

# THESIS

## EVALUATING SPATIAL, TEMPORAL, AND CLEAR-SKY ERRORS IN SATELLITE CO<sub>2</sub> MEASUREMENTS

Submitted by

Katherine Corbin

Department of Atmospheric Science

In partial fulfillment of the requirements

for the degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2005

COLORADO STATE UNIVERSITY

June 3, 2005

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR  
SUPERVISION BY KATHERINE CORBIN ENTITLED EVALUATING  
REPRESENTATION, DIURNAL CYCLE, AND CLEAR-SKY ERRORS IN  
SATELLITE CO<sub>2</sub> MEASUREMENTS BE ACCEPTED AS FULFILLING IN PART  
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

Committee on Graduate Work

---

---

---

Advisor

---

Department Head

# **Abstract**

## **Evaluating Spatial, Temporal, and Clear-Sky Errors in Satellite CO<sub>2</sub> Measurements**

Variations of atmospheric CO<sub>2</sub> concentrations contain information about sources and sinks with which air interacts with as it is transported from place to place. Using atmospheric tracer transport models, inverse modelers can quantitatively estimate the strengths and spatial distribution of sources and sinks around the world from concentration data. Satellite CO<sub>2</sub> measurements have the potential to help inverse modeling studies by improving the data constraint due to their global spatial sampling and sheer data volume. The Orbiting Carbon Observatory (OCO), scheduled to launch in 2008, will retrieve global total column CO<sub>2</sub> concentrations at 1:15 pm LST with 0.5% precision. These satellite measurements can be used in inversion models to enhance our understanding of the carbon cycle; however, several errors can be introduced when using satellite measurements to optimize CO<sub>2</sub> concentrations in inverse studies: spatial representativeness errors may be introduced into inversions that compare CO<sub>2</sub> concentrations from a model grid cell to satellite concentrations sampled over only a fraction of the domain, temporal sampling errors can result from comparing OCO measurements sampled at 1:15 PM to temporally averaged concentrations in an inversion, and local clear-sky errors may exist in inversions from comparing concentrations in a grid cell that may be partially cloudy to satellite mixing ratios sampled at the same time but only over clear areas.

This study investigates these errors using a coupled ecosystem-atmospheric model, SiB2-RAMS, and continuous data. We performed two simulations centered on a tall tower site in

Wisconsin. Using the generated CO<sub>2</sub> field, we compared the clear-sky total column concentrations from an emulated satellite track to the mean total column mixing ratio over the domain. We also investigated the near-surface clear-sky errors at the WLEF tall tower near Park Falls, WI and at Harvard Forest by comparing a two-harmonic fit to a clear-sky subset at each location to the fit to the complete timeseries.

Inversions using 100 km grid cells will have very small spatial errors and local clear-sky errors, as we found that spatial representativeness errors and local clear-sky errors on a heterogeneous 100 km by 100 km domain are <0.2 ppm and are normally distributed. Both of these errors increased in magnitude as the resolution of the grid cell increased, suggesting that to minimize errors inversions should use the finest resolution possible. By far the largest errors seen in this study came from comparing satellite CO<sub>2</sub> mixing ratios to temporally averaged concentrations. We found that the main driver of the CO<sub>2</sub> variability is not the diurnal cycle of CO<sub>2</sub> from biology, but instead is synoptic variability. Over Wisconsin, the cloud-covered days had higher mixing ratios, as fronts advected high CO<sub>2</sub> from other regions on these days. Under-sampling these events introduces a bias of ~-0.5 ppm compared to the estimated 10-day temporal mean. The results from continuous observations confirm these results, indicating that there is a year-round negative bias in CO<sub>2</sub> concentrations on clear days, which is likely due primarily to advection of high CO<sub>2</sub> in cloudy conditions. To use the satellite total column CO<sub>2</sub> measurements without introducing significant errors, inverse models will have to accurately model the atmospheric synoptic-scale transport and use satellite measurements at 1:15 PM to optimize the modeled concentrations taken at the same time.

Katherine Corbin  
Department of Atmospheric Science  
Colorado State University  
Fort Collins, CO 80523  
Summer 2005



# Acknowledgements

I would like to thank Dr. A. Scott Denning, my advisor, for his help and guidance during the past few years. I would also like to thank my committee members Dr. Niall Hanan and Dr. Graeme Stephens for taking the time to review this thesis.

Thanks also to all the members in the Denning Group. Help from Ian Baker, Aaron Wang, and Lixin Lu was very important for me to get through this research and to finally get the model running. Lara Prihodko provided me with initial output to get me started and also provided help preparing data. Kevin Schaefer, my long-time officemate, always provided valuable advice and engaged me in interesting discussions.

I would also like to thank Prof. Steve Wofsy from Harvard University for the Harvard Forest data, and Prof. Ken Davis from Penn State University for the data at the WLEF tower. Thank you also to Molly Brown from NASA/GIMMS, who provided the NDVI data.

Special thanks to my family and friends who have supported and encouraged me during this time.

This research was funded by NASA Grant NCC5-621 and by a NASA Earth System Science Fellowship 53-1970.

# Table of Contents

## Chapter 1: Introduction

1.1 Atmospheric CO <sub>2</sub> Background .....	1
1.2 Terrestrial Carbon Processes .....	4
1.3 The Effects of Clouds on NEE .....	6
1.4.1 How do Clouds Affect NEE?	
1.4.2 Mechanisms that Enhance NEE	
1.4 Overview of CO <sub>2</sub> Variability .....	12
1.4.1 Local Diurnal Variability	
1.4.2 Regional Variability	
1.4.3 Global Variability	
1.5 Explaining the Interannual Variability in the CO <sub>2</sub> Growth Rate .....	17
1.6 Locating the Missing Carbon Sink .....	20
1.6.1 Inverse Modeling	
1.6.2 Partitioning the Ocean/Atmosphere Sinks Using Oxygen and Isotopes	
1.7 The Orbiting Carbon Observatory (OCO) .....	24
1.7.1 Overview	
1.7.2 Spectral Range	
1.7.3 Orbit Requirements	
1.7.4 Observation Modes	
1.7.5 Spatial Sampling	
1.7.6 Calibration and Validation	

1.7.7 Sources of Error	
1.7.7.1 Spatial Representativeness Errors	
1.7.7.2 Diurnal Errors	
1.7.7.3 Clear-Sky Errors	
1.8 Objectives of this Study .....	34

## **Chapter 2: Methods**

2.1 Data .....	38
2.1.1 WLEF Tall Tower	
2.1.2 Harvard Forest	
2.1.3 ASOS	
2.1.4 Climatic Data	
2.1.5 ECMWF Boundary Layer Heights	
2.2 Methods for Determining the Clear-Sky Bias from Observations.....	43
2.2.1 Data Preparation	
2.2.2 Sampling Times	
2.2.3 Converting the PAR measurements to Downwelling Radiation	
2.2.4 Creating the Clear Subset	
2.2.5 Calculating the Clear-Sky Bias	
2.3 Model Description: SiB2-BRAMS.....	50
2.3.1 SiB2	
2.3.2 RAMS	
2.4 Input Data for the Model.....	59
2.4.1 Vegetation Classes	
2.4.2 Soil Type	
2.4.3 NDVI	

2.4.4 Fossil Fuel Combustion	
2.4.5 Soil Moisture, Soil Stress, and Respiration Factor	
2.4.6 Meteorological Forcing	
2.5 Case Descriptions.....	66
2.5.1 Case 1: 5-Day Clear-Sky Simulation centered at WLEF	
2.5.2 Case 2: 10-Day Simulation centered at WLEF	
2.6 Methods for Simulation OCO Using SiB2-RAMS Output.....	69

### **Chapter 3: Meteorological Evaluation of Case 2**

3.1 Meteorological Analysis.....	71
3.2 Model Evaluation.....	74
3.2.1 Large-Scale Patterns	
3.2.2 Point Comparisons	

### **Chapter 4: Spatial Representativeness of Satellite CO<sub>2</sub> Measurements**

4.1 Case 1.....	90
4.1.1 Overview of Surface CO <sub>2</sub> Spatial Variability	
4.1.2 Total Column CO <sub>2</sub> Variability	
4.1.3 Satellite Track Spatial Variability	
4.1.4 Spatial Representativeness Errors	
4.2 Case 2.....	103
4.2.1 Cloud Cover	
4.2.2 NEE Variability	
4.2.3 Total Column CO <sub>2</sub> Spatial Variability	
4.2.4 Spatial Representativeness Errors	

## **Chapter 5: Temporal Representativeness of Satellite CO<sub>2</sub> Measurements**

5.1 Case 1 .....	123
5.1.1 Diurnal Cycle	
5.1.2 Diurnal Representativeness at 1 PM	
5.1.3 Hourly Diurnal Representativeness	
5.2 Case 2 .....	131
5.2.1 Diurnal Cycle	
5.2.2 Diurnal Representativeness at 1 PM	

## **Chapter 6: Clear-Sky Errors in Satellite CO<sub>2</sub> Measurements**

6.1 Investigating the Clear-Sky Bias Using Observations .....	138
6.1.1 Results	
6.1.2 Explanatory Hypothesis: What could be going on?	
6.1.3 What Does this Mean for Satellites?	
6.2 Investigating Clear-Sky Errors using SiB2-RAMS .....	150
6.2.1 Local Clear-Sky Errors	
6.2.2 Temporal Sampling Errors	

## **Chapter 7: Conclusions and Recommendations**

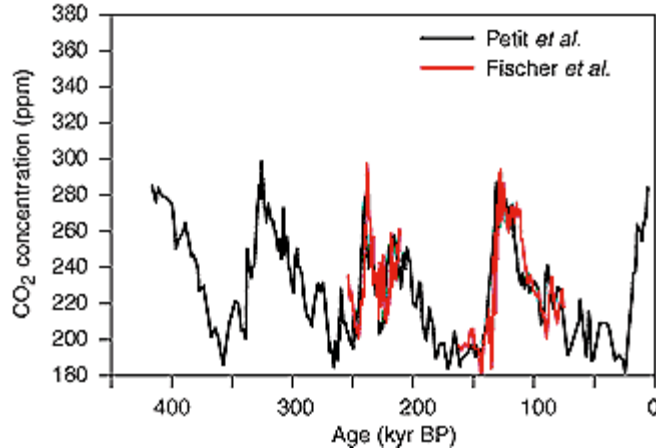
7.1 Conclusions .....	169
7.2 Recommendations for Future Work .....	173

<b>References</b> .....	174
-------------------------	-----

# Chapter 1: Introduction

## 1.1 Atmospheric CO<sub>2</sub> Background

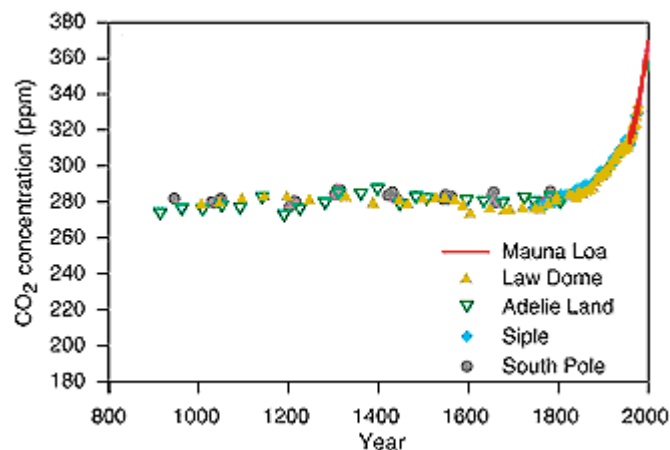
Excluding the contemporary CO<sub>2</sub> concentrations, atmospheric CO<sub>2</sub> has varied between 180 and 280 parts per million (ppm) over the past 420,000 years. Figure 1.1 shows the CO<sub>2</sub> concentration using the Vostok Antarctic ice core (Petit *et al.*, 1999; Fischer *et al.*, 1999). The atmospheric CO<sub>2</sub> concentration tracks the glacial cycles, with low atmospheric CO<sub>2</sub> during glacial periods and higher concentrations during interglacial periods. Natural processes during the glacial-interglacial cycles maintained the CO<sub>2</sub> concentrations within those bounds, despite considerable variability on multi-millennial timescales (IPCC, 2001).



**Figure 1.1:** CO<sub>2</sub> concentration in the Vostok ice core. The black line is the concentration from Petit *et al.*, while the red line is the concentration reported by Fischer *et al.*

Zooming in to a more recent timescale, over the past several thousand years the CO<sub>2</sub> concentration has remained relatively constant. According to the concentrations from the Taylor Dome ice core, natural variations in CO<sub>2</sub> during the past 11,000 years have been small, varying

only about 20 ppm (Smith *et al.*, 1999; Indermühle *et al.*, 1999). Around 8 kyr before present (BP), the atmospheric CO<sub>2</sub> concentration was at its Holocene minimum of 260 ppm, and it increased steadily towards a concentration of 280 ppm prior to the industrial revolution; however, following the industrial revolution, the CO<sub>2</sub> concentration increased dramatically. Figure 1.2 shows the CO<sub>2</sub> record for the past 1,200 years, ending in 2000. In a span of only 200 years, the atmospheric concentration of CO<sub>2</sub> has risen from close to 280 ppm in 1800 at first slowly and then progressively faster to a value of 367 ppm in 1999, echoing the increasing pace of global agricultural and industrial development (IPCC, 2001). The rate of increase over the last century is unprecedented at least in the past 20,000 years. The present atmospheric CO<sub>2</sub> concentration far exceeds the concentration over the past 420,000 years, and likely exceeds the concentration seen in the past 20 million years.



