A global evaluation of the regional spatial variability of column integrated CO$_2$ distributions

Alanood A. Alkhaled, Anna M. Michalak, S. Randolph Kawa, Seth C. Olsen, and Jih-Wang Wang

Received 9 December 2007; revised 18 April 2008; accepted 10 July 2008; published 17 October 2008.

Satellites, such as the Orbiting Carbon Observatory (OCO), are expected to provide global measurements of column-averaged carbon dioxide (CO$_2$) dry-air mole fraction ($X_{CO2}$) with the potential of improving the scientific understanding of regional carbon cycle processes and budgets. The satellite data products, however, are expected to have large data gaps due to the satellite track and geophysical limitations (e.g., clouds and aerosols). The satellite data will also be representative of the $X_{CO2}$ distribution at the spatial scale of satellite footprints, which is smaller than the resolution of typical transport or process models. Assessing the ability of the retrieved soundings to capture $X_{CO2}$ variability over different regions and times, evaluating the representation error associated with using the retrieved $X_{CO2}$ product to represent $X_{CO2}$ at typical model resolutions, and filling data gaps while providing an estimate of the associated uncertainty all require the evaluation of the spatial variability of $X_{CO2}$. In this study, the global spatial covariance structure of $X_{CO2}$ is evaluated regionally using CO$_2$ concentrations simulated using the MATCH/CASA model. Results show that regional and temporal changes in the $X_{CO2}$ distribution caused by seasonal changes in surface fluxes and transport produce a spatially and temporally variable $X_{CO2}$ covariance structure. The effects of model setup and the relatively low resolution of the MATCH/CASA model on the evaluated $X_{CO2}$ covariance structure are assessed by comparing the MATCH/CASA results to the spatial variability inferred from the higher-resolution PCTM/GEOS-4 global model, the SiB-RAMS regional model, and aircraft campaign point observations. The comparison with the higher-resolution models and aircraft data shows good agreement with MATCH/CASA results, thus indicating that the presented results provide an adequate representation of $X_{CO2}$ variability as will be measured by satellites such as OCO.


1. Introduction

[2] Predicting future changes in the Earth’s climate due to increasing carbon dioxide (CO$_2$) concentrations depends on understanding and accurately modeling the physical processes controlling the carbon cycle. Achieving these objectives requires the availability of dense, globally distributed measurements of atmospheric CO$_2$ concentrations, which could be used to infer CO$_2$ sources and sinks at regional scales. The existing CO$_2$ measurement network [GLOBALVIEW-CO2, 2005], however, is too sparse to achieve this goal. A main component of the efforts to overcome this limitation is the launch of satellite missions, such as the Orbiting Carbon Observatory (OCO) [Crisp et al., 2004; Miller et al., 2007] and the Greenhouse Gases Observations Satellite (GOSAT) [National Institute for Environmental Studies, 2006] to provide global coverage of high-density measurements of column-averaged CO$_2$ dry-air mole fraction ($X_{CO2}$) at regional precisions of 1 ppm [Crisp et al., 2004; Chevallier et al., 2007].

[3] Satellite measurements, however, are expected to have large data gaps, and to be representative of spatial scales that are much smaller than the spatial resolution of typical global models used in the estimation of CO$_2$ sources and sinks. Evaluating the information content, uncertainties, and errors associated with gap-filled satellite data products requires a quantification of the spatial and temporal scales of variability of $X_{CO2}$, and an analysis of how this variability itself varies between regions and across
seas only 5% of the total number of soundings. Understanding of the spatial variability of XCO2, which quantifies XCO2 variability, is necessary for evaluating the ability of the retrieved fraction of soundings to capture the underlying XCO2 distribution. The quantification of global variability at regional scales will also facilitate the use of optimal spatial interpolators (e.g., kriging) that not only provide gap-filled global XCO2 maps, but also provide a measure of gap-filling uncertainty. These maps are important for validating transport models and can play a key role in the validation of estimated CO2 sources and sinks.

Similarly, knowledge of XCO2 regional spatial variability would be necessary for identifying satellite soundings that provide the best characterization of the underlying XCO2 distribution, by focusing more resources on areas with higher variability. Such sounding selection may be required in early stages of the OCO mission, to manage the high computational costs associated with processing the expected massive data volumes.

The evaluation of representation errors associated with using satellite measurements to represent the XCO2 distribution at coarser resolution also requires the evaluation of XCO2 variability at regional scales. These errors result from the mismatch in spatial scale between the satellite measurement resolution (e.g., OCO soundings will have an approximately 3 km² footprint) and the resolution of typical atmospheric transport models or general circulation models (i.e., 100–1000 km) used in determining CO2 sources and sinks [Miller et al., 2007; Corbin et al., 2008]. Estimates of the regional spatial variability of XCO2 can be used to statistically quantify representation errors [Alkhaled et al., 2008]. For example, over areas with low spatial variability, retrieved soundings will be more representative of the average XCO2 over a model grid cell, and will therefore have lower representation errors relative to areas with high spatial variability.

The overall objective of this paper is to analyze the current understanding of the spatial variability of XCO2, as simulated by current models, at global and regional scales. The presented analysis quantifies monthly XCO2 variability using spatial covariance parameters that represent the global and regional spatial variability of simulated XCO2 fields. Preliminary results presented by Miller et al. [2007] established the need for such analysis. This paper quantifies the spatial variability of XCO2, and attributes observed seasonal and regional differences to variability in surface fluxes and seasonal changes in global transport. Results of this study provide an understanding of the expected information content of soundings retrieved by future satellites.

The paper is organized as follows. Section 3 describes the examined models and analysis methods. The results of the spatial variability analysis are presented in section 4. On the global scale, section 4.1 presents daily spatial variability parameters to investigate the overall seasonality of XCO2 variability. Following the global-scale analysis, monthly regional-scale spatial covariance parameters are quantified in section 4.2 to identify the local variability over different regions. The effects of using low-resolution XCO2 simulations to infer XCO2 spatial variability are evaluated in section 4.3 through comparison with higher-resolution models and aircraft data. A summary of the main conclusions of the study are presented in section 5.

2. Background

Contrary to the relatively extensive literature analyzing the variability of surface CO2 concentrations [Conway et al., 1994; Randerson et al., 1997; Geels et al., 2004; Nicholls et al., 2004; Lu et al., 2005; Wang et al., 2007], the spatial and temporal variability of XCO2 is the focus of only a limited number of studies. In general, these studies have focused on two main areas: (1) determining the type of information that XCO2 measurements can provide relative to surface concentrations and, therefore, their value to carbon cycle research and (2) evaluating the ability of retrievals from existing satellites to capture the spatial and temporal variability of XCO2, by comparing satellite retrievals to model simulations and/or surface-based Fourier Transform Infrared Spectroscopy (FTS) measurements.

Studies focusing on the first area show that XCO2 has lower spatial and temporal variability, and delayed response to surface disturbances, relative to surface concentrations, with delays reaching several weeks [Olsen and Randerson, 2004; Warneke et al., 2005]. More specifically, these studies show that column-averaged volume mixing ratios reflect the spatial and temporal variability of surface CO2 concentrations diluted by less variable concentrations beyond the Planetary Boundary Layer (PBL) [Kawa et al., 2004; Olsen and Randerson, 2004]. The low variability of column-averaged volume mixing ratios is caused by the vertical and horizontal mixing of CO2 concentrations throughout the column, which smoothes surface flux signals, thus leading to high precision requirements for XCO2 if they are to be useful in carbon cycle studies.

