Manuscript prepared for Atmos. Chem. Phys.
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 28 January 2016

Ability of the 4-D-Var analysis of the GOSAT BESD XCO_2 retrievals to characterize atmospheric CO_2 at large and synoptic scales

S. Massart[1], A. Agustí-Panareda[1], J. Heymann[2], M. Buchwitz[2], F. Chevallier[3], M. Reuter[2], M. Hilker[2], J. P. Burrows[2], N. M. Deutscher[2,9], D. G. Feist[4], F. Hase[5], R. Sussmann[6], F. Desmet[7], M. K. Dubey[8], D. W. T. Griffith[9], R. Kivi[10], C. Petri[2], M. Schneider[5], and V. A. Velazco[9]

[1]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
[2]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[3]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, IPSL, Gif sur Yvette, France
[4]Max Planck Institute for Biogeochemistry, Jena, Germany
[5]Karlsruhe Institute of Technology, IMK-ASF, Karlsruhe, Germany
[6]Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany
[7]University of Antwerp, Antwerp, Belgium
[8]Earth and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, United States
[9]Centre for Atmospheric Chemistry, School of Chemistry, University of Wollongong, Wollongong, Australia
[10]Finnish Meteorological Institute, Arctic Research, Sodankylä, Finland

Correspondence to: S. Massart (sebastien.massart@ecmwf.int)

Abstract. This study presents results from the European Centre for Medium-Range Weather Forecasts (ECMWF) carbon dioxide (CO_2) analysis system where the atmospheric CO_2 is controlled through the assimilation of column-averaged dry-air mole fractions of CO_2 (XCO_2) from the Greenhouse gases Observing Satellite (GOSAT). The analysis is compared to a free run simulation (without assimilation of XCO_2) and they are both evaluated against XCO_2 data from the Total Carbon Column Observing Network (TCCON). We show that the assimilation of the GOSAT XCO_2 product from the Bremen Optimal Estimation Differential Optical Absorption Spectroscopy (BESD) algorithm during the year 2013 provides XCO_2 fields with an improved mean absolute error of 0.6 parts per million (ppm) and an improved station-to-station bias deviation of 0.7 ppm compared to the free run (1.1 ppm and 1.4 ppm, respectively) and an improved estimated precision of 1 ppm compared to the GOSAT BESD data (3.3 ppm). We also show that the analysis has skill for synoptic situations in the vicinity of frontal systems where the GOSAT retrievals are sparse due to cloud contamination. We finally computed the 10 day forecast from each analysis at 00:00 UTC, and we demonstrate that the CO_2 forecast shows synoptic skill for the largest scale weather patterns (of the order of 1000 km) even up to day 5 compared to its own analysis.

1 Introduction

Carbon in the atmosphere is present mostly in the form of carbon dioxide (CO_2). Its amount is relatively small compared to the amount of carbon present in other reservoirs like the ocean (Ciais et al., 2013). Being well mixed, atmospheric CO_2 is nevertheless easier to monitor by measurements than other carbon reservoirs. To improve the monitoring of atmospheric CO_2, one can combine atmospheric CO_2 measurements with a numerical model. This paper describes such a system, which has been developed for the Copernicus Atmosphere Monitoring Service (CAMS).

Rather than using the relatively sparse network of the surface air-sample measurements, here we explore the measurements from satellite sounders in order to have a more global picture of the atmospheric CO_2. To extract information on the CO_2 content in the atmosphere, passive atmospheric remote sounders measure in the thermal infrared (TIR) or in the near infrared / short wave infrared (NIR/SWIR).

The Atmospheric Infrared Sounder (AIRS), measuring in the TIR, detects thermal radiation emitted by the Earth's surface and the atmosphere (Chédin et al., 2003). The assimilation of the AIRS observed radiances was developed by Engelen et al. (2009) at the European Centre for Medium-Range Weather Forecasts (ECMWF) using a four-dimensional variational (4-D-Var) data assimilation scheme. Their results showed the potential of data assimilation to constrain atmospheric CO_2. They also showed the limitations of the assimilation of AIRS radiances, in particular due to the vertical sensitivity of the sounder. Due to the low thermal contrast between the Earth's surface and the air masses above, AIRS measurements have limited or no sensitivity to the lower troposphere and higher sensitivity to the middle atmosphere. Because the signals of the CO_2 surface sources and sinks are the largest in the near-surface and lower troposphere than in the middle atmosphere, AIRS measurements were not able to capture these signals.

In contrast, column-averaged dry-air mole fractions of CO_2 (or XCO_2) with a high near-surface sensitivity are retrieved from NIR/SWIR measurements based on scattered and back-scattered solar radiation; however, the NIR/SWIR measurements also have their limitations. They need sunlight and are therefore limited to daytime observations. Sufficiently cloud-free conditions and a low aerosol optical depth are also needed for accurate XCO_2 retrievals.

The aim of this study is to document the assimilation of XCO_2 products from NIR/SWIR measurements in order to constrain atmospheric CO_2 and to document how the assimilation impacts the simulated atmospheric CO_2 concentration. For that purpose, we assimilated the XCO_2 products derived from the NIR/SWIR spectra of the Greenhouse gases Observing Satellite (GOSAT, Kuze et. al, 2009). The assimilation system is based on the ECMWF system of Engelen et al. (2009), which has lately evolved for CAMS in order to assimilate retrieved products instead of observed radiances (Massart et al., 2014).

The assimilation system provides an analysis of the atmospheric CO_2 concentration that is then integrated in time using a forecast model. The CO_2 forecast model used in this study is documented

by Agustí-Panareda et al. (2014). In this model, the production and loss of CO_2 at the surface is based on surface fluxes that are partially prescribed and partially modelled. These CO_2 surface fluxes are not directly constrained by observations and they may deviate from reality. The accumulation of surface fluxes errors then leads to biases in the atmospheric CO_2. On the other hand, the strength of the CO_2 forecast model is its ability to provide a realistic CO_2 synoptic variability. The first objective of this study is to determine the quality of the XCO_2 fields resulting from the assimilation of GOSAT XCO_2 data with a CO_2 forecast model where the CO_2 surface fluxes are not constrained.

The atmospheric CO_2 synoptic variability on a regional scale is related to the passage of frontal systems (Wang et al., 2007). These events are difficult to capture with the GOSAT measurements as the availability of the data is limited due to cloud contamination. Therefore, the second objective of this study is to document if the assimilation helps improve the simulation of atmospheric CO_2 for synoptic events despite the lack of measurements nearby frontal systems.

Within CAMS, ECMWF is providing a CO_2 analysis based on the assimilation of GOSAT XCO_2 data with a delay of 5 days behind real time. A ten day forecast is then issued from the analysis in order to provide the atmospheric CO_2 field in real time and for the next few days. The last objective of this study is to assess the quality of this forecast. The forecast quality as a function of the lead time and the season is evaluated against the analysis.

This paper is structured as follows. Section 2 introduces the data sets used in this study. Section 3 describes our atmospheric CO_2 simulations with and without assimilation of the GOSAT XCO_2 data, and how we compared them with independent measurements. Sections 4 to 6 present the global evaluation of our simulations, a case study and the evaluation of the CO_2 forecast based on the analysis. Finally, Sect. 7 presents our conclusions.

2 Data sets

In this study, we used two sets of data. The first one is the measurements from the GOSAT's Fourier transform spectrometer and the XCO_2 product retrieved from these measurements by the University of Bremen (UoB) and described in Sect. 2.1. The second one is the collection of measurements provided by the Total Carbon Column Observing Network (TCCON) and described in Sect. 2.2.

2.1 GOSAT XCO_2

The GOSAT satellite is a joint effort between the Japanese Aerospace Exploration Agency (JAXA), the National Institute for Environmental Studies (NIES) and the Japanese Ministry of the Environment (MOE) as part of the Global Change Observation Mission (GCOM) programme of Japan. The GOSAT satellite was launched on 23 January 2009 and carries the Thermal And Near-infrared Sensor for carbon Observations, which consists of a Fourier Transform Spectrometer (TANSO-FTS) and a Cloud and Aerosol Imager (TANSO-CAI).

In this study, we used XCO_2 retrieved from TANSO-FTS measurements of the upwelling radiance at the top of the atmosphere by the Bremen Optimal Estimation DOAS – Differential Optical Absorption Spectroscopy – (BESD) algorithm of UoB. The BESD algorithm was initially developed to retrieve XCO_2 from nadir measurements of the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) remote sensing spectrometer on the ENVIronment SATellite (ENVISAT, Reuter et al., 2010, 2011). The BESD algorithm has been modified to also retrieve XCO_2 from GOSAT measurements. A detailed description of the GOSAT BESD algorithm can be found in Heymann et al. (2015). In brief, the algorithm uses three fitting windows, the O_2-A band (12 920–13 195 cm^{-1}), a weak CO_2 absorption band (6170–6278 cm^{-1}) and a strong CO_2 band (4804–4896 cm^{-1}) from both the medium and high-gain (respectively M-gain and H-gain) GOSAT nadir modes. An optimal estimation based inversion technique is used to derive the most likely atmospheric state from every individual GOSAT measurement using a priori knowledge. The BESD algorithm explicitly accounts for atmospheric scattering by clouds and aerosols, reducing potential systematic biases. The scattering information on cloud and aerosols is mainly obtained from the O_2-A and strong CO_2 absorption bands.

We used an inhomogeneous GOSAT BESD XCO_2 dataset in this study as the GOSAT BESD algorithm was still under development. This intermediate version of the GOSAT BESD XCO_2 data is referred to as MACC GOSAT BESD XCO_2 (MACC standing for Monitoring Atmospheric Composition and Climate, the precursor of CAMS). Nevertheless, from the beginning of 2014 onwards, we have been assimilating the current version of the GOSAT BESD data (v01.00.02, Heymann et al., 2015) in near real time.

The TANSO-FTS detector has a circular field of view of 10.5 km when projected on the Earth's surface (at exact nadir). In 2013, it measured in a mode with 3 measurements across track, and the footprints were separated by ~ 263 km across track and ~ 283 km along track. The GOSAT satellite can also operate in target mode resulting in a finer sampling distance. For these specific situations, we further thinned the observations on a $1^\circ \times 1^\circ$ grid by removing all the observations but one chosen at random. This procedure avoids having several measurements in the same model grid cell during the assimilation. This thinning, plus the characteristics of the instrument (measurement only during sunlit periods) and the processing of the level-2 data procedure (retrievals for clear-sky conditions and only over land), reduces the number of GOSAT XCO_2 data to about 100 per day. The assimilation window being 12 hours, this means that about 50 GOSAT XCO_2 data points are assimilated during each time window.

