities one would conclude that the air masses in these partial columns did not experience the same transport history and, therefore, are not representative of each other. In section 2.5 we outline a more objective method to distinguish which partial columns experienced similar transport histories.

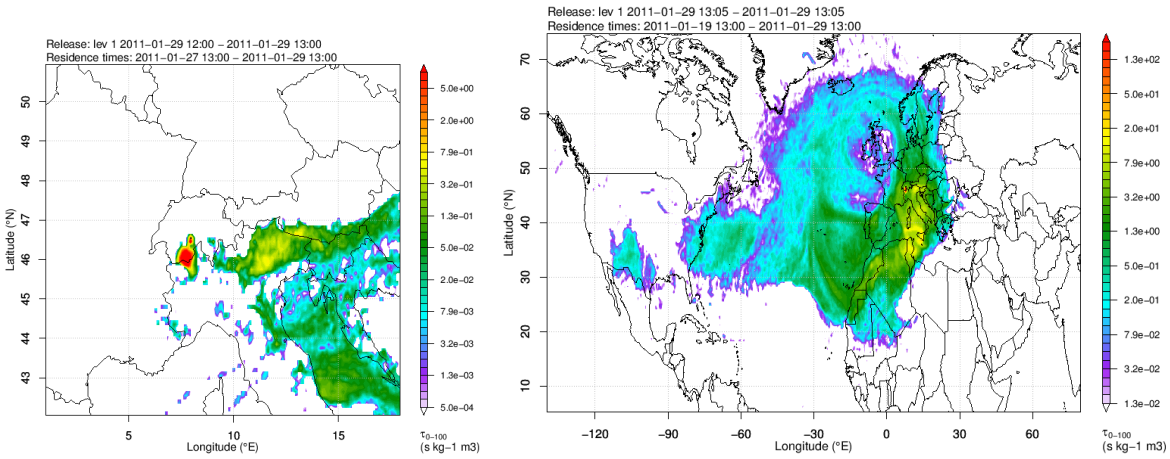

Figure 4: Example of source sensitivities obtained for lowest partial column of the CO FTIR retrieval at JFJ on 2013-01-29 13:00 UTC. (left) zoom centred over the Alps for FLEXPART-COSMO during the first 48 hours of the simulation, (right) Atlantic region combined from FLEXPART-COSMO and FLEXPART-ECMWF during the complete 240 hours of the simulation.

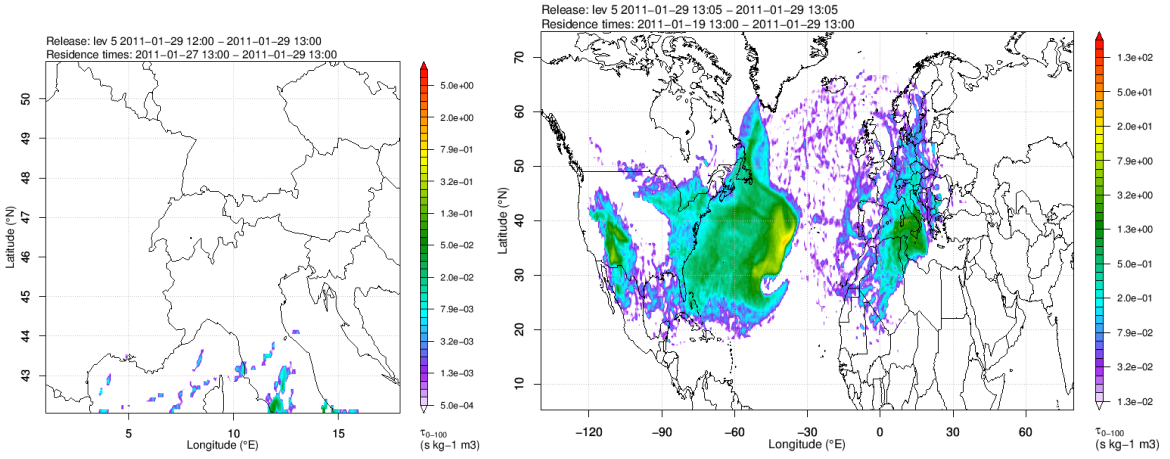

Figure 5: Example of source sensitivities obtained for the partial column reaching from 6.4 to 7.1 km asl of the CO FTIR retrieval at JFJ on 2013-01-29 13:00 UTC. (left) zoom centred over the Alps for FLEXPART-COSMO

during the first 48 hours of the simulation, (right) Atlantic region combined from FLEXPART-COSMO and FLEXPART-ECMWF during the complete 240 hours of the simulation.

2.4. Global scale model input

Global scale model data of atmospheric mole fractions of the target gases are used in two ways. Firstly, they serve as background (initial) conditions for the backward dispersion calculations as described above. Secondly, their data is used to construct the stratospheric part of the reference profiles for which no FLEXPART simulations were carried out. Since the focus here is on tropospheric data integration and the surface in-situ data are not linked to stratospheric observations this approximation seems to be justified. However, in situations when the averaging kernels for the tropospheric partial columns are very broad and reach into the stratosphere, the origin of the reference profile needs to be kept in mind.

2.4.1. MACC re-analysis

The MACC global re-analysis product contains, amongst others, the mole fractions of the NORS target trace species CO, CH_4, O_3 and NO_x. The re-analysis originally spanned the period 2003-2010 [*Inness, et al.*, 2013] but is extended with about a 6 month delay to the presence (https://www.gmes-atmosphere.eu/services/gac/reanalysis/). The MACC re-analysis is constrained by the assimilation of different satellite products, but does not incorporate any surface observations of atmospheric composition. Re-analysis fields were obtained on the native reduced Gaussian grid (T128, approximately 0.7 x 0.7 at the Equator) with a 6-hourly temporal resolution. Mole fractions were interpolated onto particle position using linear interpolation in the vertical, bilinear interpolation in the horizontal and next neighbour interpolation in time. A global bias correction of +70 ppb was applied to the MACC CH_4 mole.

2.4.2. FLEXCTM

FLEXCTM is a FLEXPART based global scale, Lagrangian transport model that can accommodate pseudo first order degradation chemistry. A simulation of CO and CH_4 spanning the period 2000 to 2008 [*Henne, et al.*, 2013a; *Henne, et al.*, 2013c] was extended within the context of NORS for the years 2009 to 2011. The model did not assimilate any observational data. It is purely constrained by the applied emissions and an OH climatology that governs the destruction of the emitted species. Model output was generated on a regular longitude-latitude grid with 2° x 2° horizontal and daily temporal resolution. As for the MACC data, mole fractions were interpolated onto particle position using linear interpolation in the vertical, bilinear interpolation in the horizontal and next neighbour interpolation in time. A global bias correction of -28 ppb was applied to the FLEXCTM CH_4 mole fractions before interpolating them onto the particle positions.

2.4.3. TM5

In addition, output from the TM5-4DVAR inverse modelling system for CH_4, that assimilated NOAA surface flask observations in order to optimise CH_4 emissions were used [*Bergamaschi, et al.*, 2013]. The model's horizontal resolution was 6° x 4° on 25 vertical layers. Output was available as daily averages. Mole fractions at particle positions were obtained as for the other global scale models.

2.5. Representative profile regions

To adjust the simulated profiles by the observed surface values, it is crucial to take the representativeness of the surface values into account. Instead of using a constant scaling function that reduces the influence of the surface observation with vertical distance to the profile point, we

developed a method that identifies different regions within the simulated profile connected with the recent transport history of the observed air masses. This history was described by the time and intensity of an air mass having contact with the Earth's surface and was taking up emissions.

To this extent, three hourly values of total source sensitivity, CO uptake, CH_4 uptake and NO_x uptake were evaluated along the track of the 10-day backward simulations. Hence, for each release volume, we obtained a set of 4x80 = 320 variables that serve as a fingerprint of the transport history. We then applied a clustering algorithm to all release volumes and the obtained variables. The clustering was done separately for each FTIR observation. It provides groups of release volumes for which the transport history of the sampled air mass was similar.