The utility of XCO2 to carbon cycle science has been demonstrated by a number of studies. Rayner and O’Brien [2001] noted that the variability characteristics of XCO2 over high-convection tropical regions can be useful for determining CO2 fluxes because the rapid vertical mixing reduces the spatial smearing of surface fluxes. Moreover, a few studies have pointed to the role of global transport in determining XCO2 variability in the middle to upper troposphere [Tiwari et al., 2006] and the lower troposphere, particularly in the Southern Hemisphere (SH) pole region [Nevison et al., 2008]. The understanding of the influence of global transport on XCO2 and the large spatial footprint of fluxes influencing local XCO2 variability, make XCO2 an important quantity for identifying the relative contribution of oceanic versus terrestrial fluxes to CO2 variability over the SH [Nevison et al., 2008].

Studies focusing on the second area evaluate the ability of satellites (e.g., AIRS, SCIAMACHY) to capture the spatial and temporal variability of XCO2. These studies show that retrieved soundings that are uncontaminated by
aerosol or clouds are capturing the general spatial patterns of \( X_{CO2} \) over examined regions [Barkley et al., 2006a, 2006b]. Twari et al. [2006] noted that there is good agreement in the amplitude of the middle to high tropospheric \( CO2 \) subcolumn observed by AIRS and predicted by models. However, there are some differences in the phase of the observed and modeled seasonal cycle. Literature analyzing SCIAMACHY retrievals shows that the monthly means of \( X_{CO2} \) retrievals (averaged over some spatial domain) have a large spread, and that the amplitude of their seasonal cycle over the Northern Hemisphere (NH) is lower than that observed using FTS measurements [Barkley et al., 2007; Buchwitz et al., 2007; Schneising et al., 2008], but higher than that represented by atmospheric models [Barkley et al., 2006b; Bösch et al., 2006]. Although reasons for the weak \( X_{CO2} \) seasonal cycle in atmospheric models were not identified in studies comparing models to satellite retrievals [Barkley et al., 2006b; Bösch et al., 2006], a number of studies attributed this underestimation to modeling uncertainties in the specifications of surface fluxes, errors in mixing parameterization, unrealistic stratospheric influence on simulated mixing ratios, and differences in meteorology [Shia et al., 2006; Stephens et al., 2007; Washenfelder et al., 2006; Yang et al., 2007].

[13] The spatial variability of \( X_{CO2} \) has been quantified using the spatial autocorrelation of aircraft measurements of partial columns of \( X_{CO2} \) [Gerbig et al., 2003a; Lin et al., 2004a], although the analysis covered only a small number of regions with limited spatial and temporal coverage. This limited knowledge about the spatial variability of \( X_{CO2} \) cannot be used to determine the information content of the global \( X_{CO2} \) measurements that will be provided by OCO, which may vary both regionally and seasonally. Similarly, the current network of high-resolution solar absorption spectrometers (FTS) (Total Carbon Column Observing Network (TCCON), http://www.tccon.caltech.edu) has only a few stations globally. Currently available satellite remote sensing data, on the other hand, have relatively low precision and large data gaps, particularly on daily time scales [Buchwitz et al., 2005; Barkley et al., 2006b; Twari et al., 2006]. These data do not provide the coverage or spatial density needed for a seasonal variability analysis on a global scale.

[14] Therefore, prior to the launch of the OCO and GOSAT satellites, the global \( X_{CO2} \) spatial variability must be estimated using simulated \( X_{CO2} \). This study provides an analysis of both the global and regional monthly \( X_{CO2} \) variability as simulated by current models.

### 3. Methods

[15] This section introduces the models used to simulate the analyzed \( X_{CO2} \) data, as well as the approach used to quantify spatial variability. Section 3.1 describes the MATCH/CASA model used for the main variability analysis. Methods used to quantify the spatial variability of \( X_{CO2} \) are presented in section 3.2. Finally, section 3.3 introduces models and aircraft data used to (1) validate results obtained using the MATCH/CASA model and (2) test the robustness of the inferred \( X_{CO2} \) variability to differences in model setup and resolution.

### 3.1. MATCH/CASA Model

[16] The analysis of the spatial variability of \( X_{CO2} \) is performed using data from the MATCH/CASA coupled biosphere transport model. The analyzed MATCH simulation has a 2-hourly temporal resolution (averaged from 30 min time steps) and a horizontal resolution of approximately \( 5.5^\circ \times 5.5^\circ \) with 26 vertical layers starting at the surface and ending at 60 km in altitude [Olsen and Randerson, 2004]. The \( CO2 \) concentrations at the different model altitude layers are pressure-averaged to obtain column-integrated \( CO2 \) concentrations (\( X_{CO2} \)). MATCH uses archived meteorological fields derived from the NCAR Community Climate Model version 3 that are representative of a climatologically average year. For the data used in this study, MATCH simulates atmospheric \( CO2 \) resulting from three types of fluxes: (1) linearly interpolated monthly means of oceanic fluxes derived from \( pCO2 \) measurements [Takahashi et al., 1999], (2) anthropogenic emissions from Andres et al. [1996] uniformly distributed throughout the year, and (3) diurnally varying net ecosystem production (NEP) fluxes based on monthly net primary production (NPP) values from the CASA model [Randerson et al., 1997] distributed diurnally according to shortwave radiation and temperature from the National Center for Environmental Prediction (NCEP) for the year 2000.

[17] \( X_{CO2} \) data corresponding to 1300 local time are selected to approximate the spatial variability as would be observed by OCO [Crisp et al., 2004; Miller et al., 2007]. The regional spatial variability of \( X_{CO2} \) is evaluated for the 15th of each month, which is assumed to represent typical variability that would be observed during an individual day in each month. Note that we are interested in the variability as will be observed by OCO, which is why the analysis is performed on a series of individual days, rather than on monthly averaged \( X_{CO2} \). The representativeness of the 15th of each month is verified by quantifying the daily spatial variability at 1300 local time for eight consecutive days during the month with the highest observed global-scale change in variability, as will be further described in section 4.2.3.

### 3.2. Spatial Variability

#### 3.2.1. Semivariogram Model

[18] The spatial variability of \( X_{CO2} \) is quantified by modeling the semivariogram of the \( X_{CO2} \) distribution, which describes the degree to which two \( X_{CO2} \) values are expected to differ as a function of their separation distance \( (h) \). To evaluate this relationship, the raw semivariogram \( \gamma(h) \) is evaluated for all pairs of \( X_{CO2} \) data:

\[
\gamma(h) = \frac{1}{2} \left[ X_{CO2}(x_i) - X_{CO2}(x_j) \right]^2
\]

where the distance \( (h) \) between locations \( x_i \) and \( x_j \) is the great circle distance between these points on the surface of the Earth:

\[
h(x_i, x_j) = r \cos^{-1} \left( \sin \phi_i \sin \phi_j + \cos \phi_i \cos \phi_j \cos(\delta_i - \delta_j) \right)
\]
and where \((\phi_i, \theta_i)\) are the latitude and longitude of location \(x_i\), and \(r\) is the Earth’s mean radius. The semivariogram is used to model the spatial autocorrelation of \(X_{CO2}\) that is not explained by a deterministic trend in the data. Therefore, the \(X_{CO2}\) north–south gradient is estimated for each month using linear regression and subtracted from the data prior to the analysis.

