A simple empirical model estimating atmospheric CO$_2$ background concentrations

M. Reuter$^1$, M. Buchwitz$^1$, O. Schneising$^1$, J. Heymann$^1$, S. Guerlet$^2$, A. J. Cogan$^3$, H. Bovensmann$^1$, and J. P. Burrows$^1$

$^1$University of Bremen, Institute of Environmental Physics, Bremen, Germany
$^2$Netherlands Institute for Space Research (SRON), Utrecht, The Netherlands
$^3$University of Leicester, Department of Physics and Astronomy, Leicester, UK

Received: 22 December 2011 – Accepted: 28 January 2012 – Published: 9 February 2012

Correspondence to: M. Reuter (maximilian.reuter@iup.physik.uni-bremen.de)

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

A simple empirical CO₂ model (SECM) is presented to estimate column-average dry-air mole fractions of atmospheric CO₂ (XCO₂) as well as mixing ratio profiles. SECM is based on a simple equation depending on 17 empirical parameters, latitude, and date. The empirical parameters have been determined by least squares fitting to NOAA’s (National Oceanic and Atmospheric Administration) assimilation system CarbonTracker version 2010 (CT2010). Comparisons with TCCON (total column carbon observing network) FTS (Fourier transform spectrometer) measurements show that SECM XCO₂ agrees quite well with reality. The synthetic XCO₂ values have a standard error of 1.39 ppm and systematic station-to-station biases of 0.46 ppm. Typical column averaging kernels of the TCCON FTS, a SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY), and two GOSAT (Greenhouse gases Observing SATellite) XCO₂ retrieval algorithms have been used to assess the smoothing error introduced by using SECM profiles instead of CT2010 profiles as a priori. The additional smoothing error amounts to 0.17 ppm for a typical SCIAMACHY averaging kernel and is most times much smaller for the other instruments (e.g. 0.05 ppm for a typical TCCON FTS averaging kernel). Therefore, SECM is well-suited to provide a priori information for state of the art ground-based (FTS) and satellite-based (GOSAT, SCIAMACHY) XCO₂ retrievals. Other potential applications are: (i) quick check for obvious retrieval errors (by monitoring the difference to SECM), (ii) near real time processing systems (that cannot make use of models like CT2010 operated in delayed mode), (iii) “CO₂ proxy” methods for XCH₄ retrievals (as correction for the XCO₂ background), (iv) observing system simulation experiments especially for future satellite missions.

1 Introduction

Our current knowledge about atmospheric CO₂ concentrations and surface fluxes at regional scales over the globe comes primarily from ground-based in situ measurements.
of air sampling networks and tall towers. These measurements are used by assimilation systems like NOAA’s (National Oceanic and Atmospheric Administration) CarbonTracker (Peters et al., 2007, 2010), modeling global distributions of atmospheric CO₂ mixing ratios and surface fluxes. However, due to the sparseness of measurements, there are still large uncertainties especially on the surface fluxes (Stephens et al., 2007). Satellite and ground-based remote sensing measurements of column-average dry-air mole fractions of atmospheric CO₂ (XCO₂) are promising candidates to significantly reduce these uncertainties in the future (Rayner and O’Brien, 2001; Houweling et al., 2004).

Current satellite and ground-based XCO₂ retrieval techniques require more or less realistic estimates of true atmospheric concentrations. This information is used as a priori, first guess, and/or linearization point (e.g. Buchwitz et al., 2000; Barkley et al., 2006; Bösch et al., 2006, 2011; Washenfelder et al., 2006; Connor et al., 2008; Schneising et al., 2008; Butz et al., 2009; Reuter et al., 2010; Wunch et al., 2011; Yoshida et al., 2011). Typically, a XCO₂ retrieval’s sensitivity can deviate from unity within the atmospheric column. Broadly spoken, the retrieval “sees” only parts of the atmosphere and the “hidden” parts are complemented with the a priori knowledge. This means, one would like to use an a priori as realistic as possible because the retrieval result contains part of the a priori. For the same reason, one would like to use a simple, traceable a priori so that one can always distinguish between features coming from the measurement and from the a priori.

