Estimating atmospheric CO_2 from advanced infrared satellite radiances within an operational 4D-Var data assimilation system: Methodology and first results

Richard J. Engelen, Erik Andersson, Frédéric Chevallier,[1] Anthony Hollingsworth, Marco Matricardi, Anthony P. McNally, Jean-Noël Thépaut, and Philip D. Watts
European Centre for Medium-Range Weather Forecasts, Reading, UK

Received 15 March 2004; revised 26 May 2004; accepted 13 July 2004; published 7 October 2004.

[1] Atmospheric CO_2 concentrations have been obtained from the Atmospheric Infrared Sounder (AIRS) radiance data within the European Centre for Medium-Range Weather Forecasts data assimilation system. A subset of channels from the AIRS instrument on board the NASA Aqua platform has been assimilated providing estimates of tropospheric and stratospheric column-average CO_2 mixing ratios. Although global estimates are obtained, the information content of the tropospheric estimates at middle and high latitudes is limited, and results are therefore only presented for the tropical region. First results for February and August 2003 show considerable geographical variability compared to the background with values ranging between 371 and 380 ppmv. These CO_2 values are representative for a layer between the tropopause and about 600 hPa. The monthly mean random error is about 1%. Careful error analysis has been carried out to minimize any systematic errors. This study has demonstrated the feasibility of global CO_2 estimation using AIRS data in a numerical weather prediction data assimilation system. In the future the system will be improved to treat CO_2 as a full three-dimensional atmospheric variable, including transport. *INDEX TERMS:* 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 0365 Atmospheric Composition and Structure: Troposphere—composition and chemistry; 1610 Global Change: Atmosphere (0315, 0325); *KEYWORDS:* AIRS, carbon dioxide, data assimilation

Citation: Engelen, R. J., E. Andersson, F. Chevallier, A. Hollingsworth, M. Matricardi, A. P. McNally, J.-N. Thépaut, and P. D. Watts (2004), Estimating atmospheric CO_2 from advanced infrared satellite radiances within an operational 4D-Var data assimilation system: Methodology and first results, *J. Geophys. Res.*, *109*, D19309, doi:10.1029/2004JD004777.

1. Introduction

[2] The importance of accurate observations of atmospheric CO_2 for top-down estimates of carbon sources and sinks has been recognized for almost a decade. In this period various synthesis inversion studies have been carried out using surface flask observations of CO_2 concentrations to estimate carbon sources and sinks at the Earth's surface [*Enting et al.*, 1995; *Fan et al.*, 1998; *Rayner et al.*, 1999; *Bousquet et al.*, 1999a, 1999b; *Kaminski et al.*, 1999; *Peylin et al.*, 2000; *Gurney et al.*, 2002]. These flask measurements are highly accurate but are limited to less than 100 sites globally. This makes the inversion problem highly data limited, especially in the tropics, where there are few surface flask stations. For some years now there has been a growing interest in the use of satellite data to improve estimates of the spatial and temporal variability of atmospheric CO_2. In particular, the development of a new generation high spectral resolution sounders that observe the atmosphere in the infrared and/or the near-infrared triggered several studies on the capabilities of these instruments to provide information on atmospheric CO_2. *Rayner and O'Brien* [2001] showed that these satellite observations have the potential to improve current CO_2 inversions, if the accuracy of their monthly mean values is better than 2.5 ppmv. *Engelen et al.* [2001a] performed a simulation study to look at the capabilities of the Atmospheric Infrared Sounder (AIRS), *Chédin et al.* [2003] did similar simulations for the Infrared Atmospheric Sounding Interferometer (IASI), and *O'Brien and Rayner* [2002] studied the near-infrared option, which might be realized by the Orbiting Carbon Observatory (OCO) mission. However, so far there is only one study that has been performed with real satellite data; *Chédin et al.* [2002] used data from the Tiros Operational Vertical Sounder (TOVS) to infer atmospheric CO_2 concentrations in the tropics. Although the capabilities of the TOVS instrument are limited with respect to CO_2 [*Engelen and Stephens*, 2004], the results of *Chédin et al.* [2002] are promising.

[3] Data from the AIRS instrument have been assimilated operationally in the European Centre for Medium-Range Weather Forecasts (ECMWF) four-dimensional variational

[1] Now at Laboratoire des Sciences du Climat et de l'Environnement, Gif sur Yvette, France.

0148-0227/04/2004JD004777$09.00

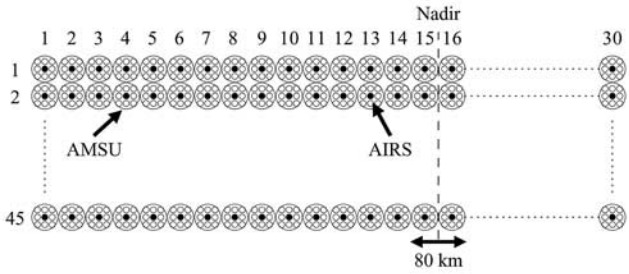

Figure 1. Field-of-view configuration of the AMSU-A and AIRS instruments on board the AQUA satellite. Solid circles are AIRS field of views currently used in the ECMWF CO_2 data assimilation.

(4D-Var) data assimilation system since October 2003. Although the main purpose of this assimilation is to improve the temperature, water vapor, and ozone fields as well as the dynamics, efforts are being made to also extract other information from the observed radiances. In particular, column CO_2 estimation has been implemented in the data assimilation system as described in this paper. While having more realistic CO_2 concentrations in the forecast model could, in principal, be beneficial for the short-term weather forecast [*Engelen et al.*, 2001b], this in itself was not the main purpose of this research. The aim is to produce global CO_2 fields from the AIRS satellite data with accuracy high enough for carbon flux inversions. It is therefore important to have an accurate characterization of the CO_2 estimates including their errors.

[4] The main benefit of estimating CO_2 mixing ratios within a numerical weather prediction (NWP) data assimilation system is that temperature and humidity, which also affect the observed infrared radiances, are well constrained by various other measurements and by the forecast model itself. For example, ECMWF assimilates radiances from three Advanced Microwave Sounding Unit A (AMSU-A) instruments, three Special Sensor Microwave Imager (SSM/I) instruments, two AMSU-B instruments, five geostationary satellites, two High Resolution Infrared Radiation Sounder (HIRS) instruments, and many radiosondes and aircraft constraining both temperature and water vapor on smaller vertical scales. These additional observational constraints on temperature and water vapor assist the extraction of atmospheric CO_2 from the AIRS observations within the NWP system.