[19] A theoretical variogram model is selected on the basis of the observed variability to represent the spatial autocorrelation structure. The theoretical variogram describes the decay in spatial correlation between pairs of \(X_{CO2}\) measurements as a function of physical separation distance between these measurements. The exponential semivariogram [e.g., Cressie, 1993] is selected here to model MATCH/CASA \(X_{CO2}\) spatial variability, based on an examination of a binned version of the raw variogram. The exponential variogram is defined as:

\[
    \gamma(h) = \sigma^2 \left(1 - \exp \left(-\frac{h}{L}\right)\right) \tag{3}
\]

where \(\sigma^2\) represents the expected variance of the difference between \(X_{CO2}\) measurements at large separation distances, and \(3L\) represents the practical correlation range between \(X_{CO2}\) measurements. These parameters also define the corresponding exponential covariance function:

\[
    C(h) = \sigma^2 \exp(-h/L).
\]

[20] The exponential model parameters are fitted to the raw semivariogram of the latitudinally detrended \(X_{CO2}\) data using nonlinear least squares. The fitted variogram parameters define the spatial covariance structure of the modeled \(X_{CO2}\) signal. The uncertainty of the least squares fit of the variance (\(\sigma^2\)) and range parameter (\(L\)) are not reported in this study because the results are based on an exhaustive sample from the simulated field, and the uncertainty resulting from limited sampling is negligible. The majority of the uncertainty associated with variogram parameters stems from assumptions about fluxes and transport, and the sensitivity to these choices is explored in sections 3.3 and 4.3.

### 3.2.2. Spatial Variability Analysis

[21] The global spatial variability is defined through semivariogram parameters fitted to the raw semivariogram. For each day, the raw semivariogram is constructed using detrended MATCH/CASA \(X_{CO2}\) at 1300 local time for all model grid cells. The analysis is repeated for each day of the model year to identify both the seasonal trends in global variability at daily resolution, and the relationships between these trends and seasonal changes in global \(CO2\) flux and transport.

[22] Regional variability in the spatial covariance structure is evaluated through localized variograms representing subareas of the global domain. This analysis requires areas (regions) large enough to capture the scales of variability within a given subdomain of the model, while at the same time small enough to reveal the characteristics of local spatial variability.

[23] A regional variability analysis with a similar methodological goal was previously adopted by Doney et al. [2003] to measure the mesoscale global spatial variability of satellite measurements of ocean color. In that study, daily anomalies from the monthly block mean of the natural log of chlorophyll concentrations were used to fit spherical variograms for nonoverlapping 5° regions globally.

[24] In the case of \(X_{CO2}\), regional covariance parameters were fit for each model grid cell, resulting in a regional spatial variability analysis at a 5.5° resolution. Because regional spatial variability may reflect global general circulation patterns as well as differences in surface fluxes between regions, correlation lengths of \(X_{CO2}\) may extend beyond individual continents or ocean basins. To account for this, the local semivariogram parameters in the current work are constructed to reflect both the local variability and its relationship to global spatial variability. First, regions are defined as overlapping 2000 km radius circles centered at each model grid cell, resulting in a total of 2048 regions covering the globe. A 2000 km radius was selected because it is sufficiently large to capture much of the variability in the vicinity of a given grid cell, while being small enough to capture regional variability in the spatial covariance structure. Second, the raw semivariogram (\(\gamma_{\text{region}}(h)\)) is constructed using pairs of points with one point always within the defined region (X_{CO2}(x_{\text{region}})) and the other either within or outside that region (X_{CO2}(x_{\text{region}} + h)) (see Figure 1). This approach focuses on the variability observed within each subregion, while also accounting for larger scales of variability:

\[
    \gamma_{\text{region}}(h) = \frac{1}{2} \left[ X_{CO2}(x_{\text{region}}) - X_{CO2}(x_{\text{region}} + h) \right]^2 \tag{4}
\]

Third, to emphasize the covariance of \(X_{CO2}\) within the analyzed region, weighted nonlinear least squares is used to fit the local semivariogram parameters, with higher weights assigned to points within a separation distance less than or equal to 4000 km. Numerically, correlation lengths are also restricted to a maximum of half the Earth’s circumference.

[25] Conceptually, a higher variance is representative of more overall variability, as is a shorter correlation length, which is indicative of more variability at smaller scales. The parameter \(h_o\) is introduced to provide a single representation of the degree of variability observed in different regions, and to merge information about both the variance and correlation lengths of \(X_{CO2}\) variability. If we consider a single sounding at a known location, \(h_o\) is defined as the maximum distance from the sounding location at which the mean squared \(X_{CO2}\) prediction error is below a preset value, \(V_{\text{max}}\). The mean squared prediction error is the uncertainty associated with using the sounding to predict the unknown value at a given distance away from the sounding location, using ordinary kriging. Ordinary kriging is a minimum variance unbiased interpolator that takes advantage of knowledge of the spatial covariance structure to interpolate available measurements while providing an estimate of the interpolation error [Chiles and Delfiner, 1999]. For an exponential variogram:

\[
    h_{o,\text{exp}} = -L_R \ln \left(1 - \frac{V_{\text{max}}}{2\sigma_R^2}\right) \tag{5}
\]

where \(\sigma_R^2\) and \(L_R\) are the fitted regional variance and range parameter, respectively.

[26] Both a higher regional variance \(\sigma_R^2\) and a shorter regional range parameter \(L_R\) lead to a decrease in the overall
spatial scale over which a given measurement is representative of the surrounding $X_{\text{CO}_2}$ values. It should be noted that no measurement error is assumed in the calculation of the regional variance $\sigma_R^2$ and range parameter $L_R$. Therefore, the resulting $h_o$ values demonstrate the overall spatial scale of the information provided by a noise-free $X_{\text{CO}_2}$ measurement over the measurement region and time.

[27] In subsequent sections of this work, variability inferred from the MATCH/CASA model is compared to other models and field data, where different theoretical variogram models are used to represent $X_{\text{CO}_2}$ spatial variability. Because parameters used to describe the variability differ between variogram models, the $h_o$ parameter also provides a convenient universal metric that can be compared across models. The equivalent $h_o$ parameters for the other variogram models used in this study are presented in subsequent sections.

[26] Conceptually, the $h_o$ parameter can also be thought of as a measure of the expected relative spatial density of retrieved soundings that would be required to capture the spatial variability of $X_{\text{CO}_2}$ over different regions. The choice of $V_{\text{max}}$ is somewhat flexible, but should represent a level of interpolation uncertainty that is relevant to potential applications of the data. In the presented results, $V_{\text{max}}$ is chosen to be 0.25 ppm$^2$ ($\sqrt{V_{\text{max}}} = 0.5$ ppm). This level is comparable to the 1 ppm regional-scale uncertainty described as a goal for OCO [Chevallier et al., 2007]. It should be noted that $V_{\text{max}}$ represents the interpolation uncertainty assuming no measurement error. Thus, the lower variance was chosen to compensate for the additional uncertainty that would be contributed by measurement errors and other sources of error.

3.3. Comparison to Other Models and Aircraft Data

[26] The regional spatial covariance parameters inferred from the MATCH/CASA model are compared to higher-resolution models and aircraft data. The objective of this comparison is to assess the effect of model setup and resolution on inferred spatial variability, particularly at spatial scales comparable to the measurement footprint of future satellites, and in comparison to actual $X_{\text{CO}_2}$ variability as observed by aircraft profiles. The regional covariance structure inferred from MATCH/CASA is compared to the covariance structures predicted by the Parameterized Chemistry and Transport global Model PCTM/GEOS-4 [Kawa et al., 2004], a high-resolution regional model for North America (SiB-RAMS) [Wang et al., 2007], and $X_{\text{CO}_2}$ aircraft data over North America and the Pacific Ocean [Lin et al., 2004a].

3.3.1. PCTM/GEOS-4 Global Model

[30] Regional spatial variability is evaluated using $X_{\text{CO}_2}$ simulated using the PCTM/GEOS-4 model [Kawa et al., 2004] for the months of January, April, July and October of 2002. The objective of this comparison is to assess the impact of differences in model resolution, model winds, and transport on the observed spatial variability of modeled $X_{\text{CO}_2}$.

