

This discussion paper is/has been under review for the journal Atmospheric Measurement Techniques (AMT). Please refer to the corresponding final paper in AMT if available.

# A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data and its application to ACOS-GOSAT and TCCON

H. Nguyen<sup>1</sup>, G. Osterman<sup>1</sup>, D. Wunch<sup>2</sup>, C. O'Dell<sup>3</sup>, L. Mandrake<sup>1</sup>, P. Wennberg<sup>2</sup>,  
B. Fisher<sup>1</sup>, and R. Castano<sup>1</sup>

<sup>1</sup>Jet Propulsion Laboratory, Pasadena, CA, USA

<sup>2</sup>California Institute of Technology, Pasadena, CA, USA

<sup>3</sup>Colorado State University, Fort Collins, CO, USA

Received: 15 January 2014 – Accepted: 16 January 2014 – Published: 14 February 2014

Correspondence to: H. Nguyen (hai.nguyen@jpl.nasa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.

**A method for  
colocating satellite  
 $X_{\text{CO}_2}$  data to  
ground-based data**

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Abstract

Satellite measurements are often compared with higher-precision ground-based measurements as part of validation efforts. The satellite soundings are rarely perfectly coincident in space and time with the ground-based measurements, so a colocation methodology is needed to aggregate “nearby” soundings into what the instrument would have seen at the location and time of interest. We are particularly interested in validation efforts for satellite-retrieved total column carbon dioxide ( $X_{\text{CO}_2}$ ), where  $X_{\text{CO}_2}$  data from Greenhouse Gas Observing Satellite (GOSAT) retrievals (ACOS, NIES, RemoteC, PPDF, etc.) or SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCHIACHY) are often colocated and compared to ground-based column  $X_{\text{CO}_2}$  measurement from Total Carbon Column Observing Network (TCCON).

Current colocation methodologies for comparing satellite measurements of total column dry-air mole fractions of  $\text{CO}_2$  ( $X_{\text{CO}_2}$ ) with ground-based measurements typically involve locating and averaging the satellite measurements within some latitudinal, longitudinal, and temporal window. We examine a geostatistical colocation methodology that takes a weighted average of satellite observations depending on the “distance” of each observation from a ground-based location of interest. The “distance” function that we use is a modified Euclidian distance with respect to latitude, longitude, time, and mid-tropospheric temperature at 700 hPa. We apply this methodology to  $X_{\text{CO}_2}$  retrieved from Greenhouse Gas Observing Satellite (GOSAT) spectra by the ACOS team, cross-validate the results to TCCON  $X_{\text{CO}_2}$  ground-based data, and present some comparison between our methodology and standard existing colocation methods showing that in general geostatistical colocation produces smaller mean-squared error.

AMTD

7, 1495–1533, 2014

### A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 1 Introduction

Carbon dioxide (CO<sub>2</sub>) is an important anthropogenic greenhouse gas, and quantifying the exchange of CO<sub>2</sub> between the atmosphere and the Earth's surface is a critical part of the global carbon cycle and an important determinant of future climate (Gruber et al., 2009). One important measure of CO<sub>2</sub> is total column carbon dioxide ( $X_{\text{CO}_2}$ ), which is available from ground-based Total Carbon Column Observing Network (TCCON: Wunch et al., 2011a) and from space-based satellite instruments such as the Greenhouse gases Observing Satellite (GOSAT: Yokota et al., 2004; Hamazaki et al., 2005) and the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY: Bovensmann et al., 1999).

Ground-based total column CO<sub>2</sub> measurements tend to be more precise and accurate than space-based measurements, but ground-based stations often are sparsely located around the globe, and areas such as Siberia, Asia, Africa, South America, and the oceans have particularly poor coverage. Satellite instruments have much better coverage and are able to sample the entire globe in a matter of days or weeks. Together, the ground-based and space-based CO<sub>2</sub>-observing instruments provide a complementary ensemble of high-precision sparse-coverage and lower-precision global-coverage measurements. An important component of satellite retrieval assessment is validation relative to independent in-situ ground-based sources of data in order to assess important metrics such as bias and variability relative to the underlying true process. These bias and variability assessments can in turn be used to improve the retrieval algorithm to reduce spurious error resulting from factors such as: limited understanding of the instrument's calibration, uncertainties in the O<sub>2</sub> and CO<sub>2</sub> absorption cross sections, and subtle errors in the implementation of the retrieval algorithm (Crisp et al., 2012).

Often, there are spatial and temporal mismatches in the observed locations of the remote sensing instrument and ground-based validation instrument, and some spatial (and temporal) interpolation is required in order to "colocate" the two sources

# AMTD

7, 1495–1533, 2014

## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

of data before a direct comparison is possible. We define “validation collocation” as estimating, through interpolation using nearby satellite observations, what a remote sensing instrument would have seen at some chosen location and time. In this paper, we examine a new collocation methodology that is mathematically motivated in an error-minimization framework. Specifically, we are interested in developing a collocation methodology to combine retrieved  $X_{\text{CO}_2}$  data from GOSAT spectra using the Atmospheric  $\text{CO}_2$  Observations from Space (ACOS) for comparison against TCCON  $X_{\text{CO}_2}$  data with the goal of minimizing the expected interpolation error.

Collocation methods for  $X_{\text{CO}_2}$  in the existing literature include geographical,  $T_{700}$  (Wunch et al., 2011b), and SRON/KIT (Guerlet et al., 2013) collocation. Geographical collocation typically defines some spatio-temporal neighborhood region, also known as a coincidence criterion, around the location of interest and then take a summary statistics (e.g., mean or median). Examples of geographical collocation includes averaging all same-day satellite observations falling within  $\pm 5^\circ$  of a location of interest (Inoue et al., 2013), averaging all observations falling within  $5^\circ$  and  $\pm 2$  h (Cogan et al., 2012), and taking the monthly median of all observations within  $10^\circ \times 10^\circ$  lat-lon box (Reuter et al., 2013).

More sophisticated collocation methodologies add other correlated geophysical covariates in constructing such “neighborhoods” under the principle that conditioning on these additional correlated covariates would improve the quality of the comparison. Wunch et al. (2011b)’s  $T_{700}$  collocation method takes the average of all GOSAT observations falling within  $\pm 30^\circ$  longitude,  $\pm 10^\circ$  latitude,  $\pm 5$  days, and  $\pm 2$  K with respect to mid-tropospheric temperature at 700 hPa ( $T_{700}$ ) of the TCCON location of interest. Guerlet et al. (2013)’s SRON/KIT method similarly takes the average of all same-day satellite values that fall within  $\pm 7.5^\circ$  latitude,  $\pm 25^\circ$  longitude, and  $\pm 0.5$  ppm of the 3-day averaged model  $X_{\text{CO}_2}$  data.

All the collocation methodologies above operate on an implicit assumption that observations “near” one another are more likely to be correlated, where “nearby” indicates being proximal in some coordinate space and metric. The notion of “nearness”



Sect. 4. Summary and discussion of the methodology along with possible extensions are presented in Sect. 5.