[5] In section 2 the AIRS data are briefly discussed. Section 3 describes the setup of the CO_2 data assimilation system. Section 4 presents some first results and relevant error statistics, and section 5 concludes with a summary.

2. AIRS Data

[6] The Atmospheric Infrared Sounder (AIRS) [*Aumann et al.*, 2003] was launched on board the NASA AQUA satellite in May 2002. After an initial period of testing, data were received operationally at ECMWF from October 2002 onward.

[7] AIRS is a grating spectrometer covering the 650–2675 cm^{-1} infrared spectral domain at a resolution of $\lambda/\Delta\lambda$ = 1200, giving 2378 channels. Accompanied by an AMSU-A instrument it flies onboard the Aqua satellite with equator crossing times of 1:30 am and 1:30 pm. The AIRS field of view (FOV) is 13.5 km at nadir with a 3 × 3 array of AIRS footprints falling into one AMSU-A FOV (see Figure 1).

[8] Because of bandwidth limits in the trans-Atlantic line and other operational constraints, ECMWF receives only 324 of the total 2378 channels in near-real time and only 1 out of every 9 AIRS FOVs within a AMSU-A FOV (solid black circles in Figure 1). The channel selection is based on an original selection of 281 channels by NOAA/NESDIS appended with 43 extra channels in the two main CO_2 absorption bands based on the work by *Crevoisier et al.* [2003]. In this study the number of channels was further reduced to avoid problems specific to certain spectral bands: (1) channels in the short-wave band were excluded from the analysis because our radiative transfer model currently does not model solar radiation and the effects of nonlocal thermodynamic equilibrium; (2) channels in the main water vapor and ozone bands as well as channels sensitive to the surface are excluded to minimize the effect of water vapor, ozone, and the surface on the CO_2 analysis; and (3) channels sensitive to the upper stratosphere were also excluded from the assimilation because the ECMWF model has large temperature biases in the mesosphere of the polar winters. This resulted in a set of 55 channels in the long-wave CO_2 absorption band, of which about 31 channels are sensitive to tropospheric CO_2. The actual number used in the assimilation depends on the cloud detection as described in section 3.

3. Setup of the Data Assimilation System

3.1. Four-Dimensional Variational Assimilation (4D-Var)

[9] A 4D-Var data assimilation system is a practical formulation of Bayesian estimation theory for the particular case of a (near) linear problem with unbiased Gaussian errors [*Lorenc*, 1986]. It seeks a model trajectory that is statistically consistent with the information provided by the observations $\mathbf{y}^o$ available for the analysis time window $[t_0, t_n]$ and the information provided by an a priori model state $\mathbf{x}^b$ called the background state. This background state is usually taken from a short-range forecast. The model trajectory itself is completely defined by the initial state $\mathbf{x}_0$ at time t_0 through the use of the dynamical and physical forecast model.

[10] The analysis correction ($\delta\mathbf{x}(t_0)$) to the model initial state is sought as a combination of the information from the observations and the background using an objective cost function with two terms [e.g., *Courtier et al.*, 1994]:

$$J(\delta\mathbf{x}(t_0)) = \frac{1}{2}\delta\mathbf{x}(t_0)^T\mathbf{B}^{-1}\delta\mathbf{x}(t_0) + \frac{1}{2}\sum_{i=0}^{n}[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i]^T\mathbf{R}^{-1}[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i] \quad (1)$$

the background term and the observation term. The observation departures ($\mathbf{d}_i$) are the differences between the observed radiances and the model simulated radiances

$$\mathbf{d}_i = \mathbf{y}_i^o - H_i[\mathbf{x}^b(t_i)] \quad (2)$$

where H_i is the full nonlinear observation operator in the form of the Radiative Transfer for the TIROS Operational Vertical Sounder (RTTOV) radiative transfer model. RTTOV [*Matricardi et al.*, 2004] is a fast radiative transfer model using profile-dependent predictors to parameterize the atmospheric optical depths. For the CO_2 assimilation experiments we applied the methods developed for RTIASI [*Matricardi*, 2003] to include CO_2 as a profile variable in RTTOV. $\mathbf{H}_i$, which appears in equation (1), is the tangent linear observation operator that is part of the RTTOV model. The background values at time t_i, needed for the calculation of the observation departures $\mathbf{d}_i$, are evolved according to the full nonlinear forecast model $\mathcal{M}$:

$$\mathbf{x}^b(t_i) = \mathcal{M}\left[\mathbf{x}^b(t_0)\right] \tag{3}$$

The increments themselves are evolved through time according to the tangent linear model $\mathbf{M}$:

$$\delta\mathbf{x}(t_i) = \mathbf{M}_i\delta\mathbf{x}(t_0) \tag{4}$$

Finally, $\mathbf{B}$ and $\mathbf{R}$ are the background error covariance matrix and the observation error covariance matrix, respectively.

[11] The cost function is then minimized with respect to the increments of the initial state ($\delta\mathbf{x}(t_0)$). These increments are added to the background state to obtain the analysis $\mathbf{x}(t_0)$:

$$\mathbf{x}(t_0) = \mathbf{x}^b + \delta\mathbf{x}(t_0) \tag{5}$$

[12] The advantage of a full data assimilation system is that it seeks to combine all available observations in an (near) optimal way. At ECMWF ground based and satellite based data are used to constrain the relevant fields in the forecast model. In addition to AIRS, satellite data from the HIRS, AMSU-A, AMSU-B, SSM/I, Geostationary Operational Environmental Satellite (GOES), and Meteosat instruments are assimilated. These data are thinned to reduce spatial correlations of the measurement errors and they also undergo a bias correction. This bias correction seeks to remove biases in the observations and radiative transfer modeling and depends for most instruments on air mass and viewing angle [*Harris and Kelly*, 2001]. For most satellite instruments these biases are calculated from the mean differences between the analyses and the satellite observations close to radiosonde locations in order to minimize the effect on the bias correction of forecast model bias. This approach was chosen because the analysis is likely to be the most accurate close to the radiosonde observations. For AIRS a global mean bias correction is used for each individual channel, because the air mass dependency is relatively small. In order to avoid biases dependent on the viewing angle of the instrument, we currently only use the central 24 field of views (FOVs) out of the total of 30 FOVs.