[31] PCTM/GEOS-4 has a 2° latitude by 2.5° longitude resolution, and is driven by analyzed meteorological fields from NASA’s Goddard Earth Observing System, version 4 (GEOS-4). The GEOS-4 fields are derived from meteorological data assimilation for 2002. The model run was spun up prior to 1998 and continued forward. Surface fluxes used in PCTM/GEOS-4 are very similar to those used in MATCH/CASA. The fossil fuel and ocean fluxes are those assembled for the TransCom-Continuous model intercomparison [Law et al., 2008], which are very similar to those used in MATCH (section 3.1). The terrestrial biosphere fluxes used in PCTM were also derived from CASA monthly means (section 3.1), with 3-hourly variations imposed using the same method of Olsen and Randerson [2004], but for PCTM the 3-hourly variations were created using the GEOS-4 meteorological data rather than NCEP. In a separate comparison, $X_{\text{CO}_2}$ results from two PCTM simulations using CASA monthly fluxes for the same year with 3-hourly variations created using GEOS-4 versus ECMWF (from TransCom-Continuous [Law et al., 2008]) were found to be very similar, suggesting that NCEP-driven results would produce the same variability observed using the model simulations used in this study.

[32] $X_{\text{CO}_2}$ output nearest 1300 local solar time is selected from hourly CTM fields. The similarity in surface fluxes
between MATCH/CASA and PCTM/GEOS-4 provides a basis for assessing differences in spatial variability resulting primarily from differences in model winds, transport, and resolution, with little influence from changes caused by different assumptions about fluxes.

3.3.2. SiB-RAMS Regional Model

[33] SiB-RAMS is a fully coupled biosphere-atmosphere regional model that predicts CO$_2$ spatial and temporal variations by simulating CO$_2$ sources and sinks [Denning et al., 2003]. The SiB-RAMS data used in this study are part of a 10-day simulation of a weather front event passing over North America from 11 to 21 August 2001 [Wang et al., 2007]. CO$_2$ concentrations are simulated at three nested grids with approximately 40 km, 10 km and 2 km spatial resolutions. The largest grid (40 km) consists of 150 × 100 cells covering a longitudinal range from 144.3$^\circ$W to 51.4$^\circ$W and a latitudinal range from 21.9$^\circ$N to 61.9$^\circ$N. The smaller grids are centered at the WLEF tower located in Chequamegon National Forest east of Park Falls-Wisconsin, and cover areas of 1500 km × 1500 km and 400 km × 400 km, respectively [see Wang et al., 2007, Figure 2]. The analysis was limited to the coarsest grid, because a preliminary spatial variability analysis showed that the limited area covered by the finer grids is not sufficient to characterize X$_{CO2}$ correlation lengths.

[34] The high resolution of the SiB-RAMS model provides an opportunity for comparing the X$_{CO2}$ spatial variability inferred from the coarse MATCH/CASA model [Olsen and Randerson, 2004] to the spatial variability at spatial scales closer to OCO footprint. Additionally, the SiB-RAMS simulation did not use prescribed CO$_2$ fluxes; instead the fluxes were derived from local meteorological conditions, thus providing a different representation of CO$_2$ flux variability.

[35] The SiB-RAMS simulation includes CO$_2$ concentrations at 44 vertical levels from approximately 30 m up to 21 km. X$_{CO2}$ is again obtained using pressure weighted averaging of all levels of CO$_2$ concentrations at 1300 local time. In addition to the spatial covariance analysis of the 21 km SiB-RAMS columns, another column value is constructed by giving the final SiB-RAMS vertical level (21 km) a weight that is proportional to an elevation increment from 21 km to 60 km to be comparable to the vertical extent of MATCH/CASA X$_{CO2}$. Furthermore, because of the limited latitudinal range of the SiB-RAMS simulation, the latitudinal trend is not significant and is therefore not removed.

[36] A daily raw semivariogram is calculated using X$_{CO2}$ over the entire SiB-RAMS domain. The semivariogram model parameters are then fitted to the raw semivariogram using nonlinear least squares. A Gaussian theoretical semivariogram model [e.g., Cressie, 1993] was found to provide the best fit to the experimental semivariogram of SiB-RAMS,

\[ \gamma(h) = \sigma^2 \left(1 - \exp\left(-\frac{h^2}{L^2}\right)\right) \]  

(6)

For the Gaussian variogram model, the practical correlation length is equivalent to 7/4 times the range parameter L. The $h_o$ parameter corresponding to this variogram model is defined as:

\[ h_{o, Gauss} = \sqrt{-L^2 \ln \left(1 - \frac{V_{max}}{2\sigma^2} \right)} \]  

(7)

3.3.3. Aircraft Data

[37] Lin et al. [2004a] evaluated the spatial variability of partial columns of atmospheric CO$_2$ concentrations using aircraft data from the CO$_2$ Budget and Rectification Airborne mission (COBRA) project over North America (NA) [Gerbig et al., 2003a, 2003b; Lin et al., 2004b] and from the NASA Global Tropospheric Experiment (GTE) mission for the Pacific Ocean [McNeal, 1983]. The results of this study are used as a comparison to the variability inferred using the models described in the previous sections.

[38] The Pacific Ocean measurements represent time periods that extend from August to October and from February to April of the years 1991 to 2001. CO$_2$ concentrations over North America, on the other hand, were collected only in August 2000 and June 2003. Lin et al. [2004a] used density-weighted CO$_2$ concentrations for four altitude ranges (0.15 – 3 km, 3 – 6 km, 6 – 9 km and <9 km). For North America, the top height of the first range (0.15 – 3 km) was not fixed at 3 km but varied according to the PBL height, which was determined from the characteristics of the measured CO$_2$ profile and auxiliary measurements, such as the vertical potential temperature, H$_2$O, and CO vertical profiles [Gerbig et al., 2003a].

[39] Lin et al. [2004a] used a power variogram model [e.g., Cressie, 1993] to represent the spatial covariance structure of the data. The power variogram had two parameters ($c_1$ and $\lambda$) to represent the growth in variance as a function of separation distance, as well as a parameter ($c_o$) to represent the measurement error:

\[ \gamma(h) = c_o + \left(\frac{h}{c_1}\right)^{\lambda} \]  

(8)

To reduce the effect of temporal variability, particularly for NA profiles that were collected during both morning and afternoon flights, Gerbig et al. [2003a] and Lin et al. [2004a] constructed the raw semivariograms using X$_{CO2}$ pairs sampled within a 3-h window of each other.

[40] The $h_o$ parameter corresponding to this variogram model is defined as:

\[ h_{o, power} = c_1 (V_{max} - c_o)^{\frac{1}{\lambda}} \]  

(9)

and is evaluated for the 0.15 – 3 km and <9 km column heights.

4. Results and Discussion

4.1. Global X$_{CO2}$ Variability

[41] Global spatial covariance parameters inferred from the MATCH/CASA simulation show strong seasonal variability (Figure 2) that can be interpreted given known seasonal changes in CO$_2$ fluxes and atmospheric transport.
and showed that despite the large seasonal cycle as Warneke et al., 2004; seasonal variability residuals (i.e., latitudinal of the seasonal variability estimated by previous 2005), Global spatial covariance function parameters in winter and spring to approximately 2.25 ppm CO₂ latitudinal of the SH; thus, with its at 1300 local time evaluated daily and smoothed Olsen and Randerson CO₂ Washenfelder et al. ALKHALED ET AL.: REGIONAL X sea-

Figure 2. Global spatial covariance function parameters (variance and correlation length) of MATCH/CASA simulated XCO₂ at 1300 local time evaluated daily and smoothed using a 1 week moving average.

For example, the effect of the NH growing season is demonstrated by the rapid increase in the global variance parameter starting at levels averaging between 0.3 ppm² and 0.75 ppm² in winter and spring to approximately 2.25 ppm² in July. The variance parameter then decreases gradually in August and rapidly in September and October, and reaches its average winter levels in November.