## 2 Overview of ACOS-GOSAT, TCCON, and auxiliary data sources

The Greenhouse Gases Observing Satellite (GOSAT) was launched on 23 January as a joint venture by the National Institute for Environmental Studies (NIES), the Japanese Space Agency (JAXA), and the Ministry of the Environment (MOE). It is a polar-orbiting satellite dedicated to the observation of total-column CO<sub>2</sub> and CH<sub>4</sub>, both major greenhouse gases, from space using reflected sun-light spectra from the Thermal And Near-infrared Sensor for carbon Observation Fourier Transform Spectrometer (TANSO FTS, Hamazaki et al., 2005). It flies at approximately 665 kilometers (km) altitude, and it completes an orbit every 100 min. The satellite returns to the same observation location every three days (Morino et al., 2011).

Following the failure of Orbiting Carbon Observatory (OCO) launch in February 2009, the OCO project formed the Atmospheric CO<sub>2</sub> Observations from Space (ACOS) and, under agreement with NIES, JAXA, and MOE, applied the OCO retrieval algorithm to the GOSAT spectra to compute column-averaged dry-air mole fractions of CO<sub>2</sub>. The ACOS-GOSAT data processing algorithm is based on the optimal estimation approach of Rodgers (2000) and is described in detail in O'Dell et al. (2012). It is modified from the OCO retrieval algorithm (Bosch et al., 2006; Connor et al., 2008; Boesch et al., 2011) to account for the different physical viewing geometries and properties such as instrument line shapes and noise models.

In this paper, we assess the performance of different collocation methods on ACOS-GOSAT data by comparing collocated values to the more precise and accurate TCCON data. We use the v3.3 release of ACOS-GOSAT data, available from the Goddard Data and Information Services Center spanning July 2009 to April 2013 (see “ACOS Data Access” in Bibliography for notes). GOSAT data are divided into three categories: glint (ocean) data, land High (H) gain data, and land Medium (M) gain data. The v3.3 Data

## A method for colocating satellite X<sub>CO<sub>2</sub></sub> data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



User's Guide notes that M data and ocean glint data have some deficiencies in that particular version and should only be used with heightened caution (Osterman et al., 2013). Hence in this paper we only make use of H gain land data.

Following the recommendation of the Data User's Guide, we screen the v3.3 H gain data using a set of 11 criteria to obtain data suitable for science analysis. The set of screening criteria are reproduced in Table 5. The ACOS-GOSAT data have a bias that is known to be correlated with certain other variables such as air mass, blended albedo, and posterior-prior surface pressure difference (Wunch et al., 2011b). The v3.3 Data Guide recommends a linear bias correction to ACOS-GOSAT  $X_{CO_2}$  based on the difference between the retrieved and prior surface pressure from the A-band cloudscreen and the ratio of the signal in the strong  $CO_2$  band to that of the  $O_2A$  band (Osterman et al., 2013). However, since such bias correction was done through comparison of v3.3 ACOS retrievals with models and TCCON retrievals, we refrain from applying the v3.3 bias correction to avoid potential "feedback" in the comparison of our colocated v3.3 ACOS and TCCON values.

The Total Carbon Column Observing Network consists of ground-based Fourier Transform Spectrometers that record direct solar spectra in the near-infrared. These spectra are then used to retrieve column-averaged abundances of atmospheric constituents including  $CO_2$ ,  $CH_4$ ,  $N_2O$ , HF, CO, and  $H_2O$ , which are directly comparable with the near-infrared total-column measurements from space-based instruments (Wunch et al., 2011a). Whereas GOSAT retrievals are susceptible to variability resulting from contamination by optically-thick clouds and aerosols that were missed by the cloud screening process (O'Dell et al., 2012), TCCON makes direct observation of the solar disk and hence is less sensitive to errors from scattered light (Crisp et al., 2012).

TCCON sites sample in a diverse range of atmospheric states, which include tropical and polar regions, continental and maritime, polluted and clean, providing valuable validation link between space-based measurements and the extensive ground-based in situ network (Wunch et al., 2011a). TCCON  $X_{CO_2}$  data in turn are validated against integrated aircraft profiles (Washenfelder et al., 2006; Deutscher et al., 2010;

## A method for collocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Messerschmidt et al., 2011; Wunch et al., 2010) and have a precision and accuracy of  $\sim 0.8$  ppm (Wunch et al., 2010).

We use the 2012 release version of the TCCON data (“GGG2012”) from the TCCON Data Archive (see “TCCON Data Access” in Bibliography for more information) for the following 16 locations: Bialystok, Bremen, Darwin, Eureka, Garmisch, Izana, Karlsruhe, Lamont, Lauder (both 120HR and 125HR), Ny Alesund, Orleans, Park Falls, Reunion, Sodankyla, Tsukuba (both 120HR and 125HR), and Wollongong. At each TCCON location, we use all available data that fall within the period July 2009 to April 2013. A map of the TCCON locations are shown in Fig. 1.

Atmospheric variability of  $X_{\text{CO}_2}$  has been shown to be correlated to the free-tropospheric potential temperature, which can be considered as a proxy for equivalent latitude for  $X_{\text{CO}_2}$  in the Northern Hemisphere (Keppel-Aleks et al., 2011). In this paper, we follow Wunch et al. (2011b) in making use of mid-tropospheric temperature as one of the covariates along with latitude, longitude, and time. Specifically, we use the mid-tropospheric temperature field at 700 hPa from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) reanalysis product, which uses a frozen state-of-the-art analysis/forecast system and performs data assimilation using past data (Kalnay et al., 1996). The mid-tropospheric temperature field at 700 hPa should be directly proportional to the potential temperature at 700 hPa for the range of temperature of interest, and its inclusion as a covariate should allow us to construct better colocation metrics.

## 2.1 Averaging kernel correction

To compare two observations obtained through optimal estimation properly, the retrievals must be computed about a common a priori profile and the averaging kernels must be applied to account for the effect of smoothing (Rodgers and Connor, 2003). A detailed exposition on applying the averaging kernel correction for a ACOS-GOSAT and TCCON comparison is given in Sect. 4 and Appendix A of Wunch et al. (2011b).

## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Typically, to compare retrieval results from two different instruments with different viewing geometrics, retrieval algorithms, a priori profiles ( $x_a$ ), and averaging kernel ( $\mathbf{A}$ ), we need an common ensemble profile ( $x_c$ ) and covariance matrix ( $\mathbf{S}_c$ ), which represent the mean and variability of the atmosphere at the common comparison location.

As an alternative, we can use one observing system to try to retrieve what the other system would have produced as its retrieved total column (Rodgers and Connor, 2003). Given that TCCON retrievals are considered more precise and accurate, we smooth the TCCON value with the ACOS-GOSAT column averaging kernels to produce what ACOS-GOSAT would have produced at the TCCON location given the TCCON profile as “truth”.