The geographic distribution of these data is dependent on the season and the atmospheric conditions as illustrated by Fig. 1. For example, in July 2013 GOSAT BESD data are available up to 75° N, and in October 2013 they are available only up to 60° N. The reason for this is the solar geometry and the filtering of measurements under high solar zenith angle (SZA) conditions where XCO_2 is more challenging to retrieve as the impact of atmospheric scattering becomes larger compared to

low SZA conditions. Other data gaps are due to the strict cloud filtering and other filtering like the ones based on the quality of the spectral fits, on scattering parameters, on the meteorological state, and on the measurement geometry.

The MACC GOSAT BESD XCO_2 dataset have been bias corrected using the TCCON data. As this dataset is delivered in near real time and the TCCON data are delivered with a delay of few months, it was not possible to directly compare the two data sets. Instead, the TCCON data from the previous year were used and they were corrected assuming a $2\,\mathrm{ppm}$ global atmospheric growth of CO_2. A global offset was then computed and applied to the MACC GOSAT BESD XCO_2 based on the comparison between this dataset and the corrected TCCON dataset of the previous year. Moreover, with this procedure the TCCON data used in this study (same year as for the MACC GOSAT BESD XCO_2 dataset) can be considered as independent data.

For the assimilation, the observation error covariances have to be specified. In this study, we assumed that the observation errors are not correlated in space and time. For the standard deviation of the observation error, we used the uncertainty of the BESD XCO_2 product provided together with the data. The BESD XCO_2 uncertainty product accounts for the various sources of uncertainty of the retrieval process. It varies in time and space around an average value of 2 parts per million (ppm). We furthermore established that the specified observation error based on the XCO_2 uncertainty globally matches the expected observation error using diagnostics posterior to the analysis (not shown).

2.2 TCCON XCO_2

The TCCON is a network of ground-based Fourier Transform Spectrometers recording direct solar spectra in the near infrared spectral region (http://tccon.ornl.gov/). The column-averaged dry-air mole fractions of CO_2 are retrieved from these spectra together with other chemical components of the atmosphere (Wunch et al., 2011a). In 2014, the version GGG2014 of the TCCON data was released. The errors on the retrieved XCO_2 are documented to be below 0.25% ($\sim 1\,\mathrm{ppm}$) until the solar zenith angles are larger than 82° (http://dx.doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662).

When we downloaded the GGG2014 data in November 2015, 20 TCCON stations were providing data within the time period we are interested in (year 2013). Not all the stations were used in this study. First we removed JPL 2011 (USA), Pasadena/Caltech (USA) and Tsukuba (Japan), as they are not background stations and are associated with significant representativity errors. We also removed Edwards (USA). This station started to retrieve data from the middle of the year 2013, and we assumed that this was not long enough to provide information on the seasonal variation of the error in our simulations. Additionally, we removed Eureka (Canada) from the list of stations as the site was providing data during only three days in 2013. This selection of the TCCON stations left 16 stations for the study (Table 1).

Orléans (France) had a specific treatment compared to the other stations. The averaging kernels were not specified in the GGG2014 release. So we decided to use the same information as for Lamont (USA) as advised in the previous release of the TCCON data (version GGG2012).

3 Experimental setup

We ran two model simulations for the year 2013. The first is similar to the operational CAMS CO_2 forecast (Agustí-Panareda et al., 2013) and is referred to as the "free run". This simulation is used as the reference to assess the impact of the assimilation of the GOSAT BESD XCO_2 data. The second simulation is the analysis in which the GOSAT XCO_2 data are assimilated and is referred to as the "analysis". The configuration of both simulations is described in Sect. 3.1. The simulations were evaluated against each other and also against the TCCON data. Section 3.2 introduces the methodology used in comparison of simulations and the TCCON data.

3.1 Model simulations

The global simulations of atmospheric CO_2 are performed within the Numerical Weather Prediction (NWP) framework of the Integrated Forecasting System (IFS). The CO_2 mass mixing ratio is directly transported within IFS as a tracer and is affected by surface fluxes. The transport is computed online and is updated each $12\,\mathrm{h}$ benefiting from the assimilation of all the operational observations within the IFS 4-D-Var assimilation system. The terrestrial biogenic carbon fluxes are also computed online by the carbon module of the land surface model (Carbon-TESSEL or CTESSEL, Boussetta et al., 2013) while other prescribed fluxes are read from CO_2 surface fluxes inventories (see Agustí-Panareda et al., 2014 for more details).

The ability to assimilate retrieval products from GOSAT was included in IFS and is detailed in Massart et al. (2014) for the assimilation of methane data. The system used in this study is similar to the one of Massart et al. (2014) and is based on fixed background errors derived from the National Meteorological Center (NMC) method (Parrish and Derber, 1992). The standard deviation of the background error is constant for each model level and slowly increases from the upper troposphere to the lower troposphere with values from about 1 to about $5\,\mathrm{ppm}$, and then rapidly increases to reach a value of about $40\,\mathrm{ppm}$ at the surface. The correlation of the background errors varies over the whole domain and vertically with a representative length scale of about $250\,\mathrm{km}$. The system does not account for the spatial or temporal correlation between the errors of the observations.

We chose in this study to have a horizontal resolution of TL255 on a reduced Gaussian grid ($\sim 80\,\mathrm{km} \times 80\,\mathrm{km}$), and 60 vertical levels from the surface up to $0.1\,\mathrm{hPa}$. This resolution is sufficient for resolving the large and synoptic scale horizontal structures ($\sim 1000\,\mathrm{km}$) of the atmospheric CO_2 fields.

3.2 Comparison with TCCON

To evaluate the quality of the model simulations (free run and analysis), we have extensively used the TCCON data in this study. The comparison is performed in the TCCON space using the TCCON a priori and averaging kernel information (see Appendix A for more details). In order to have a decomposition of the errors of the model column-averaged CO_2 against the TCCON measurement, we computed for each TCCON station k for $k \in [1, N]$, the mean difference (or bias) δ_k and the standard deviation of the difference (or scatter) σ_k over the M_k times t_i for $i \in [1, M_k]$ when we have a TCCON observation for the station k. If $\hat{c}_k^o(t_i)$ for $i \in [1, M_k]$ is the observed TCCON XCO_2 time series for the station k and if $\hat{c}_k(t_i)$ for $i \in [1, M_k]$ is the model equivalent time series, then the bias δ_k and scatter σ_k are defined by

$$\delta_k = \frac{1}{M_k} \sum_{i=1}^{M_k} \left[\hat{c}_k(t_i) - \hat{c}_k^o(t_i)\right],$$
$$\sigma_k = \sqrt{\frac{1}{M_k - 1} \sum_{i=1}^{M_k} \left[\hat{c}_k(t_i) - \hat{c}_k^o(t_i) - \delta_k\right]^2}. \tag{1}$$

Additionally, we computed the correlation coefficient r_k between $\hat{c}_k(t_i)$ and $\hat{c}_k^o(t_i)$ for $i \in [1, M_k]$.

Following Heymann et al. (2015), we also computed the model offset δ, the mean absolute error (MAE) Δ, the station-to-station bias deviation σ and the model precision π for the N TCCON stations

$$\delta = \frac{1}{N} \sum_{k=1}^{N} \delta_k, \qquad \Delta = \frac{1}{N} \sum_{k=1}^{N} |\delta_k|,$$
$$\sigma = \sqrt{\frac{1}{N - 1} \sum_{k=1}^{N} \left[\delta_k - \delta\right]^2}, \qquad \pi = \frac{1}{N} \sum_{k=1}^{N} \sigma_k. \tag{2}$$

The statistics for the comparisons of the simulations against the TCCON data have some gaps in time due to gaps in the availability of the TCCON data. They are also valid only where the TCCON sites are located, i.e. 16 points distributed over the globe. To have a more global overview of the model bias and scatter against the TCCON data, we smoothed these statistics in time and space (see Appendix B for more details). In summary, for the bias we averaged all the model–measurement differences for each TCCON site using a 1-week time bin. We then fit the time evolution of the weekly bias with a function that combines a linear and a harmonic component for each station. The second step is an extrapolation in space. For each week, the weekly biases of every station are extrapolated using a quadratic function of latitude. This results in a Hovmöller diagram of the bias as a function of time and latitude. A similar process is applied for the scatter (see Figs. 2 and 3).

4 Global evaluation of the analysis

In this section we first present the characteristics of the XCO_2 derived from the free run simulation when compared to the TCCON data. Second, we present the impact of the assimilation of the MACC GOSAT BESD XCO_2 comparing the XCO_2 from the analysis against the XCO_2 from the free run. Then, we discuss if the analysis represents an improvement compared to the free run in terms of statistics against the TCCON data. Finally, we discuss the merits of the analysis compared to the MACC GOSAT BESD data using the TCCON data as a reference.

4.1 Free run simulation vs. TCCON

When compared with the TCCON data, the free run simulation has a mean offset δ of $-0.36\,\mathrm{ppm}$ and a mean absolute error Δ of $1.08\,\mathrm{ppm}$ (Table 2). However, the individual station bias δ_k spans a range from $2.3\,\mathrm{ppm}$ at Ascension Island (Saint Helena, Ascension and Tristan da Cunha) to $-2.9\,\mathrm{ppm}$ at Białystok (Poland). The station-to-station bias deviation σ of the free run simulation has then a value of 1.27 ppm.

The variations of the bias as well as the seasonal cycle of the bias are highlighted in the Hovmöller diagram displayed in Fig. 2a. First, it shows that the initial condition of the free run has a positive bias of about $2\,\mathrm{ppm}$ over the tropical region (region between 23° S and 23° N) when compared to the TCCON data. This bias is reduced during the spring and reappears the next summer. It reaches its highest values in autumn with more than $2\,\mathrm{ppm}$. These results are slightly different from those of Agustí-Panareda et al. (2014) where the model bias was found to be more constant in the tropical region when comparing the background CO_2 in the marine boundary layer with the National Oceanic and Atmospheric Administration (NOAA) GLOBALVIEW-CO2. Here, the evaluation of the bias in the tropics is driven by the comparison with XCO2 measurements from the TCCON station of Ascension Island. For this station, the values of the bias from July to September result from the interpolation process as no measurements were reported during this period (Fig. S1 of the Supplement).

In contrast to the situation at the tropics, the initial condition of the free run has a negative bias at northern mid-latitudes (region between 23° N and 66° N) and reaches almost $4\,\mathrm{ppm}$ at the latitude of Sodankylä (Finland, 67° N) when compared to the TCCON XCO_2. This value is the result of the smoothing process as we do not have data for that period (Fig. 4a). The negative bias at these mid-latitudes is nevertheless confirmed by the comparison with other stations, like Karlsruhe (Germany) and Park Falls (USA), where we have some data at the beginning of the year (Figs. 4b and 4c). The negative bias at northern mid-latitudes remains high during the whole year, with an absolute value generally greater than $1\,\mathrm{ppm}$ at the end of spring, and in June and December. This can be explained by the fact that the model does not release enough CO_2 before and after the growing season, i.e.,

March to May and October to December, and by the fact that, in the model, the onset of the CO_2 sink associated with the growing season starts too early in the season (Agustí-Panareda et al., 2014).