The seasonal cycle of the variance is coupled with a similar cycle for the correlation length that follows with some lag. The correlation length cycle starts with short values during the NH winter, the values then increase during the NH spring and fall, and reach a maximum in July and September. The correlation length starts to decrease again in October to reach a minimum in November. Features of the correlation length cycle indicate that this parameter is affected by changes not only in surface fluxes, but also in seasonal transport. A clear indication of this effect is demonstrated by the sharp drop in the correlation length values in June and November. Although these drops can be related to biospheric flux changes in the NH due to the onset of NH summer and fall, changes in transport pathways and the gradual variance changes that occur around the same period indicate a possible role of transport. More specifically, this sharp decrease in both months can be attributed to seasonal changes in the location of the Intertropical Convergence Zone (ITCZ) and the associated seasonal changes in transport, particularly of the Asian and European transport pathways [Stohl et al., 2002].

The ITCZ separates the high-variability XCO₂ of the NH from the lower-variability XCO₂ of the SH; thus, with its northward movement during the NH summer, a larger part of the Earth shows low variability. Furthermore, the movement of the ITCZ affects the seasonal transport of the XCO₂ signal from high-flux areas over south Asia, and to a lower degree over Africa. In winter, a large portion of the south Asian emissions are transported toward the ITCZ, while another portion is transported along with European emis-
sions to the Arctic [Stohl, 2004]. These seasonal changes in transport affect the shape of the XCO₂ latitudinal gradient, particularly during transition months (i.e., October–November and June–July), thus affecting the seasonal cycles of both the variance and correlation length.

Other factors affecting the shape of XCO₂ latitudinal gradient are the strong CO₂ drawdown over the NH biospherically active areas during the NH summer, coupled with the relatively high XCO₂ values around the ITCZ regions. These factors create a nonlinear XCO₂ gradient with a maximum either south or north of the equator depending on the location of the ITCZ, and a minimum that starts over the high latitudes in June and moves toward the midlatitudes in October. The resulting nonlinearity in the latitudinal gradient creates XCO₂ residuals (i.e., latitudinally detrended XCO₂) with strong spatial continuity, evident from the variances and long correlation lengths observed in June through September.

The seasonal variability of the global covariance structure, as shown in Figure 2, indicates that as expected, the NH summer months exhibit the most spatial variability. As a result, representation errors are generally expected to be higher for these months. Furthermore, discounting any differences in geophysical limitations, the uncertainty associated with gap-filled products will be higher for the summer months, and the spatial variability will be more difficult to capture. Unexpectedly, however, high variability is also observed during a NH winter month such as November, stemming from short correlation lengths, which implies that variability may be more difficult to capture during this month as well.

Previous studies assessed XCO₂ seasonal variability by calculating global, zonal or point peak-to-peak seasonal amplitudes, found to be between 6 ppm and 11 ppm at NH mid/high latitudes [Olsen and Randerson, 2004; Washenfelder et al., 2006]. Warneke et al. [2005], Bösch et al. [2006] and Washenfelder et al. [2006] indicated that models can produce a NH seasonal cycle similar to the seasonal cycle observed by FTS measurements, but with lower amplitudes. Bösch et al. [2006] also showed that there is a good qualitative agreement between the XCO₂ seasonal cycle observed in model simulations, FTS measurements, and SCIAMACHY retrievals. More recently, Schneising et al. [2008] showed that despite the large scatter of SCIAMACHY retrievals, and taking retrieval biases into consideration, monthly averages over reasonably large spatial domains showed good agreement with the amplitude and phase of the NH XCO₂ seasonal cycle as captured by FTS instruments.

Although the global covariance parameters presented in this section clearly show the seasonal cycle of XCO₂, XCO₂ seasonal variability estimated by previous studies cannot be compared directly to the results presented in this section. Unlike the seasonal peak-to-peak amplitude, the global variance parameters presented here are a measure of the spatial variability expected for a given day or month, not temporal variability between months.
4.2. Regional $X_{CO2}$ Variability

[48] The goal of the regional analysis of the spatial structure of $X_{CO2}$ is to reveal patterns of local variability that will be observed by future satellites, and to relate these patterns to differences in the strength of surface fluxes and to seasonal differences in global transport characteristics.

4.2.1. Regional Variance

[49] The regional variance parameters (Figure 3a) show both spatial and temporal variability. The temporal variability is demonstrated by the seasonal change in the magnitude of the regional variances, which follows a seasonal cycle similar to that of the global variance parameters, but with a larger range of spatial variability within each month.

[50] The spatial patterns detected by the regional variances include large areas with relatively low variances, representative of background levels of $X_{CO2}$ variability. These variances are approximately 0.5 ppm during the NH winter and spring (November–May), and reach a maximum of 2.5 ppm during the NH summer (June–October).

[51] Areas with higher variance occur over regions with highly variable surface fluxes (e.g., a maximum of 11 ppm over boreal forests in July), and over regions affected by seasonal changes in atmospheric transport (e.g., Arctic and northern ocean during the NH winter). A less prominent role of transport is also apparent during the NH summer when the relatively high variances over the northern tropical Atlantic and North Pacific oceans are caused by flux variability in other regions. In this case, the northern tropical Atlantic is most probably affected by CO$_2$ variability originating in tropical Africa and southern Europe, and the North Pacific is influenced by CO$_2$ variability from the southern parts of North America and mid-American continent, which, in both cases, is consistent with the transport pathways established by Stohl et al. [2002].

[52] Overall, $X_{CO2}$ regional variance parameters inferred from MATCH/CASA exhibit variability between regions and reflect seasonal fluctuations in regional fluxes and transport. This is clearly shown in the collocation of high variances with, or downwind from, biospherically and anthropogenically active regions. These results support the conclusion that the information scale of retrieved soundings will vary both geographically and seasonally. While some soundings will reflect $X_{CO2}$ over a large region, others will be representative of local variability.

4.2.2. Regional Correlation Lengths

[53] In contrast to the regional variances, the seasonal variability of regional correlation lengths does not show the same seasonal cycle as the global parameters. Figure 3b shows the prevalence of shorter $X_{CO2}$ correlation lengths in the NH for all months (between 200 km and 3000 km on average), relative to the SH. This difference is caused by the limited mixing between the hemispheres and the high spatial variability of the terrestrial fluxes in the NH. The location of the large contrast in correlation lengths reflects the seasonal movement of the ITZC and is most prominent in June, July and November. The long correlation lengths in the SH are somewhat shorter in August and September, most likely because of the terrestrial tropical fluxes in Africa and South America that introduce spatial variability and cause a drop in correlation lengths over the tropics and the tropical/South Atlantic.

[54] Although regional correlation lengths largely reflect the variability of surface fluxes, their spatial patterns are not as distinct as those of the regional variances. Correlation lengths do seem to reflect large transport-related effects, however. This is apparent in the relatively short correlation lengths in the Polar Regions in general, and in particular the SH. Olsen and Randerson [2004] and Nevison et al. [2008] point to the potential role of poleward transport of CO$_2$-enriched or depleted air in elevating the variability in the South Pole region. This is shown in the presented results (Figure 3b) by the decrease in correlation lengths around the South Pole starting in December, becoming shortest in January, and then gradually increasing through April. This reduction corresponds to an increase in the variability of tropical fluxes, thus reinforcing the conclusion that correlation lengths are reflecting large-scale transport. In May, the increasing trend of the SH correlation length is interrupted by a sudden decrease over most of the SH, which marks the transition to the NH summer. This decrease is due to a global reduction in the spatial variability of terrestrial fluxes, which causes a decrease in $X_{CO2}$ variances and correlation lengths.