Fortunately, ACOS v3.3 a priori profiles and TCCON a priori profiles are very close to one another, and hence we only need to account for differences in the averaging kernels. We use the following averaging kernel equation from Appendix A of Wunch et al. (2011b):

$$\hat{z}_{12} = z_a + (\gamma - 1) \sum_j h_j a_{1j} x_{aj}, \quad (1)$$

where  $z_a$  is the a priori  $X_{\text{CO}_2}$ ,  $\hat{z}_i$  is the retrieved  $X_{\text{CO}_2}$  and  $i = 1$  for ACOS and  $i = 2$  for TCCON,  $h$  is the  $X_{\text{CO}_2}$  pressure weighting function, and  $a$  is the  $X_{\text{CO}_2}$  averaging kernel norm. The term  $\gamma$  is a scaling factor that produces the best fit of the TCCON output to the spectrum, and it is approximated as a ratio between the retrieved TCCON  $X_{\text{CO}_2}$  and the a priori  $X_{\text{CO}_2}$ ,

$$\gamma \approx \frac{\hat{z}_2}{z_a}.$$

We apply Eq. (1) to TCCON data between July 2009 and April 2013 at the 16 chosen TCCON locations. For each TCCON observation, we obtain the corresponding ACOS a priori information using the colocation neighborhood region defined in Wunch et al. (2011b); see Sect. 4 for more detail. Figure 2 displays a plot of the relationship between

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the original TCCON retrievals vs. the averaging-kernel-corrected TCCON data. In effect, the averaging kernel correction tends to pull TCCON observations closer to the ACOS a priori  $X_{\text{CO}_2}$ ; TCCON values that are higher than the ACOS a priori value tend to be pulled downwards, while TCCON values lower than the ACOS a priori  $X_{\text{CO}_2}$  tends to be pulled upwards. The standard deviation of the difference between non-corrected and corrected TCCON values, aggregated over all TCCON sites, is 0.24 ppm.

Having done averaging kernel correction to put the TCCON and ACOS retrievals on the same footing, we now describe our methodology for optimally collocating ACOS-GOSAT observations to any TCCON location in the next section. The collocated values will then be compared to averaging-kernel-corrected TCCON data in Sect. 4.

### 3 Geostatistical colocation

Our colocation methodology exists within a geostatistical framework, which is a part of the broader area of spatial statistics. Here, we briefly review that framework, give some necessary notation, and present basic derivations for estimation in a spatial context.

Let  $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$  be a hidden, real-valued spatial process on a multi-dimensional domain. In the application of ACOS-GOSAT and TCCON, we let  $\mathbf{s} = (s_{\text{lat}}, s_{\text{lon}}, s_t, s_T)'$  be a 4-dimensional vector specifying the latitude, longitude, time, and mid-tropospheric temperature at 700 hPa ( $T_{700}$ ) respectively and we assume  $Y(\mathbf{s})$  is the  $X_{\text{CO}_2}$  process at location  $\mathbf{s}$ . We assume that the  $Z(\mathbf{s})$ ,  $X_{\text{CO}_2}$  retrieval at location  $\mathbf{s}$ , is a sum of the true  $X_{\text{CO}_2}$  process and a retrieval-error term. That is,

$$\begin{aligned} Z(\mathbf{s}) &= Y(\mathbf{s}) + \epsilon(\mathbf{s}) \\ &= t(\mathbf{s}) + \nu(\mathbf{s}) + \epsilon(\mathbf{s}), \end{aligned} \quad (2)$$

where  $t(\mathbf{s})$  is a large-scale deterministic trend term that accounts for seasonal and yearly trends,  $\nu(\mathbf{s})$  is a small-scale variability term that accounts for spatial correlation, and  $\epsilon(\mathbf{s})$  is the retrieval-error term. We also assume that we have a variogram function

$2\gamma(\mathbf{s}_i, \mathbf{s}_j)$  which describes the degree of spatial dependence between any two locations  $\mathbf{s}_i$  and  $\mathbf{s}_j$  as in the following definition,

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) = \text{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = E \left( |(Z(\mathbf{s}_i) - t(\mathbf{s}_i)) - (Z(\mathbf{s}_j) - t(\mathbf{s}_j))|^2 \right); \mathbf{s}_i, \mathbf{s}_j \in D. \quad (3)$$

Let  $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N))'$  be the vector of satellite observations taken at  $N$  footprints around an interpolation point  $\mathbf{s}_0$ , and let  $\mathbf{T} = (t(\mathbf{s}_1), t(\mathbf{s}_1), \dots, t(\mathbf{s}_N))'$  be the corresponding  $N$ -dimensional vector of trend terms. We wish to find an estimate of  $Y(\mathbf{s}_0)$  as a linear combination of the detrended retrieved  $X_{\text{CO}_2}$  vector  $\mathbf{D} = (\mathbf{Z} - \mathbf{T})$  and an unknown vector of coefficients  $\mathbf{a}'_{\mathbf{s}_0}$ . That is,

$$\hat{Y}(\mathbf{s}_0) = t(\mathbf{s}_0) + \mathbf{a}'_{\mathbf{s}_0} \mathbf{D}, \quad (4)$$

such that we minimize the expected mean-squared error

$$\min_{\mathbf{a}_{\mathbf{s}_0}} E \left( (Y(\mathbf{s}_0) - \hat{Y}(\mathbf{s}_0))^2 \right), \quad (5)$$

subject to the unbiasedness constraint  $\mathbf{a}'_{\mathbf{s}_0} \mathbf{1} = 1$ .

The solution to the constrained minimization problem in Eqs. (4) and (5) can be found using the method of Lagrange multipliers. Cressie (1993) gives the following equation for the solution  $\mathbf{a}_{\mathbf{s}_0}$  that satisfies Eqs. (4) and (5),

$$\begin{pmatrix} \mathbf{a}_{\mathbf{s}_0} \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_1, \mathbf{s}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_N, \mathbf{s}_N) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_0) \\ 1 \end{pmatrix}, \quad (6)$$

where  $\lambda$  is the scalar Lagrange multiplier and  $\mathbf{a}_{\mathbf{s}_0}$  is the vector of kriging coefficient.

**A method for colocating satellite  $X_{\text{CO}_2}$  data to ground-based data**

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



An attractive property of the geostatistical approach is that the semivariogram function can be used to calculate the expected estimation error at the interpolation location. The expression for the interpolation error is as follows,

$$\hat{\sigma}(\mathbf{s}) = (\mathbf{a}'_{s_0} \lambda) \begin{pmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_0) \\ 1 \end{pmatrix}. \quad (7)$$

Estimating the fully general semivariogram model  $\gamma(\mathbf{s}_N, \mathbf{s}_0)$  is a difficult problem that is prone to robustness issues when the data are sparse. To make the problem more tractable, we assume that the variogram structure of  $Z$  is isotropic under certain distance metrics. That is, we assume that the semivariogram capturing the spatial dependence between any two locations  $\gamma(\mathbf{s}_i, \mathbf{s}_j)$  is only dependent on its “distance” as in the following equation,

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \gamma(|\mathbf{s}_i - \mathbf{s}_j|_B), \quad (8)$$

where  $|\cdot|_B$  is a modified Euclidean distance given by

$$|\mathbf{s}_i - \mathbf{s}_j|_B = \sqrt{\left( \frac{(s_{i,\text{lat}} - s_{j,\text{lat}})^2}{B_1} + \frac{(s_{i,\text{lon}} - s_{j,\text{lon}})^2}{B_2} + \frac{(s_{i,t} - s_{j,t})^2}{B_3} + \frac{(s_{i,T} - s_{j,T})^2}{B_4} \right)} \quad (9)$$

$$= \sqrt{((\mathbf{s}_i - \mathbf{s}_j)' \mathbf{B} (\mathbf{s}_i - \mathbf{s}_j))}, \quad (10)$$

and  $\mathbf{B}$  is a diagonal  $4 \times 4$  matrix whose diagonal elements  $(B_1, B_2, B_3, B_4)$  represent the scaling parameters along each of the coordinate direction: latitude, longitude, time, and

$T_{700}$ .