The precision π of the free run measured by the average scatter between the simulation and the TCCON data is $1.4\,\mathrm{ppm}$ (Table 2). Similarly to the bias, the scatter varies in time and space as highlighted by the Hovmöller diagram of the scatter (Fig. 3a). The scatter has its highest values of more than $1\,\mathrm{ppm}$ at the northern mid-latitudes during May–June–July. This increase in the scatter is driven by the behaviour of the free run at Sodankylä. There, the simulation has a larger variability than the measurements. For example, end of June, the simulation presents a decrease of about $7\,\mathrm{ppm}$ in 36 hours whereas the measurements show a decrease of about $4\,\mathrm{ppm}$ (Fig. 4a). Elsewhere, there is also an increase of the scatter between May and July which is during the northern hemisphere growing season. This increase could be explained by the difficulty for CTESSEL to model the terrestrial biogenic carbon fluxes during the growing season, which leads to higher variability of the simulated atmospheric CO_2.

4.2 Analysis vs. free run

To assess the impact of the assimilation of the MACC GOSAT BESD XCO_2, we compared the evolution of XCO_2 from the analysis with XCO_2 from the free simulation. Figure 5 presents the Hovmöller diagram (time vs. latitude) of this difference. It shows that the first region where the analysis impacts XCO_2 is the tropics. There, compared to the free run, the analysis continuously decreases XCO_2, by up to $1\,\mathrm{ppm}$ in June and by more than $2\,\mathrm{ppm}$ from September to December. The assimilation of the GOSAT data consequently causes an improvement as the free run has a positive bias in this region in autumn compared to the TCCON data.

The analysis also decreases XCO_2 over the southern extra tropics (region between 23° S and 66° S) when compared to the free run (Fig. 5). The decrease extends to the southern high latitudes ($\geq$ 66° S) even when no GOSAT data were assimilated in this region. This decrease results mainly from the transport of CO_2 from the equatorial region and southern mid-latitudes towards southern high latitudes. Unfortunately, there are no independent XCO_2 data available at southern high latitudes to assess the merits of the analysis there.

Despite the fact that some GOSAT data are assimilated in the northern mid-latitudes during the first months of the simulation, the analysis only starts to differ significantly from the free run from March onwards. In this region, north of 30° N, the analysis has higher values of XCO_2 than the free run, with a difference of more than $2\,\mathrm{ppm}$ during the northern summer. Again, the assimilation of the GOSAT data improves the simulated XCO_2 as the free run shows a strong negative bias there. Similar to the behaviour discussed for the southern high latitudes, the change in the CO_2 concentration at northern mid-latitudes is transported northward to higher latitudes. There is, nevertheless, a difference between the two hemispheres. For the Northern Hemisphere we have more data at high

latitudes, especially during the summer when the northernmost GOSAT measurements cover goes up to 80° N.

4.3 Analysis vs. TCCON data

When compared with the TCCON data, the GOSAT BESD XCO_2 analysis has an offset δ of $-0.34\,\mathrm{ppm}$ and a mean absolute error Δ of $0.57\,\mathrm{ppm}$ (Table 2). The offset is similar to that of the free run ($-0.36\,\mathrm{ppm}$), but the mean absolute error is improved ($1.08\,\mathrm{ppm}$ for the free run). The individual station bias is moreover more constant in time for the analysis compared to the free run. For example, the trend of the free run bias is $2.08\,\mathrm{ppm\,yr^{-1}}$ for Lauder (New Zealand) (Table S1 of the Supplement), and it improves to $0.47\,\mathrm{ppm\,yr^{-1}}$ for the analysis (Table S2 of the Supplement and Fig. 4c).

By increasing XCO_2 in the northern mid-latitudes as discussed before, the analysis considerably reduces the bias. A residual seasonal cycle in the bias is still present, with values usually in the range of 0 to $3\,\mathrm{ppm}$ (Fig. 2b). This could be explained by the fact that we correct the atmospheric state of CO_2 and not the CO_2 fluxes. During the seasons when the CO_2 fluxes are the main driver of the atmospheric CO_2, the optimisation of the atmospheric state only may not be enough.

The analysis has a more constant bias in time than the free run. It is also more accurate in space, with a station-to-station bias deviation σ that is largely reduced compared to the free run with a value of $0.61\,\mathrm{ppm}$ against $1.27\,\mathrm{ppm}$ (Table 2). The assimilation of the MACC GOSAT BESD XCO_2 thus helps significantly improve the accuracy of the model. The assimilation also helps improve the precision π, with the mean scatter improved by $15\,\%$, reduced to a value of $1.22\,\mathrm{ppm}$. The scatter of the analysis is reduced for all TCCON stations compared to the free run except for Garmisch (Germany) where the scatter remains essentially unchanged. The Hovmöller diagram of the scatter shows that the main reduction is in the northern high latitudes in May (Fig. 3). In particular, the analysis shows less spurious variability than the free run at Sodankylä (Fig. 4a).

4.4 Analysis vs. MACC GOSAT BESD data

The analysis is much more accurate and more precise than the free run when compared to the TCCON data. The analysis also fills the gaps in time and space of the MACC GOSAT BESD data. In this section, we evaluate the analysis against the MACC GOSAT BESD data once more using the TCCON data as a reference.

The MACC GOSAT BESD data were compared to the TCCON data using a geolocation criterion of 5° in space and a time window of $\pm 2\,\mathrm{h}$. Before computing the difference between each GOSAT/TCCON pair, following Dils et. al (2014), we added a correction to the GOSAT retrieved value in order to account for the use of different a priori CO_2 profiles in the two products. Moreover, we only kept the stations where more than 30 GOSAT/TCCON pairs were found in order to have more robust statistical results. This procedure removes Izaña (Spain), Ascension Island, Réunion

Island (France) and Lauder from the list of the used TCCON stations in the comparison and reduces the number of stations to 12 (Table 3).

For each GOSAT/TCCON pair, we extracted the CO_2 profile from the analysis at the same location and time as the GOSAT measurement before computing the difference between the model and the TCCON data. In this way, we have a fair comparison between the analysis and the MACC GOSAT BESD data with respect to the TCCON data.

The resulting subset of the analysis minus TCCON differences has a different offset than the full dataset but a similar mean absolute error, station-to-station bias deviation and precision (Tables 2 and 3). The difference in the offset is mainly due to a difference in the sampling between the subset and the full dataset over the Northern Hemisphere. Due to few or no pairs occurring in spring for the subset, the sampling misses the negative bias of the analysis there. Missing the negative bias of the analysis results in an increased offset. In that respect, the mean absolute error is less sensitive to the used dataset (subset or full dataset).

The analysis has a lower mean absolute error Δ than the one from the MACC GOSAT BESD data ($0.65\,\mathrm{ppm}$ vs $1.\,\mathrm{ppm}$, Table 3), a station-to-station bias deviation σ almost half of the one from GOSAT data ($0.7\,\mathrm{ppm}$ vs $1.3\,\mathrm{ppm}$) and has an improved precision π ($1\,\mathrm{ppm}$ vs $3.3\,\mathrm{ppm}$). The mean correlation coefficient is also higher in the analysis than in the satellite data with a value of 0.8 compared to 0.5. The statistics of the MACC GOSAT BESD data found here are different than those of Heymann et al. (2015) who used a more recent version of the GOSAT BESD product. With the successive improvements in the BESD algorithm, the latest version has a station-to-station bias deviation of $\sim 0.4\,\mathrm{ppm}$ and a precision of $\sim 2\,\mathrm{ppm}$.

The better precision (lower value of π) and the lower value of the mean absolute error Δ and station-to-station bias deviation σ of the analysis compared to the MACC GOSAT BESD dataset shows that the analysis is capable of smoothing the scatter of the satellite data. Moreover, the analysis is able to fill the gaps of the satellite data in time and space.

5 Case study of a cold front over Park Falls

The CO_2 concentration could be strongly affected by frontal systems. As an illustration, such a situation occurred at the end of May 2013, close to the TCCON station of Park Falls, Wisconsin, USA, when a cold front came from the North-West. On 31 May, the XCO_2 dropped from $398.62\,\mathrm{ppm}$ at 08:15 LT (Local Time) to $395.97\,\mathrm{ppm}$ at 12:53 LT (Fig. 6, top panel). This sudden decrease of $2.65\,\mathrm{ppm}$ in less than $5\,\mathrm{h}$ occurs after the arrival of a cold front, which is associated with a decrease of the surface pressure and a decrease of the temperature at $500\,\mathrm{hPa}$ (Fig. 6, lower panel).

The free run is able to capture the sudden decrease in XCO_2, highlighting the skill of the model for such a situation (Fig. 6, upper panel). The flow during this period is mainly a descent of cold air from Canada towards the Midwestern and Eastern US. This cold air mass is depleted in CO_2 relative

to the background (Figs. 7e and f). When it moves towards Park Falls, it results in decreasing XCO_2 as observed and simulated, but the decrease in the free run is too strong by 2 to 3 ppm compared to the measurements.

We investigated whether the assimilation of the GOSAT data helps improve the simulated evolution of the CO_2 concentration for such situations even if the number of BESD GOSAT data is limited in the vicinity of a frontal system due to the strict cloud filtering. Frontal systems are associated with clouds formed when moist air between the cold and warm fronts is lifted.

On May, 30 we have a few GOSAT measurements over the North and North-East region of North America (Fig. 7a). These measurements have the effect of increasing the XCO_2 in this region (Figs. 7b–d). The cold air mass is then richer in CO_2 in the analysis compared to the free run, and when it moves towards Park Falls, the decrease is weaker and closer to the observed decrease. The assimilation of the GOSAT data helps improve the simulation by correcting the large scale structure upstream and by improving the large scale atmospheric XCO_2 horizontal gradient.

The XCO_2 decrease continues the next day on 1 June in both simulations as the cold front continued its descent. Unfortunately, likely due to the presence of clouds, no TCCON measurements are available during this period to corroborate the simulated XCO_2 decrease.

6 Forecast based on the analysis

Within CAMS, we are receiving the GOSAT BESD data for a given day with a delay of 5 $days$ behind real time. The analysis for this day is run as soon as the data are received. A 10 day forecast is then subsequently run based on the resulting analysis.