Before applying the clustering, the involved variables were normalised to assure that they enter the clustering with the same weight and an approximately normal distribution. The normalisation was done for each variable group and all release volumes at the same time. Since source sensitivities and emission uptakes are approximately log-normally distributed, log-normalisation was applied as follows

$$x_n = \frac{ln(x) - mean\big(ln(x)\big)}{sd\big(ln(x)\big)} \tag{6}$$

where sd is the standard deviation. Figure 6 shows an example of the clustering variables, displayed as a two-dimensional array with the individual cluster variables on the x-axis and the results for different sampling volumes on the y-axis.

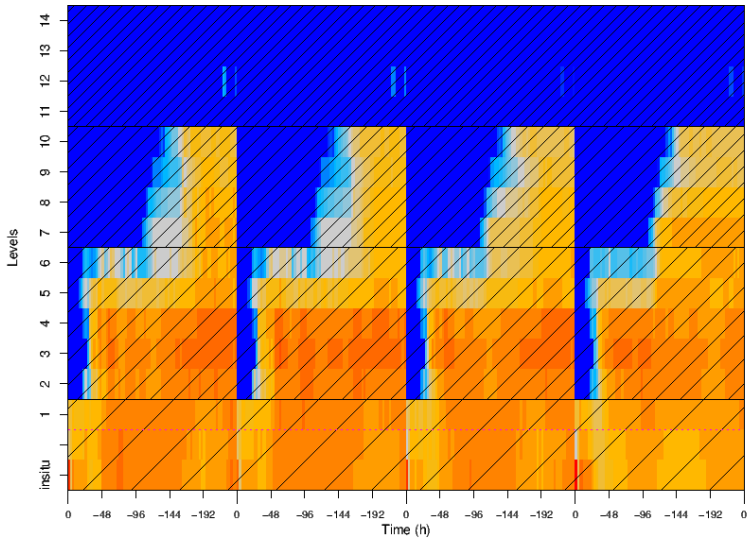

Figure 6: Two-dimensional representation of the variables used for clustering of the sampling volumes of FTIR CO observation at JFJ on 2011-01-29 13:00. On the x-axis 4 different groups of normalized cluster variables and their development over time along the backward simulation are displayed (from left to right: source sensitivities, CO emission uptake, CH_4 emissions uptake, NO_x emission uptake). On the y-axis the different vertical layers of the RS retrieval and at the bottom the surface in-situ simulations are given. Red colours correspond to large values, blue values to small values. Horizontal bars that exhibit a similar pattern identify layers that experience a similar air mass history in terms of total surface contact and emission update. The different shading shows such layers as they were identified by the cluster algorithm.

The cluster variables were processed by hierarchical clustering [*Kaufman and Rousseeuw*, 1990] using a Euclidean distance measure. The number of clusters was obtained with the inter-cluster distance method. The inter-cluster distance gives the average distance between members of the same cluster and decreases towards 0 for large numbers of clusters (see Figure 7). Depicted here is

the change of inter-cluster distance from one number of clusters to the next. The number of clusters to retain is selected as the number of clusters for which the change of inter-cluster distance remains smaller than a threshold value for all larger number of clusters. Inspecting the decrease in inter-cluster distance offers the possibility to detect situations where the finally selected number of clusters is sensitive to the selected threshold. The lower the threshold value the finer the clustering result will be. A threshold value of 10 % was chosen here based on sensitivity tests using threshold values between 5 and 20 %. Visual inspection of the clustering results for a set of FTIR observations at JFJ indicated that a realistic separation of different air mass histories could be obtained in most cases when using a threshold of 10%. Note, however that the selection of the threshold remains subjective.

An alternative approach would have been to analyse the complete time series of simulated partial columns and the calculate correlations between different layers and from that deduce a vertical correlation lengths scale. However, such an approach would only yield a temporal average estimate of representative, while the current approach can be used for an individual simulation. As can be seen in Figure 14 to Figure 19, estimated representativeness of the surface measurements varied strongly with time. Hence, the presented method was preferred to one that is using an average estimate of vertical correlations length scales.

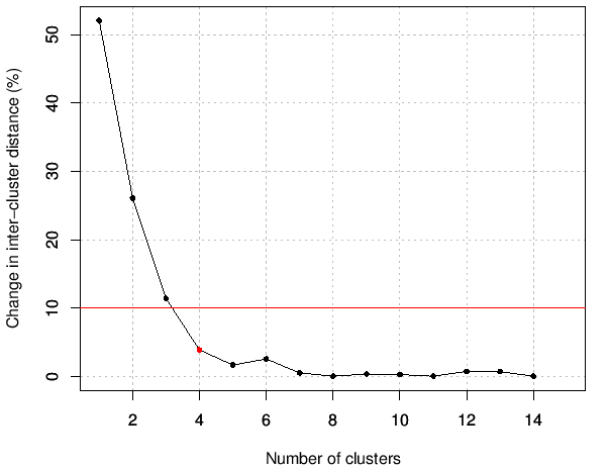

Figure 7: Example of inter-cluster distance for the sampling volumes of FTIR CO observation at JFJ on 2011-01-29 13:00. In this case the inter-cluster distance does not decrease by more than 10 % beyond 4 clusters. Hence, 4 clusters are retained by the algorithm.

2.6. Adjustment of 'in-situ' profiles

The adjustment of the model profiles as obtained by the backward dispersion simulation to bring them in close agreement with the observed surface in-situ values proceeds differently for different trace gas species. For CO and CH_4 explicit emissions were simulated during the period of advection by the backward dispersion simulation. Furthermore, the considered transport time was short enough to neglect degradation of these two species during the 10 day transport. Hence, it was possible to separate their simulated mole fractions into background and pollution signal. For NO_2 emissions were simulated, but its average chemical lifetime is smaller than the transport time. We used an average lifetime of 1 day to simulate the degradation of NO_2 as a first order process for both background and pollution signal. The uncertainty associated with this mimicked degradation is relatively large but should be dealt with by the in-situ adjustment of the simulated profiles. For O_3, which is formed secondarily in the atmosphere, no emission contribution was simulated and only the background was treated in the simulation. Ozone's lifetime in the free troposphere is long enough to

treat it passively. However, ozone is destroyed at the Earth's surface due to deposition and can be photo-chemically formed in fresh emission plumes. Neglecting these important production and loss processes within the last 10 days of transport may lead to considerable uncertainties in the FLEXPART derived profiles. Nevertheless, the adjustment to surface in-situ concentrations should compensate this deficiency to some extent.

The adjustment was only applied to partial columns for which the analysis of representativeness indicated that they had a similar air mass history as the surface in-situ sampling volume; or in other words: they belonged to the same cluster as the surface in-situ observations. In addition, we requested that all partial columns below the one of interest were also in the same cluster. Hence, partial columns, that belonged to the surface cluster but were separated from the surface by partial columns with different cluster membership, were not corrected. For all RS partial columns that were not in the surface cluster no adjustment of the FLEXPART simulated profile was done. We refer to the range of the profile to which adjustments were applied as "surface region".

In the following ψ is used for simulated mole fractions, χ for observed mole fractions (in-situ) and γ for adjusted mole fractions. An index s indicates surface in-situ observations and simulations; while an index i indicates the mole fractions for the remotely sampled partial columns. Indices b and p indicate background and pollution contribution, respectively.

For simulated profiles that were split into background and pollution the adjustment in the surface region was done in a two-step procedure. In the first step, the background fraction was adjusted with a linear bias correction that is scaled with a factor g describing the range of the background across the surface region

$$b = \psi_{s,b} - \chi_{s,b} \tag{7}$$

$$g_i = \left|\psi_{i,b} - \psi_{t,b}\right| / \left(\psi_{s,b} - \psi_{t,b}\right) \tag{8}$$

$$\gamma_{i,b} = \psi_{i,b} - b \cdot g_i \tag{9}$$

where $\psi_{t,b}$ is the simulated background mole fraction for the layer t directly above the topmost layer that belonged to the surface cluster. As can be seen, the factor g will be equal to unity for the surface simulation, forcing the simulated surface value to the observed surface value.

In the second step, the pollution fraction was adjusted with a scaling factor

$$f_i = \chi_{s,p} / \psi_{s,p} \tag{10}$$

and the final adjusted mole fraction is derived as the sum of background and pollution contribution

$$\gamma_i = \gamma_{i,b} + f_i \gamma_{i,p}. \tag{11}$$

If the observed in-situ mole fraction was smaller or equal to the background mole fraction a slightly different adjustment was performed. The above step 1 was performed followed by

$$\gamma_i = \gamma_{i,b} + \gamma_{i,p} - \gamma_{s,p} \tag{12}$$

This treatment sets the simulated pollution contribution at the surface to zero, but assures that frequently occurring elevated pollution layers are retained in the adjusted profiles.