[55] In general, $X_{CO2}$ regional correlation length parameters reflect more zonal response to changes in surface fluxes relative to the variance parameter. Moreover, correlation lengths show features that reflect global transport of CO$_2$ variability to the Poles and the lack of mixing between the hemispheres.

4.2.3. Submonthly Temporal Variability

[56] The discussion presented in sections 4.2.1 and 4.2.2 is based on an analysis of the 15th day of each month. The representativeness of this single day of the variability that would be expected for any given day within a certain month is tested by analyzing eight consecutive days in November (13–20 November), which is the month that exhibits the largest within-month change in the global parameters (Figure 2). Results show that the changes in the spatial patterns and magnitudes of the fitted regional parameters are minimal within a 4-day time window, but noticeable differences do occur in the location of the maximum regional variance parameters beyond 4 days. This result indicates that the 15th of each month is an adequate representation of individual days for low-variability months, but that covariance parameters vary on sub-monthly scales for seasonal transition months. As such, the regional spatial variability analysis may need to be repeated for multiple days within these transition months to capture the temporal changes in the regional spatial variability. This analysis could also be performed using retrieved soundings after the launch of OCO, although some regions may be more difficult to examine because of expected data gaps.

4.2.4. Overall Variability

[57] The parameter $h_o$ provides a single representation of the regional variability of $X_{CO2}$, as simulated using MATCH-CASA, by translating the seasonal variability in fluxes and transport captured by the regional covariance parameters into regional variability in the spatial scale over which a sounding is representative of local $X_{CO2}$. Assuming no measurement error or sampling limitations, $h_o$ can also be interpreted as a measure of the relative sounding densities that would be required for achieving global
Figure 3a. Regional variance of MATCH/CASA simulated $X_{CO2}$ at 1300 local time on the 15th of each month. Note differences in color scales. Regions are defined as 2000 km radius circles centered at the 2048 MATCH/CASA model grid cells.

$X_{CO2}$ coverage with relatively uniform uncertainty. Figure 4 shows monthly $h_o$ values for a 0.5 ppm uncertainty threshold ($\sqrt{V_{max}}$). In general, $h_o$ varies both spatially and seasonally, and reflects important features in surface flux and transport.

The seasonal changes in the overall variability are most noticeable by observing the maximum regional $h_o$, which peaks in November and December. $h_o$ values gradually decrease during the following months, reaching a minimum...
in July and August before increasing again in September and October.

[59] The overall variability tends to be higher over continental regions north of the ITCZ (shorter $h_o$), lower over continental regions south of the ITCZ, and variable over oceans. More specifically, regional $h_o$ values over the NH continents are short for fall, winter and spring. During the NH summer, very high variances cause $h_o$ values to further drop to lengths that are less than half of the MATCH/CASA resolution (i.e., less than 250 km). These
results are consistent with the findings of Karstens et al. [2006], who analyzed high-resolution (90 km and 55 km) monthly averaged simulated CO$_2$ concentrations over Europe. This earlier study looked at the distance at which the correlation coefficient between time series of CO$_2$ concentrations at different locations fell to 0.7. Results of Karstens et al. [2006] showed maximum separation distances between 170 km and 500 km in the summer, and increasing to 1000 km in the winter.

[60] In contrast to the NH continental regions, Australia shows large $h_o$ values that are more affected by the SH correlation length patterns. Tropical Africa and South

---

Figure 4. Information scale $h_o$ based on MATCH/CASA regional spatial covariance structure for 0.5 ppm uncertainty level.
America show short $h_o$ values, which reflect continental fluxes for seasons when the tropics have high variances relative to surrounding regions. During other seasons, $h_o$ values are strongly affected by the location of the ITCZ, with longer $h_o$ values south of the ITCZ.

Oceanic $h_o$ varies greatly between ocean basins, as well as seasonally. The Arctic Ocean shows persistently high variability and therefore low $h_o$, while the southern Pacific shows persistently lower variability and higher $h_o$. The north and tropical Atlantic oceans show some seasonality with low $h_o$ comparable to continental values for most of the year, except in December when $h_o$ is much higher. On the other hand, the North Pacific Ocean is highly seasonal with short $h_o$ values varying between values similar to continental regions in June though August, and very long $h_o$ values in December. The Indian Ocean has the strongest seasonality, with $h_o$ ranging between 350 km and 2,000 km during fall and spring, dropping to very low values from July to September and increasing to much longer values in November and December. Finally, the Southern Ocean spatial variability does not show large seasonality, but shows a longitudinal distribution with high $h_o$ values everywhere except between 60°W and 90°E where it shows higher spatial variability.

Although continental $X_{CO2}$ variability is mainly controlled by surface fluxes, variability over the oceans is less controlled by oceanic fluxes because the mixing and transport of the highly variable land fluxes in the averaged $CO_2$ column cause strong land fluxes to control oceanic $X_{CO2}$ variability near land regions. Furthermore, elevated variability is particularly apparent over coastal regions that include both land and ocean influences. Therefore, overall $X_{CO2}$ variability over oceans does not reflect the variability of underlying ocean fluxes as reported by studies such as McKinley et al. [2004]. This study noted that the highest flux variability originates from the Pacific followed by the Southern Ocean, and that the Atlantic exhibits the least flux variability. For the spatial variability analysis presented in this study, the highest overall variability over oceanic regions is over the Atlantic Ocean (particularly the north and tropical), because of the influence of fluxes from North America, Asia, and Europe. The Pacific Ocean shows some seasonality that mostly reflects the seasonal change in the location of the ITCZ and seasonal changes in transported fluxes from Asia, North and South America. These characteristics also support the conclusion that land and not ocean fluxes are the primary control on the variability of $X_{CO2}$ over oceans.

In general, the $h_o$ parameter is used to merge the spatial covariance parameters presented in sections 4.2.1 and 4.2.2 and provides a single representation of $X_{CO2}$ variability. Because this parameter can be defined for any variogram or covariance function, it also serves as an ideal basis for comparison to other models and data presented in the following section. The seasonality of $CO_2$ surface fluxes and changes in global transport cause the overall $X_{CO2}$ variability over land and ocean areas itself to display significant spatial and temporal variability; therefore indicating that the local $X_{CO2}$ covariance structure must be taken into account when evaluating the spatial representativeness of individual soundings and the uncertainty associated with gap-filled $X_{CO2}$ maps. Finally, the factors controlling this variability go beyond those that control the variability in local surface fluxes, with a strong observed influence of changes in global transport patterns.

4.3. Comparison to Other Models and Aircraft Data

This section tests the robustness of the modeled $X_{CO2}$ spatial covariance structure to (1) changes in $X_{CO2}$ introduced by differences in model setup, transport and $CO_2$ fluxes and (2) differences between the spatial resolution of MATCH/CASA (5.5°) and the sampling footprints of future satellites (e.g., 3 km² for OCO).

4.3.1. PCTM/GEOS-4 Global Simulation

The spatial variability of $X_{CO2}$ fields modeled using PCTM/GEOS-4 are compared to those from MATCH/CASA to examine whether a higher-resolution model will predict more $X_{CO2}$ variability, and whether the use of assimilated meteorological fields will have a substantial impact on the specific regions exhibiting high variability.

Figure 5 shows regional covariance parameters obtained from the PCTM/GEOS-4 model for January, April, July and October 2002. The range of variances and correlation lengths are very similar to those estimated using the MATCH/CASA model for all four months, implying a relatively low sensitivity of inferred $X_{CO2}$ variability to the increased resolution of PCTM/GEOS-4. This result supports the contention that the MATCH/CASA model is able to represent $X_{CO2}$ variability that is representative of observations taken at finer scales than the model grid. In other words, it begins to point toward the idea that the majority of the variability in $X_{CO2}$ occurs at scales that can be captured by relatively coarse global models.