In this section, we aim to evaluate the forecast as a function of its lead time by comparing the forecast to the analysis valid for the same time. This comparison informs us about how long the information provided by the analysis remains in the forecast. Assuming perfect transport and perfect surface fluxes, the analysis and the forecast (valid for the same time) should be similar given that the analysis accurately corrects the atmospheric concentration of CO_2. In practice, the differences observed between the analysis and the forecast could either come from the transport, the surface fluxes or the analysis.

To compare a forecast with the analysis valid for the same time, we computed the anomaly correlation coefficient (ACC) for XCO_2 (see Appendix C for more details). The ACC can be regarded as a skill score relative to the climatology: the higher the ACC, the better the forecast. In the framework of NWP, an ACC reaching 50 $\%$ corresponds to forecasts for which the error is the same as for a forecast based on a climatological average. An ACC of about 80 $\%$ indicates valuable skill in forecasting large-scale synoptic patterns.

We computed the ACC for each month individually as we know that the surface fluxes, drivers of the difference between the forecast and the analysis, have a strong seasonal cycle. We also computed

it for different domains (globe, tropics and mid to high latitudes) and for several forecast lead times, from 12 h up to 10 days. We found that the ACC is globally more than 90 % for day 3 and almost always more than 85 % for day 5 for each single month (Fig. 8a). This means that the forecast for today based on the analysis of 5 days ago shows the same large-scale synoptic XCO_2 patterns as the analysis. The information of the analysis therefore lasts long enough in the forecast to provide a good quality 5-day forecast for today (compared to the analysis). The information lasts longer in the tropics than in the Northern Hemisphere and slightly longer in the Northern Hemisphere than in the Southern Hemisphere (Fig. 8b to d). This difference between the two hemispheres may reflect the fact that the CO_2 variability is much weaker in the Southern Hemisphere.

For forecasts longer than 5 days, globally, there are two particular months for which the ACC decreases faster than the others, i.e. July and December. For example, for these two months the ACC at day 5 is similar to the ACC at day 10 for October. This means that for July and December, the medium range XCO_2 forecast (between 5 and 10 days) should be used more carefully. For July, the drop in skill occurs mainly over the Northern Hemisphere. The main reason is that the CO_2 fluxes are an even more important driver of the CO_2 concentration than the initial CO_2 concentration for this month. To better understand the impact of the surface fluxes, let us assume that in July we have too little release or, similarly, too much uptake of CO_2 in the atmosphere in the model over the Northern Hemisphere (as confirmed by Fig. 2a). This induces a negative bias of the CO_2 surface fluxes in the model. In the meantime, the analysis increases the CO_2 concentration helped by the GOSAT BESD data (Fig. 5). However, the next 12-hour short-term forecast (used as the background for the next analysis) will not increase enough the CO_2 concentration due to the negative bias of the CO_2 fluxes. This opposition between the analysis and the short-term forecast explains the reduction in skill during the periods when the surfaces fluxes are the most important driver of the CO_2 concentration in the atmosphere.

The global drop in skill for December is not directly related to a particular region as for July. It is nonetheless the second worst month for the tropics (after January) and the third worst for the Northern Hemisphere (together with September). Over the tropics during the winter, the reduction in skill is due to the opposite effect as for July over the Northern Hemisphere: the CO_2 fluxes are important and there is a positive bias in the fluxes (too much release or too little uptake of CO_2 in the atmosphere) in the model. For these situations when the CO_2 fluxes are the main driver of the atmospheric CO_2, the only solution to improve the skill would be to optimise the CO_2 fluxes together with the CO_2 initial conditions.

7 Conclusions

The Copernicus Atmosphere Monitoring Service (CAMS) greenhouse gases data assimilation within the Numerical Weather Prediction (NWP) framework of the Integrated Forecasting System (IFS) is

designed to correct the atmospheric concentration of CO_2 instead of the surface fluxes in order to constrain the atmospheric CO_2. This requires the use of a short assimilation window so to neglect the model errors of the short-term forecast (lasting the length of the assimilation window). In the case of atmospheric CO_2, model errors are related to potentially inaccurate surface fluxes or transport.

This article demonstrates the benefit of the assimilation of XCO_2 data derived from the Greenhouse gases Observing Satellite (GOSAT) by intermediate versions of the Bremen Optimal Estimation DOAS (BESD) algorithm of the University of Bremen (UoB). The assimilation of the GOSAT BESD XCO_2 provides a CO_2 analysis that was compared to a free run forecast where the CO_2 concentration is not constrained by any CO_2 observation. The comparison was one year long (year 2013) and both simulations (analysis and free run) were evaluated against measurements from the Total Carbon Column Observing Network (TCCON). We showed that the free run has a negative bias at northern mid-latitudes and a large positive bias in the tropical region with strong seasonal variations in both regions. These results are consistent with the biases documented by Agustí-Panareda et al. (2014) and mainly associated with biogenic fluxes.

The analysis significantly reduces these biases without completely removing them with a remaining mean offset of $-0.34\,\mathrm{ppm}$ and a mean absolute error of $0.57\,\mathrm{ppm}$ compared to the TCCON data. However, the accuracy estimated with the station-to-station bias deviation is $0.61\,\mathrm{ppm}$. This represents a large improvement compared to the free run for which the accuracy is $1.27\,\mathrm{ppm}$. The precision of the analysis estimated with the mean scatter is $1.22\,\mathrm{ppm}$, slightly better than for the free run with a value of $1.43\,\mathrm{ppm}$.

The analysis produced in this paper was compared to the assimilated MACC GOSAT BESD data using TCCON data as a reference. This comparison showed that the analysis has a lower station-to-station bias deviation than the assimilated data ($0.7\,\mathrm{ppm}$ compared to $1.3\,\mathrm{ppm}$). The precision is much better for the analysis, with a scatter of $1\,\mathrm{ppm}$, while the assimilated data have a scatter of $3.3\,\mathrm{ppm}$. The precision of the analysis is also better than the documented precision of other GOSAT XCO_2 products. The precision of the NIES product extracted from Yoshida et. al (2013) is $1.8\,\mathrm{ppm}$. The precision of the University of Leicester product and of the SRON Netherlands Institute for Space Research product is respectively $2.5\,\mathrm{ppm}$ and $2.37\,\mathrm{ppm}$ (Dils et. al, 2014). The CO_2 analysis is consequently an alternative to the standard XCO_2 GOSAT products as it provides a lower or similar station-to-station bias deviation and a better precision XCO_2 product compared to TCCON. Moreover, it has a uniform spatio-temporal resolution.

The pre-operational CAMS CO_2 analysis is similar to the analysis presented in this paper having nevertheless a higher horizontal resolution (TL511 on a reduced Gaussian grid, $\sim 40\,\mathrm{km} \times 40\,\mathrm{km}$), and a higher vertical resolution with 137 vertical levels. It currently assimilates the most recent version of the GOSAT BESD data presented by Heymann et al. (2015) in near real time. These data have an improved bias deviation ($\sim 0.4\,\mathrm{ppm}$) and an improved precision ($\sim 2\,\mathrm{ppm}$) compared to

those used in this study. The near real time CAMS CO_2 analysis should therefore have an improved station-to-station bias deviation and precision than the analysis presented in this paper.

We corrected the atmospheric concentration by only constraining the atmospheric concentration and not the surface fluxes. When and where the surface flux is a significant driver of the atmospheric concentration and if the assimilated data are not good enough or not numerous enough (in time and space), then constraining only atmospheric CO_2 does not compensate for the error in the surface flux. The next step is to further improve the carbon module CTESSEL in order to reduce the bias of the model. Another long term solution would be to constrain the surface flux at the same time as the concentration.

One strength of the CO_2 model used in this study is its ability to represent CO_2 variations associated with synoptic weather systems (Agustí-Panareda et al., 2014). By correcting the large scale XCO_2 patterns and removing part of the model bias, we showed with a case study that the analysis is able to better represent the CO_2 variations associated with these situations. The variations of the atmospheric reservoir of CO_2 are the result of changes in the surface fluxes to and from the atmosphere. If the characteristics of the analysis are found to be satisfactory in terms of bias and precision, it could be included into a flux inversion system to infer surface fluxes.

The horizontal resolution of this study is half the horizontal resolution of the pre-operational analysis and the vertical resolution of the the pre-operational analysis is also higher. One should expect an even better representation of the CO_2 variability in the pre-operational analysis. In the future, the horizontal resolution could be increased even further toward the ECMWF operational resolution of about $16\,\mathrm{km} \times 16\,\mathrm{km}$.

The quality of the analysis is considered to be sufficient to assess the quality of the forecast as a function its lead time. We showed that the forecast for day 3 and day 5, which will be the valid range for today's forecast, has an anomaly correlation coefficient of 90 and 85 %, respectively. This means that we are providing a CO_2 forecast with accurate synoptic features for today. With a good representation of the variability and a bias mostly under 1 ppm, the CAMS atmospheric CO_2 promises to become a useful product, for example, for planning a measurement campaign. It could also be used as the a priori in the satellite or TCCON retrieval algorithms or be used to evaluate the retrieval products from the Orbiting Carbon Observatory-2 (OCO-2, oco.jpl.nasa.gov).

Appendix A: Comparing the model against TCCON

For the comparison with the TCCON data, one has to account for the a priori information used in the retrieval that links $\hat{c}^o$, the TCCON retrieved XCO_2 to $\boldsymbol{x}^{\mathrm{t}}$, the true (unknown) CO_2 profile (Wunch et al., 2011b),

$$\hat{c}^o = c^b + \boldsymbol{a}^{\mathrm{T}} \left(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^b \right) + \varepsilon \,, \tag{A1}$$

where $\boldsymbol{x}^b$ is an a priori profile of CO_2, $\boldsymbol{a}$ is a vector resulting from the product of the averaging kernel matrix with a dry-pressure weighting function vector (for the vertical integration), c^b is the column-averaged mixing ratio computed from $\boldsymbol{x}^b$, and ε is the error in the retrieved column-averaged mixing ratio. This error includes the random and systematic errors in the measured signal and in the retrieval algorithm.

To compare the model with the TCCON retrieved value, we used the same a priori information, so that the model profile $\boldsymbol{x}$ is converted to a column-averaged mixing ratio $\hat{c}$ by

$$\hat{c} = c^b + \boldsymbol{a}^{\mathrm{T}} \left(\boldsymbol{x} - \boldsymbol{x}^b \right) . \tag{A2}$$

The comparison between the simulation and TCCON occurs in the observation space with the difference between the model column-averaged mixing ratio $\hat{c}$ of Eq. (A2) and the TCCON column-averaged mixing ratio $\hat{c}^o$ of Eq. (A1),

$$\hat{c} - \hat{c}^o = \boldsymbol{a}^{\mathrm{T}} \left(\boldsymbol{x} - \boldsymbol{x}^{\mathrm{t}} \right) - \varepsilon . \tag{A3}$$

Let us define $\eta = \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^{\mathrm{t}})$ as the model error in terms of the column-averaged mixing ratio. It accounts for numerous errors, for example, the errors directly linked to the model processes like the transport, the errors in the surface fluxes, the representativity error and the error due to the assimilation of the GOSAT XCO_2 data for the analysis. The difference between the smooth model column-averaged mixing ratio $\hat{c}$ and the TCCON column-averaged mixing ratio $\hat{c}^o$ is, therefore, the sum of the model error η and the error in the retrieved column-averaged mixing ratio ε.