**A method for  
colocating satellite  
 $X_{\text{CO}_2}$  data to  
ground-based data**

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



### 3.1 Application to ACOS-GOSAT and TCCON data

Computation of the colocated values and their corresponding interpolation error in Eqs. (4) and (7) requires that we know the trend  $t(\cdot)$ , the scaling diagonal matrix  $\mathbf{B}$ , and the semivariogram function  $\gamma(\cdot, \cdot)$ . In practice, these terms are unknown. For our application, we will assume certain parametric forms for the trend  $t(\cdot)$  and the semivariogram function  $\gamma(\cdot, \cdot)$  and estimate the corresponding parameters along with  $\mathbf{B}$  from the retrieved data.

Unfortunately, ACOS-GOSAT data are quite sparse once we pass the radiances through a cloud-filter, retrieval selection criteria, and post-retrieval data quality filters. The relative global sparseness of the ACOS-GOSAT data makes it difficult to obtain robust estimates of the scaling matrix  $\mathbf{B}$  and the semivariogram function  $\gamma(\cdot, \cdot)$ . We address this problem by using CarbonTracker model  $X_{\text{CO}_2}$  data to estimate these spatial-temporal dependence parameters. Our assumption here is that CarbonTracker and ACOS-GOSAT share the same medium-to-large scale spatio-temporal dependence structure for  $X_{\text{CO}_2}$ . For instance, the dynamics in CarbonTracker should reasonably approximate synoptic and large-scale dependence in moderately homogeneous areas such as the Southern Hemispheres while very fine-scale, local effects may not be captured. Since we are trying to assess larger-scale agreement, this mid-to-large scale approximation should be sufficient. Note that this assumption is not as restrictive as the assumption that CarbonTracker and ACOS-GOSAT have the same expected value of  $X_{\text{CO}_2}$  at every location.

CarbonTracker is a  $\text{CO}_2$  assimilation system developed by the National Oceanic and Atmospheric Administration (NOAA) to keep track of the global  $\text{CO}_2$  emissions and uptake. The model combines surface air samples collected around the globe and from tall towers and small aircrafts in North America with an atmospheric transport model coupled with a Kalman filter to produce estimates of atmospheric  $\text{CO}_2$  mole fractions on a global grid (Peters et al., 2007). The model  $X_{\text{CO}_2}$  data are regularly gridded at  $1^\circ \times 1^\circ$  daily resolution, making it particularly convenient for use in spatial parameter

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





### 3.3 Parameter estimation

Having modeled and estimated the trend term  $t(\cdot)$ , we now estimate the scaling matrix  $\mathbf{B}$  and the semivariogram function  $\gamma(\cdot, \cdot)$ . We model the semivariogram function with the spherical semivariogram, which has the form

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) \equiv \gamma(h) = (s - n) \left( \left( \frac{3h}{2r} - \frac{h^3}{2r^3} \right) \mathbf{1}_{(0,r)}(h) + \mathbf{1}_{[r,\infty)}(h) \right) + n, \quad (12)$$

where  $h = |\mathbf{s}_i - \mathbf{s}_j|_B$  is the modified Euclidean distance between  $\mathbf{s}_i$  and  $\mathbf{s}_j$ ,  $\mathbf{1}_S(h)$  is 1 if  $h \in S$  and 0 otherwise, the term  $n$  is the nugget, which denotes the height of the semivariogram at the origin where  $h = 0$ , the term  $s$  is called the sill, which is the limit of the semivariogram as  $h \rightarrow \infty$ , and  $r$  is the range, which is the distance at which the difference between the semivariogram and the sill is negligible (see Cressie, 1993, for more detail).

Given the scaling matrix  $\mathbf{B}$ , we can estimate the semivariogram parameters  $\{n, s, r\}$  by constructing the robust empirical semivariogram estimator discussed in Cressie (1980) and Cressie (1993, Sect. 2.4), from the CarbonTracker data as follows. For  $h > 0$ , define

$$2\bar{\gamma}(h) \equiv \frac{\left\{ \frac{1}{|N(h)|} \sum_{N(h)} |D(\mathbf{s}_m) - D(\mathbf{s}_n)|^2 \right\}^4}{0.457 + \frac{0.494}{|N(h)|}}, \quad (13)$$

where  $D(\mathbf{s})$  is the detrended CarbonTracker value at location  $\mathbf{s}$ , and  $N(h)$  is the set of observation pairs that are separated by distance of  $h$ ,

$$N(h) \equiv \{(\mathbf{s}_n, \mathbf{s}_m) : |\mathbf{s}_n - \mathbf{s}_m|_B = h; \quad m, n = 1, \dots, N\}.$$

In practice, the set  $N(h)$  is defined using a small tolerance interval around  $h$ , since it may not be possible to find pairs of locations that are exactly distance  $h$  apart (Cressie, 1993, p. 70). The term  $|N(h)|$  denotes the number of unique elements in  $N(h)$ .

## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



We assume that the scaling matrix  $\mathbf{B}$  and the semivariogram function  $\gamma(\cdot, \cdot)$  are constant with respect to each hemisphere and with respect to time. To estimate the scaling matrix  $\mathbf{B}$ , we construct empirical semivariograms using data pairs that only approximately differ in one of the four coordinates. This effectively sets three of the four terms in the right-hand side of Eq. (9) to zero (or very close to zero), allowing us to estimate the single remaining scaling parameter  $B_i$ . For instance, to estimate the scaling parameter along the longitudinal direction, we search through the CarbonTracker data for the corresponding hemisphere to obtain pairs of observations that share the same date, latitude, and  $T_{700}$  but different longitudes. We then calculate the robust semivariogram estimator in Eq. (13), compute the corresponding semivariogram fit to the spherical model in Eq. (12), and set the scaling parameter  $B_i$  equal to the resulting range  $r$ .

Table 2 contains the list of scaling parameters for the Northern Hemisphere and the Southern Hemisphere. In general, the scaling parameters agree fairly well with the coincidence windows given in Wunch et al. (2011b)'s  $T_{700}$  methodology. For the Northern Hemisphere, the scaling parameters for latitude, longitude, time, and  $T_{700}$  are 15.1, 24.5, 3, and 3.1, respectively; the corresponding parameters for the Southern Hemisphere are 11.6, 19.3, 3, and 2.3. This indicates that in general the Northern Hemisphere has longer spatial correlation range than the Southern Hemisphere.

Having estimated the scaling parameters  $\mathbf{B}$ , we construct a set of empirical semivariogram values using Eq. (13) for each of the hemispheres and estimate the semivariogram parameters  $\{n, s, r\}$  using an iterative, Gauss–Newton fitting algorithm to fit to the chosen spherical semivariogram model to the empirical semivariogram estimates (Cressie, 1985). The resulting nugget, sill, and range parameters for the northern and Southern Hemisphere are in Table 2.

The nugget and the sill parameters indicate that in general the Northern Hemisphere has higher variability and that the spatial correlation structure is weaker in the Southern Hemisphere. Having estimated the parameters for  $\mathbf{B}$  and  $\gamma(s_j, s_j)$ , we can compute the collocated ACOS-GOSAT value at any TCCON location using Eqs. (4) and (6).



within  $\pm 30^\circ$  longitude,  $\pm 10^\circ$  latitude,  $\pm 5$  days, and  $\pm 2$  K in  $T_{700}$ . The longitudinal constraint is reduced to  $\pm 10^\circ$  for the Tsukuba TCCON site to avoid inordinate influence from ACOS-GOSAT retrievals over China. In this section, we assess the performance of Wunch et al. (2011b)'s collocation criterion and the geographical method relative to geostatistical collocation methodology between the period July 2009 and April 2013.