Figure 5 also shows that for both models, the location and timing of high regional variances correspond to areas with high variability in the surface fluxes due to either fossil fuel emissions or biospheric activity. Nevertheless, there are differences in the magnitude and spatial extent of high-variance regions. The most prominent differences in regional variances occur over (1) eastern Europe in January and October, where MATCH/CASA indicates high variances whereas PCTM/GEOS-4 exhibits lower variances, and (2) Asia, where MATCH/CASA indicates high variances during April and October, while PCTM/GEOS-4 shows high variances over a larger area in January. Although large differences are also found over tropical South America in April, the variability observed in the PCTM/GEOS-4 simulation appears to be dependent on the time of day (results not shown).

Given the similarity of the fluxes prescribed in the two models, differences in the observed location and timing of high-variability regions are mostly attributable to differences in mixing and transport. Nevertheless, some of the observed differences may also be due to differences in submonthly variability in the fluxes imposed in MATCH/CASA versus PCTM/GEOS-4. These conclusions were further validated by analyzing the spatial variability of the prescribed biospheric and fossil fuel fluxes, as well as the spatial variability of $X_{CO2}$ simulations using PCTM/GEOS-4 including only biospheric or fossil fuel fluxes (results not shown).

Correlation lengths of MATCH/CASA and PCTM/GEOS-4 are even more consistent than the variance parameters. The most noticeable differences are the shorter correlation lengths inferred by PCTM/GEOS-4 over the
Figure 5. PCTM/GEOS-4 X<sub>CO2</sub> (a) regional variance, (b) regional correlation length, and (c) overall information scale (h<sub>i</sub>) for 0.5 ppm uncertainty level. All columns are evaluated at 1300 local time on 15 January, 15 April, 15 July, and 15 October.
North Pacific and over South America in January and April. Over the North Pacific, differences can be attributed to differences in transport and possibly model resolution, while over South America differences are mostly due to differences in biospheric flux variability. PCTM/GEOS-4 also shows slightly longer correlation lengths over Antarctica in January, which again most likely reflect differences in modeled transport.

Table 1. Parameter $h_o$ for Coincident Regions and Periods for MATCH/CASA Global Simulation, PCTM/GEOS-4 Global Simulation, SiB-RAMS Regional Model, and Aircraft Measurements

<table>
<thead>
<tr>
<th>Model/Data</th>
<th>Resolution</th>
<th>Column Elevation</th>
<th>North America: June–August</th>
<th>Pacific Ocean: June–August</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATCH/CASA</td>
<td>5.5° × 5.5°</td>
<td>surface to 60 km</td>
<td>85</td>
<td>1200</td>
</tr>
<tr>
<td>PCTM/GEOS-4</td>
<td>2.5° × 2.5°</td>
<td>surface to 48 km</td>
<td>70</td>
<td>400</td>
</tr>
<tr>
<td>SiB-RAMS</td>
<td>40 km</td>
<td>30 m to 3 km (PBL)</td>
<td>60--90°</td>
<td>--</td>
</tr>
<tr>
<td>SiB-RAMS</td>
<td>40 km</td>
<td>30 m to 9 km</td>
<td>85--140°</td>
<td>--</td>
</tr>
<tr>
<td>SiB-RAMS</td>
<td>40 km</td>
<td>30 m to 21 km</td>
<td>100--150°</td>
<td>--</td>
</tr>
<tr>
<td>SiB-RAMS</td>
<td>40 km</td>
<td>30 m to 60 km</td>
<td>105--160°</td>
<td>--</td>
</tr>
<tr>
<td>Aircraft point</td>
<td></td>
<td></td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>North America: PBL</td>
<td></td>
<td></td>
<td>5--100°</td>
<td>200--750°</td>
</tr>
<tr>
<td>Aircraft point</td>
<td>150 m to 9 km</td>
<td></td>
<td>Pacific Ocean: 150 m to 3 km</td>
<td>200--750°</td>
</tr>
</tbody>
</table>

*Reported $h_o$ values correspond to different months within the indicated range because of data availability.

*Range represents $h_o$ variability over 10 day simulation period.

*Range represents uncertainty in fitted variogram parameters.

4.3.2. SiB-RAMS Regional Simulation

The SiB-RAMS simulation provides a measure of $X_{CO2}$ spatial variability under conditions that most closely resemble satellite measurements among the three examined models, because (1) the model resolution is closest to the spatial scale of future satellite footprints; (2) the analyzed simulation represents $X_{CO2}$ concentrations over North America in August before and during the passage of a weather front, which is preceded by large CO$_2$ fluctuations with surface CO$_2$ concentrations increasing by up to 40 ppm [Wang et al., 2007]; and (3) the biospheric fluxes in the SiB-RAMS simulation are generated by the model according to the meteorological and biospheric conditions of the simulated time period. Thus, the SiB-RAMS model represents an important validation step for the $X_{CO2}$ spatial variability that will be observed by future satellites.

Results for individual days of the SiB-RAMS simulation show that the weather front is preceded by an increase in CO$_2$ variability followed by a decrease during the weather front then a return to initial variability levels of about 110 km. Even during the rapid CO$_2$ fluctuations, before the passage of the weather front (between 11 and 14 August), $h_o$ values do not vary substantially, indicating that although the concentrations themselves vary in time, the scales of spatial variability remain relatively constant. Therefore, $h_o$ values of MATCH/CASA simulations for North America can capture very similar features in $X_{CO2}$ spatial variability as a model with a 40 km resolution. Note that a 40 km resolution is comparable to the 10 km swath width of the OCO instrument.
relatively coarse global model can be representative of the variability that would be observed at finer scales, and is indicative of the relative scales of variability that will need to be captured by satellites such as OCO. This smoothness relative to surface fluxes and concentrations is caused by the averaging of low variability CO$_2$ at higher altitudes.

[76] However, given that it is possible that some additional artificial smoothing is also caused by transport effects, errors in mixing parameterization, or errors in flux distributions that are consistent between the examined models, the spatial variabilities predicted by these models are also compared to those derived from in situ measurements in the next section.

4.3.3. Aircraft Data

[77] This section compares the $h_o$ values of MATCH/CASA to the $h_o$ values calculated using the covariance parameters of X$_{CO2}$ partial columns (<9 km) as reported by Lin et al. [2004a]. The comparison is limited to North America and the Pacific Ocean because of the spatial coverage of the aircraft campaigns.

[78] For an uncertainty threshold of 0.5 ppm, the ranges of the power variogram parameter reported by Lin et al. [2004a] over North America for June 2003 correspond to $h_o$ with an uncertainty range of 5 km to 100 km. The MATCH/CASA average $h_o$ for the same month and region is 85 km, which is within the uncertainty bounds of the parameter derived from the aircraft data. For the Pacific Ocean (February to April and August to October over several years), the reported covariance parameters correspond to an $h_o$ in the range of 200 km to 750 km. The average MATCH/CASA $h_o$ for the same regions is 1200 km and 900 km in March and September, respectively (Table 1).