To compute the model column-averaged mixing ratio $\hat{c}$ of Eq. (A2) equivalent to each TCCON measurement, we extracted the two model profiles that are closest to the measurement time and at the nearest grid point to the measurement. The two profiles are then interpolated in time in order to obtain the model profile at the same time as the measurement. Finally, we computed the column-averaged mixing ratio according to Eq. (A2).

Appendix B: Smoothing the statistics against TCCON

In order to have a more global view of the bias and the scatter of a simulation against the data from the TCCON network, we have developed and used a two-step algorithm. The first step consists in computing the statistics (bias and the standard deviation) for each week of 2013 and for each TCCON station when the data are available. The weekly statistics are then interpolated in time using a function described in the following Sect. B1. This allows one to fill in the gaps in time when no data are available. We therefore have a value for the bias at each station and for each week. For the second step, we compute a quadratic function of latitude that best fits the interpolated biases for each week (Sect. B2).

B1 Time smoothing

For each TCCON station k and for each week w^l for $l \in [1, 52]$, we compute the mean difference δ_k^l and the standard deviation of the difference σ_k^l between every TCCON observation during this week and the model equivalent value. The statistics are computed only when more than 10 TCCON measurements are available during the week. The averaged difference (or bias) is then interpolated in time t with the function $\tilde{b}_k(t)$ that combines a linear growth and a harmonic component,

$$\tilde{b}_k(t) = a_k\, t + b_k + \alpha_k \sin\left(\frac{t}{\tau_1} + \varphi_k\right) + \beta_k \sin\left(\frac{t}{\tau_2} + \varphi_k\right). \tag{B1}$$

a_k, b_k, α_k, β_k and φ_k are the parameters of the function $\tilde{b}_k(t)$ obtained by an optimisation procedure that minimises the distance between $\tilde{b}_k(t)$ and the series of δ_k^l for $l \in [1, 52]$. τ_1 is chosen to be 6 months and τ_2 3 months. The form of the function of Eq. (B1) thus gives a linear growing bias and allows seasonal variations. A similar function is used for the standard deviation.

B2 Spatial smoothing

The time smoothing allows us to fill in the gaps in the time series of the bias for each station, when for a given week we do not have any measurement to compare with. Following Bergamaschi et al. (2009), we then compute for each week w^l the best fit of the interpolated biases with a quadratic function of latitude $\hat{b}^l$,

$$\hat{b}^l(\phi) = a^l \phi^2 + b^l \phi + c^l, \tag{B2}$$

with ϕ is the sine of the latitude. a^l, b^l and c^l are obtained by an optimisation procedure that minimises the distance between $\hat{b}^l$ and the weekly interpolated biases δ_k^l for $k \in [1, N]$. A similar function is used for the standard deviation.

B3 Discussion

For some stations, the availability of the weekly differences is not uniform in time and the time smoothing of Eq. B1 provides spurious values. We solved this issue by fixing the coefficient α_k to a zero value (See Table S1 of the Supplement).

With a root mean square error (RMSE) mostly under 0.7 ppm and a correlation mostly over 0.8, the smoothed bias matches well with the weekly bias (Table S1 of the Supplement). The Hovmöller diagram (Fig. 2) can, thus, be considered as an accurate representation of the overall bias.

Compared to the bias, the fit between the time series of the weekly scatter and the regression is not as good for the scatter. The correlation coefficient is mostly between 0.5 and 0.7 (Table S1 of the Supplement).

Appendix C: Anomaly correlation coefficient

The anomaly correlation coefficient ACC between the forecast f and the analysis a is computed using the climatology c by

$$\mathrm{ACC} = \frac{\overline{(f-c)(a-c)}}{\sqrt{\overline{(f-c)^2}\,\overline{(a-c)^2}}}\,, \tag{C1}$$

where the overline, is the spatial and temporal average. For example, for the forecast range $24\,\mathrm{hr}$, we take the $\mathrm{XCO_2}$ fields from all the $24\,\mathrm{hr}$ forecasts for a given month, all the analyses valid for the same time, and a fixed climatology for this month.

The climatology is based on a free run simulation using the optimised $\mathrm{CO_2}$ surface fluxes from Chevallier et al. (2010) which simulated the years from 2003 to 2012. For each month, we compute the average over the $10\,\mathrm{years}$ of the simulation, rescaling the mean so that the mean is the same as for the analysis, avoiding by this procedure the issue of the increase in $\mathrm{CO_2}$ over time. The two dimensional climatology field for $\mathrm{XCO_2}$ for the month m is

$$c(m) = \frac{1}{10}\sum_{y=2003}^{2012}\frac{1}{n(y,m)}\sum_{d=1}^{n(y,m)}\left[\Sigma(y,m,d) - \overline{\Sigma}(y,m,d)\right] + \overline{\Sigma}_{\mathrm{an}}(m)\,, \tag{C2}$$

where y is the year, n the number of days for the year y and the month m, d is an index for the day, $\Sigma(y,m,d)$ is the $\mathrm{XCO_2}$ field from the simulation for the year y, the month m and the day d, $\overline{\Sigma}$ is a spatial average of Σ and $\overline{\Sigma}_{\mathrm{an}}(m)$ is the spatial and temporal average of the $\mathrm{XCO_2}$ fields from the analysis for the month m (and the year 2013).

The Supplement related to this article is available online at doi:10.5194/acp-0-1-2016-supplement.

Author contributions. S. Massart designed and carried out the experiments with the help of A. Agustí-Panareda and advice from F. Chevallier, J. Heymann and M. Buchwitz. J. Heymann, M. Reuter, M. Hilker, M. Buchwitz and J. P. Burrows have been responsible for the design and operation of the BESD GOSAT $\mathrm{XCO_2}$ retrieval algorithm. S. Massart prepared the manuscript with contributions from A. Agustí-Panareda, J. Heymann, M. Buchwitz, F. Chevallier, M. Reuter, M. Hilker, J.P. Burrows, D. G. Feist, and F. Hase. N. M. Deutscher and R. Sussmann contributed to the ACP version of the paper. F. Desmet is the co-investigator of the La Réunion TCCON station. N. M. Deutscher and C. Petri are responsible for the Białystok, Bremen and Orléans TCCON data M. Dubey is the PI of Four Corners TCCON station. D. G. Feist is the PI of the Ascension TCCON station. D. W. T. Griffith and V. Velazco are the PIs of Darwin and Wollongong stations. F. Hase is the PI of the Karlsruhe TCCON station. R. Kivi is the PI of the Sodankylä TCCON station. M. Schneider is the PI of Izaña TCCON station. R. Sussmann is the PI of the Garmisch TCCON station.

Acknowledgements. This study was funded by the European Commission under the European Union's Horizon 2020 programme. The development of the GOSAT BESD algorithm received funding from the European Space Agency (ESA) Greenhouse Gases Climate Change Initiative (GHG-CCI). TCCON data were obtained from the TCCON Data Archive, hosted by the Carbon Dioxide Information Analysis Center (CDIAC) – http://tccon.
595 onrl.gov. Garmisch work was funded in part via the ESA GHG-CCI project. Four Corners TCCON was funded by LANL's LDRD program. Darwin and Wollongong TCCON measurements are funded by NASA grants NAG5-12247 and NNG05-GD07G and the Australian Research Council grants DP140101552, DP110103118, DP0879468, LE0668470 and LP0562346. We are grateful to the DOE ARM program for technical support in Darwin, and Clare Murphy, Nicholas Jones and others for support in Wollongong. TCCON measurements in
600 Białystok and Orléans are supported by ICOS-INWIRE, InGOS and the Senate of Bremen. Nicholas Deutscher is supported by an ARC-DECRA Fellowship, DE140100178. The authors are grateful to Marijana Crepulja for the acquisition of the BESD GOSAT data at ECMWF and the preparation of the data for the assimilation. The authors would like to acknowledge Paul Wennberg, PI of the Lamont and Park Falls TCCON stations. Finally, we would like to express our great appreciation to William Lahoz, editor of this paper, for his useful comments
605 during the revision process.

References

Agustí-Panareda, A., Massart, S., Boussetta, S., Balsamo, G., Beljaars, A., Chevallier, F., Engelen, R., Peuch, V.-H., and Razinger, M.: The new MACC-II CO_2 forecast, ECMWF Newsletter, 135, 8–13, 2013.

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

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., and Levin, I.: Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, J. Geophys. Res.-Atmos., 114, D22301, doi:10.1029/2009JD012287, 2009.

Blumenstock, T., Hase, F., Schneider, M., García, O. E., and Sepúlveda, E.: TCCON data from Izana, Tenerife, Spain, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.izana01.R0/1149295, 2014.

Boussetta, S., Balsamo, G., Beljaars, A., Agustí-Panareda, A., Calvet, J.-C., Jacobs, C., van den Hurk, B., Viterbo, P., Lafont, S., Dutra, E., Jarlan, L., Balzarolo, M., Papale, D., and van der Werf, G.: Natural carbon dioxide exchanges in the ECMWF integrated forecasting system: implementation and offline validation, J. Geophys. Res.-Atmos., 118, 1–24, doi:10.1002/jgrd.50488, 2013.

Chédin, A., Saunders, R., Hollingsworth, A., Scott, N., Matricardi, M., Etcheto, J., Clerbaux, C., Armante, R., and Crevoisier, C.: The feasibility of monitoring CO_2 from high-resolution infrared sounders, J. Geophys. Res.-Atmos., 108, 4064, doi:10.1029/2001JD001443, 2003.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res.-Atmos., 115, D21307, doi:10.1029/2010JD013887, 2010.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. L., Myneni, R., Piao, S., and Thornton, P.: Climate change 2013: the physical science basis, in: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., chap. Carbon and Other Biogeochemical Cycles, Cambridge University Press, Cambridge, UK and New York, NY, USA, 465–570, 2013.

De Mazière, M., Desmet, F., Hermans, C., Scolas, F., Kumps, N., Metzger, J.-M., Duflot, V., and Cammas, J.-P.: TCCON data from Reunion Island (La Reunion), France, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.reunion01.R0/1149288, 2014.