#### 4.1 Comparison between ACOS-GOSAT and TCCON data

For the performance assessment, we use all available TCCON data from 16 locations (four in the Southern Hemisphere, 12 in the Northern Hemisphere, see Fig. 1) between the period July 2009 and April 2013. Since the three collocation methodologies compute averages over large temporal spans, applying the three methodologies to individual same-station TCCON observations (which may be spaced seconds or minutes apart from one another) would result in a scenario where many temporally-proximal TCCON observations are matched to the same ACOS-GOSAT collocated value. We avoid this problem by taking the daily median of TCCON  $X_{\text{CO}_2}$  values and use them as the standard against which we assess the outputs of the collocation methodologies.

Having taken the daily median of the TCCON  $X_{\text{CO}_2}$  values, we scan the entire TCCON dataset and locate corresponding matches using the collocation methodologies. At every TCCON location and every day for which we have a daily median TCCON  $X_{\text{CO}_2}$  value, we gather the corresponding ACOS-GOSAT values falling within the respective coincidence regions and then compute the collocated value for each of the three methodologies. Some TCCON locations may not have a corresponding collocated ACOS-GOSAT value for particular days due to the fact that they do not have any GOSAT values within their coincidence neighborhood. The neighborhood regions are constructed separately for each TCCON location, and thus some ACOS-GOSAT sounding may be used more than once in computing the collocated values for several temporally or spatially proximal TCCON daily-median values.

Figure 4 displays the collocated ACOS-GOSAT values vs. the data from the 16 TCCON sites. The  $x$  axis displays the temporal range while the  $y$  axis displays the

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



TCCON daily-median values and the collocated ACOS-GOSAT from the three collocation methodologies. In general the Wunch et al. (2011b) and the geostatistical methodology have a lot more collocated values due the fact that they use data from a large spatio-temporal neighborhood surrounding a TCCON site. From Fig. 4, all collocation methodologies indicate that ACOS-GOSAT  $X_{\text{CO}_2}$  tends to be larger than TCCON  $X_{\text{CO}_2}$  and the magnitude of this bias is between 1–1.5 ppm. Northern TCCON sites such as Eureka and Ny Alesund do not have nearby ACOS-GOSAT H-gain good-quality retrievals during the winter due to ice and snow issues.

Table 4 displays five ACOS-GOSAT/TCCON summary statistics: number of matched days ( $N$ ), mean bias, standard deviation, correlation coefficient ( $r$ ), and slope. Geostatistical and  $T_{700}$  methodologies have wider coincidence criteria than our chosen geographical methodology, and consequently have more daily-matched ACOS-GOSAT/TCCON pairs. To better examine the patterns in Table 4, we display the main statistics (bias, standard deviation, and correlation coefficient) in graphical form in Fig. 5 with the TCCON stations listed in order of decreasing latitude.

In the top panel of Fig. 5, we examine the average bias between the three collocation methodologies vs. TCCON daily-median  $X_{\text{CO}_2}$  at each of the 18 TCCON datasets. The average bias is fairly consistent between the three methodologies and range between 0.6 ppm to 2.5 ppm. In general, the  $T_{700}$  and geostatistical collocation methodologies tend to produce the same bias, while the geographical method has more pronounced variability in the estimates of mean bias at the TCCON sites. This is likely due to the fact that the geographical method has a much smaller neighborhood region, and thus do not yield enough collocated matches relative to TCCON to produce a robust bias estimate. All three methodologies tend to have high bias estimates for the three northern-most TCCON sites: Sodankyla, Eureka, and Ny Alesund. This is likely because soundings acquired over these snowy and icy surfaces have low reflectivity in the 1.61 and 2.06  $\mu\text{m}$  bands; consequently scattering by thin clouds and aerosols can constitute a larger fraction of the total signal and introduce larger uncertainties in the optical path length (Crisp et al., 2012).



other collocation methodologies in this section, is more reflective and representative of the true underlying dependence structure.

Another way to assess the fit between ACOS-GOSAT and TCCON values is through examining the correlation coefficient. Since the satellite and station instruments are both observing total-column  $X_{\text{CO}_2}$ , we expect the two to follow a linear relationship with slope 1 and  $y$  intercept equal to the mean bias. The correlation coefficient is a good tool to examine the strength and direction of the linear relationship between the ACOS-GOSAT and TCCON values, and we show the correlation estimates at each of the TCCON sites in the bottom panel of Fig. 5.

In our particular application, correlation values closer to 1 indicate stronger linear dependence. In this respect, geostatistical collocation performs marginally better than  $T_{700}$  collocation, and they both perform better than geographical collocation. The correlation estimates mostly cluster between the range 0.75 to 0.99, with the exception of being TCCON stations in the Southern Hemisphere. This likely results from the fact that the Northern Hemisphere and the Southern Hemisphere have different seasonal and synoptic variability ratio. The ACOS-GOSAT retrievals in general do quite well in capturing the overall seasonal trends in both hemisphere. However, in the Northern Hemisphere, the seasonal variability amplitude is larger than the synoptic variability, and consequently the correlation coefficient between ACOS-GOSAT and TCCON is larger. In the Southern Hemisphere, the seasonal variability amplitude is much smaller at 0.3 ppm, thus lowering the correlation coefficient.

The comparisons in this section indicate that in general geographical collocation has low matching yield and poor accuracy performance. This is likely because the geographical collocation used (same-day 500 km circle) lacks the large coincident neighborhood to reduce the individual retrieval variability through averaging. Wunch et al. (2011)'s  $T_{700}$  and geostatistical collocation both take advantage of a larger collocation neighborhood to produce more accurate ACOS-GOSAT collocated values.

While these methodologies produce roughly the same bias estimates, geostatistical collocation produces distinctly lower root mean-squared error. This improvement likely

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

comes from the fact that while the existing collocation methodologies tend to give equal weights to all satellite observations falling within the coincident window, our methodology gives different weights to the coincident satellite observations based on the distance metric defined in Eq. (10).  $X_{\text{CO}_2}$  in general tends to be a smoothly-varying field and it can be reasonably assumed to follow the geographical principle that locations close to one another are more likely to be similar than locations far apart (Tobler, 1970); therefore our methodology produces better accuracy because its spatial-dependency model is more reflective and representative of the true underlying  $X_{\text{CO}_2}$  field.

## 5 Conclusions

Validation collocation, or the practice of interpolating satellite and ground-based validation data to the same spatio-temporal coordinate, is an important part of instrument validation and assessment. All collocation methodologies are essentially data interpolation, which carries with it interpolation error. This interpolation error is an extra component in the root mean-squared error of the difference between validation and collocated data; it is important to minimize the interpolation error as much as possible in order to better assess important instrumental and operational metrics such as bias and variability relative to the validation data.

This paper examines a new collocation technique in comparing ACOS-GOSAT and TCCON data. We model the spatial dependence structure as being isotropic under a modified Euclidean distance metric. Our methodology is similar to previous collocation techniques (e.g., geographical,  $T_{700}$ , SRON/KIT) in that we assume that nearby observations are more likely to be correlated than observations far apart. However, whereas the existing methodologies define some neighborhood regions and then give all neighboring observations equal weights, our methodology weights each observation depending on the distances between the data and the interpolation location of interest.