**Figure 1.2:** CO<sub>2</sub> concentration over the past millennium (IPCC, 2001).

The recent atmospheric CO<sub>2</sub> increase is caused by anthropogenic emissions of CO<sub>2</sub>. Organic carbon buried in sediments as coal, natural gas, and oil over hundreds of millions of years is now being consumed by human activities and returned to the atmosphere as CO<sub>2</sub> on a much shorter timescale. In addition, land use changes also contribute CO<sub>2</sub> to the atmosphere (Sarmiento and Gruber, 2002). Changes in land use and management affect the amount of carbon in plant biomass and soils (IPCC, 2001). Conversion of natural vegetation to agriculture is a

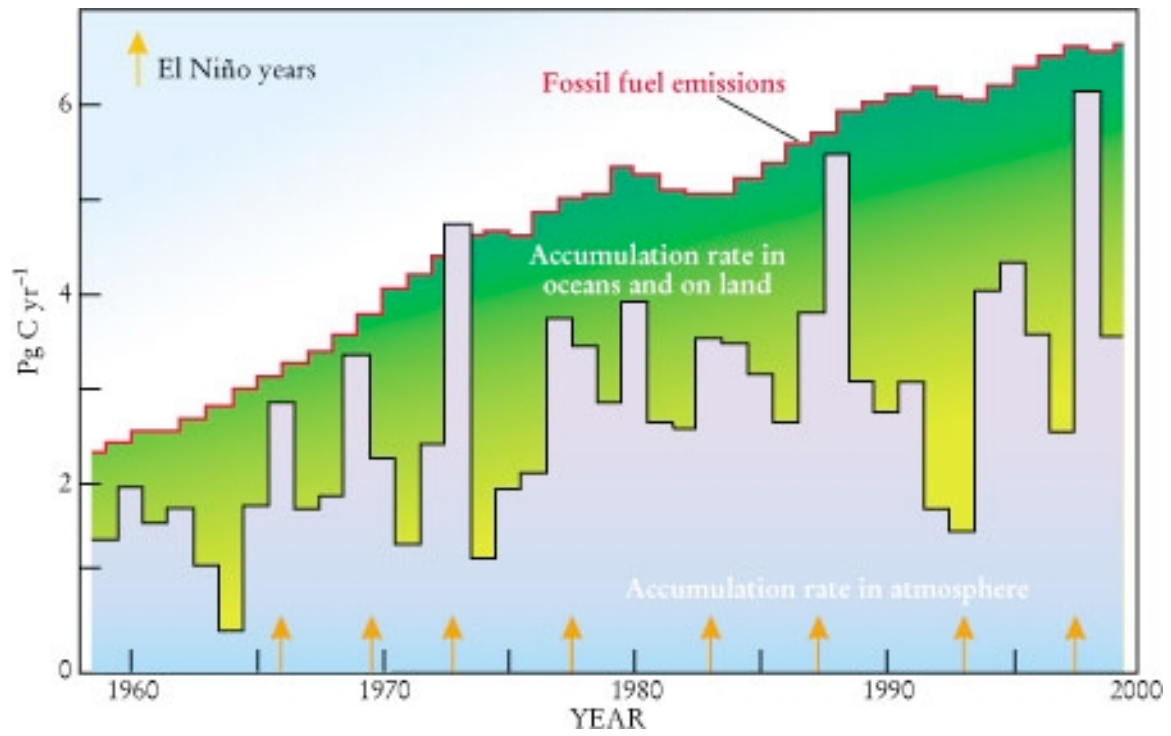
major source of atmospheric CO<sub>2</sub> not only due to losses of plant biomass, but also due to increased decomposition of soil organic matter, although the use of high-yielding plant varieties, fertilizers, irrigation, residue management, and reduced tillage can enhance the soil uptake in managed areas (IPCC, 2001). The clearing of forests for agricultural lands and the harvesting of wood are responsible for almost 90% of the estimated emissions due to land-use change since 1850, with a 20% decrease in global forest area (Houghton, 1999; IPCC, 2001; Sarmiento and Gruber, 2002).

The recent increase of atmospheric CO<sub>2</sub> has caused alarm as CO<sub>2</sub> is a non-reactive gas that plays an important role in the radiative balance of the Earth. CO<sub>2</sub> is a “greenhouse gas”: while being nearly transparent to solar short-wave radiation, it absorbs infrared radiation in the same wavelengths as the thermal radiation emitted from Earth. Although water vapor is the most important greenhouse gas, CO<sub>2</sub> is an important greenhouse gas as it absorbs at several different infrared wavelengths and partially closes the small window in water vapor absorption where heat and radiation can leave the atmosphere. Even though the effect of the increased trapping of long-wave radiation depends on a number of complex feedbacks, the strong consensus of the scientific community is that the increased trapping will lead to global warming and will play a significant role in altering the Earth’s climate (e.g. Sarmiento and Gruber, 2002).

Even though the atmospheric CO<sub>2</sub> concentration is increasing rapidly, only about half of the CO<sub>2</sub> emitted by human activities is accumulating in the atmosphere. Figure 1.3 shows the annual global fossil fuel emissions and the growth rate of atmospheric CO<sub>2</sub>. The portion of the emitted CO<sub>2</sub> not present in the atmosphere, the green area in the figure, is absorbed by “sink” processes on land or in the ocean: the CO<sub>2</sub> is either taken up by terrestrial ecosystems due to an excess of primary production (photosynthesis) over decomposition or is dissolved in sea water and mixed into the deep ocean. In addition to showing the magnitude of the carbon sinks, Figure 1.3 also shows the considerable interannual variability that exists in the atmospheric CO<sub>2</sub> growth



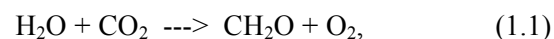
rate. Although recent studies have advanced our knowledge of the carbon cycle, scientists are still unable to fully explain the atmospheric growth rate, leaving many questions still unanswered.



**Figure 1.3:** Fossil fuel emissions and the accumulation rate in the atmosphere (Sarmiento and Gruber, 2002).

## 1.2 Terrestrial Carbon Processes

To understand the carbon cycle, first it is necessary to discuss the terrestrial flux of carbon. The discussion in this section is derived mainly from Denning [1994]. Photosynthesis is the process by which plants use sunlight to manufacture simple carbohydrates from atmospheric  $\text{CO}_2$  and water. Plants acquire  $\text{CO}_2$  by diffusion through tiny pores (stomata) into the leaves and thus to the sites of photosynthesis. When moisture is abundant, both in the soil around the plant's roots and in the air in the form of vapor, the stomata open as much as possible and photosynthetic carbon fixation is rapid. Although photosynthesis is a complicated process involving many biochemical steps, it can be represented by the simple reaction



where  $\text{CH}_2\text{O}$  represents any simple carbohydrate. The rate at which plants convert inorganic  $\text{CO}_2$  into organic compounds is called *gross primary productivity*, and the amount of  $\text{CO}_2$  that is converted from  $\text{CO}_2$  to carbohydrate during photosynthesis is known as *gross primary production* (GPP).

The processes by which living organisms rearrange organic compounds to extract energy and build tissues are collectively known as metabolism. Metabolic energy is derived from carbohydrate molecules by basically the reverse reaction of (1.1), and this energy extraction reaction is called *autotrophic respiration* (respiration by plant tissues). Plants respire carbon that they have previously fixed to provide the energy necessary to grow, reproduce, and sustain life. Plant respiration occurs constantly and dominates over photosynthesis at night and during other times (such as winter) when plants need more energy than they can produce. In most ecosystems, approximately half of the annual GPP is incorporated into new plant tissues such as leaves, roots, and wood while the other half of GPP is converted back to atmospheric  $\text{CO}_2$  by autotrophic respiration (IPCC, 2001). Annual plant growth, which is the net amount of carbon fixed from  $\text{CO}_2$  by plants, is called the *net primary production* (NPP). NPP is defined as the difference between photosynthesis (GPP) and autotrophic respiration ( $R_A$ ), which is written as

$$\text{NPP} = \text{GPP} - R_A. \quad (1.2)$$

When plants or parts of plants (like leaves and branches) die, the dead biomass falls to the ground. Once on the ground, bacteria and other microorganisms decompose the biomass in a process called *heterotrophic respiration* ( $R_H$ ). Essentially, heterotrophic respiration is the reverse reaction of (1.1), where the energy is released in the process of decomposition. The rate of  $R_H$  depends on both the chemical composition of the dead tissues as well as the environmental conditions (for example low temperatures, dry conditions, and flooding slow down decomposition). The difference between NPP and  $R_H$  is called *net ecosystem production* (NEP), which is written as

$$\text{NEP} = \text{NPP} - R_H = \text{GPP} - R_T, \quad (1.3)$$

where  $R_T = R_A + R_H$  is the total respiration in the ecosystem. NEP determines how much carbon is lost or gained by the ecosystem in the absence of disturbances that remove carbon from the ecosystem (such as harvest or fire). In mature ecosystems over large spatial scales at a steady state, the NEP is approximately zero, with the annual total of heterotrophic respiration nearly equaling the annual NPP, and the total biomass of the ecosystem remaining constant. After a major disturbance (like a forest fire or harvest) that removes biomass from an ecosystem, large-scale NEP is generally positive. In forests, NEP reaches a maximum several decades after a disturbance and declines to nearly zero after a century or two.

In atmospheric science, it is common to define fluxes into the atmosphere as positive rather than having fluxes into the biosphere be positive, which is done in NEP. To do this, atmospheric scientists have introduced a variable called *net ecosystem exchange* (NEE), which is simply the opposite sign of NEP. NEE can be written as

$$NEE = -NEP = R_H - NPP = R_T - GPP. \quad (1.4)$$

NEE is commonly measured and reported at flux towers and used in atmospheric applications rather than NEP.

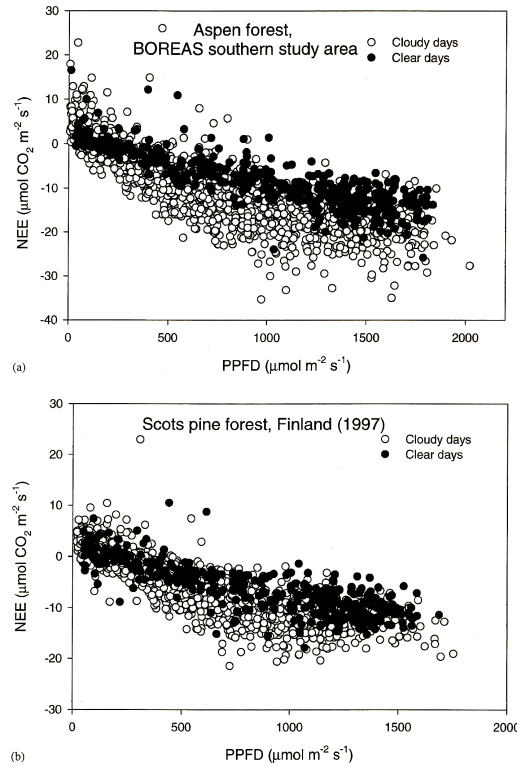
### 1.3 The Effects of Clouds on NEE

A key issue for the carbon cycle is the response of vegetation to the environment. Clouds influence the environmental conditions by altering the radiation budget: clouds reduce the global solar radiation while increasing the relative proportion of diffuse radiation at the Earth's surface. In addition to changes in surface solar radiation, the presence of clouds can cause changes in temperature and humidity. Changes in these environmental conditions can influence the rates of both photosynthesis as well as respiration: thus, clouds and aerosols can alter the NEE of  $CO_2$  between the biosphere and the atmosphere. This section will investigate recent literature to

provide an overview of the effects of clouds on carbon uptake, as well as the mechanisms that cause these effects.

### **1.3.1 How Do Clouds Affect NEE?**

Numerous studies have investigated the impact of clouds on NEE, all reaching the conclusion that moderate cloud cover enhances NEE (i.e. cause NEE to have the most negative value). One of the first studies that investigated the effects of clouds on photosynthetic uptake is a study by Price and Black in 1990. By using a modified Bowen ratio and energy balance technique at a juvenile Douglas-fir forest, the authors found that canopy CO<sub>2</sub> flux density on cloudy days was typically higher than on clear days. Also studying boreal coniferous forests, Law *et al.* [2002] and Gu *et al.* [2002] both also found that net carbon uptake was greater under cloudy conditions rather than under clear conditions. Based on measurements above an old black spruce forest, Goulden *et al.* [1997] showed that photosynthetic efficiency was higher during cloudy periods. In addition to boreal coniferous forests, boreal deciduous forests also exhibit a consistent and significant increase of NEE in the presence of clouds, with the uptake enhanced by as much as 52% (Gu *et al.*, 1999; Freedman *et al.*, 2002; Law *et al.*, 2002). Figure 1.4 shows the total carbon uptake for varying photosynthetic photon flux density (PPFD) for a coniferous and a deciduous boreal forest on both cloudy and clear days (Law *et al.*, 2002). Since the rate of NEE is greater with clouds, the figure shows that cloudy days provide more optimal conditions for NEE, particularly on days with >500  $\mu\text{mol}/\text{m}^2/\text{s}$ . For days with very low PPFD, cloudy days have suppressed NEE.



**Figure 1.4:** NEE in relation to diffuse and direct PPFD for (a) a boreal aspen forest and (b) a boreal coniferous forest (Law *et al.*, 2002).

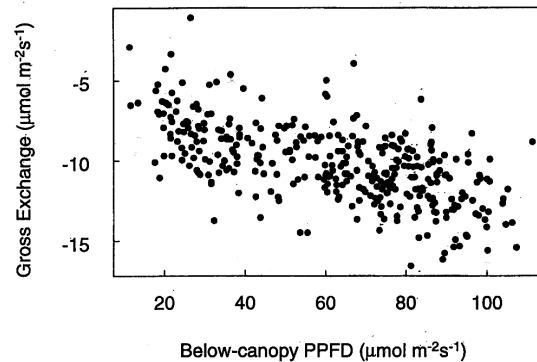
At an old temperate broadleaf evergreen forest, Hollinger *et al.* [1994] looked at specific days to discover that total daily net ecosystem CO<sub>2</sub> uptake was greater on cloudy days, despite the total PPFD on the cloudy days being from 21 to 45% lower than on the clear days. Gu *et al.* [1999], Baldocchi [1997], and Gu *et al.* [2002] report that in a temperate deciduous forest, rates of NEE are diminished by more than half when clear sky conditions are compared with cloudy conditions. Finally, Gu *et al.* [2002] conducted a study at both a native tall grass prairie and a cultivated wheat crop, finding that the canopy quantum yield of diffuse photosynthetically active radiation (PAR) is consistently higher than the canopy quantum yield of direct PAR. A synthesis of these studies shows that the largest uptake of carbon occurs under moderately cloudy conditions; and the enhanced photosynthetic uptake occurs for a wide variety of biomes in the mid and high latitudes and also for both young and old growth stands.

### 1.3.2 Mechanisms That Enhance NEE

Since field observations show that the highest rate of net ecosystem exchange occurs on cloudy days and that for a given irradiance level cloudy days have a higher NEE, it is necessary to understand the mechanisms that cause this enhancement. To explain these observations, three key mechanisms have been postulated. The first mechanism is enhancement of GPP by diffuse radiation. The importance of diffuse radiation on photosynthesis can be understood by investigating shadows: in direct radiation, since the light is concentrated in a beam, the top of the canopy receives radiation while vegetation within the canopy receives little radiation due to deep shade from large shadows; however, diffuse radiation is isotropic, allowing more light to penetrate lower into the canopy (Roderick *et al.*, 2001). Under direct beam radiation, photons are “wasted” by concentrating the light resource to only a fraction of all leaves which are already light-saturated, leading to less efficient photosynthetic use of light by plant canopies; in contrast, diffuse radiation effectively avoids the light saturation constraint by more evenly distributing radiation amongst all leaves, leading to a more efficient use of light (Gu *et al.*, 2002). Although clouds may decrease the total radiation on the canopy, the saturating response of leaf photosynthesis to light intensity typically leads to higher photosynthesis when all leaves have moderate light than when the top leaves receive bright light while leaves within the canopy are in deep shade (Farquhar and Roderick, 2003).

Price and Black [1990] state that a major factor leading to more optimal conditions for larger carbon uptake is the greater canopy penetration by diffuse short-wave radiation directly attributable to the presence of cloud cover. Goulden *et al.* [1997] used measurements at an old black spruce site in central Manitoba to show that the increase in diffuse light enhances photosynthesis. In a Black Spruce forest, the light in the canopy was 20-40  $\mu\text{mol}/\text{m}^2/\text{s}$  during clear-sky periods with incident light from 250-550  $\mu\text{mol}/\text{m}^2/\text{s}$ ; however, during cloudy periods with the same incident light, the light within the canopy increased to 60-100  $\mu\text{mol}/\text{m}^2/\text{s}$ . The increase in below-canopy irradiance is correlated with an increase in forest photosynthesis from

8-12  $\mu\text{mol}/\text{m}^2/\text{s}$ , which is shown in Figure 1.5, and this variation in photosynthesis is not associated with other aspects of the physical environment.