We will present a simple empirical CO₂ model (SECM) which addresses these needs but can be used for various other applications also (as discussed later). In the following section, a simple empirical equation estimating the global distribution of XCO₂ will be given. Afterwards, an equation will be presented to also estimate a simplified profile shape (Sect. 3). The corresponding parameterized error covariance matrix will be given in Sect. 4. In Sect. 5 SECM will be validated with TCCON (total column carbon observing network) FTS (Fourier transform spectrometer) measurements. In order to prove SECM’s usability as a priori information for state of the art satellite and ground-based
XCO₂ retrievals, we will analyze the smoothing error introduced when using SECM instead of CT2010 (CarbonTracker version 2010) (Sect. 6).

2 XCO₂

Our first step aims at finding a simple empirical description of the global XCO₂ distribution. Given the coarse assumption that the longitudinal dependency of XCO₂ can statistically be neglected, we use CT2010 as an estimate for the true XCO₂. In order to have a good estimate for background concentrations, we analyze a Pacific north/south transect being less influenced by local land sources and sinks (see Fig. 1). We then fit the parameters \(a_{00} \ldots a_{14}\) of the XCO₂ estimation function \(X_e\) so that the squared differences to the CT2010 transect are minimized:

\[
X_e(t, l) = a_{00} + a_{01} t + a_{02} \tanh (a_{03} l + a_{04}) + S(t, l).
\]  

(1)

As one can see, \(X_e\) basically depends on the date \(t\) (in units of years since 2003) and latitude \(l\). Geophysically, \(a_{00}\) and \(a_{01}\) account for a linear year-to-year increase mainly driven by anthropogenic CO₂ emissions. \(a_{02} \ldots a_{04}\) define the background north/south gradient with typically larger values at northern latitudes due to anthropogenic emissions. \(S\) represents the seasonal component modulating the increase and north/south gradient depending on date and latitude.

\[
S(t, l) = (a_{05} \tanh (a_{06} l + a_{07}) + a_{08} t) \sin (2 \pi t + a_{09} l)
+ (a_{10} \tanh (a_{11} l + a_{12}) + a_{13} t) \sin (4 \pi t + a_{14} l).
\]  

(2)

The seasonal component has a 12 month period with latitudinal dependent phase \((a_{09} l)\) and a 6 month period with latitudinal dependent phase \((a_{14} l)\) (see e.g. Baldocchi et al., 2001; Chamard et al., 2003). The amplitudes of both periods are defined by \(a_{05} - a_{08}\) and \(a_{10} - a_{13}\), respectively. They can vary with latitude (e.g. due to more vegetation at northern latitudes, Conway et al., 1994) and time (e.g. due to changing biospheric
activity, Keeling et al., 1995). Table 1 lists the fit results for the parameters \( a_i \) and Fig. 1 shows the comparison of CT2010 and synthetic values using Eq. (1).

In order to quantify the quality of the estimates derived with Eq. (1), we used an independent data set of 10,000 globally, randomly chosen CT2010 CO\(_2\) profiles in the period 2003–2009 from which we calculated XCO\(_2\). The standard deviation of the difference, referred to as standard error in the following, amounts to 0.99 ppm. The correlation between both data sets is 0.97.

### 3 Profile shape

In the second step we try to find a simple empirical function \( x_e \) defining the shape of a mixing ratio profile at given XCO\(_2\):

\[
x_e(p, t, l) = c_0 \left( 0.5 p^2_t - p_t \right) + c_1 S(t, l) \left( p_t - 0.5 - 0.5 p^2_t \right) + \begin{cases} c_0 p : p \leq p_t \\ S(t, l) c_1 (p - p_t) + c_0 p_t : p > p_t \end{cases}
\]