In Sect. 4, we show that our geostatistical collocation methodology has the lowest mean-squared error of the difference between collocated ACOS-GOSAT data and



## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



In this paper we assumed that that ACOS-GOSAT retrievals can be approximated and treated as zero-area points. In certain applications it may be more reasonable to assume that a satellite observation is an average of the true geophysical process  $Y(\cdot)$  over the area of the footprint plus some measurement-error term. The resulting process of inferring a spatial process at one resolution from data at another resolution, also known as the change-of-support problem, is more complex; see Gotway and Young (2002) for a review. In general, there is no analytical solution for estimating from areal data the parameters of standard variogram models (e.g., spherical, exponential, etc.), however, certain classes of spatial models provide for straightforward and seamless parameter estimation (for instance, see Spatial Random Effects model, Cressie and Johannesson, 2008; Nguyen et al., 2012).

In Sect. 4, we model the variogram parameters as temporally-constant. An extension of the methodology would be to model temporal dependence in the variogram parameters. Naturally, good models of the temporal dependence would improve the colocation performance, but there is a trade-off in the complexity of the temporal evolution models and the robustness of the parameter estimates. One possible approach would be to assume that the spatial-correlation structure is constant over a season; although care would be needed in combining data straddling different seasons. Further examination about the the trade-off between estimation robustness and temporal evolution complexity would be needed.

*Acknowledgements.* ACOS data are obtained from Goddard Earth Sciences Data and Information Services Center, operated by NASA, from the website <http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS.shtml>. TCCON data were obtained from the TCCON Data Archive, operated by the California Institute of Technology, from the website at <http://tcon.ipac.caltech.edu/>. NCEP Reanalysis data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.cdc.noaa.gov/>.

## References

ACOS Data Access: Goddard Earth Sciences Data and Information Services Center, available at: <http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS.shtml>, last access: June 2013. 1500

5 Artis, M., Clavel, J. G., Hoffmann, M., and Nachane, D.: Harmonic Regression Models: a Comparative Review with Applications, IEW – Working Papers 333, Institute for Empirical Research in Economics – University of Zurich, 2007. 1508

Boesch, H., Baker, D., Connor, B., Crisp, D., and Miller, C.: Global characterization of CO<sub>2</sub> column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 Mission, *Remote Sensing*, 3, 270–304, 2011. 1500

10 Bosch, H., Toon, G. C., Sen, B., Washenfelder, R. A., Wennberg, P. O., Buchwitz, M., de Beek, R., Burrows, J. P., Crisp, D., Christi, M., Connor, B. J., Natraj, V., and Yung, Y. L.: Space-based near-infrared CO<sub>2</sub> measurements: testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, *J. Geophys. Res.-Atmos.*, 111, D23302, doi:10.1029/2006JD007080, 2006. 1500

15 Bovensmann, H., Burrows, J., Buchwitz, M., Frederick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Geode, A. P. H.: SCIAMACHY: mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 127–150, 1999. 1497

Butz, A., Guerlet, S., Jacob, D. J., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G. C., Wunch, D., Wennberg, P. O., Deutscher, N. M., Griffith, D. W. T., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.: Toward accurate CO<sub>2</sub> and CH<sub>4</sub> observations from GOSAT, *Geophys. Res. Lett.*, 38, 2–7, doi:10.1029/2011GL047888, 2011. 1511

20 Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): comparison with ground-based TCCON observations and GEOS-Chem model calculations, *J. Geophys. Res.-Atmos.*, 117, D21301, doi:10.1029/2012JD018087, 2012. 1498, 1511

25 Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: inverse method and prospective error analysis, *J. Geophys. Res.-Atmos.*, 113, D05305, doi:10.1029/2006JD008336, 2008. 1500

30 Cressie, N.: M-estimation in the presence of unequal scale, *Stat. Neerl.*, 34, 19–32, 1980. 1509

# AMTD

7, 1495–1533, 2014

## A method for colocating satellite X<sub>CO<sub>2</sub></sub> data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

5659 of Society of Photo-Optical Instrumentation Engineering (SPIE) Conference Series, 5659, 73–80, doi:10.1117/12.581198, 2005. 1497, 1500

Inoue, M., Morino, I., Uchino, O., Miyamoto, Y., Yoshida, Y., Yokota, T., Machida, T., Sawa, Y., Matsueda, H., Sweeney, C., Tans, P. P., Andrews, A. E., Biraud, S. C., Tanaka, T., Kawakami, S., and Patra, P. K.: Validation of  $X_{CO_2}$  derived from SWIR spectra of GOSAT TANSO-FTS with aircraft measurement data, *Atmos. Chem. Phys.*, 13, 9771–9788, doi:10.5194/acp-13-9771-2013, 2013. 1498, 1511

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., and Reynolds, R.: The NCEP/NCAR 40-year Reanalysis Project, *B. Am. Meteorol. Soc.*, 77, 437–471, 1996. 1502

Keppel-Aleks, G., Wennberg, P. O., and Schneider, T.: Sources of variations in total column carbon dioxide, *Atmos. Chem. Phys.*, 11, 3581–3593, doi:10.5194/acp-11-3581-2011, 2011. 1502, 1511

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., Andres, R. J., Blavier, J.-F., Connor, B., Davis, K. J., Desai, A. R., Messerschmidt, J., Notholt, J., Roehl, C. M., Sherlock, V., Stephens, B. B., Vay, S. A., and Wofsy, S. C.: The imprint of surface fluxes and transport on variations in total column carbon dioxide, *Biogeosciences*, 9, 875–891, doi:10.5194/bg-9-875-2012, 2012. 1511

Messerschmidt, J., Geibel, M. C., Blumenstock, T., Chen, H., Deutscher, N. M., Engel, A., Feist, D. G., Gerbig, C., Gisi, M., Hase, F., Katrynski, K., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Ramonet, M., Rettinger, M., Schmidt, M., Sussmann, R., Toon, G. C., Truong, F., Warneke, T., Wennberg, P. O., Wunch, D., and Xueref-Remy, I.: Calibration of TCCON column-averaged  $CO_2$ : the first aircraft campaign over European TCCON sites, *Atmos. Chem. Phys.*, 11, 10765–10777, doi:10.5194/acp-11-10765-2011, 2011. 1502

Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., Toon, G. C., Wunch, D., Roehl, C. M., Notholt, J., Warneke, T., Messerschmidt, J., Griffith, D. W. T., Deutscher, N. M., Sherlock, V., Connor, B., Robinson, J., Sussmann, R., and Rettinger, M.: Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra, *Atmos. Meas. Tech.*, 4, 1061–1076, doi:10.5194/amt-4-1061-2011, 2011. 1500, 1511

Nguyen, H., Cressie, N., and Braverman, A.: Spatial statistical data fusion for remote sensing applications, *J. Am. Stat. Assoc.*, 107, 1004–1018, 2012. 1518