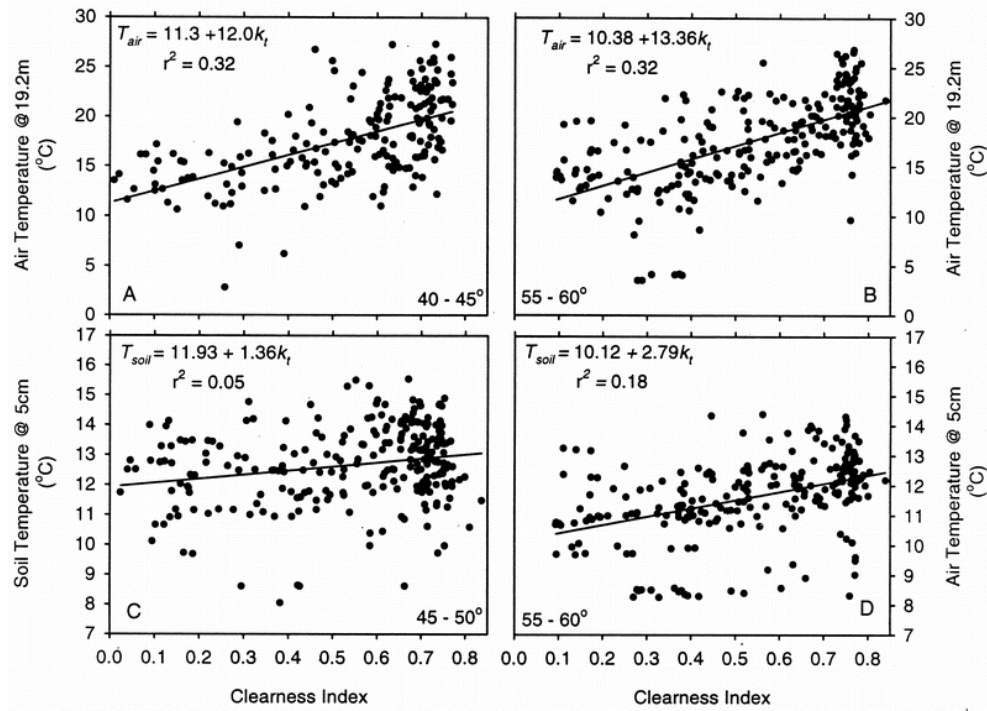


**Figure 1.5:** Whole-forest gross exchange at a black spruce forest as a function of below-canopy PPFD (Goulden *et al.*, 1997).

Although plants within the canopy may have less photosynthetic capacity due to lower time-mean light, based on their research, Goulden *et al.* state that one-third to two-thirds of forest photosynthesis occurs on shaded surfaces. Additionally, Hollinger *et al.* [1994] and Law *et al.* [2002] indicate that the diffuse fraction of PPFD is a significant factor causing the high rates of net  $\text{CO}_2$  uptake on cloudy days. Since plant canopies have a higher light use efficiency for diffuse PAR than for direct PAR, the gain of canopy photosynthesis due to an increase in the diffuse PAR can exceed the loss due to a decrease in the direct PAR [Gu *et al.*, 1999].

In addition to enhanced photosynthesis from diffuse light, another mechanism that may contribute to increased NEE is decreases in both leaf and soil respiration. Respiration predominantly depends on temperature: a decrease in temperature will cause an exponential decrease in respiration. Using a clearness index to measure cloudiness, which decreases with increasing cloud cover, Gu *et al.* [1999] show that both air and soil temperatures decrease as cloudiness increases. This result is shown in Figure 1.6. The decreases in temperatures reduce both leaf and soil respirations (Price and Black, 1990; Gu *et al.*, 1999). Baldocchi [1997] shows that on clear days leaves at the top of the canopy become light saturated and warmer than leaves in the lower, shadier canopy; consequently, sunlit leaves experience enhanced respiration. By

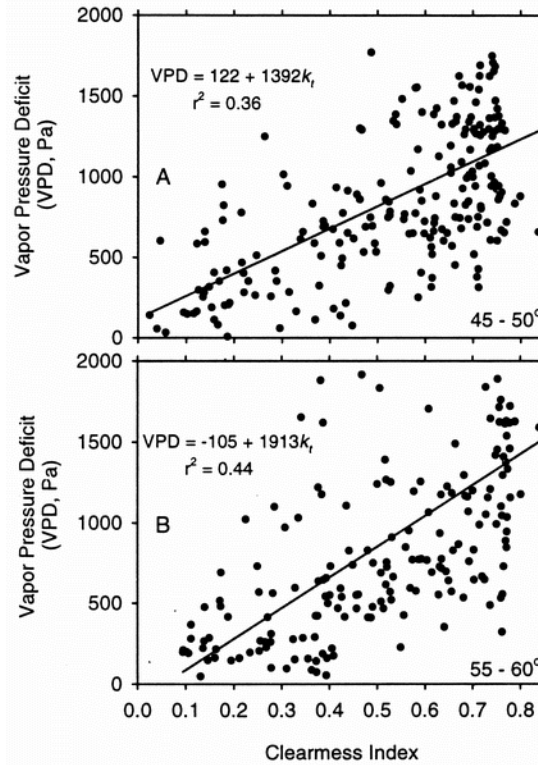
increasing diffuse light and cooling temperatures, clouds will decrease respiration on sunlit leaves (Baldocchi, 1997).



**Figure 1.6:** Changes of (a and b) air and (c and d) soil temperatures with the clearness index for selected solar elevation angles at a boreal aspen forest in 1994 (Gu *et al.*, 1999).

The third factor that can enhance canopy photosynthesis is the general trend of a decrease in vapor pressure deficit (VPD) associated with cloudy conditions (Gu *et al.*, 1999). VPD is defined as the difference between the amount of moisture in the air and how much moisture the air can hold when it is saturated. VPD combines the effects of both humidity and temperature into one value, and thus gives a good indication of plant well-being. If VPD is high, then the humidity is low and the stomata on the leaves tend to close in order to limit transpiration and prevent wilting. Closing of the stomata also limits the rate of CO<sub>2</sub> uptake and photosynthesis. In general, studies have found that as cloudiness increases VPD decreases. Figure 1.7 shows the decrease in VPD from the study by Gu *et al.* in 1999. Decreased VPD induces stomatal openness and thus enhances leaf photosynthesis (Law *et al.*, 2002; Gu *et al.*, 1999).





**Figure 1.7:** Changes of vapor pressure deficit (VPD) with the clearness index for selected intervals of solar elevation angles for the boreal aspen forest in 1994 (Gu *et al.*, 1999).

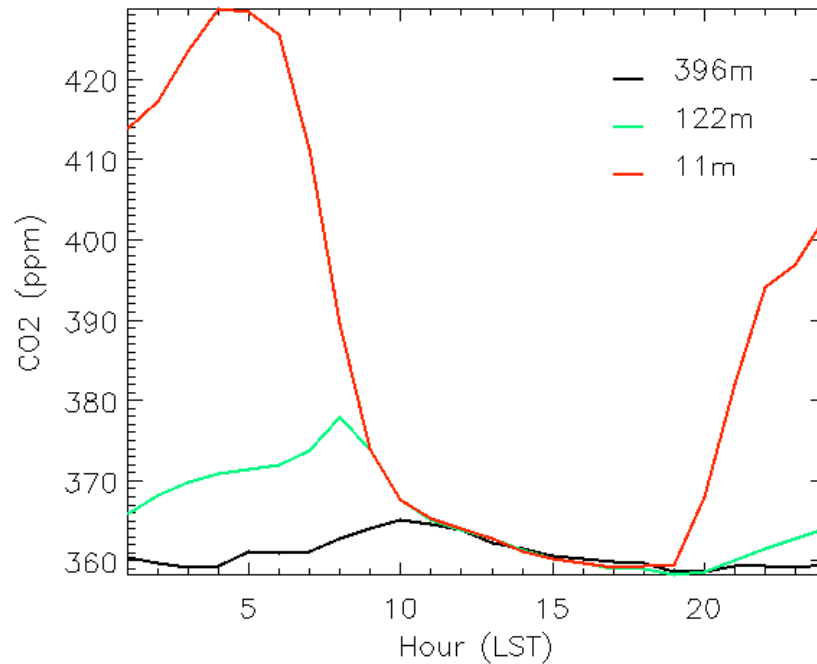
The studies investigated here show that clear skies do not provide the ideal environmental conditions for carbon uptake; instead, cloudiness, which reduces total solar radiation incident on the forest canopy, can enhance forest carbon absorption: NEE is greater on slightly overcast days rather than on sunny days, and for a given irradiance level cloudy days have a higher NEE than clear days. Increases in diffuse radiation and decreases in water stress enhancing photosynthesis and decreases in air and soil temperatures suppressing respiration all play roles in the enhancement of carbon uptake under cloudy conditions.

## 1.4 Overview of CO<sub>2</sub> Variability

Terrestrial fluxes of carbon, combined with anthropogenic sources and oceanic fluxes, lead to substantial variability of CO<sub>2</sub> on various spatial and temporal scales. This section will

discuss local atmospheric CO<sub>2</sub> concentrations, scaling up through regional mixing ratio distributions to a discussion of global CO<sub>2</sub> variability.

#### 1.4.1 Local Diurnal Variability



**Figure 1.8:** Average diurnal cycle of CO<sub>2</sub> in July at the WLEF tall tower. The CO<sub>2</sub> concentration at 396 m is shown in black, the concentration at 122 m is the red line, and the green line displays the average July diurnal cycle at 11 m.

The diurnal cycle of CO<sub>2</sub> near the land surface has been extensively explored and is driven by the diurnally varying biological source/sink and by dynamics of the planetary boundary layer (PBL) (e.g. Bolin and Keeling, 1963; Leith, 1963; Keeling *et al.*, 1976; Bakwin *et al.*, 1998; Denning *et al.*, 2003). Figure 1.8 displays the average diurnal cycle in July from continuous measurements taken at the WLEF tall tower site near Park Falls, WI. Further information regarding the tower is presented in section 2.1.1. The figure shows the CO<sub>2</sub> concentration at 396 m (black line), 122 m (red line), and 11 m (green line).

During the day, photosynthetic uptake of CO<sub>2</sub> dominates over respiration, and the PBL is well mixed by convection and wind shear to typically a 1-2 km depth (Bakwin *et al.*, 1998, Denning *et al.*, 2003). Due to photosynthesis, the CO<sub>2</sub> concentration is low in the PBL during the

day; however, the CO<sub>2</sub> depleted air at the surface is mixed into a deeper boundary layer, which increases the surface concentration. In contrast, at night the PBL is stably stratified and shallow, and respiration provides a source of CO<sub>2</sub> (Bakwin *et al.*, 1998; Denning *et al.*, 2003). Above the PBL, the CO<sub>2</sub> concentration remains relatively constant at night due to the decoupling from the surface; however, the CO<sub>2</sub> concentrations build up to very high levels beneath the nocturnal temperature inversion (Bakwin *et al.*, 1998; Denning *et al.*, 2003). In Figure 1.8, the 11 m sensor at WLEF (the green line) shows the high CO<sub>2</sub> concentrations near the surface at night, while the 396 m measurements (black line) remain close to the daytime concentrations at night because they become decoupled from the surface. Just after sunrise, a layer of very high CO<sub>2</sub> concentration remains trapped under the temperature inversion, while the air above remains at lower concentrations. As the morning progresses, the turbulent boundary layer begins to grow and the 11 m measurements show the CO<sub>2</sub> concentrations being diluted by entrainment from the residual layer as the boundary layer grows. While the PBL is growing, photosynthesis also continues to remove CO<sub>2</sub>, creating lower concentrations. As the boundary layer grows, the higher CO<sub>2</sub> is mixed with the residual CO<sub>2</sub> above it, creating an increase in the CO<sub>2</sub> measurements. In Figure 1.8, we can see this increase when the boundary layer grows past the CO<sub>2</sub> sensor at both 122 m and 396 m. The increase in CO<sub>2</sub> at 122 m is around 9 am, a few hours after sunrise, while the increase at 396 m is an hour later. By mid-morning, the depth of the turbulence exceeded the top of the tower, and the concentrations display a weak vertical gradient for the rest of the day (Denning *et al.*, 2003). The July mean diurnal cycle at the WLEF tower shows that during the summer the surface concentrations of CO<sub>2</sub> vary considerably from night to day. During the winter, the diurnal cycle of CO<sub>2</sub> is much weaker; and the main variability is not due to the biosphere, but rather due to weather and synoptic scale variability.

### **1.4.2 Regional Variability**

Due to the sources and sinks of CO<sub>2</sub> at the Earth's surface, CO<sub>2</sub> at the surface is highly variable. Heterogeneity in land cover, which includes both mixtures of land and water coverage

as well as variability in vegetation density and type, contributes to the variability in CO<sub>2</sub> concentrations. To demonstrate the CO<sub>2</sub> variability over a domain including both land and water, a study by Sun *et al.* [1998] investigates the CO<sub>2</sub> concentrations over Candle Lake in Saskatchewan Canada, which is about 90 km NNE of Prince Albert. This study investigates the effects of nocturnal land breezes and daytime lake breezes using aircraft and boat data collected during the Boreal Ecosystem Atmosphere Study (BOREAS). During the day, the air over the lake has a high CO<sub>2</sub> concentration compared to the concentration over land, which results from the downward motion induced by a daytime lake breeze combined with the lack of CO<sub>2</sub> uptake at the surface. At night, Sun *et al.* showed that the CO<sub>2</sub> concentration over the lake was highly variable near the surface. Over the center of the lake the air is warm and moist with a low CO<sub>2</sub> concentration, due to both the negligible respiration rate and the strong upward transport of high CO<sub>2</sub>; but the edges experience effects from the nocturnal land breezes. A weak eastward land breeze transports high CO<sub>2</sub> from the west coast of the lake over the western portion of the lake, where the high CO<sub>2</sub> is then transported upward through vertical mixing. The eastern edge of the lake experiences a sharp horizontal decrease of CO<sub>2</sub> due to substantial mixing associated with buoyancy-induced turbulence and strong advection. Despite the land being the source of CO<sub>2</sub> at night, the vertically integrated CO<sub>2</sub> concentration in the air over the lake is higher than the CO<sub>2</sub> concentrations over the land due to the high CO<sub>2</sub> air being horizontally advected by the nocturnal land breeze and then vertically vented over Candle Lake: at night the lake acts as a “chimney” to transport the CO<sub>2</sub>-rich air to higher levels.