This equation estimates the mixing ratio for the pressure (height) \( p \) given in fraction of surface pressure \( p_s \), i.e. \( p \in [0, 1] \). The parameters \( c_0 \) and \( c_1 \) are determined similarly as \( a_{00}–a_{14} \) by least squares fitting to CT2010 mixing ratio profiles (see Table 1). At the pressure \( p_t = 0.2 \) (also given in fraction of surface pressure), the simplified atmosphere is split into two differently handled parts (approximately troposphere and stratosphere). The first two lines of Eq. (3) only account for preserving XCO\(_2\) while the profile shape is defined in the last part of Eq. (3). The idea is to have a linear decrease (with decreasing pressure) in the stratosphere \( (p \leq p_t) \). This accounts for slow mixing processes resulting in “older” air (with lower CO\(_2\) mixing ratios) towards top of atmosphere.

Within the troposphere, Eq. (3) approximates the profile also with a linear relation having a continuous transition to the stratosphere. In contrast to the stratosphere, the slope in the troposphere depends on the seasonal component \( S \). This results
in increasing values (with height) in the growing season where lowest values can be expected near the surface. Figure 2 shows the estimated profiles for three examples and corresponding CT2010 profiles. Obviously, Eq. (3) can reproduce the CT2010 profile shape to some extent but especially in the lower boundary layer close to regional sources and sinks distinct differences between SECM and CT2010 can be observed (see also Fig. 3).

4 Error covariance matrix

We again use the randomly chosen data set of 10 000 CT2010 CO₂ profiles to derive the error covariance matrix of SECM in comparison to CT2010. Figure 3 shows the error correlation matrix and the corresponding profile of the standard deviation of the difference between SECM and CT2010. We now use a simple correlation model (also used as an example in the textbook of Rodgers, 2000) to parameterize the correlation matrix C:

\[ C_{i,j} = e^{-|p_i - p_j|/\xi}. \]  

Here, \( p_i \) and \( p_j \) are the normalized pressure values of layer \( i \) and \( j \), \( \xi \) is the correlation length. Least squares fitting of the measured and parameterized error correlation matrix results in an optimal correlation length of \( \xi = 0.30 \). The profile of standard deviations \( \sigma_i \) was parameterized with (see Fig. 3):

\[ \sigma_i = \left( 1.25 + 1.75 p_i^5 \right) \text{ ppm}. \]  

The elements of the parameterized error covariance matrix \( S \) can now be calculated with:

\[ S_{i,j} = C_{i,j} \sigma_i \sigma_j. \]  

The parameters of Eq. (5) have been chosen in a way that they subjectively fit the profile of standard deviations. Additionally, the chosen parameters ensure that the
XCO$_2$ variance which can be calculated from $S$ is consistent with the variance directly calculated from the XCO$_2$ difference between SECM and CT2010.

It should be kept in mind, that the parameterized covariance matrix only describes errors of SECM in respect to CT2010. However, the total error consists of an additional part because of differences between CT2010 and “true” atmospheric profiles. This means, the parameterized covariance matrix can only be a reasonable approximation of the total error if the total error is dominated by the differences between SECM and CT2010. Otherwise a comparison of measured (“true”) CO$_2$ mixing ratios and corresponding simulated values would be required (Eguchi et al., 2010).

5 Comparison with TCCON

From Sect. 2 we already know, that the synthetic XCO$_2$ generated with SECM follows CarbonTracker quite well statistically. In this section, synthetic XCO$_2$ values are compared with TCCON measurements. From the results we can estimate how well SECM reproduces reality. For each “good” flagged TCCON measurement in the period 2006–2010 we computed a corresponding SECM value (Fig. 4). SECM agrees with an average standard error of 1.39 ppm with TCCON (even though SECM has no diurnal component). This agrees reasonably well with the 0.99 ppm error obtained in comparison with CT2010 (Sect. 2) given the fact that TCCON measurements have a single measurement precision of about 0.6 ppm (Toon et al., 2009). The station-to-station bias (standard deviation of all station biases) amounts to 0.47 ppm which is comparable to the TCCON accuracy (1 $\sigma$) of about 0.4 ppm (Wunch et al., 2010).