For species for which no emissions were explicitly calculated (only O_3 in the context of NORS) the adjustment was done as follows. No distinction between background and pollution could be made for the simulated species directly, but a similarity factor, h_i, based on the simulated pollution contributions for CO was calculated.

$$h_i = 1 - \left|CO_{i,p} - CO_{s,p}\right| / \left(CO_{i,p} - CO_{s,p}\right) \tag{13}$$

h_i becomes 1 for the surface in-situ simulation or in situations when simulated pollution contributions are similar for the surface and elevated layers. h_i tends to zero if simulated pollution contributions are different between surface and elevated layers. The final adjustment then uses the scaling factor h_i to project the difference between surface observation and simulation onto the other sampling volumes

$$\gamma_i = \psi_i + (\chi_s - \psi_s) \cdot h_i. \quad (14)$$

Note again, that the adjustment was only applied to layers i that were part of the surface region.

Using this procedure, the FLEXPART simulated mole fractions for the surface in-situ volume was adjusted to exactly match the observed mole fractions. For the simulated partial columns the surface influence decreased and the amount of adjustment reduced.

Figure 8 shows two examples for the surface adjustment of simulated CO profiles at JFJ. On 2011-01-27 09:00 UTC the surface in-situ observations were detected to be only representative for a shallow layer above the site. Hence, the simulated profile was only adjusted towards the surface observation at the lowest FTIR retrieved partial column. In addition, it is noteworthy that an elevated layer of increased CO was simulated at an altitude of about 8 km. Such elevated layers were frequently simulated for JFJ and originate from North American emissions. The second example is from 2011-01-29 15:00 UTC. Here the influence of the surface observations reaches higher and simulated mole fractions were adjusted up to a height of 7.5 km. In both cases, the simulated mole fraction in the lowest FTIR partial column was adjusted to almost match the surface in-situ observation. The remaining difference stems from the fact that the simulation for the surface in-situ sampling volume and the lowest RS partial column are not identical.

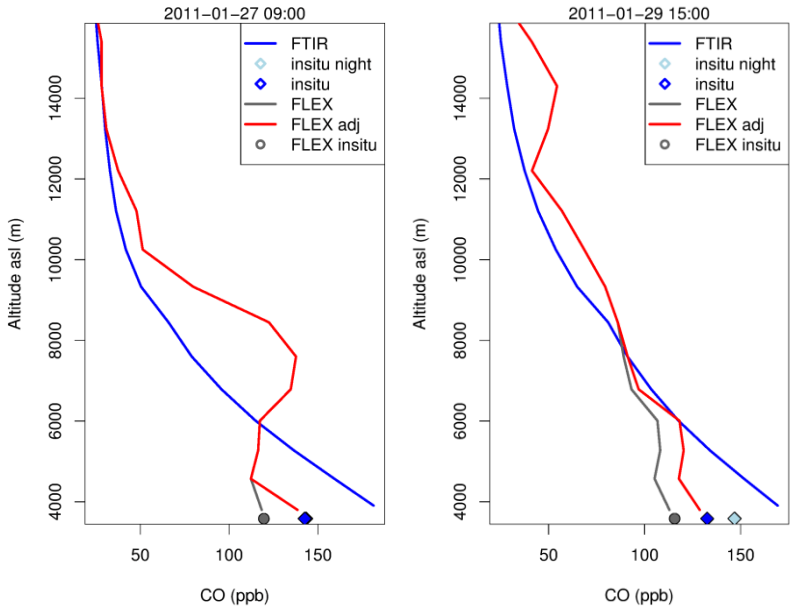

Figure 8: Examples of simulated and RS profiles for two days in January 2011. The dark blue diamond represents the surface in-situ measurement at the time of the FTIR observation, while the light blue diamond is the night-time observation bracketing the day-time FTIR observation. The simulated surface mole fraction is shown as a grey circle. The simulated profile is given as a grey line, while the adjusted profile is given as a red line. Where no grey line is visible it is hidden behind the red line, hence, no adjustment was applied for these levels. In addition, the FTIR retrieved profile is given in blue.

2.7. Comparison with remote sensing profiles

In order to compare the reference profiles with the RS profiles the averaging kernel (AVK) of the remote sensing technique needs to be taken into account. One can apply the AVK to the reference profiles 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 he retrieval a-priori, a

$$\phi = \chi_a + \boldsymbol{A}(\gamma - \chi_a) \quad (15)$$

to yield the smoothed profile ϕ.

3. Results

In the following the results of the FLEXPART simulations for FTIR observations are discussed in terms of their performance when compared to the in-situ observations. From this we deduce recommendations for the use of the generated reference profiles and their uncertainties. Adjusted profiles are presented here, while the actually produced data sets are described in the deliverable report D5.2. Further information on individual profiles is available through the website: http://lagrange.empa.ch/NORS_browser/.

3.1. Validation of model simulations

In order to document the quality of the FLEXPART simulations, their results for the surface sampling volume were compared to surface in-situ observations (Figure 9 to Figure 13).

For CH_4 at JFJ, three different simulated data sets were available, based on different initial conditions as obtained from the GCTMs (see section 2.4). Linear regression analysis between the simulated in-situ and observed in-situ mole fractions (Figure 9) show that the performance of the FLEXPART simulation strongly depends on these initial conditions. The slope of the linear regression was obtained using weighted total least-square regression [*Krystek and Anton*, 2007], which takes uncertainties in both variables into account. The best simulation results (in terms of correlation coefficient, RMSE and linear regression slope) were obtained when input was taken from bias corrected FLEXCTM. Using the bias corrected MACC re-analysis data resulted in relatively large scatter and the model slightly underestimated the slope. Using initial conditions from TM5, the comparison shows that the model had difficulties to reproduce the lowest observed mole fractions and systematically overestimated these. This is even more evident when directly comparing the TM5 output with the surface observations (not shown). At IZO the model performance was not as good as for JFJ in terms of overall correlation but was similar in terms of RMSE and slope (Figure 10). Again, the largest scatter was observed when using MACC input. Similarly to JFJ, TM5 showed difficulties to reproduce the lowest part of observed surface mole fractions. One of the reasons for a weaker correlation between simulation and observation may be seen in the generally smaller variability observed at IZO where CH_4 mole fractions mainly ranged between 1800 and 1850 ppb, while at JFJ mole fractions between 1800 and 1900 ppb were observed.

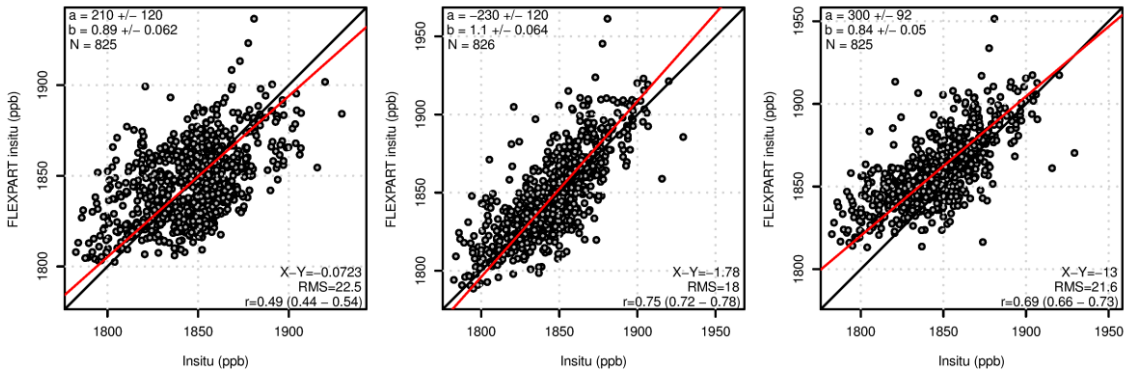

Figure 9: FLEXPART simulated CH_4 mole fractions compared to in-situ observations at JFJ. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM, TM5.