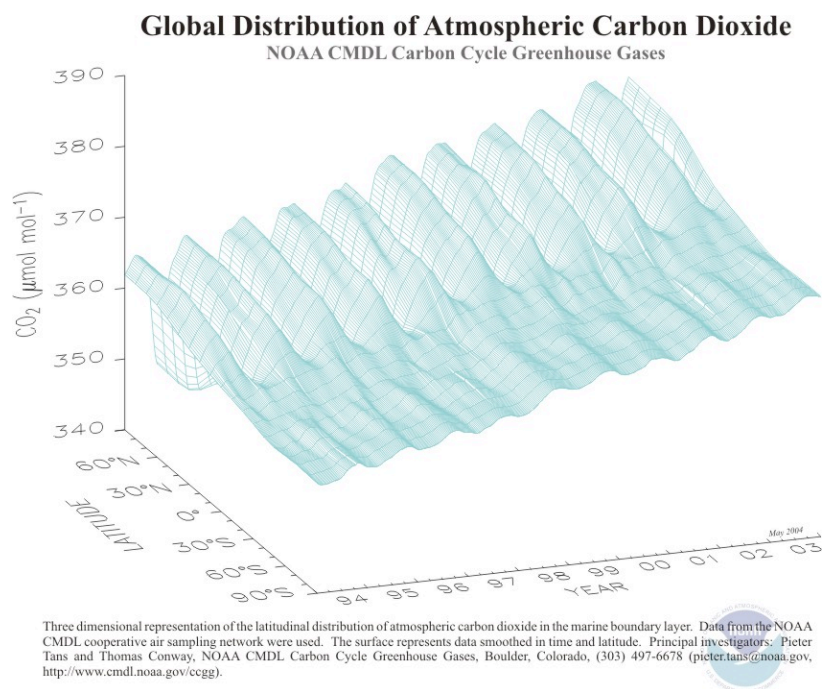
The modeling study by Nicholls *et al.* [2004] also shows that surface heterogeneity, primarily due to a combination of land and water coverage, contributes to substantial CO<sub>2</sub> spatial variability. During that simulation, the CO<sub>2</sub> surface concentrations vary by up to 70 ppm at any one particular point in time. Well-mixed, low daytime CO<sub>2</sub> air in the boundary layer is advected by the ambient westerly flow and the lake breeze over Lake Superior and Lake Michigan, where it then subsides due to the downward vertical motion from the divergence caused by the lake

breezes. Due to the lake breezes, the minimum concentration over the lakes occurs near midnight. Through the night, the higher CO<sub>2</sub> air advects over the lakes, causing the lakes to reach their highest concentration just after daybreak.

The studies discussed above illustrate the regional and local variability of CO<sub>2</sub>. The high variability in the surface mixing ratio is caused by heterogeneity in the land surface, which interacts with atmospheric motion. The studies illustrate that bodies of water, even relatively small lakes, have a significant effect on the CO<sub>2</sub> concentrations and spatial variability.

### **1.4.3 Global Variability**

Moving from a regional scale to a global scale, CO<sub>2</sub> is generally regarded as being well mixed in the atmosphere. Due to its long lifetime, the surface variability mixes with the relatively homogeneous CO<sub>2</sub> concentrations in the upper troposphere, diluting the high CO<sub>2</sub> variability near the surface and creating the well-mixed concentrations on the global scale; however, systematic variations are known to occur both latitudinally and seasonally (Conway *et al.*, 1988). Figure 1.9 shows the variations of atmospheric CO<sub>2</sub> concentration by time and latitude for 1994 through 2003, based on flask measurements collected by the National Oceanic and Atmospheric Administration Climate Monitoring and Diagnostics Laboratory (NOAA CMDL). The pronounced seasonal cycle in the northern hemisphere is due primarily to the terrestrial biosphere and follows the seasonal growth and decay of vegetation. Vegetation in the tropics experiences much weaker seasonal variability than higher latitudes and vertical mixing by convection is strong in the tropics, causing the amplitude of the surface seasonal cycle to be weaker around the equator. In the southern hemisphere, the seasonal cycle is out of phase with that in the northern hemisphere, which is expected because of the different growing period. The amplitude of the southern hemisphere is also much weaker than the northern hemisphere, which is due to the larger areal fraction of ocean.



**Figure 1.9:** Atmospheric CO<sub>2</sub> distribution by time and by latitude, from NOAA-CMDL.

## 1.5 Explaining the Interannual Variability in the CO<sub>2</sub> Growth Rate

Now we can use our knowledge about terrestrial fluxes of CO<sub>2</sub> and about the spatial variability of CO<sub>2</sub> to investigate the variability in the global growth rate of atmospheric CO<sub>2</sub>. Existing measurements and models cannot fully explain why there is considerable interannual variability in the accumulation rate of atmospheric CO<sub>2</sub> while the fossil fuel emission rates are steadily rising, although several different hypotheses exist. The variability of the atmospheric growth rate of CO<sub>2</sub> appears to be largely caused by circulation anomalies and associated changes in temperature and precipitation. One well-known circulation anomaly is the El Niño-Southern Oscillation (ENSO) phenomenon. During normal conditions, there is a huge accumulation of deep warm water in the western Pacific Ocean, while the east Pacific has cool sea-surface temperatures (SSTs) and a shallow thermocline. In contrast, El Niño years are characterized by a relaxation of the trade winds, allowing the warm pool to flow eastward. The warm SSTs, accompanied by a deeper thermocline, shift to the central and eastern Pacific. The eastward shift

in SSTs shifts convection and rainfall to the east: Asia typically experiences a drought while more rain falls in the Americas. Following El Niño events, the carbon dioxide accumulation rate increases, indicating that the strength of the land and ocean sinks decreases (Sarmiento and Gruber, 2002).

Early evidence for the connection between ENSO and the interannual variability of the carbon balance has been provided by Bacastow [1976] and Keeling *et al.* [1989]. The increase of the CO<sub>2</sub> growth rate corresponding to ENSO is thought to be due to the influence of ENSO on physiological processes in the terrestrial biosphere throughout the world [e.g. Hansen *et al.*, 1999] and to the increase of emissions due to fires brought on by drought conditions during El Niño [e.g. Jones *et al.*, 2001]. Current observations now suggest that the CO<sub>2</sub> growth rate increases due to ENSO events; however, the cause for this increase is still under investigation.

Another circulation anomaly is the northern annular mode (NAM) (e.g. Barnston and Livezey, 1987; Wallace and Thompson, 2002). The northern annular mode, also known as the North Atlantic Oscillation (NAO), is one of the most prominent patterns of atmospheric variability and is characterized by a seesaw of atmospheric mass alternating between polar and subtropical regions. These changes in mass and pressure fields lead to variability in the strength and pathway of storm systems crossing the Atlantic from the east coast of the United States to Europe. A positive NAO index phase depicts a stronger than usual subtropical high-pressure center and a deeper than normal Icelandic low. This increased pressure difference results in more and stronger winter storms crossing the Atlantic Ocean on a more northerly track: warm and wet winters in Europe, mild and wet winter conditions in the eastern US, and cold and dry winters in northern Canada and Greenland. A negative NAO index phase indicates a weak subtropical high and a weak Icelandic low, leading to a reduced pressure gradient and weaker winter storms crossing on a more west-east pathway: moist air into the Mediterranean, cold weather in Europe, more frequent cold outbreaks and snowier winter conditions on the east coast of the US, and milder winter temperatures in Greenland. A study by Russell and Wallace [2003] shows that

winters preceding anomalously high drawdown seasons exhibit patterns characteristic of the high index of the NAO. In addition, Buermann *et al.* [2003] found that features of the positive phase NAO signal include enhanced warm and green conditions over large regions in Europe and Asian Russia, which also supports enhanced drawdown. Schaefer *et al.* [2004 and 2005] confirms that winters with a high NAO index are followed by more productive growing seasons, showing that winters with a high NAO index have earlier spring snowmelt in the northern hemisphere due to increased temperatures and decreased temperatures. The earlier snowmelt is followed by earlier spring leaf-out, allowing the growing season to be longer and more productive.

One final global climate perturbation discussed in the literature that influences the growth of atmospheric CO<sub>2</sub> is volcanic eruptions. On June 15, 1991 Mt. Pinatubo erupted, injecting twenty megatons of SO<sub>2</sub> into the stratosphere, which was converted into sulfuric acid and encircled the Earth in approximately 22 days (McCormick *et al.*, 1995). Scattering of solar radiation by sulphuric acid aerosols not only increases the planetary albedo, but also increases the diffuse fraction of incident light (Molineaux and Ineichen, 1996). Although these aerosols absorb longwave radiation causing stratospheric warming, the shortwave effects dominate causing cooling of the surface (McCormick *et al.*, 1995). Sarmiento [1993] first reported a decrease of the atmospheric CO<sub>2</sub> growth rate from volcanic aerosols following the eruption of Mt. Pinatubo. He concluded that since emissions of CO<sub>2</sub> from fossil fuels had not reduced, the cause for the decline in CO<sub>2</sub> growth must be due to volcanic aerosols altering processes in the ocean or in the terrestrial biosphere. Studies using inversions of atmospheric CO<sub>2</sub>, O<sub>2</sub>/N<sub>2</sub> ratios, and carbon isotope measurements indicate that an enhanced terrestrial carbon sink explains the decline (Battle *et al.*, 2000; Bousquet *et al.*, 2000; Gu *et al.*, 2003).

Recent literature suggests that a combination of two different mechanisms accounts for the decrease in the atmospheric CO<sub>2</sub> growth rate following volcanic eruptions: enhanced photosynthesis from the increase in diffuse light caused by volcanic aerosols and reduced respiration due to cooler temperatures. The dominating mechanism appears to vary



geographically depending on the ecosystem. In regions that may not be light-limited or where other effects such as reduced growing season may cancel the enhanced photosynthesis, such as the mid-latitudes, the decreased respiration may be the dominant sink of CO<sub>2</sub> (Lucht *et al.*, 2002; Jones and Cox, 2001; Krakauer and Randerson, 2003). However, in areas that are light limited, such as the tropics or dense forests with significant undergrowth, it is reasonable that increased photosynthesis may become important (Roderick *et al.*, 2001; Reichenau and Esser, 2003; Jones and Cox, 2001). While studies investigating the influence of volcanic eruptions on the CO<sub>2</sub> growth rate, in addition to studies about the effects of ENSO and the NAM on atmospheric CO<sub>2</sub>, enhance our understanding of the sources and sinks of CO<sub>2</sub>, more research needs to be conducted before the variability in the CO<sub>2</sub> growth rate can be completely explained.

## **1.6 Locating the “Missing” Carbon Sink**

In addition to being unable to explain the interannual variability of the atmospheric growth rate of CO<sub>2</sub>, scientists are currently unable to isolate the regions responsible for the carbon sink. It is difficult to isolate the carbon sinks responsible for the “missing” carbon in the atmosphere because of the large natural fluxes of carbon. In the ocean, approximately 90 GtC/yr is exchanged with the atmosphere, while ~120 GtC/yr is exchanged between the biosphere and the atmosphere. In contrast to the large background fluxes, human activity produces 7 GtC/yr and the “missing” carbon is only 3 GtC/yr, making the missing carbon sink difficult to locate. Inverse models, isotope measurements, and oxygen concentrations are all being used to investigate the missing sink.

### **1.6.1 Inverse Modeling**

Variations in observed atmospheric CO<sub>2</sub> concentrations contain information about regional carbon sources and sinks that can be revealed via inverse modeling with atmospheric tracer transport models (Gurney *et al.*, 2002). The spatial and temporal distribution of CO<sub>2</sub>

sources and sinks produces small gradients in the observed CO<sub>2</sub> concentration distribution (Sarmiento and Gruber, 2002). In the inversion method, land and ocean surfaces are first divided into regions. Next, atmospheric transport models are used to estimate the concentration footprint that would result if a given region emitted a unit flux of CO<sub>2</sub>. Finally, statistical optimization is used to determine the weighted sum of footprints that best reproduces the observations, with the weights representing the flux contributions from each region (Sarmiento and Gruber, 2002).

The largest and most robust signal from the atmospheric CO<sub>2</sub> distribution is the north-south gradient of about four ppm between the two poles. This gradient exists because almost 90% of the global fossil fuel emissions occur north of the tropics. However, the gradient is substantially smaller than expected if the carbon sinks were evenly distributed around the world. The discrepancy between the observed gradient and the calculated gradient suggests a CO<sub>2</sub> sink located in the northern hemisphere.

To explain the northern hemisphere sink, scientists have proposed two hypotheses. Keeling *et al.* [1989] presented the first explanation, stating that the northern hemisphere sink may reflect a natural source-sink pattern of oceanic CO<sub>2</sub> fluxes. They postulated that the North Atlantic takes up CO<sub>2</sub> and transports it south as part of the large-scale ocean circulation. As the waters upwell in the southern ocean, they give the CO<sub>2</sub> back to the atmosphere. This view is supported by the early atmospheric CO<sub>2</sub> data from the 1960s (Bolin and Keeling, 1963), which do not show a clear latitudinal gradient despite the fossil fuel emissions occurring at that time (IPCC, 2001).

In contrast, Pieter Tans, Inez Fung, and Taro Takahashi showed that the northern hemisphere sink must be on land (Tans *et al.*, 1990). Using observations of the difference in the partial pressure of CO<sub>2</sub> between the atmosphere and the surface ocean in the North Atlantic and in other oceanic locations as additional constraints in their atmospheric inversions, they concluded that the Northern Hemisphere sink cannot be attributed to the ocean and thus must be driven by the terrestrial biosphere.

Investigating the two strongly diverging hypotheses on the location of the northern hemisphere sink using inverse models dominated much of the carbon cycle research in the 1990s. In these inverse modeling studies, the CO<sub>2</sub> concentrations are based on samples collected by the NOAA CMDL flask network, which collects CO<sub>2</sub> concentrations once a week at over 100 different locations primarily in remote areas. The sparse measurements forced inverse modelers to solve for the CO<sub>2</sub> flux from large regions, and the resulting uncertainty in the obtained fluxes was quite large. In addition, these studies required the use of prior estimates and prescribed patterns in both space and time, and the results were quite dependent on the values specified.

Using sixteen transport models and model variants, the TransCom 3 project investigated the estimates of surface CO<sub>2</sub> fluxes from an inter-comparison of atmospheric CO<sub>2</sub> inversion models. This study, reported in Gurney *et al.* [2002], found a northern land carbon sink that is distributed almost evenly among the continents of the Northern Hemisphere, although this result showed substantial sensitivity to the transport. The TransCom results and other recent analyses of new oceanic and atmospheric observations indicate that the northern hemispheric sink is primarily driven by the terrestrial biosphere, with a small contribution by the ocean (Sarmiento and Gruber, 2002).

Atmospheric inversion studies have raised additional questions regarding the carbon cycle. The first question concerns the Southern Ocean. Model simulations suggest that the southern ocean is a strong sink (Sarmiento and Gruber, 2002); however, the TransCom inter-comparison project found the southern oceans to be a small sink, with the CO<sub>2</sub> uptake in the southern extra-tropical ocean roughly half that estimated from ocean measurements (Gurney *et al.*, 2002). This conclusion was not sensitive to transport models or methodological approaches, raising the question of the magnitude of the southern ocean sink. Another question raised by inversion studies is the flux of carbon in the tropics. Although inverse studies have found the tropical land to be a source of CO<sub>2</sub> primarily from deforestation, due to the lack of data constraint in the tropics using the current flask network studies, these regions have large uncertainties

(Sarmiento and Gruber, 2002; Gurney *et al.*, 2002). Extended inversion models combined with an improved observational network will help answer these questions.

### **1.6.2 Partitioning the Ocean/Atmosphere Sinks Using Oxygen and Isotopes**

In addition to investigating the location of carbon sinks using inverse modeling, oxygen measurements can help partition carbon sinks. Oxygen measurements can be used to investigate the partitioning of the carbon sink between the land and the ocean because oxygen and CO<sub>2</sub> are exchanged in relatively fixed stoichiometric ratios during the burning of fossil fuels as well as during photosynthesis and respiration by plants, animals, and bacteria. In contrast, the exchange of CO<sub>2</sub> between the ocean and the atmosphere is independent of the exchange of oxygen. Since the exchange of CO<sub>2</sub> into the ocean does not involve O<sub>2</sub>, only uptake of CO<sub>2</sub> by the terrestrial biosphere will leave an imprint on atmospheric oxygen. Measuring the changes in atmospheric oxygen and knowing the fossil-fuel emissions and the exact values of the stoichiometric ratios allow the ocean uptake to be separated from the uptake in the terrestrial biosphere (Sarmiento and Gruber, 2002).