[13] The 4D-Var data assimilation system currently uses clear radiance data only. A detection algorithm was developed specifically for AIRS [*McNally and Watts*, 2003]. This scheme detects which AIRS channels are affected by clouds and removes those channels from the assimilation, while keeping the channels that are not affected by clouds. This allows use of AIRS data even where the view is cloudy. If there is high cloud, only stratospheric information will be assimilated, but, if there are low clouds only, a significant amount of tropospheric information can be used. Finally, some channels were removed from the analysis either because of instrumental problems or because of unaccounted errors in the observation operator. Main example of the latter is the removal of the short-wave 4.2 μm band, which is affected by solar radiation not modelled in the current version of RTTOV.

3.2. CO_2 as a Column Variable

[14] Within the 4D-Var system, CO_2 is currently estimated as an independent column variable. This means that CO_2 is not a tracer variable in the transport model and is estimated only at the observation locations. Background error correlations between CO_2 and the other assimilation variables are neglected. In practice, this means that while the forecast model variables (e.g., temperature and water vapor) appear in the control vector ($\delta\mathbf{x}(t_0)$) as three-dimensional fields, CO_2 appears as a vector of column variables at the observation locations. The link between the initial state and the states at observation locations and times, as represented by equations (3) and (4), is absent for CO_2. Another important difference between the CO_2 variable and the regular analysis variables (temperature, water vapor, etc.) is that the background field of the regular variables is based on a forecast that uses the previous analysis as its initial state. For CO_2, each analysis uses a climatological background state of CO_2. CO_2 information is thus not carried over from one analysis to the next.

[15] This methodology was used as a first step to implement CO_2 in the data assimilation system making use of the current operational architecture. It makes full use of the accurate temperature and water vapor analysis fields constrained by all available observations, but there are also some limitations. First, the individual CO_2 estimates are not constrained by the model transport during the 12-hour assimilation window. During this 12-hour time span the model transport is usually accurate and can help to advect information from one place to another. Second, by assimilating column CO_2 values instead of full profiles a hard constraint is applied to the analysis in the form of a fixed profile shape. This removes some of the flexibility in the adjustments and can lead to errors if the used profile shape is far from the truth. This hard constraint also means that all vertical levels are fully correlated and any adjustments in the stratosphere will therefore also adjust the troposphere. In case of many stratospheric radiance channels and only few tropospheric radiance channels this leads to a dominant stratospheric signal in the estimated CO_2 column value.

[16] On the basis of first results (not shown here) indicating that the column variable was indeed dominated by the large amount of stratospheric AIRS channels, the column variable was split into a tropospheric column and a stratospheric column. These two columns act as independent variables without any error correlation in the analysis. The tropopause height that separates the two columns is estimated from the background temperature profile using an algorithm based on lapse rates, and varies with location. This ensured that the tropospheric analysis results were not dominated by the stratosphere. However, any potentially

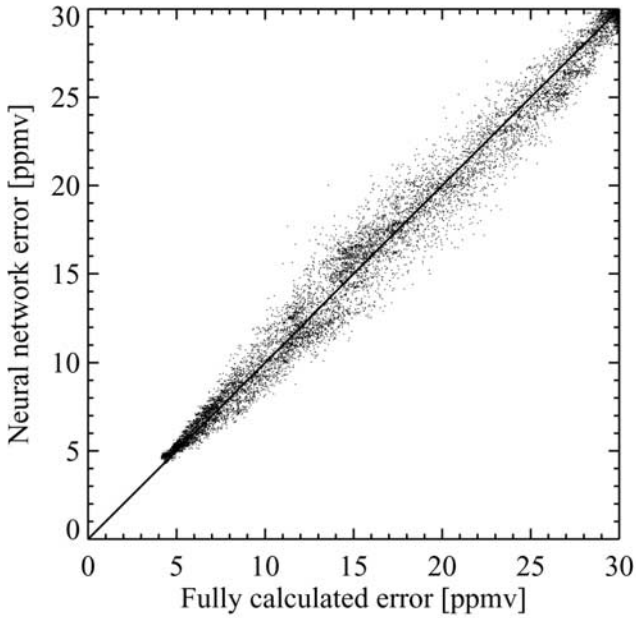

Figure 2. Scatter diagram of the estimated analysis error using an artificial neural network versus the analysis error estimated from Bayesian theory.

useful correlations between stratospheric CO_2 and tropospheric CO_2 are disregarded. A drawback is that the tropospheric column is quite variable in the vertical. Depending on the tropopause height and the cloud top height, the column varies from shallow to deep allowing respectively less or more channels to be used in the tropospheric analysis. As shown in section 3.3, the number of channels used in the analysis is an important determining factor for the analysis error.

3.3. Analysis Error Estimation

[17] It is crucial to have an estimate of the individual analysis errors to properly interpret the analysis results. Initially, the analysis error (σ_a) within the 4D-Var system was calculated from the background error (σ_b), the observation error covariance (**R**), and the CO_2 Jacobians (**H**):

$$\sigma_a^2 = \left[\sigma_b^{-2} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \tag{6}$$

The errors in the temperature, water vapor, and ozone profiles that enter the radiative transfer equation are taken into account by inflating the observation error covariance (**R**) based on the sensitivity of the radiative transfer to perturbations defined by the respective background covariance matrices. Although this is a simplification of the real error model, it is sufficient for our purposes. Besides, the CO_2 analysis itself is part of a multivariate minimization problem that takes into account all relevant error sources.

[18] Within a variational assimilation system, equation (6) requires separate calculations of the Jacobians **H** for each observation, which amounts to significant extra computer time. Because the analysis error is in first approximation a function of the tropospheric temperature lapse rate and the number of assimilated AIRS channels peaking in the troposphere (determined by the tropopause height and the cloud top height), a nonlinear regression (artificial neural network [e.g., *Rumelhart et al.*, 1986]) was used to relate the analysis error to these two variables. Analysis errors based on equation (6) were calculated for data from 1 to 7 March 2003, on which the neural network was trained. The network was then tested with data from August, as shown in Figure 2. The performance of the neural network is remarkably good considering the little information given to the network.