6 Smoothing error

The column averaging kernel (vector) of a XCO$_2$ retrieval describes its height (or pressure) dependent sensitivity to the true CO$_2$ mixing ratio. A perfect retrieval would have
an averaging kernel which is unity in every height under every measurement condition. Unfortunately reality is different and averaging kernels vary from unity. This results in the so called smoothing error which is non zero if the retrieval’s a priori CO\textsubscript{2} profile differs from the true profile. In the following, we calculate the smoothing error profile $\Delta x$ which would be introduced when using SECM ($x_{secm}$) instead of CT2010 ($x_{ct}$) as a priori profile.

$$\Delta x = (A - I) (x_{ct} - x_{secm}).$$  \hspace{1cm} (7)

Here $A$ is the diagonal column averaging kernel matrix which is defined by the retrieval’s column averaging kernel (vector). The column average smoothing error $\Delta X$, i.e. the XCO\textsubscript{2} smoothing error, can be derived by integration of Eq. (7) over all (dry-air) pressure intervals $\Delta p$:

$$\Delta X = \sum \Delta x_i \Delta p_i.$$ \hspace{1cm} (8)

Figure 5 shows typical averaging kernels of three state of the art satellite-based full physics retrievals and the TCCON FTS retrieval algorithm (Washenfelder et al., 2006; Wunch et al., 2011). The satellite retrievals are SCIAMACHY BESD (Bremen optimal estimation DOAS, Reuter et al., 2010), GOSAT RemoTeC (developed at SRON, Butz et al., 2009), and GOSAT UOL-FP (University of Leicester Full Physics algorithm, Connor et al., 2008; Bösch et al., 2011). The averaging kernels depend not only on the instrument but also on the retrieval technique. This explains the differences between the averaging kernels of GOSAT RemoTeC and GOSAT UOL-FP.

We used the averaging kernels to calculate the smoothing error $\Delta X$ which would have been introduced when using SECM instead of CT2010 as a priori profiles. For this purpose, we analyzed the 10 000 profiles of the randomly chosen data set (used before) and corresponding SECM profiles. The results are summarized in Table 2 which also shows the smoothing error introduced by a constant 380 ppm mixing ratio profile (as benchmark). Not surprisingly, the results show that it is always better to use SECM instead of a constant profile. In cases where the averaging kernel has larger variations from unity the difference is quite pronounced.
Reuter et al. (2011) estimated the single measurement precision of BESD with 2.5 ppm; they found station-to-station biases having a standard deviation of about 0.8 ppm. This means the smoothing error of 0.83 ppm resulting from constant a priori profiles is comparable to BESD’s accuracy. In contrast to this, SECM reduces the smoothing error to 0.17 ppm being distinctively lower than BESD’s accuracy and precision.

The averaging kernel of the GOSAT UOL-FP retrieval is similar to BESD’s averaging kernel. Consequently, the resulting smoothing errors are very similar (0.62 ppm and 0.15 ppm for the constant and the SECM a priori profile, respectively).

Compared to BESD and UOL-FP, the TCCON FTS retrieval and also the GOSAT RemoTeC retrievals have averaging kernels which are closer to unity. For this reason, the observed improvement by using SECM instead of a constant a priori profile is less pronounced. All corresponding smoothing error values are equal or less 0.08 ppm and, therefore, distinctively lower than the FTS instrument’s accuracy (0.4 ppm, Wunch et al., 2010) and precision (0.6 ppm, Toon et al., 2009).

However, averaging kernels change e.g. with the solar zenith angle so that the effect can be more pronounced under other viewing geometries. In all four cases (Table 2), the SECM introduced smoothing error is significantly lower than the estimated model transport error of about 0.5 ppm (Houweling et al., 2010). This becomes important when doing surface flux inverse modeling.