A method using oxygen to partition the carbon sink between the land and the ocean was developed by Ralph Keeling of the Scripps Institute of Oceanography (Keeling and Shertz, 1992). To make this method practical, Ralph Keeling developed an interferometric technique that allowed oxygen to be measured with sufficient precision; however, one downfall of this method is that it may have somewhat larger uncertainties than previously thought because it assumes no net exchange of oxygen between the ocean and the atmosphere (Sarmiento and Gruber, 2002). Michael Bender, who is currently at Princeton University, also developed a method for measuring the O<sub>2</sub>/N<sub>2</sub> ratio of air to a standard error of  $\pm 6$  per meg ( $\pm 0.006\%$ ) for a sample analyzed in quadruplicate, corresponding to  $\pm 1.2$  ppm O<sub>2</sub> in air out of 210,000 [Bender *et al.*, 1994]. This method involves measuring the ratio of <sup>16</sup>O<sub>2</sub> to <sup>15</sup>N<sup>14</sup>N by isotope ratio mass spectrometry. Both of these labs continue to make these measurements.

Similar to oxygen concentrations, measurements of two stable isotopes, carbon-13 and carbon-12, can also be used to separate the sink into land and ocean contributions. During photosynthesis, CO<sub>2</sub> molecules that contain the lighter <sup>12</sup>C atoms diffuse more readily into plant leaves, and the plant enzymes that facilitate the processing of CO<sub>2</sub> into organic carbohydrates have a greater affinity for <sup>12</sup>CO<sub>2</sub> (Denning, 1994). Because of these two factors, carbon that has undergone photosynthesis is depleted in <sup>13</sup>C relative to <sup>12</sup>C (Denning, 1994). Since plants derived the carbon contained in fossil fuel, atmospheric CO<sub>2</sub> from fossil-fuel consumption is slightly depleted in <sup>13</sup>C and is causing a long-term decrease of the <sup>13</sup>C/<sup>12</sup>C ratio of atmospheric CO<sub>2</sub> (Denning, 1994; Sarmiento and Gruber, 2002).

Partitioning the sink between the land and ocean reservoirs indicates that the net oceanic uptake is larger than the land uptake. In a study by Battle *et al.* [2000], recent measurements of atmospheric O<sub>2</sub> show that the land biosphere and world oceans annually sequester 1.4 ±0.8 and 2.0 ±0.6 gigatons of carbon, respectively, between mid-1991 and mid-1997. These results are slightly different from the estimates obtained using the inverse modeling approach, where the uptake was relatively evenly distributed between the land and the ocean [Gurney *et al.*, 2002]. Although we have gained knowledge of the carbon cycle by using both inverse models and oxygen and isotope measurements, it is evident that future work is required to determine the location of the carbon sinks.

## **1.7 The Orbiting Carbon Observatory (OCO)**

### **1.7.1 Overview**

Global simulations with source-sink synthesis inversion models indicate that uncertainties in the atmospheric CO<sub>2</sub> balance could be reduced substantially if data from the existing ground-based CO<sub>2</sub> network were augmented by spatially-resolved, global measurements of the column-integrated dry air mole fraction (X<sub>CO2</sub>) with precisions of ~1ppm (0.3% of 370

ppm) (Rayner and O'Brien, 2001; Pak and Prather, 2001; Engelen *et al.*, 2001; Houweling *et al.*, 2004). This information would also facilitate monitoring compliance with future CO<sub>2</sub> emissions treaties that offer credits for CO<sub>2</sub> sequestration as well as emission reductions. Due to the global coverage, space-based measurements will provide the greatest benefit in regions that are poorly sampled by existing ground-based CO<sub>2</sub> monitoring networks, and their high spatial density may also contribute to carbon cycle process studies. The Orbiting Carbon Observatory (OCO), scheduled to launch in September 2008, will make the first global, space-based measurements of atmospheric carbon dioxide with the precision, resolution, and coverage needed to characterize CO<sub>2</sub> sources and sinks on regional scales, and OCO will serve as a pathfinder for future long-term CO<sub>2</sub> monitoring missions.

The measurement approach and instrument specifications were determined through an analysis of existing carbon cycle data and a series of observing system simulation experiments. During its two-year mission, OCO will fly in a 1:15 PM sun-synchronous orbit with a 16-day ground-track repeat time. OCO will carry a single instrument that incorporates three bore-sighted high-resolution spectrometers designed to measure reflected sunlight in the 0.76  $\mu\text{m}$  O<sub>2</sub> A-band and in the CO<sub>2</sub> bands at 1.61 and 2.06  $\mu\text{m}$ . Simultaneous soundings in these three bands will be used to retrieve the column-averaged CO<sub>2</sub> dry air mole fraction ( $X_{\text{CO}_2}$ ), and a comprehensive validation program ensures that the space-based  $X_{\text{CO}_2}$  measurements have precisions of  $\sim 0.3\%$  (1 ppm) on regional scales at monthly intervals. The following sections, derived mainly from Crisp *et al.* [2004], discuss more details of the OCO measurement approach.

### **1.7.2 Spectral Range**

The OCO satellite will collect high-resolution spectra of reflected sunlight in the 0.76  $\mu\text{m}$  O<sub>2</sub> A-band and the CO<sub>2</sub> bands at 1.61  $\mu\text{m}$  and 2.06  $\mu\text{m}$ . A single sounding will consist of simultaneous, bore-sighted observations from all three bands. Each sounding will be analyzed with an algorithm that incorporates an atmospheric radiative transfer model, an instrument simulator model, and a retrieval algorithm that adjusts the assumed atmospheric state to better

match the measurements. High-resolution spectroscopic observations of near infrared CO<sub>2</sub> absorption bands in reflected sunlight were selected because they provide high sensitivity near the surface, where the CO<sub>2</sub> sources and sinks are located. The weak CO<sub>2</sub> band near 1.61  $\mu\text{m}$  was selected for three reasons. First of all, that spectral region is relatively free of absorption by other gases. Second, few of the spectral lines in that band saturate for the range of observing conditions considered in the mission, so the absorption at this band increases nearly linearly with CO<sub>2</sub> abundance and path length. Finally, thermal emissions from the atmosphere and instrument are negligible at these wavelengths, which simplifies the instrument design and radiometric calibration.

Radiance measurements at two frequencies in the CO<sub>2</sub> absorption band around 1.61  $\mu\text{m}$  alone in principle allow the CO<sub>2</sub> column to be estimated precisely in a clear atmosphere; however, in the presence of thin cloud and aerosol, measuring only CO<sub>2</sub> will not provide sufficient accuracy (O'Brien and Rayner, 2002). O'Brien and Rayner [2002] found that simulations based on a simple radiance model show that precision better than a few percent is unlikely even when the optical thickness is as low as 0.02, because scattering by cirrus shortens the mean path lengths of photons reflected to space and hence biases the estimate of the CO<sub>2</sub> column low. In the presence of modest amounts of cloud and aerosol, simultaneous measurements of O<sub>2</sub> and CO<sub>2</sub> can make the measurements more accurate. Measurements of the A-band will allow the mean scattering level to be estimated and the O<sub>2</sub> measurements can be used to characterize the vertical distribution of clouds and aerosols in each sounding, allowing the CO<sub>2</sub> column to be determined with precision approaching 0.5%. Soundings with a scattering optical depth  $\tau > 0.3$  will be rejected.

The spectral range for each channel includes the complete molecular absorption band as well as some nearby continuum, which minimizes biases due to uncertainties in atmospheric temperature and provides constraints on the optical properties of the surface albedo and aerosols. The spectral resolving power for each channel was selected to maximize the sensitivity to

variations in the column abundances of CO<sub>2</sub> and O<sub>2</sub> and to minimize the impact of systematic measurement errors (Johnson and Crisp, 2003). A spectral resolving power,  $\lambda / \Delta\lambda \sim 21,000$  separates individual CO<sub>2</sub> lines in the 1.61 and 2.06  $\mu\text{m}$  regions from weak H<sub>2</sub>O and CH<sub>4</sub> lines and from the underlying continuum. For the O<sub>2</sub> A-band, a resolving power of 17,500 is used to distinguish the O<sub>2</sub> doublets. With these resolving powers, the OCO retrieval algorithm can characterize the surface reflectance throughout the band and solve for the wavelength dependence of the aerosol scattering, minimizing X<sub>CO2</sub> retrieval errors contributed by uncertainties in the continuum level (Johnson and Crisp, 2003).

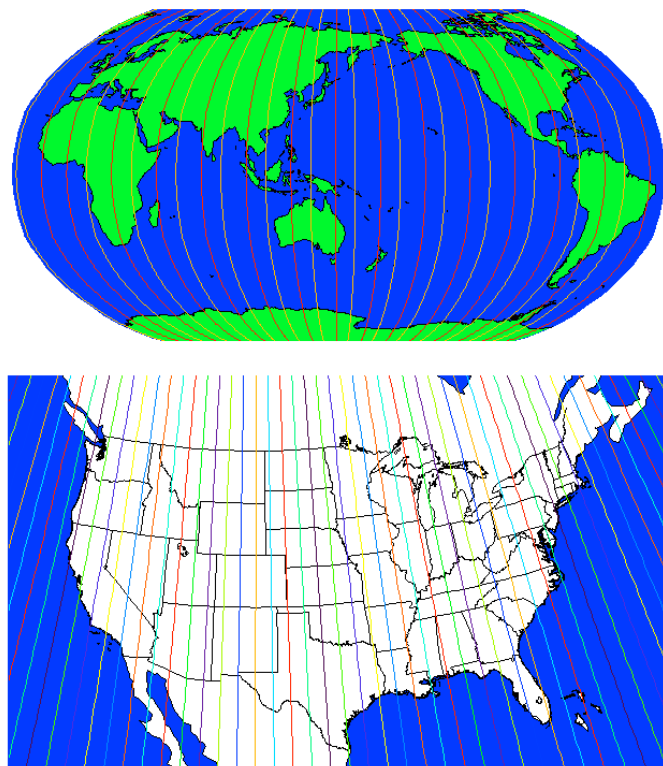
### 1.7.3 Orbit Requirements

The OCO satellite will fly in a polar, sun synchronous orbit, providing global coverage with a 16-day repeat cycle. Figure 1.10 shows a sample OCO orbit track. The top panel shows global OCO measurements from two days, with the first day's measurements in red and the second day's measurements in orange; and the bottom panel shows the coverage from the entire 16-day cycle over the United States.

OCO will fly just ahead of Earth Observing System (EOS) Afternoon Constellation (A-Train), with a 1:15 PM equator crossing time. A fixed equator crossing time samples all regions of the sunlit hemisphere of the Earth at approximately the same local time of day, which yields identical Sun-Earth-satellite observing geometry along any given latitude circle and minimizes east-west biases along a given latitude circle contributed by variations in viewing geometry. The equator crossing time of 1:15 PM was selected for four reasons: the sun is high in the sky, maximizing the signal to noise of the X<sub>CO2</sub> measurements; *in situ* data show CO<sub>2</sub> concentrations are near their diurnally averaged values at that time; the planetary boundary layer is deep and slowly varying, facilitating efforts to validate the space-based CO<sub>2</sub> observations with ground-based, flux tower, and aircraft measurements; and measurements from other satellites in the A-



Train or from satellites with similar equator-crossing times (e.g. AIRS, CloudSat, MODIS) have the potential to help improve retrievals.

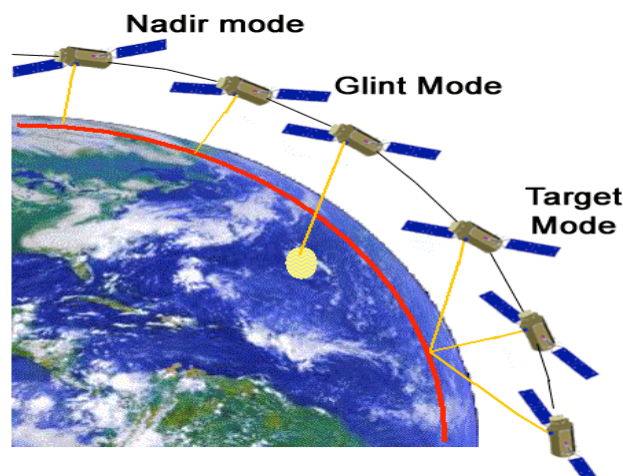


**Figure 1.10:** Sample OCO orbit track. The top panel shows the global satellite coverage after two days and the bottom panel displays the full 16-day global coverage over the United States.

#### 1.7.4 Observation Modes

OCO will use three science observation modes, which are shown in Figure 1.11. In nadir mode, the satellite will point the instrument at local nadir to collect data along the ground track. Although this mode provides the highest spatial resolution, it may not provide adequate signal to noise over dark oceans. To compensate for this problem, OCO has a glint mode: the spacecraft will point the instrument toward the bright “glint” spot, where solar radiation is specularly reflected from the surface. Glint measurements will provide a much larger signal to noise ratio over oceans. OCO will switch from nadir to glint modes on alternate 16-day global track repeat cycles. Finally, to help with validation, OCO has a target mode, which will be used to track

specific surface targets as the satellite flies overhead. Target mode will provide up to 24,000 samples per orbit over sites that include ground-based calibration assets.

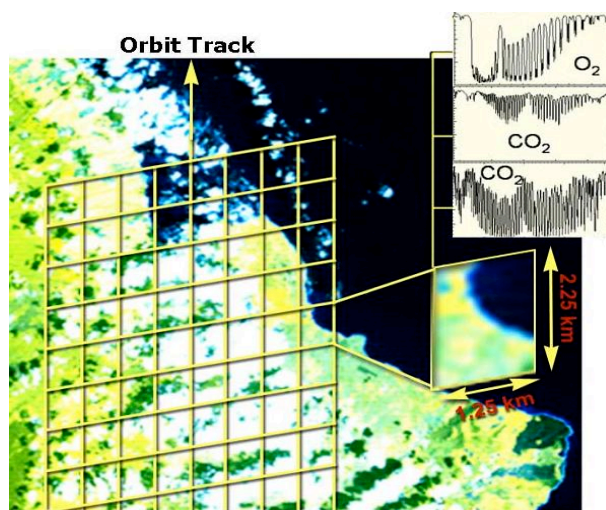


**Figure 1.11:** Illustration of the three OCO observation modes.

### 1.7.5 Spatial Sampling

While many soundings must be collected on regional scales to adequately characterize regional variations, contiguous spatial sampling is not required because CO<sub>2</sub> diffuses over a large area as it is mixed through the column (Johnson and Crisp, 2003). However, to provide useful constraints on surface CO<sub>2</sub> sources and sinks, OCO will sample the full atmospheric column. To obtain an adequate number of soundings on regional scales even in the presence of patchy clouds, each OCO spectrometer will have a 10 km-wide cross-track field of view (FOV) at nadir. The FOV is divided into eight 1.25-km wide samples, as shown in Figure 1.12. OCO will collect spectral soundings at a rate of 4.5 Hz as the satellite moves along its ground track at 6.78 km/sec, yielding a 2.25 km down-track resolution at nadir. Per degree of latitude along the orbit track, OCO will collect ~740 soundings, of which it is estimated that on average 10-20% will be clear. Although the footprint size can vary in glint and target modes, it will not exceed 10 km<sup>2</sup> even at large spacecraft nadir angles (~60°). By combining the large number of samples on regional scales, OCO will return X<sub>CO2</sub> estimates with precisions of 0.3% (1ppm). A study by Rayner *et al.*

[2002], demonstrated that this sampling approach would yield adequate constraints on  $X_{\text{CO}_2}$  even in regions occupied by patchy clouds.