4. Results

[19] Some first results of the CO_2 data assimilation scheme are presented here to illustrate the capabilities of the system. The tropospheric background values used in the assimilation were zonal mean monthly averaged mixing ratios based on surface flask observations from the previous year [*GLOBALVIEW-CO_2*, 2003]. These averaged flask observations are based on maritime air samples, and a constant value of 2 ppmv was added to compensate for the annual trend. For the stratospheric background a constant value of 375 ppmv was used. The background error was set to 30 ppmv and was deliberately taken large to minimize the contribution of the background to the analysis in these preliminary experiments. Individual analysis values at the observation locations were gridded onto a $1^\circ \times 1^\circ$ latitude-longitude grid for a whole month. Within a grid box the data were averaged using a weighted average with the analysis errors as weights. This $1^\circ \times 1^\circ$ grid was then smoothed with a $15^\circ \times 15^\circ$ moving boxcar average. The same boxcar smoothing was used by *Chédin et al.* [2003] and is applied here to allow easier comparison with their results. Each individual grid box needed to have more than 10 observations within a month to be included in the smoothing averaging. Therefore some geographical areas have no data in the final monthly mean fields because of consistent high cloud cover.

4.1. Quality of Analysis

[20] The information content of the CO_2 estimates is highly variable due to variations in temperature lapse rate and cloudiness (see also section 3.3), and therefore the contribution of the background to the analysis estimate varies as well. The value of the analysis error relative to the background error shows how much information is gained from the observations and can be formally represented by the averaging kernel [*Rodgers*, 2000], which is for a single scalar analysis variable defined as

$$A = 1 - \frac{\sigma_a^2}{\sigma_b^2} \tag{7}$$

The averaging kernel varies between 0 and 1, where 0 means that we retrieve the background value back in the analysis, while 1 means that we have an analysis independent of the used background. With a constant background error (as currently used), there is no fundamental difference between the averaging kernel and the

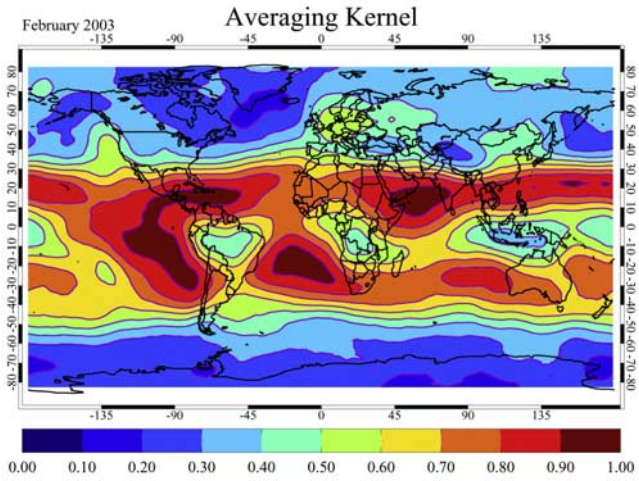

Figure 3. Averaging kernel averaged for February 2003 on a $1^\circ \times 1^\circ$ grid.

analysis error itself, but when the background error varies geographically, the averaging kernel is much easier to interpret. Also, the absolute values of the averaging kernel largely depend on the value of the background error. We will therefore only use the relative values of the averaging kernel, because our assumption of a very large background error causes the averaging kernel to be unrealistically close to one. Figure 3 shows the mean averaging kernel for February 2003. It is immediately clear that the information content of the analysis is highest in the tropics, but degrades quickly at higher latitudes. This is caused by the shallower troposphere (lower tropopause) and the smaller temperature lapse rate at higher latitudes. Also, tropical convective areas have a much smaller mean averaging kernel value due to high cloud top levels. However, in these cloudy tropical areas there are still many occasions where the satellite instrument sees clear areas or areas with low clouds, even in the convective regions. Therefore it is still possible to calculate monthly mean CO_2 concentrations for most of these areas. On the basis of Figure 3 we will show only analysis results for the region between 30°S and 30°N. CO_2 is estimated outside this region, but the results depend significantly on the assumed background values.

4.2. Monthly Mean CO_2 Distribution

[21] The CO_2 analysis results are shown in Figure 4 for February 2003 and in Figure 5 for August 2003. Figures 4 (left) and 5 (left) show the background values and Figures 4 (right) and 5 (right) show the actual analysis results. The background field is not entirely zonal, because the individual observations are not homogeneously distributed over the averaging grid and each individual background value was interpolated in latitude from the GLOBALVIEW zonal means.

[22] Both Figures 4 and 5 show that the analysis adds structure to the zonal background field. Although the main north-south gradient remains, meridional variability is produced by the analysis. In the equatorial region the analysis tends to have more CO_2 in the convective areas, especially in the west Pacific. Another feature can be observed over the southern part of North America. A careful analysis was done using AMSU-A data to see if these features were caused by biases in the temperature analysis. This seems indeed to be the case for the high values over southern North America in February, where a cold bias is observed in the temperature analysis field compared to AMSU-A measurements. This could cause a positive bias in the CO_2 field. However, for the other regions such a cold analysis bias is not present. Also, plots of AIRS first-guess departures (the difference between the observed brightness temperatures and the model simulated brightness temperatures from the 6-hour forecast) that drive the analysis show the same patterns as the CO_2 analysis field. These patterns are very dissimilar from the AMSU-A first-guess departures and can therefore not be explained completely by errors in the temperature forecast. To further illustrate the point, we have plotted in Figure 6 the monthly mean CO_2 increments (analysis minus background) as a function of the observation departures for AIRS channels 193 and 213 (both sensitive to midtropospheric CO_2), AMSU-A channel 7