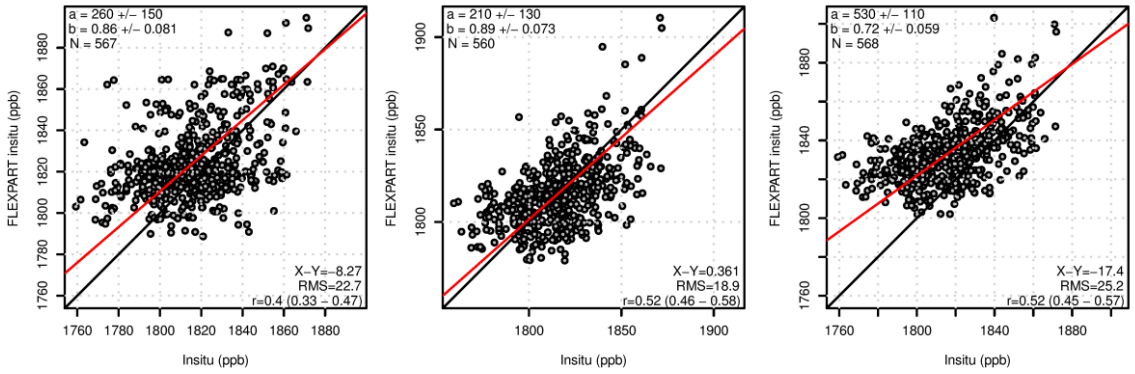

Figure 10: FLEXPART simulated CH_4 mole fractions compared to in-situ observations at IZO. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM, TM5.

In the case of CO, MACC re-analysis and FLEXCTM simulations were used as initial conditions for the FLEXPART simulations. The results of the comparison with the surface in-situ observations at JFJ were very satisfactory and similar in terms of correlation and RMSE for both sets of initial conditions (Figure 11). However, the slope was significantly closer to 1 for FLEXCTM. For IZO, the comparison between simulations and surface in-situ observations was less satisfactory as compared to JFJ in terms of correlation and slope (Figure 12). As for CH_4, CO RMSE were similar at both sites. Results for the two initial condition data sets were similar, with MACC showing a slightly better correlation.

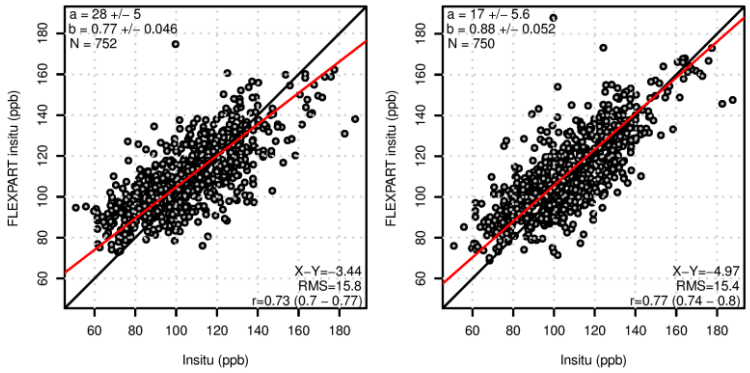

Figure 11: FLEXPART simulated CO mole fractions compared to in-situ observations at JFJ. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM.

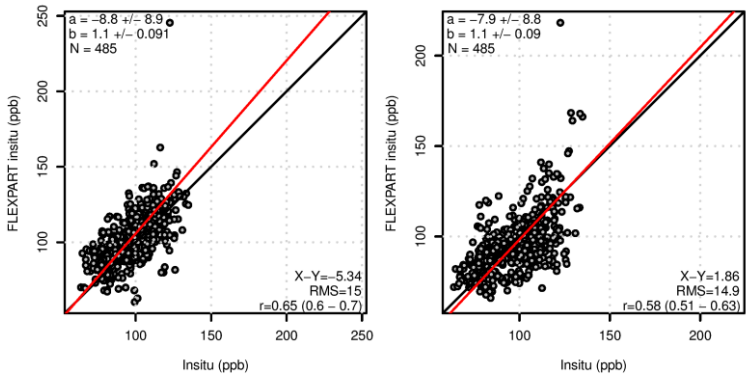

Figure 12: FLEXPART simulated CO mole fractions compared to in-situ observations at IZO. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM.

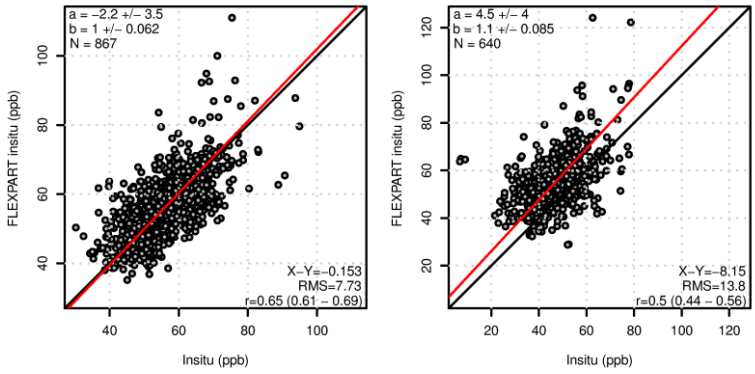

Figure 13: FLEXPART simulated O_3 mole fractions compared to in-situ observations at (left) JFJ and (right) IZO. Initial conditions were taken from the MACC re-analysis.

For ozone initial conditions were solely taken from MACC re-analysis. For JFJ the model simulations were well correlated with the observations (r=0.65) and the linear regression revealed a slope close to 1 (Figure 13). For IZO both correlation coefficient (r=0.5) and regression slope (b=1.1) were not as satisfactory as for JFJ. The comparison did not improve when comparing the model simulations to the night-time observations. This was tested to avoid possible day-time influence of up-slope flow on O_3 [*Cuevas, et al.*, 2013] that could not be captured by the transport model.

Table 1: Performance statistics of simulated surface mole fractions as compared to surface in-situ observations. The model realization (in terms of initial conditions) with the best performance is highlighted in bold italics. R: correlation coefficient, RMSE: root mean square error, N: number of observations. Rows given in bold and italics indicate the initial conditions yielding the best performance.

Site	Species	Initial Cond	R	Slope	RMSE	N
JFJ	CH4	MACC	0.49	0.89	22.5	825
		FLEXCTM	***0.75***	***1.1***	***18***	***826***
		TM5	0.69	0.84	21.6	825
JFJ	CO	MACC	0.73	0.77	15.8	752
		FLEXCTM	***0.77***	***0.88***	***15.4***	***750***
JFJ	O3	***MACC***	***0.65***	***1***	***7.73***	***867***
IZO	CH4	MACC	0.4	0.86	22.7	567

		FLEXCTM	***0.52***	***0.89***	***18.9***	***560***
		TM5	0.52	0.72	25.2	568
IZO	CO	***MACC***	***0.65***	***1.1***	***15***	***485***
		FLEXCTM	0.58	1.1	14.9	485
IZO	O3	***MACC***	***0.5***	***1.1***	***13.8***	***640***

3.2. Vertical extent of profile adjustment

The vertical representativeness and the extent of the model profile adjustment is documented in Figure 14 to Figure 19, which show results for the three target species and two sites for the model realisation with the best performance for the surface in-situ observations (see also Table 1).

For CH_4 simulations at JFJ the adjustment was mostly restricted to the lowest 2 to 3 RS partial columns (Figure 14, left) as indicated by the same cluster membership of these partial columns as the surface in-situ. The adjustment was mostly negative for the first half of the comparison period and tended to be positive in the winter months of the second half of the comparison period. This may indicate a small trend in the CH_4 initial conditions provided by FLEXCTM. The Adjustment was usually in the range of -30 to +20 ppb, with few cases exceeding these limits.

Similar vertical representativeness was deduced for the CO observations at JFJ (Figure 15). Note, that the vertical discretisation of the CO and CH_4 FTIR retrievals is not the same at JFJ. For CO the vertical resolution is small in the lower part of the troposphere and larger for the upper part of the troposphere as compared to the CH_4 retrieval. This leads to a reduced extent of the surface region for the CO partial columns. Usually only the first two CO partial columns were put in the same vertical region (cluster) as the surface in-situ observation. The CO adjustment was mostly negative (model overestimating mole fractions at the site) for the years 2009 and 2011, but positive for 2010. It mainly ranged from -25 to +15 ppb.

For O_3 simulations at JFJ the adjustment ranged from -10 to +15, showing model underestimation of the observations in the years 2009 and 2010 and a slight overestimation in 2011 (Figure 16).