**Figure 1.12:** Landsat image of Hilo Bay, Hawaii illustrating the OCO spatial sampling approach.

### 1.7.6 Calibration and Validation

For OCO, both *in situ* and remote sensing measurements will be used to evaluate the  $X_{\text{CO}_2}$  retrievals. To calibrate OCO, a system of 4-way “propagation of primary standards” is being established. The core of the surface  $\text{CO}_2$  measurement program will consist of the existing NOAA CMDL network of flask stations (Conway *et al.*, 1994). A few stations will be upgraded to provide continuous measurements throughout the diurnal cycle, and the flask data will be augmented by  $\text{pCO}_2$  measurements from moorings (e.g. the Tropical Atmosphere-Ocean and Pirana buoy arrays). The flasks are used to calibrate continuous data from both *in situ* data from the FLUXNET tower network (Baldocchi *et al.*, 2001) and from aircraft. Ongoing aircraft campaigns will be expanded to capture the vertical  $\text{CO}_2$  profiles above the designated locations in the target mode. These profiles, obtained by flying in a vertical spiral from the surface to the tropopause, will be used to test the representativeness of OCO’s  $X_{\text{CO}_2}$  concentrations at spatial scales of 10’s to 100’s of km and also to calibrate ground-based solar-viewing Fourier Transform Infrared (FTIR) Spectrometers, which also measure the total column dry-air mole fraction of  $\text{CO}_2$ . FTIRs have an extremely high signal to noise ratio, and Washenfelder *et al.* [2005] has

shown FTIRs retrieve  $X_{\text{CO}_2}$  at better than 1 ppm relative to *in situ* measurements. The total column  $X_{\text{CO}_2}$  concentrations from FTIRs will be used to calibrate and validate OCO. FTIRs, optimized for  $\text{O}_2$  and  $\text{CO}_2$  measurements, will be deployed at sites where *in situ*  $\text{CO}_2$  observations are routinely performed and at locations of various latitudes. One location will be at the WLEF tall tower site near Park Falls, WI (45.9° N, 90.2° W), one FTIR will be located at Cape Grim Baseline Air Pollution Station in Tasmania, Australia (40.4° S, 144.4° E), and another FTIR will be located at the Atmospheric Radiation Monitoring (ARM) site at Lamont, OK (36.6° N, 97.5° W). Other possible locations for OCO validation sites are the Large-scale Biosphere-Atmosphere experiment (LBA) tower site at Floresta Nacional do Tapajos, Brazil (3.4° S, 54.9° W) and the ARM site in Barrow, Alaska (71.3° N, 156.8° W). The OCO Project also plans to upgrade up to four existing FTIR spectrometers that were deployed by the Network for Detecting Stratospheric Change (NDSC). The NDSC FTIRs will be upgraded to measure  $X_{\text{CO}_2}$  in the OCO spectral regions. Finally, data from other satellite experiments will be used to place the OCO sounding in the correct chemical, spatial, and temporal context.

### **1.7.7 Sources of Error**

The  $X_{\text{CO}_2}$  fields retrieved by OCO will be used as inputs to synthesis inversion and data assimilation models. The concentrations from OCO have the potential to help reduce the uncertainty in the flux estimates from inverse models; however, to utilize OCO  $X_{\text{CO}_2}$  measurements, inversion models and data assimilation systems must correctly account for several types of sampling errors. Incorrectly accounting for sampling errors leads to misleading discussions about the value of the observations in the best case, and could lead to incorrect inversion results [Engelen *et al.*, 2002].

#### **1.7.7.1 Spatial Representativeness Errors**

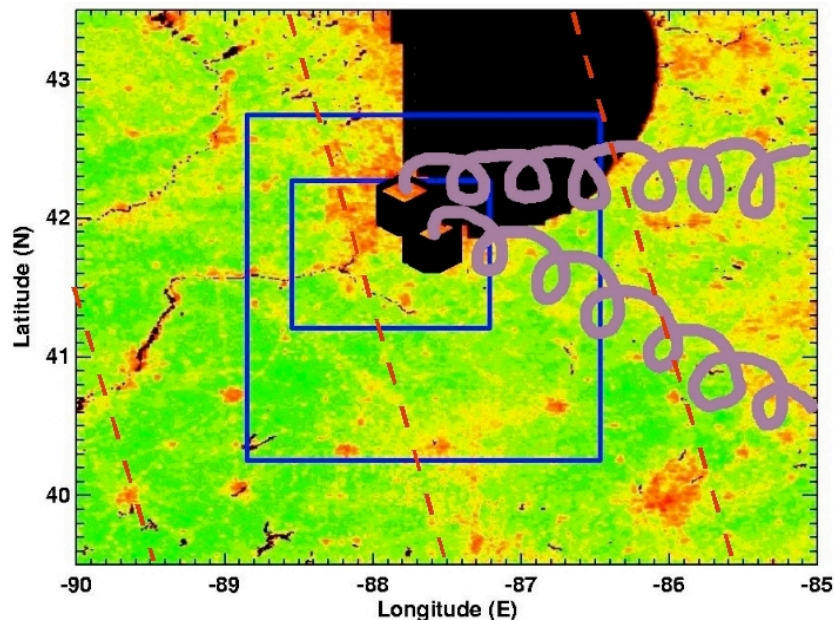
If inverse models use satellite concentrations from only a small fraction of a grid cell to optimize the concentrations in the entire grid cell, the inversion may be introducing spatial errors that can alter the resulting fluxes, as satellite mixing ratios may not be spatially representative of

larger areas. The OCO measurements, with a track width of only 10 km, will be used in inversion models with a resolution of at best 100 by 100 km. An area of this size can have considerable surface variability, as was discussed in section 1.4.2. Although combining the surface concentrations with the well-mixed concentrations in the majority of the atmosphere above the surface will dilute the surface variability, satellites will still potentially sample from an atmosphere with considerable spatial variability. If the concentrations from a single pass do not accurately represent the average concentration on the inversion model grid cell, then the OCO sampling strategy may not be representative of the grid cell and the OCO measurements will contain spatial representativeness errors.

A sample grid cell over the southern portion of Lake Michigan and the surrounding region, shown in Figure 1.13, can help illustrate potential spatial representativeness errors. The land surface in the figure is the normalized difference vegetation index (NDVI), which measures the greenness and the vigor of the vegetation. The figure shows the vegetation is quite heterogeneous in this area. The two black squares on the western side of the lake are used to represent Chicago, and the mauve spirals represent the anthropogenic emissions from the city. In the figure, two grid cells are plotted in blue. The innermost grid cell is 100 x 100 km and the outer grid cell is 2.5° x 2.5°. Satellite tracks for the full 16-day global coverage are shown by the red dashed lines.

Looking at the influence of the heterogeneity in the region on CO<sub>2</sub> concentrations, satellite tracks on the eastern portion of the domain see the lake and high fossil fuel emissions from Chicago, causing the retrieved total column CO<sub>2</sub> concentrations to be high. In contrast, tracks on the western side see green vegetation and low fossil fuel emissions, causing these concentrations to underestimate the domain-averaged mixing ratio. Inversions will adjust fluxes in this and upwind grid cells to match the satellite observations; and if the representation errors are large, inversions will have to “deweight” the satellite observations so as not to estimate significantly different fluxes depending on where the satellite track passes. Since heterogeneity

in land cover can cause considerable CO<sub>2</sub> surface variability, further investigation is required to determine if factors such as water and cities will cause significant spatial representativeness errors in OCO X<sub>CO2</sub> fields.



**Figure 1.13:** Sample grid-cells a satellite track may be representing. The NDVI is contoured on the land surface. The black squares represent Chicago, while the mauve spirals represent the anthropogenic emissions from the city. The dashed red lines are OCO tracks.

### 1.7.7.2 Temporal Errors

Another error satellite measurements might encounter is a temporal error from sampling at 1:15 PM. As we have seen in section 1.4.1, the diurnal variability near the surface can be quite large, easily reaching a difference of 60 ppm between daytime and nighttime concentrations. Although the diurnal variability in total column CO<sub>2</sub> is considerably less than the variability at the surface, the net diurnal variability of the column-averaged CO<sub>2</sub> is estimated to be ~0.5% peak-to-peak over active forested regions, with the maximum occurring approximately one hour after sunrise and the minimum about one hour before sunset (Chou *et al.*, 2002). Inversions that compare the diurnally averaged CO<sub>2</sub> concentration from each grid cell to satellite concentrations may be introducing temporal or diurnal errors because the satellite only sample the concentrations at one time. If the OCO measurements over a region are below the diurnal average and models

optimize the diurnal average to the satellite mixing ratios, then the inverse models may overestimate a sink in that region; and if the OCO measurements are above the diurnal average then either the sink from that region will be underestimated or the source from the region will be overestimated. Studies by Crisp *et al.* [2002] suggest the diurnal errors of  $X_{CO_2}$  are small and primarily negative (differences  $< 0.3$  ppm), but these systematic differences may introduce a bias that is important when integrated over an annual cycle.

### 1.7.7.3 Clear-Sky Errors

Satellite measurements of  $X_{CO_2}$  may have clear-sky errors, or sampling errors arising from retrieving  $X_{CO_2}$  only in cloud-free conditions. Two types of clear-sky errors exist: local clear-sky errors and temporal sampling errors. Local clear-sky errors will be present in inverse studies that compare concentrations in a grid cell that contains cloudy and clear regions to satellite measurements taken at the same time but sampled over only clear-sky areas, and these local clear-sky errors may be caused by changes in the vegetation behavior or by advection associated with clouds. In high latitudes during the winter, the NEE is approximately zero, so we speculate that the wintertime errors will be random; however, the clear-sky errors during the summer and over photosynthetically active regions is unknown. Recent literature that was discussed in section 1.3 shows that NEE is the greatest on slightly cloudy days, which would make the  $CO_2$  concentration the lower than average on these days. This result implies that clear-sky only retrievals will overestimate  $X_{CO_2}$  because they will not see favorable cloudy days; however, a caveat is that satellite measurements will also not see overcast days with suppressed NEE. In contrast, intuition suggests that the satellite will underestimate the average  $CO_2$  concentration during the growing season, as many ecosystems may have more active photosynthesis on sunny days than on average for all days. Although satellites may not be measuring the days with maximum  $CO_2$  uptake, they may still be systematically underestimating the total  $CO_2$  concentration, since the average concentration includes all ranges of sky conditions, from sunny days to completely overcast days when the  $CO_2$  uptake is extremely low. Since no

studies have been conducted comparing the NEE on sunny days to all days, we currently cannot explain how the ecosystem behaves on clear days compared to all days, and the clear-sky errors that OCO's  $X_{\text{CO}_2}$  measurements may experience is unknown. In addition to errors from the biology, local clear-sky errors may also be caused by advection associated with cloud cover, which also requires further investigation.

The other type of clear-sky error is temporal sampling errors, caused by under-sampling synoptic variations. Clouds are frequently associated with fronts and changes of air masses. Atmospheric transport or air mass trajectories may be systematically different on clear days, not just causing sampling errors, but actually causing a bias compared to the temporal average. Temporal sampling errors will be present in inverse studies that use satellite  $\text{CO}_2$  concentrations taken at 1:15 PM to optimize temporally averaged concentrations such as weekly or bi-monthly mixing ratios and will likely depend on the location of the measurements, as they depend on the local fluxes as well as transport. In regions with active ecosystems, the temporal sampling errors may introduce a negative bias as synoptic weather events associated with clouds, such as fronts, may systematically advect high  $\text{CO}_2$ . In contrast, regions with an upward flux of  $\text{CO}_2$  may have a positive temporal sampling bias as synoptic systems might advect low  $\text{CO}_2$  from surrounding regions. The magnitude and sign of both local clear-sky errors as well as temporal sampling errors requires further investigation.

## **1.8 Objectives of this Study**

Existing measurements and models do not currently have the capability to explain the location of the  $\text{CO}_2$  sinks, nor why there is considerable interannual variability in the accumulation rate of  $\text{CO}_2$  in the atmosphere. Due to their global spatial sampling and sheer data volume, satellite  $\text{CO}_2$  measurements have the potential to help identify sources and sinks at the regional scale. It is essential that the flux estimation procedure accurately specifies the precision



with which measurements represent the average concentration at the inversion model resolution, and systematic errors must be minimized. The previous section shows that satellite measurements have at least three potential sources of error that need to be considered before the  $X_{CO_2}$  measurements can be used in inverse modeling studies. This study will investigate these three errors introduced when using satellite measurements in inverse studies:

- 1) Spatial Representation Errors:** To what degree can one satellite track from a heterogeneous domain accurately represent the average  $CO_2$  concentration at the inversion resolution?
- 2) Temporal Representation Errors:** Will measurements at 1:15 PM accurately capture the  $CO_2$  diurnal average?
- 3) Clear-Sky Errors:** What is the sign and magnitude of the local clear-sky errors? Will the measurements have temporal sampling errors from undersampling synoptic events?

To investigate these errors, we will be using both a coupled regional ecosystem-atmosphere model (SiB2-RAMS) and continuous data. Using a model that simulates weather as well as biology produces a realistic  $CO_2$  concentration field that we can analyze. SiB2-RAMS allows us to simulate a domain of comparable size to an inversion grid cell, to sample the concentrations using the same methodology as the OCO measuring strategy, and to compare the concentrations OCO would obtain with the average concentrations over the domain. Comparing emulated satellite concentrations to concentrations on an inverse model resolution reveals the spatial, and temporal, and clear-sky errors OCO would have had if it had flown over the modeled region and contains information about the errors that can be used by inverse modelers. We will also analyze continuous *in situ* data from forests in both Wisconsin and Massachusetts to determine the clear-sky errors at those locations.

The results from these investigations will be presented in the following chapters. Chapter two covers the methods and data used in this study, including a description of the model SiB2-RAMS. Chapter three analyzes the meteorology during a SiB2-RAMS simulation and evaluates the performance of the model. Chapter four discusses the spatial errors from the model results and chapter five discusses temporal errors. Chapter six investigates both local clear-sky errors and temporal sampling errors, with the first section of the chapter discussing the results from analyzing the continuous data and the following sections of chapter five presenting the results from the SiB2-RAMS simulations. The final chapter discusses the conclusions from this study and recommendations for future work.

# Chapter 2: Methods

## 2.1 Data

To determine the clear-sky bias using continuous measurements, we will examine data from both the WLEF tall tower and from Harvard Forest. The clear days will be determined by observations from the Automated Surface Observing System (ASOS) and from the National Climatic Data Center (NCDC). In explaining the results of the clear-sky bias, we used boundary layer data from the European Centre for Medium-Range Forecasts (ECMWF). In addition, the measurements from WLEF will also be used to evaluate the model, SiB2-RAMS. This section discusses the data used in this study more in depth.

### 2.1.1 WLEF Tall Tower

The WLEF tall tower is a 450 m tall television transmission tower located in the Chequamegon National Forest, which is 24 km west of Park Falls, WI (WLEF-TV 45° 55° N, 90° 10° W). The region is in a heavily forested zone of low relief. Mixed evergreen and deciduous forests typical of the region dominate the area surrounding the tower. Much of the area was logged between 1860 and 1920, primarily for pine, and has since re-grown. Figure 2.1 shows the tower and the surrounding forest. The Climate Monitoring and Diagnostics Laboratory in the National Oceanic and Atmospheric Administration (NOAA-CMDL) as well as the Chequamegon Ecosystem Atmosphere Study (ChEAS) collect the measurements at the tower.