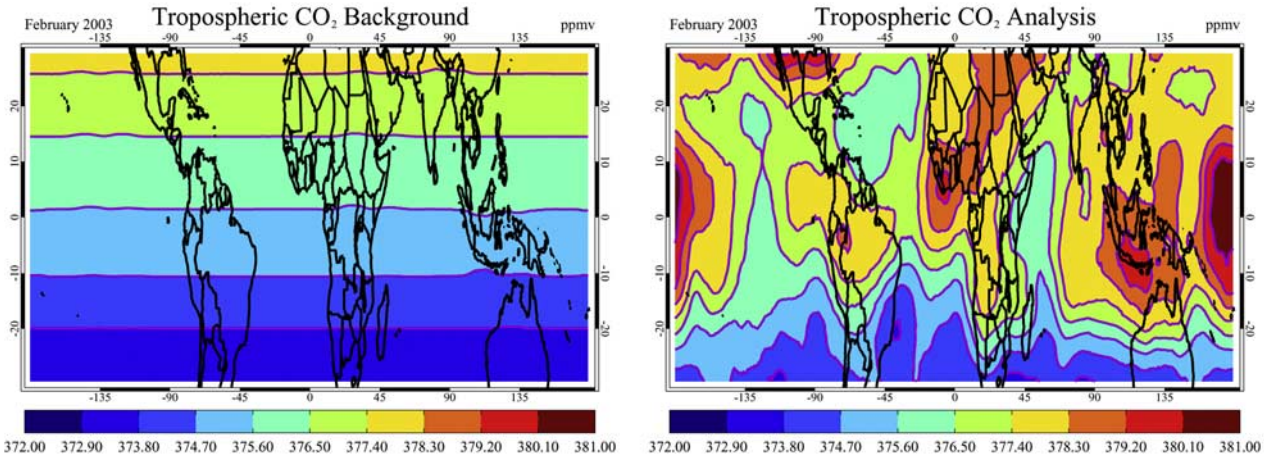

Figure 4. (left) Background and (right) analysis of CO_2 distribution averaged for February 2003.

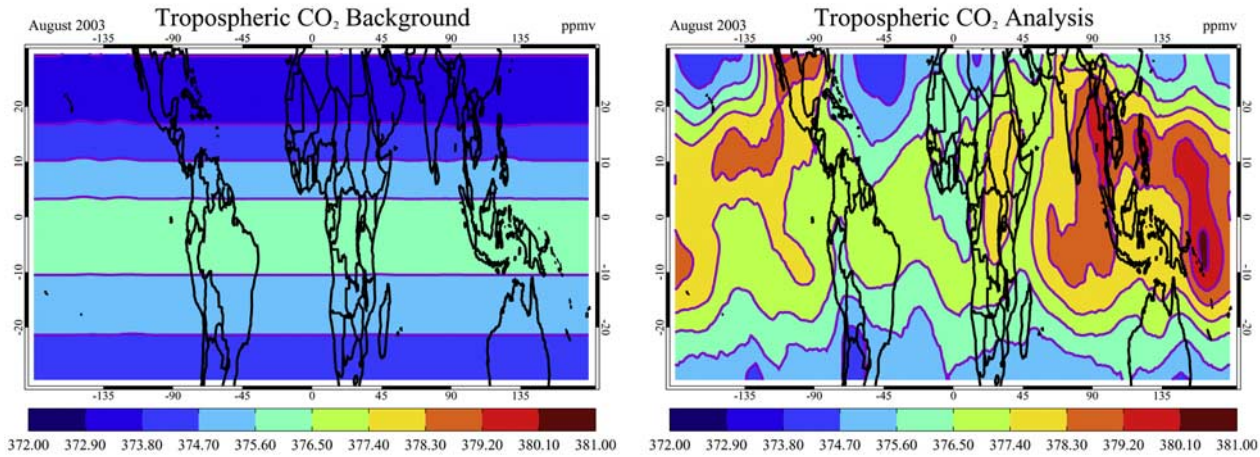

Figure 5. (left) Background and (right) analysis of CO_2 distribution averaged for August 2003.

(sensitive to midtropospheric temperature), and AMSU-B channel 4 (sensitive to tropospheric water vapor). Any biases in the model temperature and water vapor fields that are aliased in the CO_2 results would show up as correlations in the two AMSU plots. While there is a significant correlation in the two AIRS plots, such correlations are not shown in the AMSU plots. Although these results are not conclusive, they indicate that the effect of