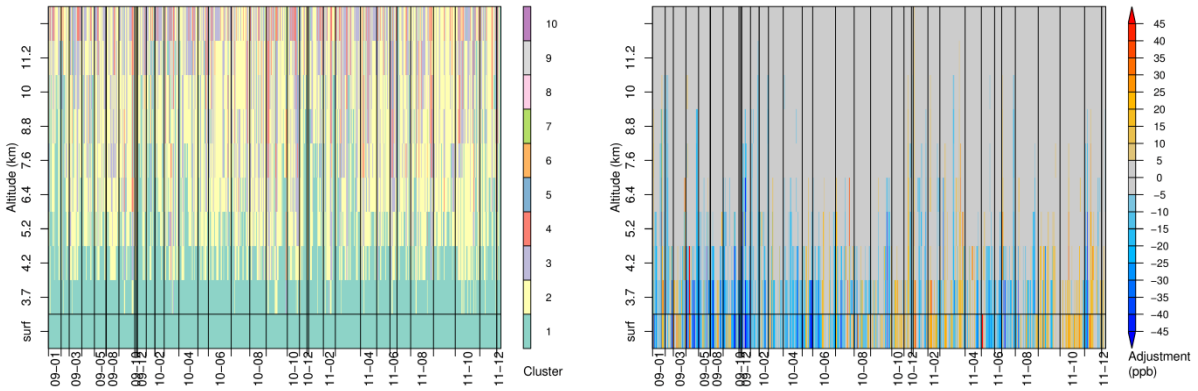

Figure 14: Time altitude cross section of (left) estimated atmospheric layer (cluster) and (right) amount of adjustment of the simulated profile (initial conditions from FLEXCTM) by the surface in-situ observation for FTIR CH_4 observations at JFJ. Note that the vertical axis is not linear and that the time axis is not continuous, but individual profiles at the time of FTIR observations were stacked together. The vertical black lines separate observations from individual months, labelled by year-month on the x-axis. The value in the bottom row refers to the simulations for the surface in-situ sampling volume.

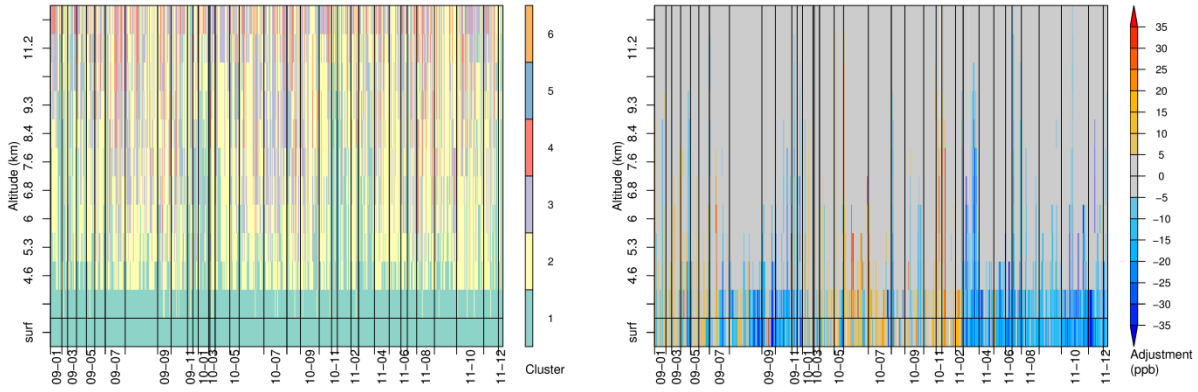

Figure 15: Same as Figure 14, but for FTIR CO observations at JFJ and the simulation using FLEXCTM initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14.

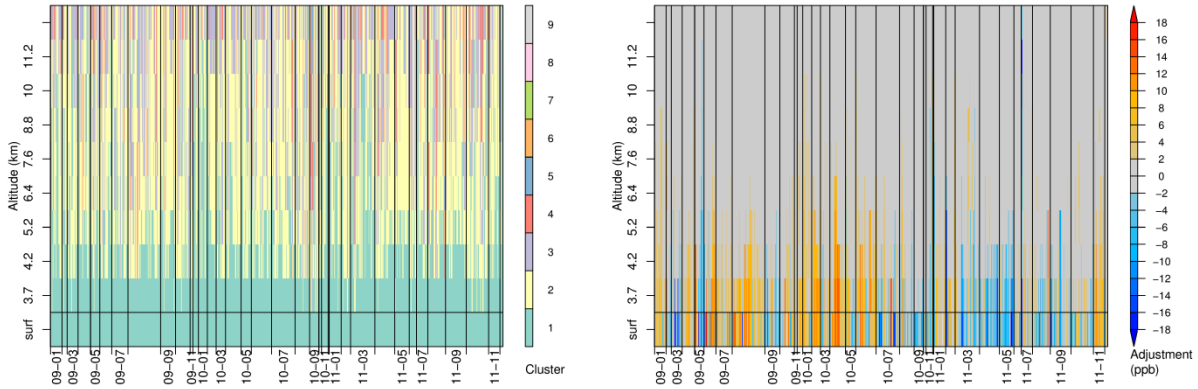

Figure 16: Same as Figure 14, but for FTIR O_3 observations at JFJ and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14.

The vertical representativeness of the surface observations at IZO is more complex than at JFJ often showing multilayer structures as depicted in Figure 22 and elevated layers that were put in the same cluster as the surface observations (Figure 17, left). Mole fractions in such elevated layers were not corrected with the surface in-situ observations (see section 2.6), often limiting the vertical influence of the adjustment. For CH_4 the adjustment was mostly negative with some periods of positive adjustment during the winter months, and restricted to the lowest RS partial column.

For CO as well as for O_3 the adjustments were generally negative with few exceptions in the winter months.

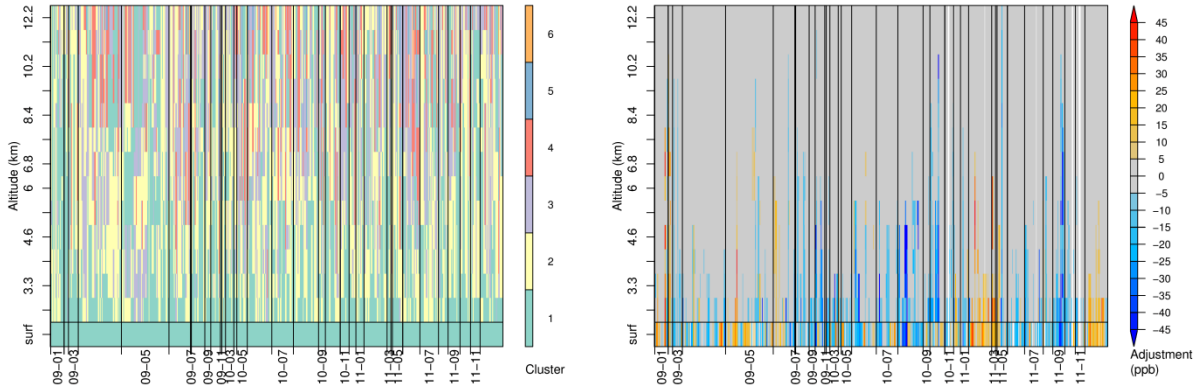

Figure 17: Same as Figure 14, but for FTIR CH_4 observations at IZO and the simulation using FLEXCTM initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14.

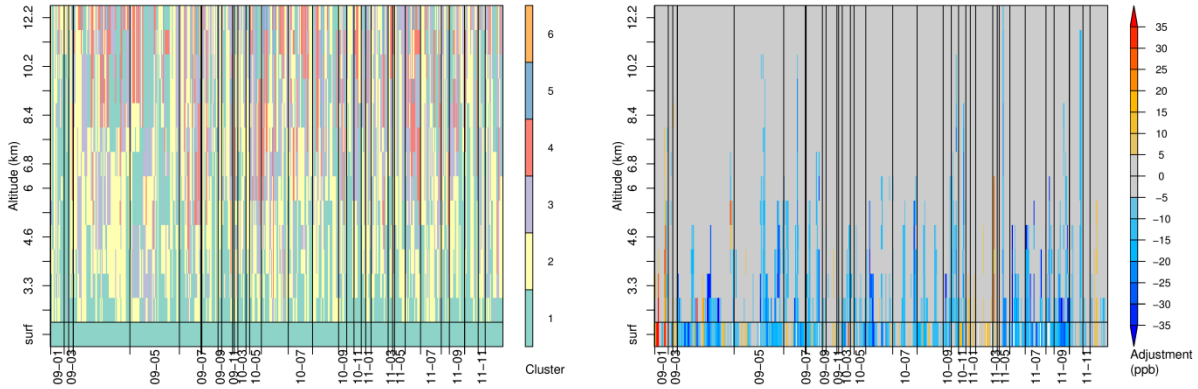

Figure 18: Same as Figure 14, but for FTIR CO observations at IZO and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14.