**Figure 2.1:** The WLEF tower and surrounding area. Photo taken by Michael L. Jensen.

Meteorological observations and measurements at the WLEF tower began in October 1994. The tower and surrounding vicinity is equipped with instruments that measure a wide variety of variables. Temperature, wind speed, wind direction, and CO<sub>2</sub> concentration are all measured at six heights on the tower (11, 30, 76, 122, 244, and 396 m above ground). At 396 m, two independent Licor CO<sub>2</sub> gas analyzers measure the CO<sub>2</sub> concentration, and the concentration at 396 m used in this study is the average between the two measurements. If one analyzer has missing data, then the concentration from the other analyzer is used, which helps to reduce data gaps. Times when both analyzers had missing data were not used in this study.

Photosynthetically active radiation (PAR), which is radiation in the visible wavelengths between .4 and .7  $\mu\text{m}$ , is also measured at the WLEF site, as is the incident downwelling radiation at the surface. The radiation measurements are made at a grassy clearing at the site that is shaded once each afternoon. In addition, surface pressure and precipitation are observed at WLEF.

The net ecosystem exchange (NEE), sensible heat flux, and latent heat flux have all been computed using eddy covariance methods from observations at the tower of both the associated turbulent fluxes (i.e. CO<sub>2</sub>, heat and water vapor, respectively) and the rate of change of storage for the respective variables below the sensor (Davis *et al.*, 1996; Davis *et al.*, 2003). The

turbulent fluxes are measured at three heights (30, 122, and 396 m). Because of the multiple level and high altitude flux and mixing ratio data, as well as a clearing around the base of the tower, ChEAS has developed an algorithm that selects one or more of the multiple flux measurements for each hourly flux value. Data are taken from 122 and 396 m under strongly unstable conditions (surface heat flux is larger than or equal to  $100 \text{ Wm}^{-2}$ ) and from 30 m under slightly unstable to stable conditions. If the preferred levels are missing, data are taken from any existing turbulent flux level; and one or two missing hours for the three fluxes are interpolated. Flux measurements at the tower are available from 1997 through 2001.

### **2.1.2 Harvard Forest**

Groups from the Atmospheric Sciences Research Center (ASRC) and Harvard University make nearly continuous atmospheric turbulence and trace gas measurements at the Environmental Measurement Site (EMS) in Harvard Forest, which is a 3,000 acre forest in north-central Massachusetts (EMS  $42^{\circ} 32' \text{ N}$ ,  $72^{\circ} 10' \text{ W}$ ). The forest is primarily mixed hardwood and conifer and contains a mixture of red oak, red maple, hemlock, and spruce. A view of the tower and the surrounding forest is shown in Figure 2.2. The first measurements at Harvard Forest were taken in 1992.



**Figure 2.2:** The tower at Harvard Forest. Photo courtesy of Harvard University.

CO<sub>2</sub> concentrations, CO concentrations, temperature wind speed, wind direction and PAR are some of the variables measured at 29 m on the tower. CO<sub>2</sub> flux as well as canopy carbon storage are also observed. For this study, we calculated NEE by subtracting the storage measurements from the turbulent CO<sub>2</sub> flux.

### **2.1.3 ASOS**

To determine if a day is clear at both Harvard Forest and WLEF, we will use data from the Automated Surface Observing System (ASOS), which has stations throughout the United States that were deployed by the National Weather Service. These stations measure cloud conditions below 12,000 ft every minute using a laser beam ceilometer operating in the near infrared. An algorithm employing Light Detection and Ranging (LIDAR) principles calculates the cloud coverage and categorizes the conditions every half hour as clear (cloud cover < 5%), few (> 5% and < 25%), scattered (> 25% and < 50%), broken (> 50% and < 88%), and overcast. The nearest ASOS station to WLEF is the Rhinelander-Oneida County Airport, which is approximately 65 km southeast of the tower (RHI 45° 38' N, 89° 28' W). Data are available at RHI from 1999-2001. At Harvard Forest, the nearest ASOS station is Orange Municipal Airport, which is only 9 km from the tower (ORE 42° 34' N, 72° 17' W). Data at ORE are available from 1997-2002.

### **2.1.4 Climatic Data**

We also used data from the National Climatic Data Center (NCDC) to determine clear-sky days, as the NCDC records the average number of clear days during each month for various cities throughout the United States. The monthly number of clear days is based on at least 40 years of data from major weather observing stations. In determining the clear days, only daytime hours were considered and a human observer categorized the day as clear if average cloud cover was less than 30%. For WLEF, the nearest cities are Green Bay, WI (232 km); Duluth, MN (175 km); and Minneapolis/St. Paul, MN (262 km). We used an average of the number of clear-sky

days from all three cities in this study. At Harvard Forest, the closest city is Worcester, MA, which is 45 km away. Although the cities are quite far away, especially at WLEF, the number of clear days at these cities still provides an accurate measure of the climatological number of sunny days in both regions. For both towers, the number of clear days per month was converted to a monthly percentage by dividing the clear days by the length of each month. Table 2.1 shows the monthly percentage of clear days from the NCDC climatological data at both WLEF and Harvard Forest.

Month	% of clear days	
	WLEF	Harvard
January	26	23
February	24	24
March	23	26
April	20	23
May	23	19
June	27	20
July	26	19
August	26	26
September	27	30
October	26	26
November	19	23
December	19	26

**Table 2.1:** The monthly percentage of clear days from the climatological data at both towers.

### 2.1.5 ECMWF Boundary Layer Heights

This study uses boundary layer heights from the ECMWF 40-year Re-Analysis (ERA-40), which was provided by ECMWF. For the horizontal resolution, ERA-40 uses a reduced Gaussian grid with approximately 125 km spacing for surface fields. This study uses the grid cell that includes the tower for both locations. Since the ERA-40 data assimilation produced four analyses per day, this study will use the boundary layer heights at 18 UTC, when the heights are closest to the maximum daytime depth. ERA-40 is available from September 1957 through August 2002. The length of the ECMWF time-series used in this study is the same as close as possible to the length of the CO<sub>2</sub> data available from each tower, which is from January 1995 through August 2002 at WLEF and from January 1993 through August 2002 at Harvard Forest.

## **2.2 Methods for Determining the Clear-Sky Bias from Observations**

### **2.2.1 Data Preparation**

Continuous data from both WLEF and Harvard Forest are used to determine the clear-sky bias. At WLEF, this study uses CO<sub>2</sub> concentrations at 396 m and PAR measurements from 1995 through 2003 and NEE measurements from 1997 through 2001. The Harvard Forest measurements we will use are the CO<sub>2</sub> concentrations at 29 m, the PAR measurements, NEE, and CO concentrations, and all these variables are available from 1993 through 2002. All outliers beyond three standard deviations of the mean of all the data were removed from each time-series, and this study requires that both CO<sub>2</sub> concentrations and PAR measurements must be valid for each hour, otherwise the measurements are not used. When investigating the NEE, CO<sub>2</sub>, PAR, and NEE all must be valid, and for CO the CO<sub>2</sub>, CO and PAR measurements must all exist. In addition, we detrended the CO<sub>2</sub> concentrations.

### **2.2.2 Sampling Times**

The first step to mimicking the OCO measuring strategy using continuous observations is to sample appropriate time periods. Since OCO will be tracking south collecting clear measurements with an equator crossing time of 1:15 PM, we analyzed two different time periods. The first time period we investigated is from 11 AM to 4 PM. For this time period, the observations were the mean of the measurements over these six hours, and this time period was selected because it represents the bias a satellite will have over areas of persistent clear skies by using time as a substitute for space. In CO<sub>2</sub> inversions, the satellite measurements will be used to represent a larger area, such as a 100 km by 100 km grid cell. As the satellite tracks south collecting measurements, these clear-sky concentrations will be averaged to create one value for the entire grid cell. Since we are at a stationary location, averaging the concentrations over six hours will sample the same air as OCO would sample in 100 km if the air over the site travels ~



4.6 m/s: taking a series of measurements over 100 km and averaging those measurements would be the same as sampling the air from the tower for 6 hours because in that amount of time the air travels 100 km.

We chose 11 AM as the start time for the six-hour time period by investigating the boundary layer. When OCO measures CO<sub>2</sub> at 1:15 PM, the boundary layer will be quite deep. A study by Zhang [2002] describes the normal diurnal cycle of the boundary layer at the WLEF tower: the boundary layer grows rapidly from 7 AM until 12 PM, remains deep and relatively unchanged during the afternoon, and finally collapses to a shallow nighttime stable layer in the evening around 6 PM. The time period from 11 AM to 4 PM should have a relatively constant boundary layer to help reduce mixing effects that may alter the concentrations at the towers compared to total column concentrations while still being centered near 1 PM.

The second time period we will analyze is 1 PM, which was selected for two reasons. First, it is close to the time OCO will make its measurements and should capture the biological activity at that snapshot in time. Second, investigating the bias at 1 PM will allow us to examine the bias OCO will have by being able to take measurements through holes in the clouds, since the sky-cover must only be clear for that one hour.

#### **2.2.4 Creating the Clear Subset**

We created a clear-sky subset by selecting mid-day values (defined as 1 PM and 11AM-4PM means) with PAR measurements higher than a threshold value defined by month for each site. The threshold clear-sky PAR values used were set as the percentile of ranked measurements from all years at each site corresponding to the percentage of clear days for each month recorded by NCDC. The monthly percentages of clear days at both WLEF and Harvard Forest are shown in Table 2.1; however, we decreased the percentages shown by 5%. This decrease helps ensure that our study does not overestimate the clear-sky CO<sub>2</sub> bias by including partly cloudy days that may have enhanced CO<sub>2</sub> uptake, as recent literature has suggested that

NEE is greater on slightly cloudy days rather than on clear days [Gu *et al.*, 1999; Freedman *et al.*, 2002; Law *et al.*, 2002].

To confirm the clear-sky subset, we analyzed the clear days from the nearby ASOS stations. Using ASOS data, 11 AM to 4 PM was considered clear if all six hours were clear; and to qualify as clear for the 1 PM analysis, only that single hour had to be clear. We correlated the clear subset using the 3-year ASOS record at WLEF and the 6-year ASOS record at Harvard Forest with the clear subsets for the corresponding years created using the climatological data. At both towers and for both times, the clear ASOS subset was significantly correlated (above the 95% level) with the previous method. Since the ASOS data does not cover the complete time-span of either tower, we did not evaluate the clear-subset created from the ASOS data; but, the fact that it is significantly correlated to the method we analyzed helps to validate the accuracy of the clear-sky subset created using monthly percentages from NCDC.

### 2.2.5 Calculating the Clear-Sky Bias

To determine the clear-sky bias in CO<sub>2</sub>, NEE, and CO, we created a smoothed function to represent the data by fitting the equation

$$a_o + \sum_{i=1}^{nharm} e^{ikt} = a_o + \sum_{i=1}^{nharm} A_k * \cos(2\pi * \frac{t}{365}) + \sum_{i=1}^{nharm} B_k * \sin(2\pi * \frac{t}{365}) \quad (1)$$

to both the complete data sets as well as the corresponding clear subsets using a linear least squares method. To simplify this calculation, if the year was a leap year, then the data for February 29 was removed.

To determine the number of harmonics needed to best fit the data, we fit equation (1) to the both CO<sub>2</sub> and NEE data from 11 AM to 4 PM at both WLEF and Harvard Forest using one, two, and three harmonics. In addition to evaluating the fit by sight, the reduced  $\chi^2$  value, which indicates the goodness of the fit, for each location was calculated using the equation

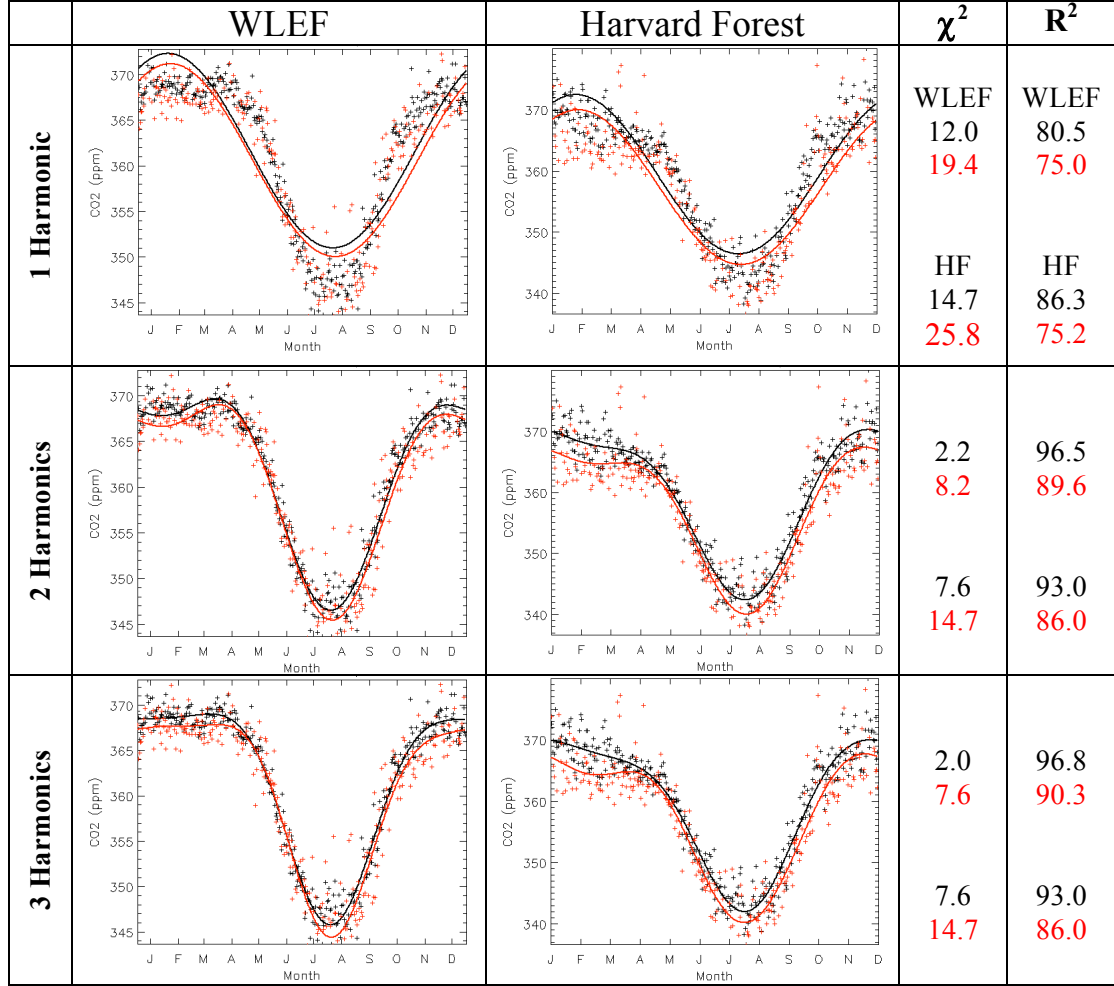
$$\chi^2 = \frac{1}{N-m} * \sum_{i=1}^N \frac{(y_i - f(x_i))^2}{\sigma^2}, \quad (2)$$

where  $y$  is the CO<sub>2</sub> dataset,  $f(x)$  is the fit to the data,  $\sigma$  is the uncertainty in the measurements,  $N$  is the total number of points in the dataset, and  $m$  is the number of parameters used in the fit. In this case, the number of parameters solved for in the fit is one plus the number of harmonics multiplied by two, and in this study  $\sigma$  is set to 1. The ideal value of the reduced  $\chi^2$  is 1, which means that the typical deviation between a data point and the fit line should be about one standard deviation. Another statistic that can help identify how goodness of the fit is the fraction of the variance explained by the fit, or  $R^2$ . The fraction of the variance explained is calculated as

$$R^2 = \sum_{i=1}^N \frac{(f(x_i) - \bar{y})}{(y_i - \bar{y})}, \quad (3)$$

where  $\bar{y}$  is the mean of the CO<sub>2</sub> measurements.  $R^2$  represents the percent of the variance that is explained by the fit and has been converted from the fractional value into a percent by multiplying by 100.