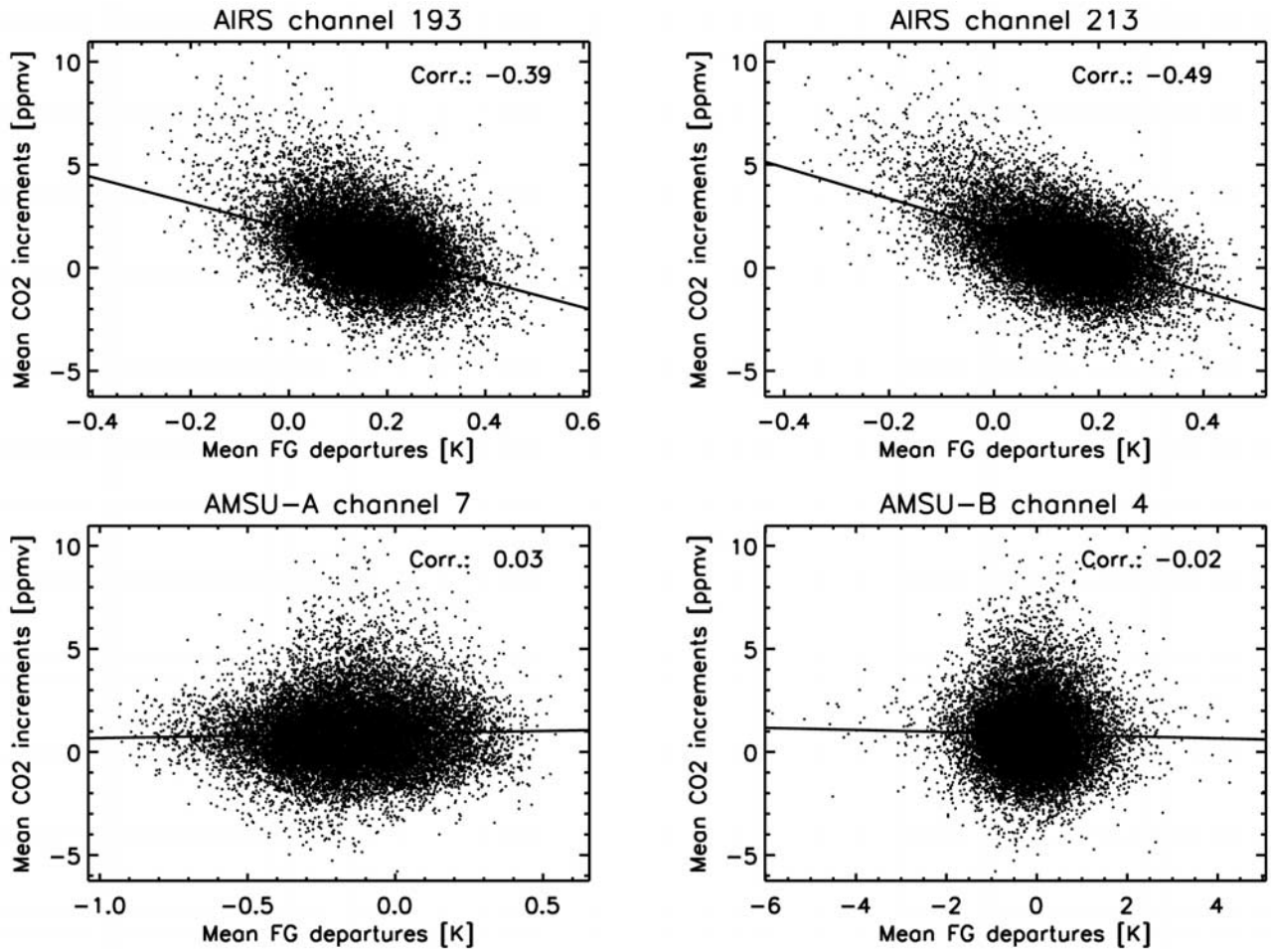

Figure 6. Monthly mean CO_2 increments as a function of observation departures for AIRS channels 193 and 213, AMSU-A channel 7, and AMSU-B channel 4.

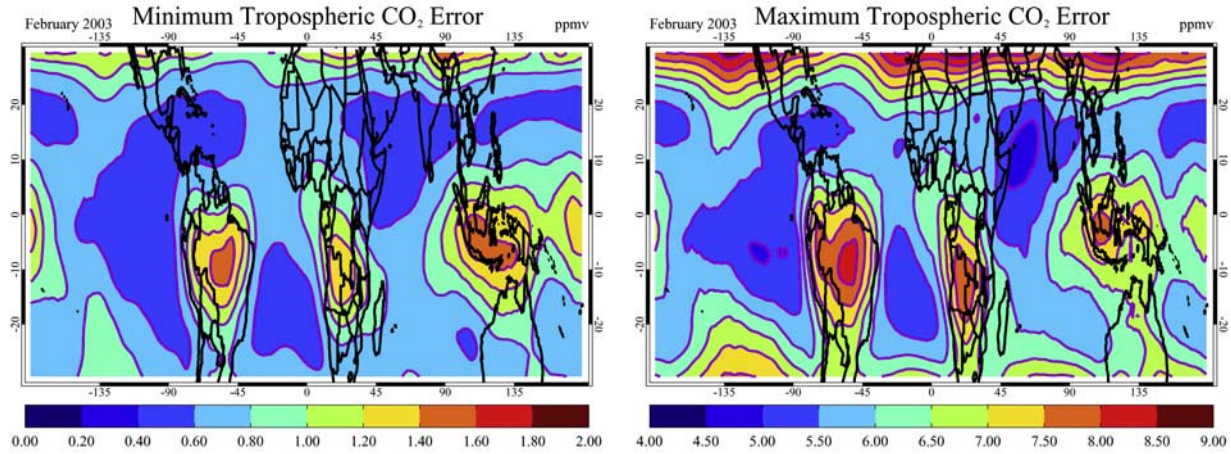

Figure 7. (left) Lower and (right) upper estimates of the error in the monthly averaged CO_2 distribution for February 2003.

model temperature and water vapor biases on the CO_2 analysis is not large.

[23] The higher CO_2 values on the west side of Africa in February could be explained by biomass burning effects. Similar patterns in the MOPITT carbon monoxide observations can be observed over that area in February 2003 (see http://www.eos.ucar.edu/mopitt/data/index.html). The high values in the western Pacific are probably more surprising. One explanation could be that anthropogenic emissions from southeast Asia are lifted up and transported to the western Pacific by the general circulation. During this part of the year there is a circular wind pattern in the middle troposphere bringing air east from the southeast Asian coast and then south to the middle of the Pacific. However, more careful analysis of the results should be carried out before drawing firm conclusions. For example, clouds are detected in our cloud detection scheme within a small error margin. Therefore it is in principle possible to have a systematic error in the lower channels due to undetected clouds resulting in a CO_2 bias of a few parts per million by volume. Also, air-mass-dependent errors in the radiative transfer (e.g., the spectroscopy) could cause systematic errors in the CO_2 analysis results on regional scales.

[24] Comparison with the results of *Chédin et al.* [2003] shows both similarities and dissimilarities. They present results from their HIRS CO_2 retrievals for March and August for the years 1987–1991. Their results for March 1990 show high CO_2 values over northern Africa and over the Pacific, similar to our February results. For August, they find high values over India and Indonesia like we do. However, their maximum spatial gradients (12 ppmv) are larger than the maximum gradients in our results (7 ppmv). It has to be kept in mind, though, that the results are 13 years apart, so differences have to be carefully interpreted.

4.3. Error Estimate

[25] To provide an indication of the error in the monthly mean CO_2 distribution, Figure 7 shows the individual analysis errors averaged under different assumptions. The minimum error ($\overline{\sigma}_{\min}$) is calculated assuming that the errors of all individual estimates are completely uncorrelated using

$$\overline{\sigma}_{\min} = \left(\sum_i^N \frac{1}{\sigma_i^2} \right)^{-1/2} \tag{8}$$

which implies a $\sqrt{N}$ reduction of the individual errors (σ_i), where N is the number of observations within a grid box. The maximum error ($\overline{\sigma}_{\max}$) is calculated assuming that the errors of all individual estimates are correlated using

$$\overline{\sigma}_{\max} = \frac{1}{N} \sum_i^N \sigma_i \tag{9}$$

Both error estimates clearly depend on the number of observations in the monthly average with cloudy areas having larger errors. Neither error estimate includes systematic errors, but their range gives a reasonable estimate of the average random error. For the tropical area the expected monthly average analysis error is therefore between 1 and 6 ppmv, which is on the order of 1%.