[79] In general, the spatial variability observed during the aircraft campaigns and that predicted by MATCH/CASA are consistent over the examined regions. Differences in $h_o$ values, particularly over the Pacific Ocean, during the sampled months can be attributed to three factors. First, in contrast to model simulations, the aircraft data are collected on relatively long time scales. Although Lin et al. [2004a] restricted the covariance function calculation to data pairs within a 3-h time window, different pairs would still represent multiple days or different times of the day. Therefore, nonstationary temporal variability captured by the aircraft data may have been interpreted as additional spatial variability. Second, X$_{CO2}$ spatial variability of aircraft data is evaluated using concentrations up to only 9 km in altitude, whereas MATCH/CASA models X$_{CO2}$ up to an elevation of 60 km, which is more comparable to the column that will be observed by future satellites. The effect of increasing the height of the measured column on X$_{CO2}$ spatial variability is demonstrated in Table 1, where the SiB-RAMS $h_o$ value for a 60 km column is about double its value for a 3 km column. The reduced X$_{CO2}$ spatial variability with increasing column elevation is also demonstrated by Lin et al. [2004a], who reported variogram parameters corresponding to $h_o$ values that increase as the height of the column increases (Table 1). Thus, the higher X$_{CO2}$ variability observed during the aircraft flights can also be attributed to the aircraft measurement of only partial columns (9 km elevation), with correspondingly higher CO$_2$ variability, relative to the column that will be measured by OCO and that was modeled by MATCH/CASA. Third, although the observed differences can be explained by the two factors described above, the higher X$_{CO2}$ spatial variability inferred from aircraft data relative to model simulations could potentially also be caused by actual variability not captured by the highest-resolution model used in this study (SiB-RAMS) or by differences between model fluxes/transport and actual aircraft sampling conditions. If this is the case, future satellites may observe higher X$_{CO2}$ variability than predicted by model simulations, but this last possibility cannot be tested using the limited available data.

[80] Overall, the comparison of X$_{CO2}$ variability inferred from the various models and aircraft data supports the conclusion that the scales of variability inferred using the MATCH/CASA model are representative of those that will be observed by satellites such as OCO. The analyses performed using PCTM/GEOS-4 and SiB-RAMS demonstrate the robustness of the estimates described in section 4.2 to a range of model resolutions and setups. The comparison to aircraft data demonstrates the consistency of the estimates with observed variability. One caveat that must be taken into account, and that cannot be resolved given the current limited availability of column-integrated measurements, is the remaining possibility that model simulations cannot reproduce some of the small-scale variability that will be observed by future satellites with small measurement footprints (e.g., 3 km$^2$ for the OCO). This is due to the high uncertainty in X$_{CO2}$ spatial variability inferred from aircraft measurements, which prevents a definite conclusion on this question. Nevertheless, the presented results provide strong evidence that the smoothness of the X$_{CO2}$ signal means that models that would be too coarse to resolve variability in surface CO$_2$ concentrations may indeed be able to adequately capture variability in the integrated X$_{CO2}$ signal. In addition, the presented results demonstrate that the predominant spatial patterns in the variability of X$_{CO2}$ are consistent between models, as well as for available field data, and that these patterns are attributable both to variability in the underlying flux distribution, and regional and seasonal variability in global transport.

5. Conclusions

[81] Understanding of the spatial variability of X$_{CO2}$, as well as the seasonality and regional differences in this variability, is necessary for making optimal use of the data that will be provided by satellites such as OCO. The evaluated spatial variability facilitates the use of spatial interpolation methods to (1) gap fill the retrieved soundings and evaluate the uncertainty associated with gap-filled data products; (2) quantify the representation errors associated with incorporating X$_{CO2}$ into atmospheric transport models or GCMs when the model resolution differs from the satellite footprint; and (3) in the case of computational limitations, design a sounding selection algorithm that captures the underlying spatial variability of the X$_{CO2}$ field.

[82] In this study, the regional X$_{CO2}$ spatial covariance structure is inferred using global X$_{CO2}$ distributions simulated using the MATCH/CASA model. The evaluated spatial variability is compared to that inferred from a second higher-resolution global model, a regional model, and aircraft measurements. Results show that the degree of observed spatial variability is consistent among the examined models,
and robust at spatial resolutions down to 40 km. Results are also consistent with variability inferred from aircraft measurements. Together, these results support the conclusion that the spatial variability inferred using the MATCH/CASA model is representative of the variability of $X_{CO2}$ as will be observed at much finer footprints. Because the $X_{CO2}$ signal is very diffuse relative to surface CO$_2$ concentrations, relatively coarse global models are able to represent the expected degree of $X_{CO2}$ spatial variability at smaller scales.

[87] The presented analysis shows that both the variance and correlation lengths of the $X_{CO2}$ field vary spatially and seasonally, and that this variability is attributable to changes in both surface fluxes and seasonal patterns in global transport. The variance parameter shows a clear cycle that reflects the NH growing season with peak values collocated with regions of highly variable CO$_2$ surface fluxes. The effect of transport on the variance parameter is clearest during the NH winter months, when highly variable fluxes from NH continents are transported to the Arctic and northern ocean. The correlation lengths of the $X_{CO2}$ field, on the other hand, do not show a distinct seasonal cycle, but clearly reflect transport and mixing effects. These effects are demonstrated by the contrast in correlation lengths between the hemispheres, and the effects of the seasonal movement of the ITCZ on the boundary between regions with low and high correlation lengths.

[84] Overall $X_{CO2}$ spatial variability is quantified using the parameter $h_x$, which represents the relative spatial scale of the information provided by a single $X_{CO2}$ observation in a given region. $h_x$ values vary between hemispheres and between ocean and continental regions. Values are lowest during the NH summer over highly active continental flux areas, and are highest over the Pacific Ocean during the NH winter. The $X_{CO2}$ variability over the oceans, particularly near continental regions, is primarily controlled by transport of CO$_2$ signals from continental regions.

[85] Results are consistent with the conclusion that a spatially and temporally variable sounding density would be required to capture the regional differences in the spatial variability of $X_{CO2}$ with a uniform precision. Moreover, the representativeness of individual soundings is expected to vary regionally and seasonally, as a function of the heterogeneity in the identified spatial variability, with soundings from high-variability regions having higher representation errors when used to represent $X_{CO2}$ in coarser models.

[86] Geophysical limitations (e.g., clouds, aerosol) and instrument characteristics (e.g., satellite track) are expected to cause large gaps in retrieved $X_{CO2}$ distributions, and computational costs may further limit the fraction of retrieved soundings in the early parts of the OCO mission. The evaluated $X_{CO2}$ variability provides important information about the ability of future satellites to capture the underlying $X_{CO2}$ distribution given these limitations. For example, given the maximum 2500 km coverage gap of consecutive orbits of OCO at the equator, and discounting all other sampling limitations, this analysis suggest gap-filling uncertainties that closely follow the patterns seen in the variance maps (Figure 3a), reaching a maximum of 4 ppm over boreal Asia in July, and with generally low values (<1 ppm) over oceans and regions with low surface flux variability (detailed results not shown).

[88] Finally, the analysis presented in this paper establishes the main patterns of $X_{CO2}$ spatial variability and how surface fluxes and transport affect these patterns as simulated by current models. Because these models reflect the current scientific understanding of surface fluxes of CO$_2$, the results also provide a baseline for evaluating the contribution of future satellites in improving the present understanding of the $X_{CO2}$ distribution and its spatiotemporal variability.

Acknowledgments. The research described in this paper was partially performed for the Orbiting Carbon Observatory Project at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Additional support was provided by NASA under grant NNX08A92G “Mapping Global CO$_2$: Development and Application of Geostatistical Algorithms for Gap Filling and Uncertainty Assessment for the Orbiting Carbon Observatory” issued through the ROSES A.3 Carbon Cycle Science Program. Kuwait University Scholarship Committee provided partial funding for the work presented in this study. The authors gratefully acknowledge the helpful input provided by Charles Miller (JPL) on early drafts of this manuscript. The PCTM work was enabled by G. J. Collatz and Z. Zhu and was supported by NASA Carbon Cycle Science Program. This paper was greatly improved as a result of detailed input provided by three anonymous reviewers.

References
Dennis, A. S., M. Nicholls, L. Pröhl, I. Baker, P. L. Vidale, K. Davis, and P. Bakwin (2003), Simulated variations in atmospheric CO$_2$ over a...