Deutscher, N., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., Grupe, P., and Katrynski, K.: TCCON data from Bialystok, Poland, Release GGG2014R0, TCCON data archive, hosted by the

Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.bialystok01.R0/1149277, 2014.

Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J. P., Butz, A., Deutscher, N. M., Frankenberg, C., Hase, F., Hasekamp, O. P., Heymann, J., De Mazière, M., Notholt, J., Sussmann, R., Warneke, T., Griffith, D., Sherlock, V. and Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON, Atmos. Meas. Tech., 7, 1723–1744, 2014.

Dubey, M., Lindenmaier, R., Henderson, B., Green, D., Allen, N., Roehl, C., Blavier, J.-F., Butterfield, Z., Love, S., Hamelmann, J., Wunch. D.: TCCON data from Four Corners, NM, USA, Release GGG2014R0. TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. doi:10.14291/tccon.ggg2014.fourcorners01.R0/1149272 2014.

Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric CO_2 using AIRS observations, J. Geophys. Res.-Atmos., 114, D03303, doi:10.1029/2008JD010739, 2009.

Feist, D. G., Arnold, S. G., John, N., and Geibel, M. C.: TCCON data from Ascension Island, Saint Helena, Ascension and Tristan da Cunha, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.ascension01.R0/1149285, 2014.

Griffith, D. W. T., Deutscher, N., Velazco, V. A., Wennberg, P. O., Yavin, Y., Aleks, G. K., Washenfelder, R., Toon, G. C., Blavier, J.-F., Murphy, C., Jones, N., Kettlewell, G., Connor, B., Macatangay, R., Roehl, C., Ryczek, M., Glowacki, J., Culgan, T., and Bryant, G.: TCCON data from Darwin, Australia, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.darwin01.R0/1149290, 2014a.

Griffith, D. W. T., Velazco, V. A., Deutscher, N., Murphy, C., Jones, N., Wilson, S., Macatangay, R., Kettlewell, G., Buchholz, R. R., and Riggenbach, M.: TCCON data from Wollongong, Australia, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.wollongong01.R0/1149291, 2014b.

Hase, F., Blumenstock, T., Dohe, S., Groß, J., and Kiel, M.: TCCON data from Karlsruhe, Germany, Release GGG2014R1, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.karlsruhe01.R1/1182416, 2014.

Heymann, J., Reuter, M., Hilker, M., Buchwitz, M., Schneising, O., Bovensmann, H., Burrows, J. P., Kuze, A., Suto, H., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kawakami, S., Kivi, R., Morino, I., Petri, C., Roehl, C., Schneider, M., Sherlock, V., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: Consistent satellite XCO_2 retrievals from SCIAMACHY and GOSAT using the BESD algorithm, Atmos. Meas. Tech. Discuss., 8, 1787–1832, doi:10.5194/amtd-8-1787-2015, 2015.

Kawakami, S., Ohyama, H., Arai, K., Okumura, H., Taura, C., Fukamachi, T., and Sakashita, M.: TCCON data from Saga, Japan, Release GGG2014R0, TCCON data archive, hosted by the Carbon

Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.saga01.R0/1149283, 2014.

Kivi, R., Heikkinen, P., and Kyro, E.: TCCON data from Sodankyla, Finland, Release GGG2014R0., TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.sodankyla01.R0/1149280, 2014.

Kuze, A., Suto, H., Nakajima, M. and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied Optics, 48, 6716–6733, 2009.

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

Notholt, J., Petri, C., Warneke, T., Deutscher, N., Buschmann, M., Weinzierl, C., Macatangay, R., Grupe. P.: TCCON data from Bremen, Germany, Release GGG2014R0. TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. doi:10.14291/tccon.ggg2014.bremen01.R0/1149275, 2014.

Parrish, D. and Derber, J.: National Meteorological Center's spectral statistical interpolation analysis system, Mon. Weather Rev., 120, 1747–1763, 1992.

Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H., and Burrows, J. P.: A method for improved SCIAMACHY CO_2 retrieval in the presence of optically thin clouds, Atmos. Meas. Tech., 3, 209–232, doi:10.5194/amt-3-209-2010, 2010.

Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri, C., Robinson, J., Schneising, O., Sherlock, V., Velazco, V., Warneke, T., Wennberg, P. O., and Wunch, D.: Retrieval of atmospheric CO_2 with enhanced accuracy and precision from SCIAMACHY: validation with FTS measurements and comparison with model results, J. Geophys. Res.-Atmos., 116, D04301, doi:10.1029/2010JD015047, 2011.

Sherlock, V. B., Connor, Robinson, J., Shiona, H., Smale, D., and Pollard, D.: TCCON data from Lauder, New Zealand, 125HR, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.lauder02.R0/1149298, 2014.

Sussmann, R. and Rettinger, M.: TCCON data from Garmisch, Germany, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.garmisch01.R0/1149299, 2014.

Wang, J.-W., Denning, A. S., Lu, L., Baker, I. T., Corbin, K. D., and Davis, K. J.: Observations and simulations of synoptic, regional, and local variations in atmospheric CO_2, J. Geophys. Res.-Atmos., 112, D04108, doi:10.1029/2006JD007410, 2007.

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N., Petri, C., Grupe, P., Vuillemin, C., Truong, F., Schmidt, M., Ramonet, M., and Parmentier, E.: TCCON data from Orleans, France, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.orleans01.R0/1149276, 2014.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., Keppel-Aleks, G., Allen, N., and Ayers, J.: TCCON data from Park Falls, Wisconsin, USA, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.parkfalls01.R0/1149161, 2014a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin, J.: TCCON data from Lamont, Oklahoma, USA, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.lamont01.R0/1149159, 2014b.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The total carbon column observing network, Philos. T. Roy. Soc. A, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011a.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Elder-ing, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO_2 total columns from space, Atmos. Chem. Phys., 11, 12317–12337, doi:10.5194/acp-11-12317-2011, 2011b.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., Roehl, C. M., Wennberg, P. O., Griffith, D. W. T., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Connor, B., Rettinger, M., Sussmann, R., Ahonen, P., Heikkinen, P., Kyrö, E., Mendonca, J., Strong, K., Hase, F., Dohe, S. and Yokota, T.: Improvement of the retrieval algorithm for GOSAT SWIR XCO_2 and XCH_4 and their validation using TCCON data, Atmos. Meas. Tech., 6, 1533–1547, 2013.

Table 1. List on the used TCCON stations ordered by latitude from North to South.

Site	Lat	Lon	Starting date	Reference
Sodankylä (sodankyla01)	67.37	26.63	6 Feb 2009	Kivi et al. (2014)
Białystok (bialystok01)	53.23	23.02	1 Mar 2009	Deutscher et al. (2014)
Bremen (bremen01)	53.10	8.85	6 Jan 2005	Notholt et al. (2014)
Karlsruhe (karlsruhe01)	49.10	8.44	19 Apr 2010	Hase et al. (2014)
Orléans (orleans01)	47.97	2.11	29 Aug 2009	Warneke et al. (2014)
Garmisch (garmisch01)	47.48	11.06	16 Jul 2007	Sussmann and Rettinger (2014)
Park Falls (parkfalls01)	45.94	−90.27	26 May 2004	Wennberg et al. (2014a)
Four Corners (fourcorners01)	36.80	−108.48	1 Mar 2011	Dubey et al. (2014)
Lamont (lamont01)	36.60	−97.49	6 Jul 2008	Wennberg et al. (2014b)
Saga (saga01)	33.24	130.29	28 Jul 2011	Kawakami et al. (2014)
Izaña (izana01)	28.30	−16.48	18 May 2007	Blumenstock et al. (2014)
Ascension Island (ascension01)	−7.92	−14.33	22 May 2012	Feist et al. (2014)
Darwin (darwin01)	−12.43	130.89	28 Aug 2005	Griffith et al. (2014a)
Réunion Island (reunion01)	−20.90	55.49	6 Oct 2011	De Mazière et al. (2014)
Wollongong (wollongong01)	−34.41	150.88	26 Jun 2008	Griffith et al. (2014b)
Lauder 125HR (lauder02)	−45.05	169.68	2 Feb 2010	Sherlock et al. (2014)

Table 2. Statistics of the XCO_2 difference between the simulations (free run and analysis) and the average hourly TCCON data (model-TCCON): bias (δ_k, in ppm), scatter (σ_k, in ppm) and correlation coefficient (r_k). Also shown are the mean, the mean absolute error (MAE) and the deviation of the stations bias (respectively δ, Δ and σ, in ppm), the mean scatter (π, in ppm) and the mean r (last three rows). The second column (N) is the number of data used for computing the statistics.

		Free run			Analysis		
Site	N	Bias	Scatter	r	Bias	Scatter	r
Sodankylä	20441	−1.59	1.35	0.91	−0.55	1.35	0.92
Białystok	16063	−2.68	1.96	0.81	−1.66	1.80	0.77
Bremen	4883	−1.62	1.52	0.79	−0.41	1.27	0.82
Karlsruhe	4201	−1.26	1.72	0.80	−0.25	1.54	0.82
Orléans	8444	−0.38	1.36	0.85	0.09	1.21	0.91
Garmisch	10371	−0.92	1.59	0.82	−0.29	1.62	0.80
Park Falls	27991	−1.69	2.06	0.81	−0.60	1.45	0.90
Four Corners	19872	0.69	1.76	0.58	0.57	1.43	0.74
Lamont	43731	−0.20	2.09	0.59	−0.04	1.35	0.80
Saga	10349	−1.19	1.61	0.75	−0.64	1.33	0.83
Izaña	4463	0.27	0.80	0.90	0.40	0.62	0.94
Ascension Island	7111	2.31	1.29	0.24	0.72	1.27	0.21
Darwin	29194	1.57	1.12	0.78	−0.02	1.04	0.79
Réunion Island	18880	0.56	0.73	0.76	−0.77	0.60	0.78
Wollongong	27562	0.30	1.05	0.71	−1.08	1.06	0.65
Lauder	53500	0.01	0.83	0.86	−0.97	0.59	0.85
Mean	16	−0.36	1.43	0.75	−0.34	1.22	0.78
MAE	16	1.08	–	–	0.57	–	–
Deviation	16	1.27	–	–	0.61	–	–

Table 3. Statistics of the XCO_2 differences between the MACC GOSAT BESD dataset and the average hourly TCCON data (left block, GOSAT-TCCON) or the analysis and the average hourly TCCON data (right block, model-TCCON): bias (δ_k, in ppm), scatter (σ_k, in ppm) and correlation coefficient (r_k). The analysis has been sampled similarly to the GOSAT dataset in time and space. Also shown are the mean, the mean absolute error (MAE) and the deviation of the stations bias, the mean scatter (all in ppm) and the mean r (last three rows). The second column (N) is the number of data points used for computing the statistics.