4.4. Tropospheric Layer Definition

[26] As described by *McNally and Watts* [2003], the cloud detection algorithm searches for channels that are unaffected by clouds within the error margin of the observations and radiative transfer modeling. Within the algorithm the channels are ranked in vertical space by assigning them a "trip" level that represents the height of an opaque cloud needed to affect the specific channel by more than 1%. The 1% number is somewhat arbitrary, but is only used to rank the channels; it does not represent any error threshold in the remaining channel radiances. After the cloud detection the lowest trip level of the remaining channels then approximates the lowest level of the observation sensitivity. If there is a cloud, this corresponds to the cloud top height; if there is no cloud, it corresponds to the lowest level of the clear column

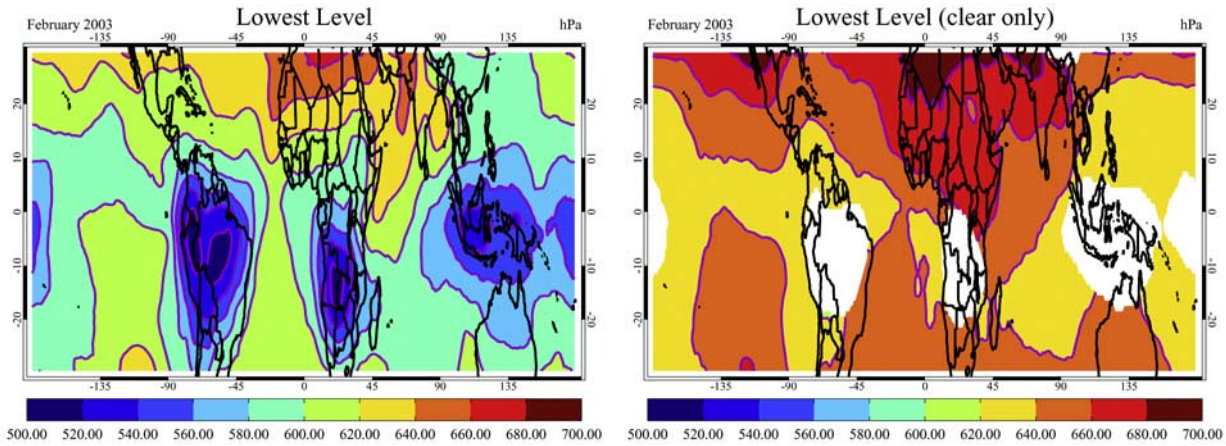

Figure 8. Lowest level of CO_2 sensitivity (left) using all data and (right) using cloud-free FOVs only for February 2003.

where the observation is still sensitive to CO_2. The actual error of undetected clouds in the observed brightness temperatures is of the order of 0.2 K as shown by *McNally and Watts* [2003].

[27] The thickness of the layer between this trip level and the tropopause determines the part of the troposphere that is represented by the measurement. Figure 8 shows the monthly mean trip level for February 2003 (Figure 8, left) as well as the monthly mean trip level when only cloud-free FOVs are considered (Figure 8, right). Although it is possible to have an estimate of CO_2 in the convective areas, the layer represented by this estimate is much shallower than in areas with low clouds or no clouds at all. This could in principle make results harder to interpret. When only cloud-free FOVs are used, the thickness of the column if more uniform as shown in Figure 8 (right). For comparison, Figure 9 shows the CO_2 analysis results for February 2003 using clear FOVs only. The CO_2 values are slightly lower than the values shown in Figure 5, especially over the Indonesian area, but the geographical patterns remain very similar. However, if the thickness of the representative layer as well as the estimated analysis error are taken into account, using all available data has the advantage of better spatial and temporal cover. This could be a significant benefit in surface flux inversion studies.

5. Summary

[28] Global estimates of CO_2 concentrations have been obtained from AIRS radiance data. A subset of channels from the AIRS instrument on board the NASA Aqua platform has been assimilated in the ECMWF data assimilation system providing estimates of tropospheric and stratospheric CO_2 mixing ratios. Currently, CO_2 is not included as a tracer in the transport model, but treated as a column variable estimated at the time and location of each AIRS observation entering the system. This setup has enabled first CO_2 assimilation experiments, but has the disadvantage that it lacks the transport constraint and the adjustment flexibility in the vertical. The analysis errors have been estimated using a artificial neural network that relates the CO_2 analysis error to the number of assimilated channels sensitive to tropospheric CO_2 and the tropospheric temperature lapse rate based on earlier simulations that estimated the analysis error using Bayesian theory.

[29] First results for February and August 2003 are presented showing considerable geographical variability compared to the background. Various quality checks were carried out to exclude as many potential error sources as possible. Careful analysis is needed to guarantee the validity of results considering the small CO_2 signal compared to these various error sources. Although the results presented in this paper have been quality checked, more error analysis and validation will be carried out in the near future.

[30] The current results show CO_2 values ranging between 373 ppmv and 380 ppmv in the tropics with an estimated error of about 3 ppmv for the monthly average. These values are representative for a layer between the tropopause and about 650 hPa if only observations from cloud-free field of views are used. The lower boundary of the representative layer varies between 500 and 700 hPa, when also observations are used where some channels have been removed by the cloud detection algorithm.

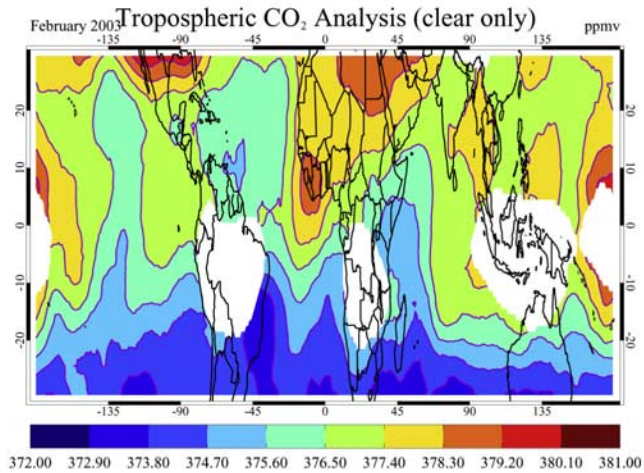

Figure 9. CO_2 analysis distribution averaged for February 2003 using clear field of views only.