**A method for  
colocating satellite  
X<sub>CO<sub>2</sub></sub> data to  
ground-based data**

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO<sub>2</sub> retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, 5, 99–121, doi:10.5194/amt-5-99-2012, 2012. 1500, 1501
- Oshchepkov, S., Bril, A., Yokota, T., Morino, I., Yoshida, Y., Matsunaga, T., Belikov, D., Wunch, D., Wennberg, P. O., Toon, G. C., O'Dell, C. W., Butz, A., Guerlet, S., Cogan, A., Boesch, H., Eguchi, N., Deutscher, N. M., Griffith, D., Macatangay, R., Notholt, J., Sussmann, R., Rettinger, M., Sherlock, V., Robinson, J., Kyrö, E., Heikkinen, P., Feist, D. G., Nagahama, T., Kadyrov, N., Maksyutov, S., Uchino, O., and Watanabe, H.: Effects of atmospheric light scattering on spectroscopic observations of greenhouse gases from space: validation of PPDF-based CO<sub>2</sub> retrievals from GOSAT, *J. Geophys. Res.*, 117, 1–18, 2012. 1511
- Osterman, G., Eldering, A., Avis, C., O'Dell, C., Martinez, E., Frankenberg, C., Fisher, B., and Wunch, D.: ACOS Level 2 Standard Product Data User's Guide v3.3, Revision Date: Revision G, 13 June 2013, available at: [http://oco.jpl.nasa.gov/files/oco/ACOS\\_v3.3\\_DataUsersGuide.pdf](http://oco.jpl.nasa.gov/files/oco/ACOS_v3.3_DataUsersGuide.pdf) (last access: 15 September 2013), 2013. 1501, 1528
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *P. Natl. Acad. Sci. USA*, 104, 18925–18930, 2007. 1507
- Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P., O'Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S., Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to deliver satellite retrieved atmospheric CO<sub>2</sub> concentrations for surface flux inversions: the ensemble median algorithm EMMA, *Atmos. Chem. Phys.*, 13, 1771–1780, doi:10.5194/acp-13-1771-2013, 2013. 1498, 1511
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, River Edge, NJ, 2000. 1500
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.-Atmos.*, 108, 4116, doi:10.1029/2002JD002299, 2003. 1503

TCCON Data Access: TCCON Data Archive, available at: <http://tcon.ipac.caltech.edu/>, last access: June 2013. 1502

Tobler, W.: A computer movie simulating urban growth in the Detroit region, *Econ. Geogr.*, 46, 234–240, 1970. 1516

5 Washenfelder, R. A., Toon, G. C., Blavier, J.-F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M., and Daube, B. C.: Carbon dioxide column abundances at the Wisconsin Tall Tower site, *J. Geophys. Res.-Atmos.*, 111, D22305, doi:10.1029/2006JD007154, 2006. 1501

10 Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, *Atmos. Meas. Tech.*, 3, 1351–1362, doi:10.5194/amt-3-1351-2010, 2010. 1502

15 Wunch, D., Toon, G., Blavier, J., Washenfelder, R., Notholt, J., Connor, B., Griffith, D., Sherlock, V., and Wennberg, P.: The Total Carbon Column Observing Network, *Philos. T. R. Soc. A*, 369, 2087–2112, 2011a. 1497, 1501, 1511

20 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., Eldering, 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<sub>2</sub> total columns from space, *Atmos. Chem. Phys.*, 11, 12317–12337, doi:10.5194/acp-11-12317-2011, 2011b. 1498, 1501, 1502, 1503, 1510, 1511, 1512, 1513

25 Yokota, T., Oguma, H., and Inoue, G.: A nadir looking SWIR FTS to monitor CO<sub>2</sub> column density for Japanese GOSAT project, *Proc. Twenty-fourth Int. Sympo. on Space Technol., and Sci. (Selected Papers)*, 887–889, 2004. 1497

30

## AMTD

7, 1495–1533, 2014

### A method for colocating satellite X<sub>CO<sub>2</sub></sub> data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 1.** Annual and seasonal trend coefficients.

	Intercept (ppm)	Slope (ppm yr <sup>-1</sup> )	Amplitude (ppm)	Phase Shift (radian)
Northern Hemisphere	385.7900	2.6061	3.2040	0.1556
Southern Hemisphere	383.5127	2.4878	0.3099	4.0978

## A method for collocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 2.** Scaling coefficients and semivariogram parameters for Eq. (9).

	$B_1$	$B_2$	$B_3$	$B_4$	nugget ( $n$ )	sill ( $s$ )	range ( $r$ )
Northern Hemisphere	15	25	3	3	0.3	2.3	1.98
Southern Hemisphere	11	19	3	2	0.21	0.73	1.7





## A method for colocating satellite $X_{\text{CO}_2}$ data to ground-based data

H. Nguyen et al.

**Table 5.** Advanced screening criteria for ACOS v3.3 L2 H-gain data (Osterman et al., 2013, Sect. 2.5.2).

Variable	Comment	Criteria
RetrievalResults/outcome_flag	Flag indicating full physics outcome	1 or 2
RetrievalResults/aerosol_total_aod	Retrieved total column integrated aerosol optical depth for all aerosol types	0.01 to 0.02
SoundingGeometry/sounding_altitude_stddev	Standard deviation of the measure of altitude of the surface within the sounding	< 200
IMAPDOASPreprocessing/CO2_ratio_idp	Ratio of retrieved CO <sub>2</sub> column (no scattering code) in weak and strong CO <sub>2</sub> band	0.995 to 1.015
IMAPDOASPreprocessing/H <sub>2</sub> O_ratio_idp	Ratio of retrieved H <sub>2</sub> O column (no scattering code) in weak and strong CO <sub>2</sub> band	0.92 to 1.05
ABandCloudScreen/surface_pressure_delta_cld	Difference between surface pressure and a priori surface pressure	−825 to 575
SpectralParameters/reduced_chi_squared_O <sub>2</sub> _fph	The reduced $\chi^2$ value of the O <sub>2</sub> A-band clear-sky fit used in determine the presence or absence of cloud	< 1.5
RetrievalResults/albedo_slope_strong_CO2	Retrieved spectral dependence of Lamberion component of albedo within strong CO <sub>2</sub> channel	> $-10.0 \times 10^{-5}$
RetrievalResults/albedo_slope_o2	Retrieved spectral dependence of Lamberion component of albedo within O <sub>2</sub> channel	< $-1.3 \times 10^{-5}$
Blended Albedo	A mixture of two albedo terms	< 0.08

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

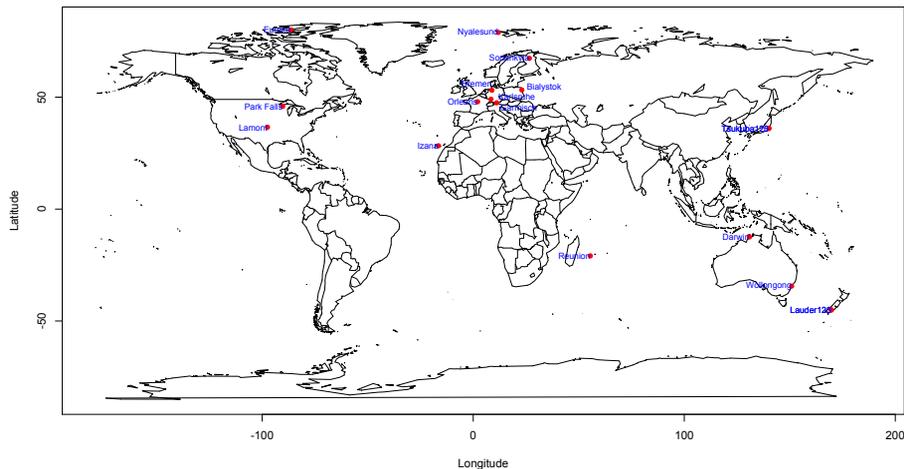
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.