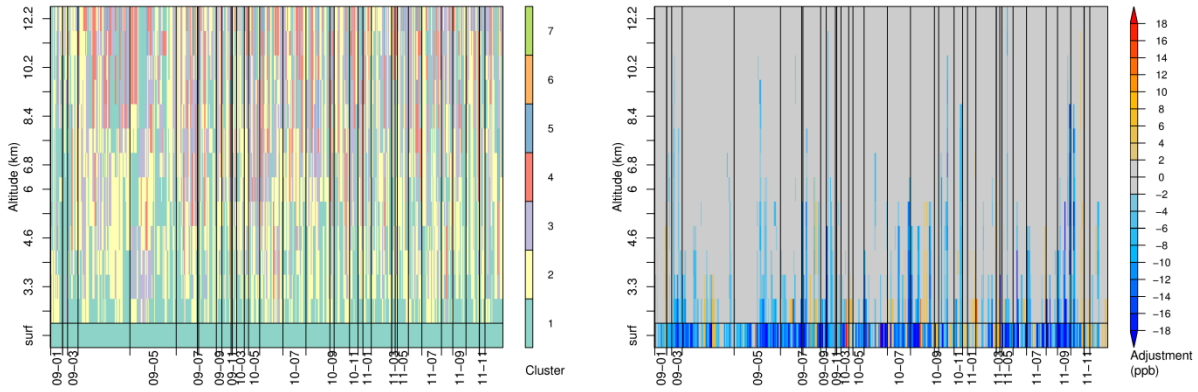

Figure 19: Same as Figure 14, but for FTIR O_3 observations at IZO and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14.

3.3. Validation of adjusted profiles

The adjustment of the simulated profiles (RS partial columns) was done with the aim of creating a reference data set to which the RS observation can be compared. Since the reference dataset is a secondary product merging the in-situ with model data, it should be tested against other reference datasets, such as airborne in-situ observations, as well. In lack of such dataset in the vicinity of the demonstration sites, a possible check of the reference profile is the comparison of the adjusted simulated values of the lowest RS partial column with the surface in-situ observation itself. Assuming

that the surface in-situ observation is in general also representative for the lowest RS partial column the adjustment should produce perfect agreement between the two time series. While the applied adjustment does not make this assumption but evaluates the representativeness range of the surface observations, one can still expect that the adjustment will bring the simulations in the lowest partial column towards the surface observation. Hence, such a comparison can be seen as a validation of the adjustment method.

The linear regression between the surface in-situ data and the adjusted lowest RS partial column is shown in Figure 20 and Figure 21 for the model realisation with the initial conditions that gave the best performance when surface simulations were compared. At JFJ the adjusted simulations in the lowest RS partial column were in close agreement to the surface observations. This can be seen by correlation coefficients of ~0.9 and slopes that were not significantly different from 1. It is also noteworthy that the adjustment worked well for the extreme values of the observed distribution. These are situations, in which models usually perform the least well. The comparison supports the conclusion that the surface measurements are often representative of the lowest RS partial column, the adjustment works correctly and removes most of the simulation uncertainty by incorporating the surface in-situ observations. Hence, the adjusted datasets for the RS sampling volumes can be seen as a valid extrapolation of the surface reference data.

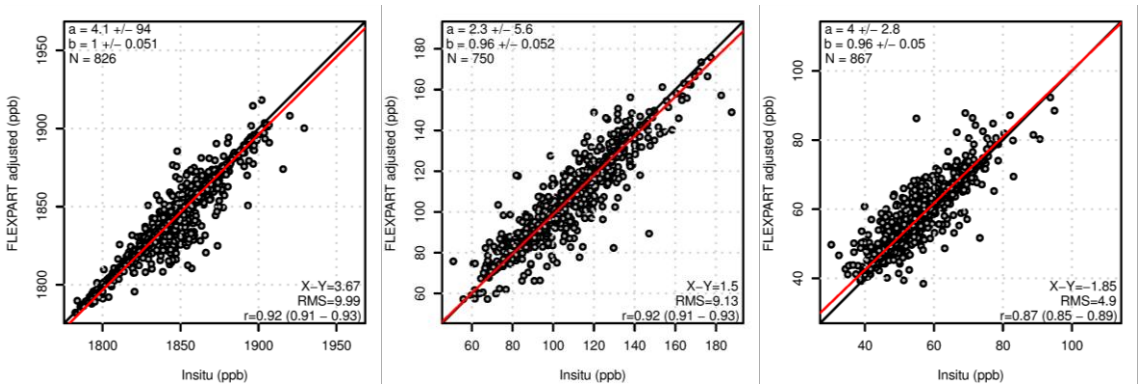

Figure 20: Linear regression of adjusted mole fraction in the lowest RS partial column versus surface in situ observations for JFJ. (left) adjusted CH_4 simulations with initial conditions from FLEXCTM, (middle) adjusted CO simulations with initial conditions from FLEXCTM, and (right) adjusted O_3 simulations with initial conditions from MACC.

The same comparison for the site IZO did not show a similarly close connection between the surface observations and the adjusted simulations for the lowest RS partial column (Figure 21). Next to the generally worse model performance at IZO this is caused by the fact that on average the vertical range, for which the in-situ observations were estimated to be representative, was smaller at IZO than at JFJ. This is somewhat surprising since in a more remote location like IZO one would expect horizontal and vertical concentration gradients to be small. On the other hand, IZO is located in the sub-tropical high pressure belt, with generally sinking motion and increased vertical stability. The latter hinders vertical exchange and mixing, resulting in long-lived vertical layer structures. These are often identified as multiple layers by the cluster approach described above, restricting the vertical influence of the surface measurements. An example of such a complex vertical layering is given in Figure 22. In contrast, JFJ is situated in the mid-latitudes where transient low pressure systems and frontal passages often lead to strong vertical mixing, extending the vertical representativeness of surface observations. The question remains if the vertical clustering approach as it is used here is too sensitive to the detection of such long lived elevated layers at IZO and if their distinction is also justified in terms of atmospheric composition.

Nevertheless, the created reference data set for the FTIR observation at IZO is still thought to be useful for the validation of the FTIR retrievals. The uncertainty associated to the reference data set, though, is larger than for JFJ and this should be taken into account when using the data.

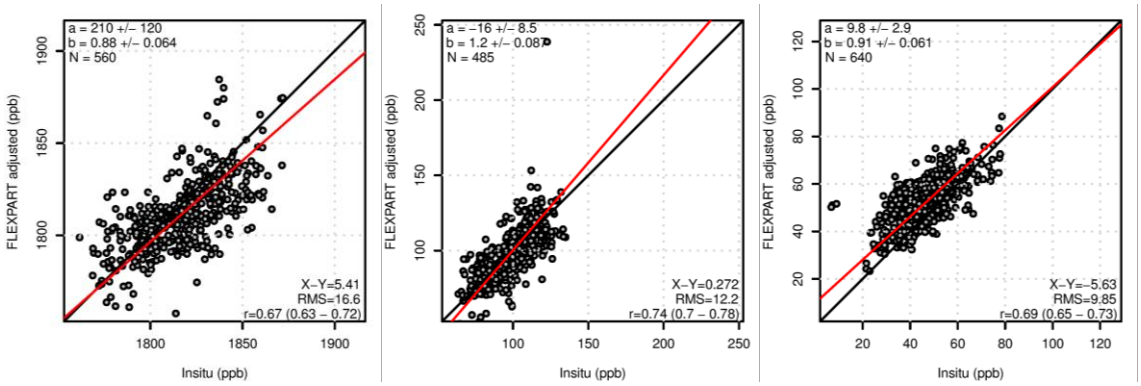

Figure 21: Linear regression of adjusted mole fraction in the lowest RS partial column versus surface in situ observations for JFJ. (left) adjusted CH_4 simulations with initial conditions from FLEXCTM, (middle) adjusted CO simulations with initial conditions from MACC, and (right) adjusted O_3 simulations with initial conditions from MACC.