[31] This study has demonstrated the feasibility of global CO_2 estimation using AIRS data in an NWP data assimilation system. In the future the system will be improved to treat CO_2 as a full three-dimensional atmospheric variable, including transport.

[32] **Acknowledgments.** The work described in this paper was financially supported by the EC project COCO (EVG1-CT-2001-00056). The authors wish to thank two anonymous reviewers for their very useful comments.

References

Aumann, H. H., et al. (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, *41*, 253–264.

Bousquet, P., P. Ciais, P. Peylin, M. Ramonet, and P. Monfray (1999a), Inverse modeling of annual atmospheric CO_2 sources and sinks: 1. Method and control inversion, *J. Geophys. Res.*, *104*, 26,161–26,178.

Bousquet, P., P. Ciais, P. Peylin, M. Ramonet, and P. Monfray (1999b), Inverse modeling of annual atmospheric CO_2 sources and sinks: 2. Sensitivity study, *J. Geophys. Res.*, *104*, 26,179–26,194.

Chédin, A., A. Hollingsworth, N. A. Scott, S. Serrar, C. Crevoisier, and R. Armante (2002), Annual and seasonal variations of atmospheric CO_2, N_2O and CO concentrations retrieved from NOAA/TOVS satellite observations, *Geophys. Res. Lett.*, *29*(8), 1269, doi:10.1029/2001GL014082.

Chédin, A., R. Saunders, A. Hollingsworth, N. Scott, M. Matricardi, J. Etcheto, C. Clerbaux, R. Armante, and C. Crevoisier (2003), The feasibility of monitoring CO_2 from high-resolution infrared sounders, *J. Geophys. Res.*, *108*(D2), 4064, doi:10.1029/2001JD001443.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth (2004), A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, *120*, 1367–1387.

Crevoisier, C., A. Chedin, and N. A. Scott (2003), AIRS channel selection for CO_2 and other trace-gas retrievals, *Q. J. R. Meteorol. Soc.*, *129*, 2719–2740.

Engelen, R. J., and G. L. Stephens (2004), Information content of infrared satellite sounding measurements with respect to CO_2, *J. Appl. Meteorol.*, *43*, 373–378.

Engelen, R. J., A. S. Denning, K. R. Gurney, and G. L. Stephens (2001a), Global observations of the carbon budget: 1. Expected satellite capabilities for emission spectroscopy in the EOS and NPOESS eras, *J. Geophys. Res.*, *106*, 20,055–20,068.

Engelen, R. J., G. L. Stephens, and A. S. Denning (2001b), The effect of CO_2 variability on the retrieval of atmospheric temperatures, *Geophys. Res. Lett.*, *28*, 3259–3262.

Enting, I. G., C. M. Trudinger, and R. J. Francey (1995), A synthesis inversion of the concentration and $d^{13}C$ of atmospheric CO_2, *Tellus*, *Ser. B*, *47*, 35–52.

Fan, S.-M., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi, and P. Tans (1998), A large terrestrial carbon sink in North America implied by atmospheric data and oceanic carbon dioxide data and models, *Science*, *282*, 442–446.

GLOBALVIEW-CO_2 (2003), Cooperative Atmospheric Data Integration Project–Carbon Dioxide [CD-ROM], NOAA Clim. Monit. and Diagn. Lab., Boulder, Colo. (Also available via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW)

Gurney, K. R., et al. (2002), Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, *Nature*, *415*, 626–630.

Harris, B. A., and G. Kelly (2001), A satellite radiance-bias correction scheme for data assimilation, *Q. J. R. Meteorol. Soc.*, *127*, 1453–1468.

Kaminski, T., M. Heimann, and R. Giering (1999), A coarse grid three-dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO_2 in the 1980s, *J. Geophys. Res.*, *104*, 18,555–18,581.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, 1177–1194.

Matricardi, M. (2003), RTIASI-4, a new version of the ECMWF fast radiative transfer model for the infrared atmospheric sounding interferometer, *Tech. Memo. 425*, Eur. Cent. for Medium -Range Weather Forecasts, Reading, U. K.

Matricardi, M., F. Chevallier, G. Kelly, and J.-N. Thépaut (2004), An improved general fast radiative transfer model for the assimilation of radiance observations, *Q. J. R. Meteorol. Soc.*, *130*, 153–173, doi:10.1256/qj.02.181.

McNally, A. P., and P. D. Watts (2003), A cloud detection algorithm for high-spectral-resolution infrared sounders, *Q. J. R. Meteorol. Soc.*, *129*, 3411–3423, doi:10.1256/qj.02.208.

O'Brien, D. M., and P. J. Rayner (2002), Global observations of the carbon budget: 2. CO_2 column from differential absorption of reflected sunlight in the 1.61 μm band of CO_2, *J. Geophys. Res.*, *107*(D18), 4354, doi:10.1029/2001JD000617.

Peylin, P., P. Bousquet, P. Ciais, and P. Monfray (2000), Differences in CO_2 flux estimates based on a time-independent versus a time-dependent inversion method, in *Inverse Methods in Global Biogeochemical Cycles*, *Geophys. Monogr. Ser.*, vol. 114, edited by P. Kasibhatla et al., pp. 811–841, AGU, Washington, D. C.

Rayner, P., and D. O'Brien (2001), The utility of remotely sensed CO_2 concentration data in surface source inversions, *Geophys. Res. Lett.*, *28*, 175–178.

Rayner, P., I. Enting, R. Francey, and R. Langenfelds (1999), Reconstructing the recent carbon cycle from atmospheric CO_2, d13C, and O_2/N_2 observations, *Tellus*, *Ser. B*, *51*, 213–232.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding. Theory and Practice*, World Sci., River Edge, N. J.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986), Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Macrostructure of Cognition 1*, edited by D. E. Rumelhart and McClelland, pp. 318–362, MIT Press, Cambridge, Mass.

E. Andersson, R. J. Engelen, A. Hollingsworth, M. Matricardi, A. P. McNally, J.-N. Thépaut, and P. D. Watts, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK. (richard.engelen@ecmwf.int)

F. Chevallier, Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France.