		MACC GOSAT dataset			Analysis		
Site	N	Bias	Scatter	r	Bias	Scatter	r
Sodankylä	90	−0.26	4.50	0.39	0.24	1.41	0.92
Białystok	58	−0.28	3.45	0.32	1.06	1.99	0.17
Bremen	41	1.19	2.34	0.53	0.54	0.86	0.81
Karlsruhe	91	1.45	2.74	0.52	0.89	0.74	0.88
Orléans	52	0.20	2.44	0.34	1.29	0.57	0.84
Garmisch	76	1.64	3.10	0.55	1.17	1.06	0.77
Park Falls	63	1.50	3.22	0.71	−0.08	1.03	0.95
Four Corners	102	−0.00	3.79	0.64	0.65	0.81	0.89
Lamont	340	−1.01	4.05	0.57	0.05	1.01	0.91
Saga	61	0.40	2.95	0.76	0.14	0.88	0.90
Darwin	234	−1.27	3.37	0.42	−0.11	0.81	0.84
Wollongong	221	−3.03	3.86	0.31	−1.54	1.07	0.74
Mean	12	0.04	3.32	0.50	0.36	1.02	0.80
MAE	12	1.02	–	–	0.65	–	–
Deviation	12	1.31	–	–	0.74	–	–

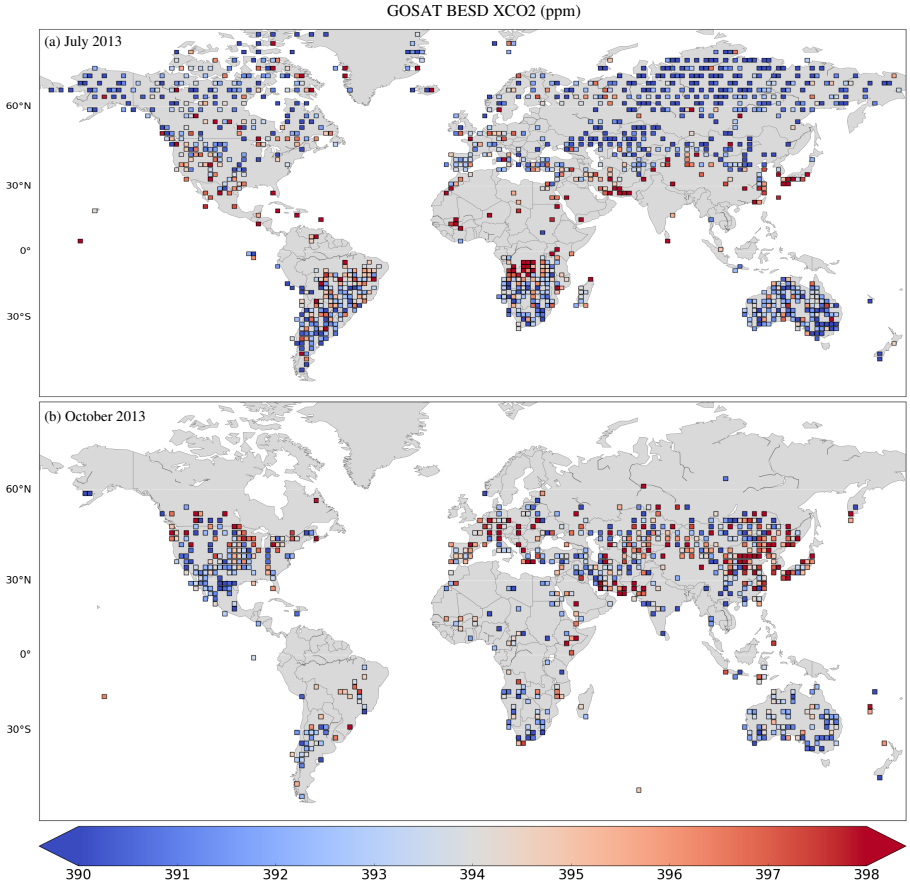

Figure 1. Example of the distribution of the assimilated GOSAT BESD XCO_2 data: July 2013 (top panel, about 3400 retrievals) and October 2013 (bottom, about 1270 retrievals). The monthly data are here aggregated on a $2^\circ \times 2^\circ$ grid and averaged. The blue/red represents the low/high averaged XCO_2 values in ppm.

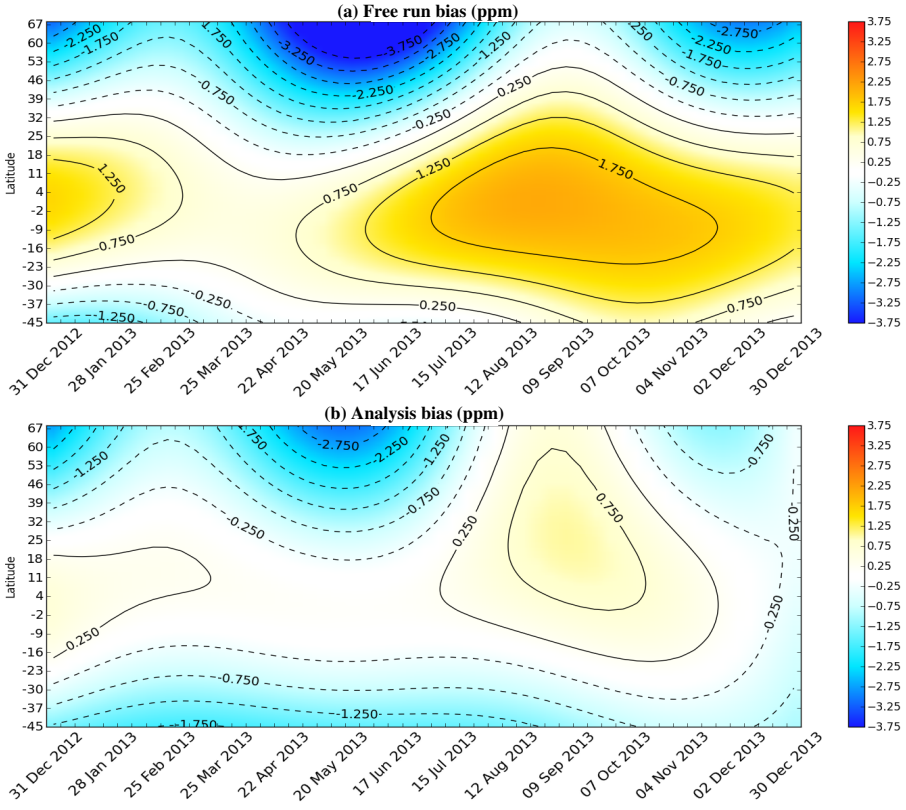

Figure 2. Hovmöller diagram (latitude vs. time) of the smoothed bias (in ppm, negative/positive in blue/red) of the simulated XCO_2 against the data of the TCCON network, from 1 January to 31 December, 2013. Top: free run simulation. Bottom: analysis.

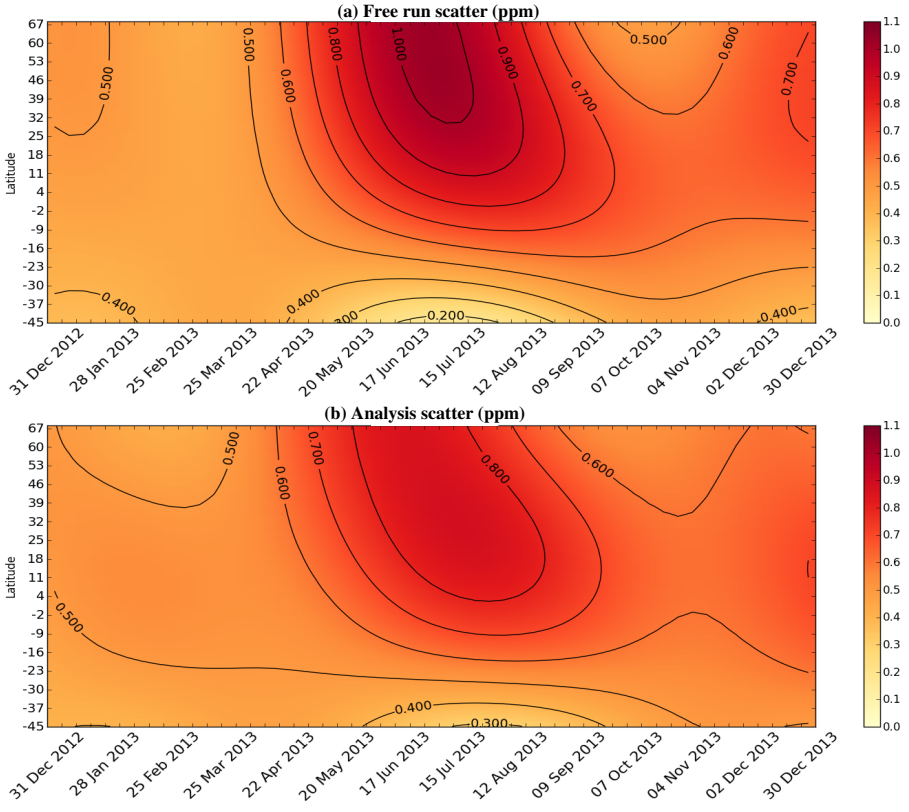

Figure 3. Same as Fig. 2 but for the standard deviation and yellow/red for low/high values.

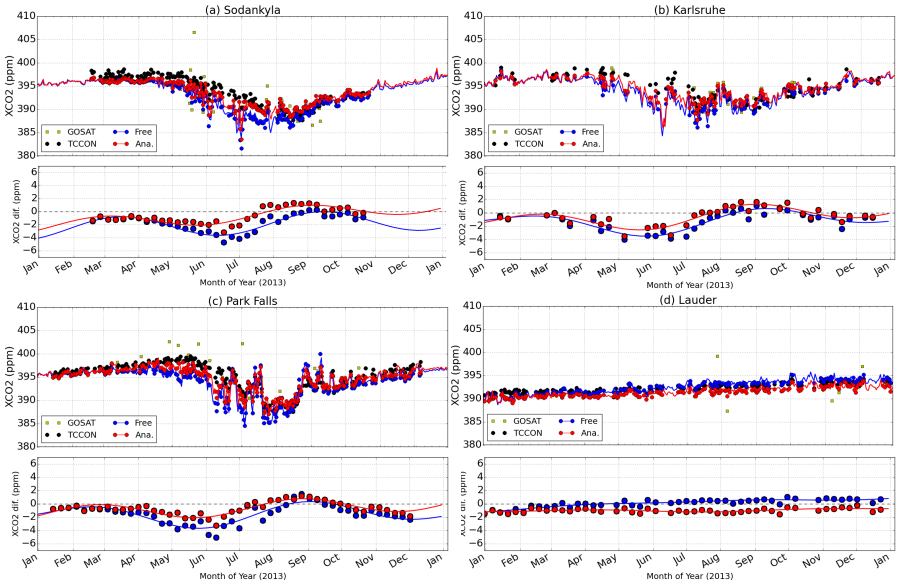

Figure 4. Time series of XCO_2 (in ppm) at **(a)** Sodankylä, Finland, **(b)** Karlsruhe, Germany, **(c)** Park Falls, USA and **(d)** Lauder, New Zeland, between 1 January and 31 December, 2013. For each station, the top panel presents the daily averaged data from TCCON (black dots), the daily averaged data from GOSAT co-located in time and space with the station (yellow squares), the simulated XCO_2 (solid lines) and the daily averaged simulated XCO_2 in the observation space (coloured dots). The bottom panel presents the weekly averaged bias of the simulated XCO_2 against the TCCON data (coloured dots) and the smoothed bias (solid lines). The blue colour is for the free run while the red colour is for the analysis.