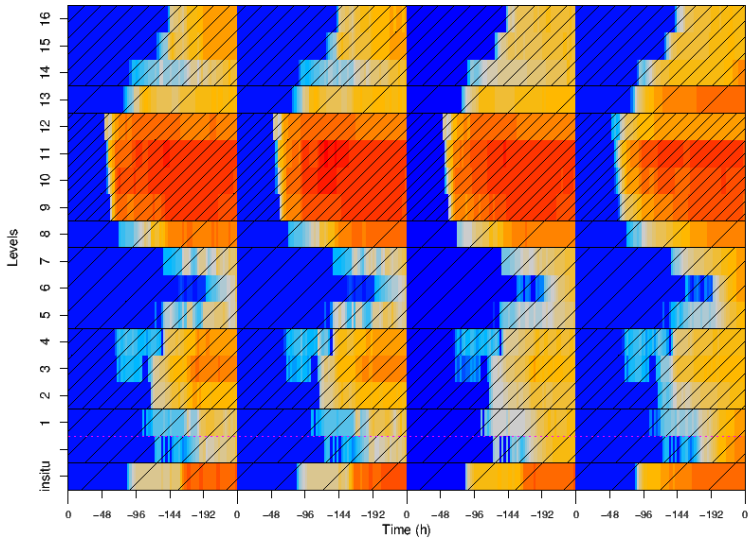

Figure 22: Same as Figure 6 but for the sampling volumes of the FTIR CO observation on 2011-07-11 16:00 at IZO.

3.4. Uncertainties of adjusted profiles

A complete analysis of the uncertainties of the adjusted model profiles, which serve as a reference for the FTIR observations, is not feasible for different reasons. Firstly, the model simulation itself does not provide an estimate of its uncertainty, given the deterministic character of the used model. Second, the error propagation of the adjustment process is not straight forward and cannot be covered in the context of this study.

In lack of a better estimate, we suggest to use the combined uncertainty of the in-situ observations and the RMSE as obtained from the in-situ simulation versus observation comparison as a proxy for the overall uncertainty of the adjusted profiles for partial columns that were adjusted.

$$\sigma_\gamma = \sqrt{f\sigma_\chi^2 + (1-f)RMSE^2} \quad (16)$$

where f is a factor between 0 and 1 describing the influence of the observed values on the adjusted profile. Using the RMSE as a proxy for the model uncertainty is a very conservative estimate of the uncertainty, since it was derived for the model's performance with respect to surface observations. In the free troposphere, where the variability of trace species is usually smaller, we expect the model to perform better. It is important to note that the uncertainty σ_γ is the uncertainty of an individual partial column and not of the total tropospheric column. It remains unclear how the factor f should be derived. f is not proportional to the amount of adjustment of the simulated profiles, since a profile that is not adjusted because it is close to the observations should still be assigned a small uncertainty. The simple approach that we followed here was to set f to 1 for all partial columns that were clustered into the surface region and set f to 0 elsewhere. This procedure may underestimate the uncertainty in the surface region because it assumes that the surface observation is fully representative for the surface region. It properly overestimates uncertainty for free tropospheric partial columns for reasons given above.

4. Conclusions

A method for surface in-situ measurement extrapolation was presented that is tailored towards individual remote sensing observations. The approach uses detailed Lagrangian backward transport simulations of individual remote sensing partial columns and of in-situ observations to analyse representativeness and generate volume specific trace gas simulations. The obtained information is finally used to extrapolate the surface in-situ observations and produce reference datasets. These serve to validate the remote sensing observation taking the particularities of the retrieval, as expressed by the averaging kernels and a-priori profiles, into account.

Here, the method was applied to the FTIR observations of CO, CH_4, and O_3 at the NORS demonstration sites Jungfraujoch and Izaña for the period 2009 to 2011. The method is neither limited to this remote sensing method nor to the parameters or sites. However, it is required that the remote sensing measurement volume can be reasonably well described, because the applied transport simulations are specific to these.

Simulated surface trace gas mole fractions were compared to surface in-situ observations, showing generally very satisfactory performance of the transport simulations. The use of different initial model conditions allowed for the selection of the model realisation with the best performance. The use of the respective reference dataset for remote sensing validation is recommended (see Table 1).

As a general test of the performance of the in-situ data extrapolation, the lowest remote sensing partial column of the generated reference datasets was compared to the in-situ observations as well. For Jungfraujoch this test showed very close agreement with correlation coefficients of ~0.9 and slopes that were not significantly different from unity. For Izaña the test showed somewhat reduced comparison statistics, but this was attributed to the more limited vertical representativeness of the in-situ observations as compared with Jungfraujoch.

The generated reference profiles will further serve for the validation of the currently available FTIR retrieved tropospheric trace gas partial columns. In a further step, the existing FTIR data will be re-analysed and the retrieval may be tuned to improve the comparability to the reference datasets.

Acknowledgements

The authors thank P. Bergamaschi for making TM5 CH_4 simulations and E. Cuevas for making surface in-situ observations from Izaña available to the NORS project.

Acronyms and abbreviations

asl	above sea level
AVK	Averaging Kernel
CO	Carbon Monoxide
CLD	Chemiluminescence Detectors
CH_4	Methane
FTIR	Fourier transform infrared spectroscopy
GAW	Global Atmosphere Watch
GEOMS	Generic Earth Observation Metadata Standard
GC	Gas Chromatograph
LPDM	Lagrangian Particle Dispersion Model
MAXDOAS	Multi axis differential optical absorption spectroscopy
NIST	US National Institute for Standards and Technology
NDACC	Network for the Detection of Atmospheric Composition Change
IZO	Izaña observatory
JFJ	Jungfraujoch observatory
NO_2	Nitrogen Dioxide
NWP	Numerical Weather Prediction
O_3	Ozone
RMSE	Root mean square error
RS	Remote Sensing
SRP	Standard Reference Photometer
WMO	World Meteorological Organisation

References

Barret, B., M. De Mazière, and E. Mahieu (2003), Ground-based FTIR measurements of CO from the Jungfraujoch: characterisation and comparison with in situ surface and MOPITT data, *Atmos. Chem. Phys.*, *3*, 2217-2223, 10.5194/acp-3-2217-2003.

Bergamaschi, P., S. Houweling, A. Segers, M. Krol, C. Frankenberg, R. A. Scheepmaker, E. Dlugokencky, S. C. Wofsy, E. A. Kort, C. Sweeney, et al. (2013), 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, 10.1002/jgrd.50480.

Cuevas, E., Y. González, S. Rodríguez, J. C. Guerra, A. J. Gómez-Peláez, S. Alonso-Pérez, J. Bustos, and C. Milford (2013), Assessment of atmospheric processes driving ozone variations in the subtropical North Atlantic free troposphere, *Atmos. Chem. Phys.*, *13*, 1973-1998, 10.5194/acp-13-1973-2013.

Dils, B., J. Cui, S. Henne, E. Mahieu, M. Steinbacher, and M. De Maziere (2011), 1997-2007 CO trend at the high Alpine site Jungfraujoch: a comparison between NDIR surface in situ and FTIR remote sensing observations, *Atmos. Chem. Phys.*, *11*, 6735-6748, 10.5194/acp-11-6735-2011.

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

Henne, S., D. Brunner, D. Folini, S. Solberg, J. Klausen, and B. Buchmann (2010), Assessment of parameters describing representativeness of air quality in-situ measurement sites, *Atmos. Chem. Phys.*, *10*, 3561-3581.

Henne, S., D. Brunner, J. Klausen, C. Schnadt-Poberaj, and B. Buchmann (2013a), A Lagrangian particle dispersion model for simulations of global carbon monoxide and methane: Part I setup and evaluation against surface observations, *Geosci. Model Dev.*, *in preparation*.

Henne, S., P. Kaufmann, M. Schraner, and D. Brunner (2013b), Coupling FLEXPART to the regional scale numerical weather prediction model COSMO: Implementation, evaluation and first results, paper presented at EGU General Assembly, European Geophysical Union, Vienna, Austria.