Note: (i) Statistically, the smoothing error is not necessarily a systematic error because the averaging kernel as well as the difference between SECM and “truth” can vary from measurement to measurement. (ii) The smoothing error becomes less important if XCO₂ retrievals are used in an inverse modeling framework because it will be removed from the retrieval within the assimilation process. However, in this case, the retrieval still profits from a well chosen first guess linearization point which typically results in better convergence behavior.
7 Conclusions

We presented a simple empirical model SECM which can be used to simulate atmospheric CO₂ background concentrations in form of mixing ratio profiles and XCO₂. We assumed that CT2010 represents our current knowledge about the global distribution of XCO₂ which can be gained (mainly) from surface-based flask measurements. Therefore, we used CT2010 to determine the free parameters of the proposed empirical model. SECM is able to reproduce CT2010 with a standard error of 0.99 ppm and a correlation of 0.97. In other words, SECM explains more than 94 % of CT2010’s variability, i.e. of our current knowledge.

The atmospheric CO₂ profiles simulated by SECM have a linear pressure dependency with different slopes in troposphere and stratosphere. The standard error profile has values between 1 ppm and 2 ppm over large parts of the atmosphere which means that SECM is able to roughly reproduce the profile shape. Larger deviations are found especially near the surface where the influence of local sources and sinks is largest. In addition to SECM estimating XCO₂ and CO₂ profiles, we proposed a simple parameterization of the error covariance matrix so that SECM can be used as a priori knowledge in an optimal estimation framework without additional external information.

We compared SECM XCO₂ not only with CT2010 but also with TCCON FTS measurements. The average standard error of 1.39 ppm agrees reasonably well with 0.99 ppm found when comparing SECM with CT2010. The standard deviation of all station-to-station biases amounts to 0.47 ppm which is consistent with TCCON’s accuracy of about 0.4 ppm.

The TCCON comparison goes one year beyond the fitting period 2003–2009. As we found no obvious problems in 2010, we conclude that SECM is also (at least to some extend) able to extrapolate into the future. In the case of extrapolating into a farer future or past, it would be advantageous to replace the linear increase of Eq. (1) by an exponential term. This, however, could require a longer fitting period to produce stable results.
We analyzed the smoothing error introduced by using SECM instead of CT2010 in order to assess the usability of SECM as a priori profiles. For this purpose, we used typical averaging kernels of four state of the art XCO\textsubscript{2} retrieval algorithms. Our analysis basically shows two things: (i) using SECM instead of constant a priori profiles reduces the smoothing error; (ii) the smoothing error due to SECM is distinctively lower than the expected retrieval error and typical model transport errors. Therefore, one can conclude that SECM is well-suited to be used as a priori information for the analyzed (or comparable) retrieval techniques. Using SECM also as first guess linearization point has furthermore the potential to enhance the convergence behavior of an iterative retrieval.

Of course, SECM cannot compete with physics-based models like CarbonTracker because it is only a coarse statistical description of the past. Under no circumstances will it be able to capture any event deviating from this statistic, i.e. it is not possible to learn anything new from SECM. However, SECM has some distinct benefits: (i) SECM is extremely simple and can be implemented with minimal effort; (ii) SECM results are easily reproducible without the need for significant disk space or computing power; (iii) SECM is always available.

Beyond the application for a priori information, SECM can be used for several other applications. SECM can be used to identify obvious retrieval errors (by monitoring the difference between retrieval and SECM). Due to its availability, SECM can be used in a near real time environment or for observing system simulation experiments especially for future satellite missions (e.g. Bovensmann et al., 2010). Its accuracy meets the requirements to be used as XCO\textsubscript{2} background in “CO\textsubscript{2} proxy” methods for XCH\textsubscript{4} retrievals (e.g. Frankenberg et al., 2005; Schneising et al., 2009).

Moreover, it is remarkable, how well (more than 94\% explained variance) a simple empirical equation (depending only on date and latitude) can reproduce atmospheric CO\textsubscript{2} concentrations.
Acknowledgements. This work was in part funded by ESA/ESRIN (GHG-CCI), EU FP7 (MACC-II), DLR (SADOS), and the State and the University of Bremen. We thank NOAA for making available the CarbonTracker CO$_2$ fields and TCCON for making available the XCO$_2$ data and averaging kernels. TCCON is funded by the NASA terrestrial carbon cycle program, grant NNX08AI86G. We further thank our reviewers (in advance) for their helpful and valuable comments to improve this work. Many thanks are given to S. Pfeifer and J. Reuter for proofreading the manuscript.