**Fig. 1.** Locations of the 16 TCCON locations for which we perform GOSAT/ACOS and TCCON colocation comparison.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

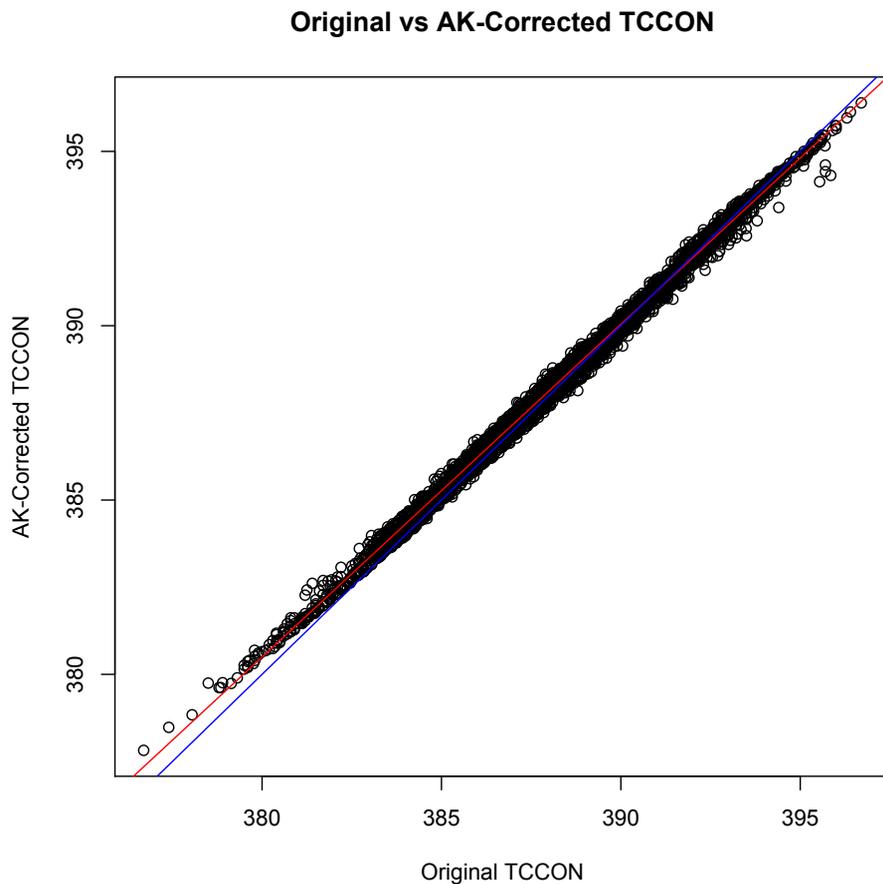
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Fig. 2.** Matched TCCON original daily median observations vs. averaging-kernel-corrected TCCON values. The red line is the linear regression line, while the blue is the 1–1 line.

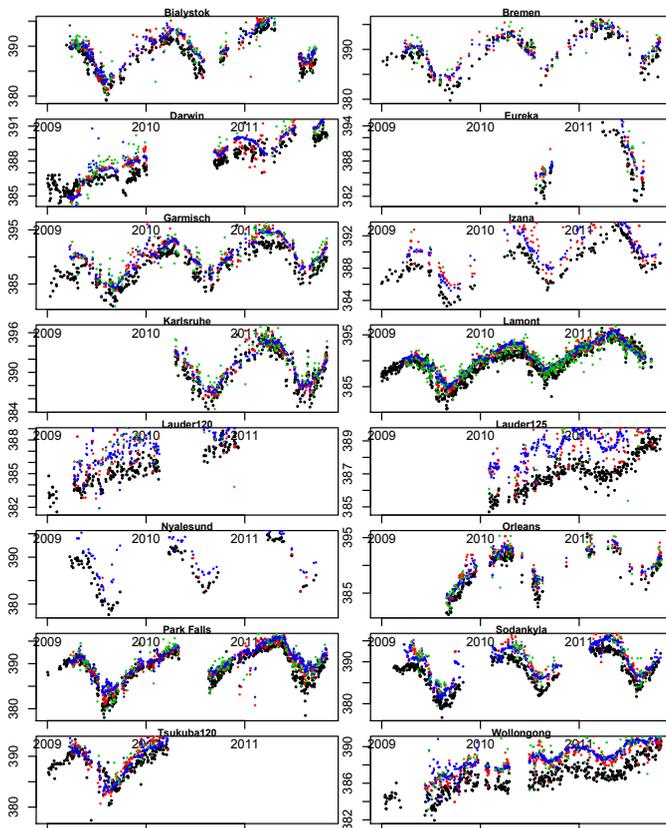
## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	







**Fig. 4.** TCCON daily median (black) vs.  $T_{700}$  (red), geostatistical (blue), and geographical (green) collocation values. Time in days is displayed on the x axis, while  $X_{CO_2}$  concentration in ppm is displayed on the y axis. Reunion and Tsukuba125 datasets are omitted due to the low number of observations.

**A method for colocating satellite  $X_{CO_2}$  data to ground-based data**

H. Nguyen et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

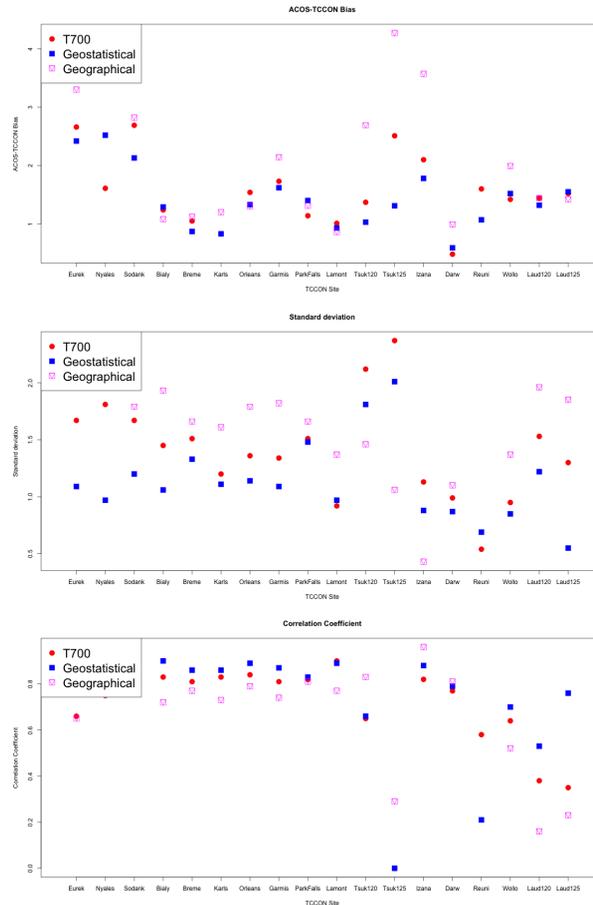
Printer-friendly Version

Interactive Discussion



## A method for colocating satellite $X_{CO_2}$ data to ground-based data

H. Nguyen et al.



**Fig. 5.** Summary statistics for the comparison between ACOS and TCCON using 3 colocation methodologies (top panel – bias; middle panel – standard deviation; bottom panel – correlation coefficients). TCCON stations are listed in order of decreasing latitude.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