Henne, S., C. Schnadt Poberaj, S. Reimann, and D. Brunner (2013c), Global-scale tropospheric Lagrangian particle models with linear chemistry, in *Lagrangian Modeling of the Atmosphere*, edited by J. C. Lin, C. Gerbig, et al., pp. 235-250, AGU, Washington, DC.

Inness, A., F. Baier, A. Benedetti, I. Bouarar, S. Chabrillat, H. Clark, C. Clerbaux, P. Coheur, R. J. Engelen, Q. Errera, et al. (2013), The MACC reanalysis: an 8 yr data set of atmospheric composition, *Atmos. Chem. Phys.*, *13*, 4073-4109, 10.5194/acp-13-4073-2013.

Kaufman, L., and P. J. Rousseeuw (1990), *Finding Groups in Data. An Introduction to Cluster Analysis*, 342 pp., John Wiley and Sons, New York.

Krystek, M., and M. Anton (2007), A weighted total least-squares algorithm for fitting a straight line, *Measurement Science & Technology*, *18*, 3438-3442, 10.1088/0957-0233/18/11/025.

Olivier, J. G. J., and J. J. M. Berdowski (2001), Global Emissions Sources and Sinks, in *The Climate System*, edited by J. Berdowski, R. Guicherit, et al., pp. 33-78, Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, The Netherlands.

Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *Journal of Geophysical Research: Atmospheres*, *108*, 4116, 10.1029/2002JD002299.

Ruckstuhl, A. F., S. Henne, S. Reimann, M. Steinbacher, M. K. Vollmer, S. O'Doherty, B. Buchmann, and C. Hueglin (2012), Robust extraction of baseline signal of atmospheric trace species using local regression, *Atmos. Meas. Tech.*, *5*, 2613-2624, 10.5194/amt-5-2613-2012.

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

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

Zellweger, C., C. Huglin, J. Klausen, M. Steinbacher, M. Vollmer, and B. Buchmann (2009a), Inter-comparison of four different carbon monoxide measurement techniques and evaluation of the long-term carbon monoxide time series of Jungfraujoch, *Atmos. Chem. Phys.*, *9*, 3491-3503, 10.5194/acp-9-3491-2009.

Zellweger, C., J. Klausen, B. Buchmann, and H. E. Scheel (2009b), System and Performance Audit for Surface Ozone, Global GAW Station Izana, Spain, 53 pp, Empa-WCC, Dübendorf, Switzerland.

List of Tables

Table 1: Performance statistics of simulated surface mole fractions as compared to surface in-situ observations. The model realization (in terms of initial conditions) with the best performance is highlighted in bold italics. R: correlation coefficient, RMSE: root mean square error, N: number of observations. Rows given in bold and italics indicate the initial conditions yielding the best performance. 21

List of figures

Figure 1: Time series of surface in-situ observations that are collocated with FTIR observations (not shown) for (pale dots) JFJ and (crosses) IZO. 10

Figure 2: Example of the horizontal position (rectangles) of the tropospheric partial columns of the FTIR observation at JFJ (blue dot in centre) on 2013-01-27 09:00 UTC. The colour scale on the right refers to the height of the topography, while the colour of the rectangles gives the altitude of the partial column centre between the surface and 16 km altitude (scale not shown). 11

Figure 3: COSMO domains used as input for the LPDM simulations. Courtesy of MeteoSwiss. 12

Figure 4: Example of source sensitivities obtained for lowest partial column of the CO FTIR retrieval at JFJ on 2013-01-29 13:00 UTC. (left) zoom centred over the Alps for FLEXPART-COSMO during the first

48 hours of the simulation, (right) Atlantic region combined from FLEXPART-COSMO and FLEXPART-ECMWF during the complete 240 hours of the simulation..... 13

Figure 5: Example of source sensitivities obtained for the partial column reaching from 6.4 to 7.1 km asl of the CO FTIR retrieval at JFJ on 2013-01-29 13:00 UTC. (left) zoom centred over the Alps for FLEXPART-COSMO during the first 48 hours of the simulation, (right) Atlantic region combined from FLEXPART-COSMO and FLEXPART-ECMWF during the complete 240 hours of the simulation..... 13

Figure 6: Two-dimensional representation of the variables used for clustering of the sampling volumes of FTIR CO observation at JFJ on 2011-01-29 13:00. On the x-axis 4 different groups of normalized cluster variables and their development over time along the backward simulation are displayed (from left to right: source sensitivities, CO emission uptake, CH_4 emissions uptake, NO_x emission uptake). On the y-axis the different vertical layers of the RS retrieval and at the bottom the surface in-situ simulations are given. Red colours correspond to large values, blue values to small values. Horizontal bars that exhibit a similar pattern identify layers that experience a similar air mass history in terms of total surface contact and emission update. The different shading shows such layers as they were identified by the cluster algorithm..... 15

Figure 7: Example of inter-cluster distance for the sampling volumes of FTIR CO observation at JFJ on 2011-01-29 13:00. In this case the inter-cluster distance does not decrease by more than 10 % beyond 4 clusters. Hence, 4 clusters are retained by the algorithm. 16

Figure 8: Examples of simulated and RS profiles for two days in January 2011. The dark blue diamond represents the surface in-situ measurement at the time of the FTIR observation, while the light blue diamond is the night-time observation bracketing the day-time FTIR observation. The simulated surface mole fraction is shown as a grey circle. The simulated profile is given as a grey line, while the adjusted profile is given as a red line. Where no grey line is visible it is hidden behind the red line, hence, no adjustment was applied for these levels. In addition, the FTIR retrieved profile is given in blue..... 18

Figure 9: FLEXPART simulated CH_4 mole fractions compared to in-situ observations at JFJ. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM, TM5..... 20

Figure 10: FLEXPART simulated CH_4 mole fractions compared to in-situ observations at IZO. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM, TM5..... 20

Figure 11: FLEXPART simulated CO mole fractions compared to in-situ observations at JFJ. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM..... 20

Figure 12: FLEXPART simulated CO mole fractions compared to in-situ observations at IZO. Initial conditions were taken from different global scale chemistry transport models: (left to right) MACC, FLEXCTM..... 21

Figure 13: FLEXPART simulated O_3 mole fractions compared to in-situ observations at (left) JFJ and (right) IZO. Initial conditions were taken from the MACC re-analysis. 21

Figure 14: Time altitude cross section of (left) estimated atmospheric layer (cluster) and (right) amount of adjustment of the simulated profile (initial conditions from FLEXCTM) by the surface in-situ observation for FTIR CH_4 observations at JFJ. Note that the vertical axis is not linear and that the time axis is not continuous, but individual profiles at the time of FTIR observations were stacked together. The vertical black lines separate observations from individual months, labelled by year-

month on the x-axis. The value in the bottom row refers to the simulations for the surface in-situ sampling volume. 22

Figure 15: Same as Figure 14, but for FTIR CO observations at JFJ and the simulation using FLEXCTM initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14. ... 23

Figure 16: Same as Figure 14, but for FTIR O_3 observations at JFJ and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14. ... 23

Figure 17: Same as Figure 14, but for FTIR CH_4 observations at IZO and the simulation using FLEXCTM initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14. ... 24

Figure 18: Same as Figure 14, but for FTIR CO observations at IZO and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14. ... 24

Figure 19: Same as Figure 14, but for FTIR O_3 observations at IZO and the simulation using MACC initial conditions. Note that the colour scale of the atmospheric layer is different from Figure 14. ... 24

Figure 20: Linear regression of adjusted mole fraction in the lowest RS partial column versus surface in situ observations for JFJ. (left) adjusted CH_4 simulations with initial conditions from FLEXCTM, (middle) adjusted CO simulations with initial conditions from FLEXCTM, and (right) adjusted O_3 simulations with initial conditions from MACC. 25

Figure 21: Linear regression of adjusted mole fraction in the lowest RS partial column versus surface in situ observations for JFJ. (left) adjusted CH_4 simulations with initial conditions from FLEXCTM, (middle) adjusted CO simulations with initial conditions from MACC, and (right) adjusted O_3 simulations with initial conditions from MACC. 26

Figure 22: Same as Figure 6 but for the sampling volumes of the FTIR CO observation on 2011-07-11 16:00 at IZO. 26