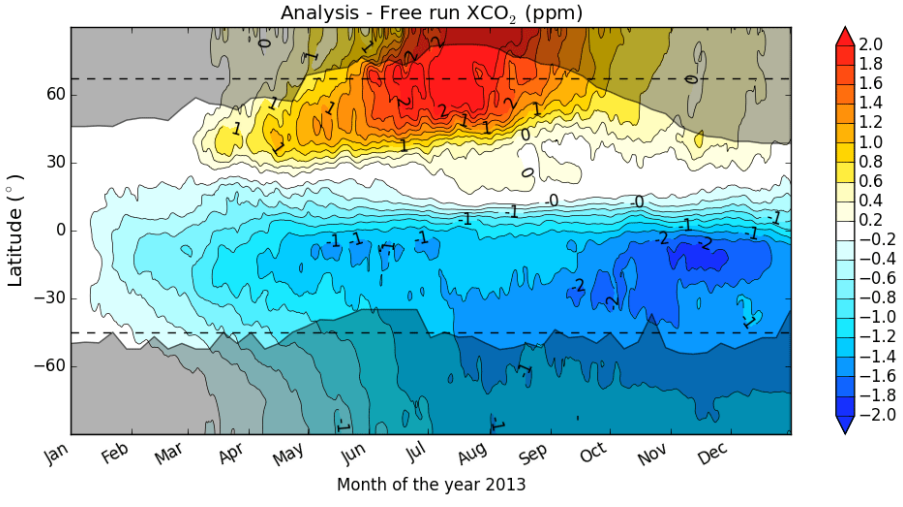

Figure 5. Hovmöller diagram (latitude vs. time) of the difference in ppm (negative/positive in blue/red) between XCO_2 from the analysis and from the free run simulation, from 1 January to 31 December, 2013. The horizontal dotted lines represent the latitude of the northernmost and the southernmost TCCON station respectively. The grey shaded areas are where GOSAT does not provide observations.

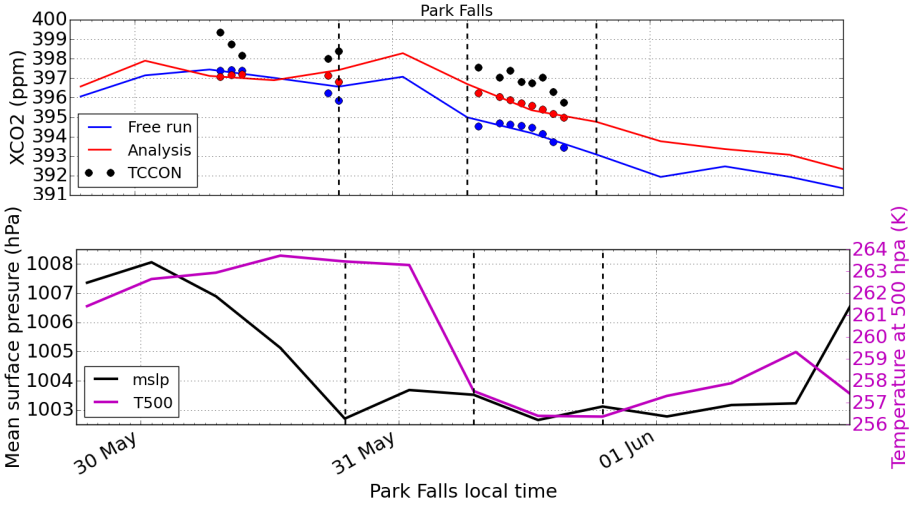

Figure 6. Situation over Park Falls (USA) between 30 May and 2 June. Top panel: evolution of XCO_2 (in ppm) from hourly averaged TCCON data (black dots), the free run (blue line and dots) and the analysis (red line and dots). The dots are the values of the model in the observation space. Lower panel: evolution of the mean sea level pressure (in hPa, black line) and the temperature at 500 hPa (in K, magenta line). The vertical dotted lines represent 31 May, at 00:00 UTC at 12:00 UTC, and the 1 June, at 00:00 UTC.

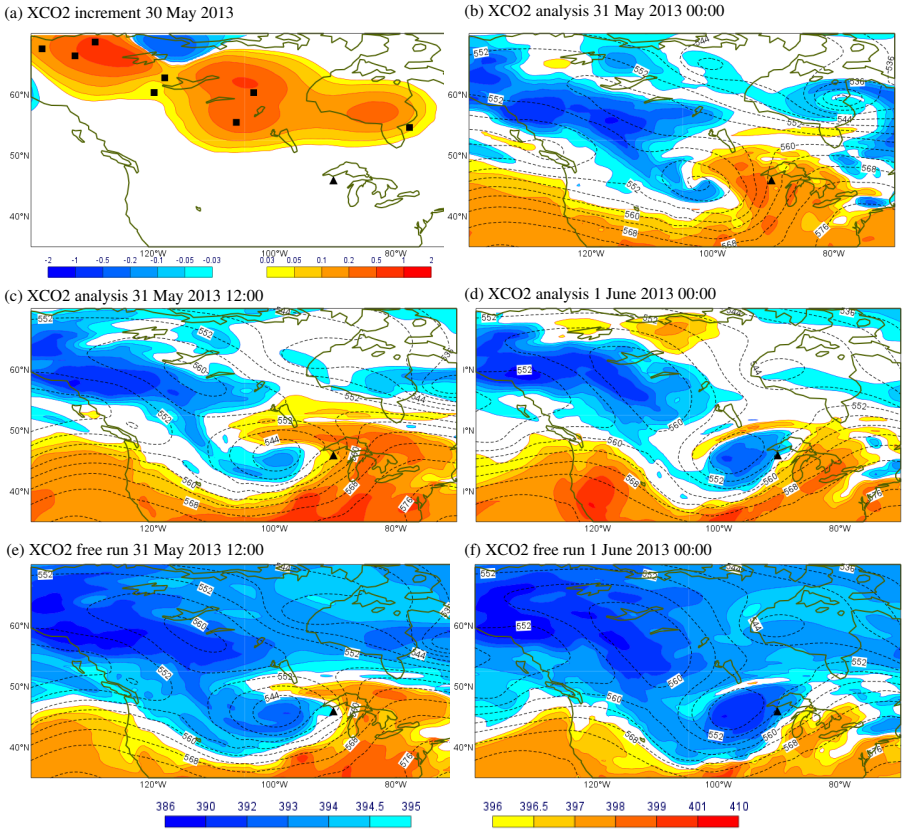

Figure 7. Situation around Park Falls (black triangle), Wisconsin, USA, end of May 2013. **(a)** average increment in terms of XCO_2 (in ppm, negative/positive in blue/red) on 30 May 2013 (contours) and location of the GOSAT measurements during this day (black rectangles). **(b–d)** XCO_2 (in ppm) respectively on 31 May at 00:00 UTC, at 12:00 UTC and on 1 June at 00:00 UTC from the analysis. **(e, f)** XCO_2 (in ppm) on 31 May at 12:00 UTC and on 1 June at 00:00 UTC from the free run (below/above background value in blue/red). For **(b)** to **(f)** the dark contours are the values of the geopotential at 500 hPa.

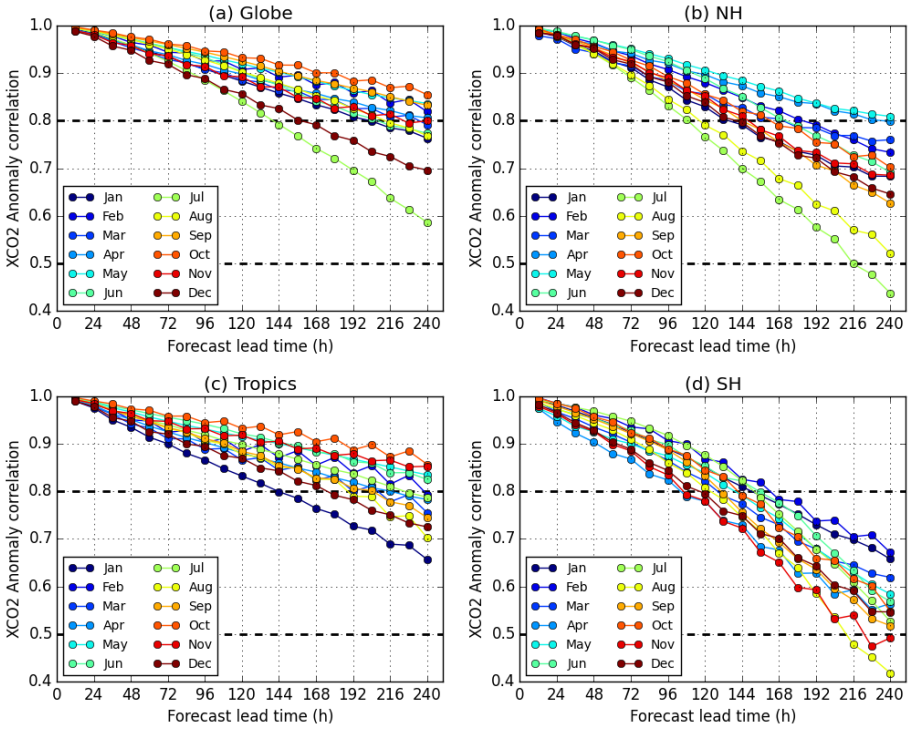

Figure 8. Anomaly correlation coefficient (ACC) of the forecast compared to its own analysis as a function of the forecast lead time and for each month: **(a)** global ACC, **(b)** ACC for the Northern Hemisphere (20° N–90° N), **(c)** ACC for the tropics (20° S–20° N), **(d)** ACC for the Southern Hemisphere (90° S–20° S). Each month is represented by a different colour (see inset legends).