References


A simple empirical CO₂ model

M. Reuter et al.

Abstract

Introduction

Conclusions

References

Tables

Figures

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

doi:10.5194/acp-10-9981-2010, 2010. 1301


Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macanta-
gay, R., and Warneke, T.: Three years of greenhouse gas column-averaged dry air mole frac-


Table 1. Least squares fit results for parameters $a_i$ (Eqs. 1 and 2) and $c_i$ (Eq. 3).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{00}$</td>
<td>373.09 ppm</td>
<td>$a_{01}$</td>
<td>1.923 ppm a$^{-1}$</td>
</tr>
<tr>
<td>$a_{02}$</td>
<td>1.605 ppm</td>
<td>$a_{03}$</td>
<td>0.033°</td>
</tr>
<tr>
<td>$a_{04}$</td>
<td>0.623</td>
<td>$a_{05}$</td>
<td>5.470 ppm</td>
</tr>
<tr>
<td>$a_{06}$</td>
<td>0.00602°</td>
<td>$a_{07}$</td>
<td>0.179</td>
</tr>
<tr>
<td>$a_{08}$</td>
<td>0.133 ppm a$^{-1}$</td>
<td>$a_{09}$</td>
<td>0.00351°</td>
</tr>
<tr>
<td>$a_{10}$</td>
<td>−7.254 ppm</td>
<td>$a_{11}$</td>
<td>0.00166°</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.0711</td>
<td>$a_{13}$</td>
<td>0.0125 ppm a$^{-1}$</td>
</tr>
<tr>
<td>$a_{14}$</td>
<td>−0.00239°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_0$</td>
<td>25.93 ppm</td>
<td>$c_1$</td>
<td>2.259</td>
</tr>
</tbody>
</table>
Table 2. Standard smoothing error (in ppm) when using a constant 380 ppm or SECM profile instead of CT2010 based on 10,000 globally, randomly chosen CT2010 profiles in the period 2003–2009.

<table>
<thead>
<tr>
<th></th>
<th>SCIAMACHY</th>
<th>GOSAT</th>
<th>GOSAT</th>
<th>FTS</th>
<th>TCCON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BESD</td>
<td>UOL-FP</td>
<td>RemoTeC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const380</td>
<td>0.83</td>
<td>0.62</td>
<td>0.08</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>SECM</td>
<td>0.17</td>
<td>0.15</td>
<td>0.03</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. Pacific transect (−150° E, 12:00 LT – local time) for CT2010 (left panel) and SECM (right panel).
Fig. 2. Exemplary CT2010 and SECM profiles: northern hemispheric winter (left panel), northern hemispheric summer (middle panel), and tropical (right panel).
Fig. 3. Error correlation matrix (SECM vs. CT2010) of 10,000 globally, randomly chosen profiles in the period 2003–2009 (left panel), parameterized error correlation matrix (middle panel), and corresponding standard error profiles (right panel).
Fig. 4. XCO₂ time series of various TCCON sites and corresponding SECM values together with per station statistics: bias $d$ (in ppm), standard error $s$ (in ppm), and correlation coefficient $r$. There are two Lauder data sets because a new instrument have been put into operation in 2010.
Fig. 5. Typical column averaging kernels of four different XCO$_2$ retrieval systems: FTS TCCON, SCIAMACHY BESD, GOSAT RemoTeC, and GOSAT UOL-FP. The FTS TCCON column averaging kernel is typical for a solar zenith angle for 50°. The averaging kernels of SCIAMACHY BESD and GOSAT RemoTeC represent global mean averaging kernels for August 2009. The averaging kernel of GOSAT UOL-FP is the global mean averaging kernel for September 2009.