Figure 2.3 shows the results from both the complete timeseries and the clear-sky subset from 11 AM to 4 PM: the fit and statistics for one harmonic at both WLEF and Harvard Forest is shown in the first row, the results using two harmonics is in the middle row, and the bottom row displays the results using three harmonics. Both the total timeseries and the clear-sky subset were condensed into a single year with only one value per day to make the plots easier to see. The black symbols depict the CO<sub>2</sub> values from the complete timeseries, the black line shows the fit to the total timeseries, the red symbols depict CO<sub>2</sub> values from the clear-sky subset, and the red line shows the fit to the clear-sky subset. Both the  $\chi^2$  and the  $R^2$  values are worse for the clear-sky dataset than for the complete timeseries, which is most likely due to a combination of the reduced number of datapoints in the subset combined with high variability on clear days; however, the overall fits and the improvement of the fits between using 1, 2, or 3 harmonics is very similar for both the total data and the clear subset at both towers.



**Figure 2.3:** The fit and residual to the complete CO<sub>2</sub> timeseries (black) and to the clear-sky subset (red) from 11 AM to 4 PM for both WLEF (first column) and Harvard Forest (second column). The third column shows the  $\chi^2$  values for both WLEF (top numbers) and Harvard Forest (bottom numbers), while the fourth column displays the variance explained at WLEF (top) and Harvard Forest (bottom).

Looking at the fits more closely, using one harmonic only captures the seasonal cycle at both locations, and we can see from the figure that the fits deviate substantially in the spring and fall. In addition, the  $\chi^2$  values are quite high and the percent of the variance that is explained is relatively low. Moving on to two harmonics, shown in the middle row, the fits have improved substantially over only using one harmonic. The functions now capture the minimum concentrations in the summertime, as well as the concentrations through the transition periods in spring and fall and the residuals to this fit are primarily random. Both the diagnostic statistical values have also improved significantly: the  $\chi^2$  values have decreased at both locations and

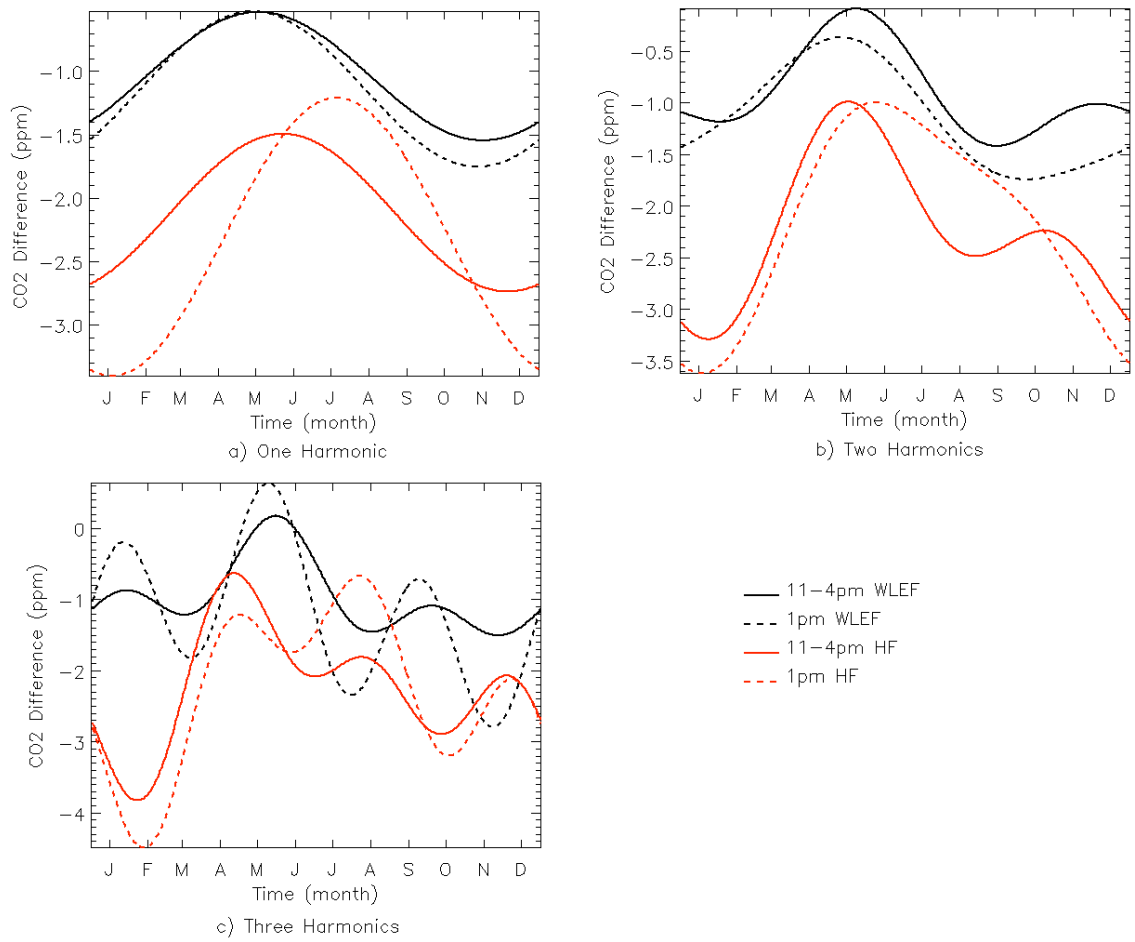
explained variance is now above 90% at both towers for the complete 11-4 PM CO<sub>2</sub> dataset. Finally, using three harmonics yields only minor improvement over the fit obtained from using two harmonics. Changes in the fit and in the residual are not detectable by sight, and the  $\chi^2$  and  $R^2$  values improved only slightly.

We also performed a similar analysis looking at the fit to the NEE data using one, two, and three harmonics. These results were identical: one harmonic only captured the seasonal cycle and the residuals had a sinusoidal shape indicating that the fit was not capturing important variability, using two harmonics improved the fit substantially, and using three harmonics offered only minor improvement over using only two harmonics. Based on these analyses, two harmonics is the optimum number of harmonics at both locations to capture the dominant variability in the time-series.

Once a function was fit to the data, we calculated the clear-sky bias by subtracting the fit to the complete dataset from the fit to the clear-sky subset. The clear-sky bias for CO<sub>2</sub>, NEE, and CO at both towers and at both times was calculated in this manner.

To extend our analysis of the preferred number of harmonics, we investigated the clear-sky CO<sub>2</sub> bias at both WLEF and Harvard Forest using one, two, and three harmonics. Figure 2.4 displays the resulting clear-sky CO<sub>2</sub> bias. The figure illustrates that the results remain similar no matter how many harmonics are used to calculate the fits: at both towers the CO<sub>2</sub> clear-sky bias is greatest in the winter and least in the summer. Using one harmonic produces a simplified sinusoidal shape with only one minimum bias in the summer and one maximum bias that are located in the winter months. Using two harmonics only adds minor “wiggles” in the 11 AM to 4 PM biases, and the overall shape seen from using only one harmonic is maintained. However, using two harmonics begins to alter the magnitudes of the bias, as the winter bias becomes greater, particularly in March and October, while the summertime bias shrinks even more. Looking at the biases generated fitting the data to three harmonics, again the overall shape is still present but additional wiggles have been added, altering the magnitudes even more. Using three

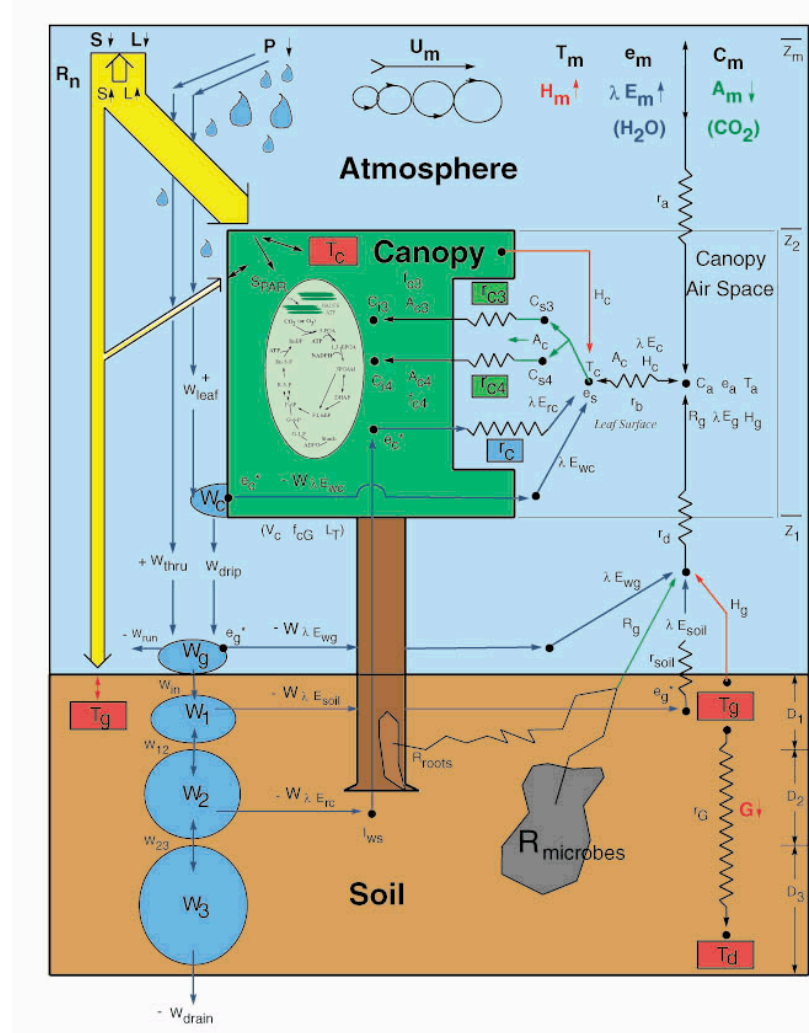
harmonics, severe biases exist around February and October, the remainder of the fall and spring months has only a moderate bias, and the bias in the summer becomes negligible. Since three harmonics appears to add considerable noise to the biases, investigating the biases produced with different numbers of harmonics confirms that using two harmonics to represent the data is reasonable and does not loose any information provided by additional harmonics. Both the fit analysis and the bias investigation indicate using a fit with two harmonics is suitable, so this study will fit the total data and the clear-sky subsets with equation (1) using two harmonics.



**Figure 2.4:** The clear-sky CO<sub>2</sub> bias from fitting the data using one harmonic (left panel), two harmonics (middle panel), and three harmonics (right panel). The solid black line is the bias clear-sky bias from 11 AM to 4 PM at WLEF and the dashed black line is the clear-sky bias at 1 PM at WLEF. The solid red line displays the bias at Harvard Forest from 11 AM to 4 PM and the dashed red line displays the bias at Harvard Forest at 1 PM.

## 2.3 Model Description: SiB2-BRAMS

### 2.3.1 SiB2



**Figure 2.5:** Depiction of the SiB2 structure.

The biosphere model used in this study is the Simple Biosphere (SiB) Model developed by Sellers *et al.* [1986]. SiB has undergone substantial modification and is now referred to as SiB2 (Sellers *et al.* 1996a,b; Denning *et al.*, 1996, 2003; Schaefer *et al.*, 2002, Nicholls *et al.*, 2004). SiB2 calculates the transfer of energy, mass, and momentum between the atmosphere and the vegetated surface of the earth. The structure of SiB2 is shown in Figure 2.7. SiB2 has a single vegetation canopy layer and three soil layers. The surface soil layer, w1, is a thin layer

from which there can be a significant rate of withdrawal of water by direct evaporation into the air when the pores of the soil are at or near saturation. The roots access the soil moisture from the second layer, w2. The third layer, w3, acts as the underlying recharge layer within which the transfer of water is governed only by gravitational drainage and hydraulic diffusion. Each soil layer has an associated soil moisture store, w, and the three stores are represented in the figure by the blue circles in the soil.

The vegetation in SiB2 is represented by one of thirteen cover types or biomes, while the soil texture in SiB2 is one of twelve categories based on the soil properties within the United States Department of Agriculture (USDA) texture classes. Table 2.2 displays the biome types and soil classes in SiB2. The morphological parameters used in SiB are given in Table 3 and Table 5 of Sellers *et al.* [1996b].

Vegetation/Biome Types			Soil Classes			
Class	Type	Name	Class	Name	% Clay	% Sand
1	C3	Tall Broadleaf-Evergreen Trees	1	Sand	3	92
2	C3	Tall Broadleaf-Deciduous Trees	2	Loamy Sand	5	82
3	C3	Tall Broadleaf and Needleleaf Trees	3	Sandy Loam	10	65
4	C3	Tall Needleleaf Trees	4	Silt Loam	13	22
5	C3	Tall Needleleaf-Deciduous Trees	5	Silt Loam	7	7
6	C4	Short Vegetation/Pasture	6	Loamy Sand	18	42
7	C4	Short Veg. with Maize Optical Properties	7	Sandy Clay Loam	28	58
8	C4	Short Veg. with Maize Optical Properties	8	Sandy Clay Loam	40	52
9	C3	Short Broadleaf Shrubs with Bare Soil	9	Clay Loam	39	32
10	C3	Short Ground Cover (Tundra)	10	Silty Clay Loam	39	10
11	C4	No Vegetation (Low Latitude Desert)	11	Silty Clay Loam	41	7
12	C3	Agriculture (wheat) and Grasslands	12	Clay Loam	65	19
13		Ice				

**Table 2.2:** SiB2 classification of vegetation and soil.

The yellow arrow on the left hand side of Figure 2.5 represents the net radiation. To describe the interception, reflection, transmission, and absorption of radiation by vegetation and soil, SiB uses the two-stream approximation model originally described by Coakley and Chylek [1975] and extended by Sellers [1985]. SiB2 requires the magnitudes of five components of the downward radiation flux: visible direct beam radiation, visible diffuse radiation, near infrared



direct beam radiation, near infrared diffuse radiation, and thermal infrared diffuse radiation. Using the radiation input at the surface from the radiative transfer parameterization in RAMS, SiB solves the two-stream approximation model and calculates the radiation absorbed by the canopy and soil from each incident component. The reflected shortwave and emitted thermal infrared fluxes are returned to RAMS, which then calculates the rate of surface emissions through the atmospheric column

Fluxes of water vapor, sensible heat, and CO<sub>2</sub> are expressed as differences in potentials divided by resistances (Sellers *et al.*, 1997). The right side of Figure 2.5 displays the transfer pathways for temperature, water vapor, and carbon dioxide. The figure shows how the sensible heat fluxes from the canopy and the ground must traverse the aerodynamic resistances  $r_b$  or  $r_d$  and  $r_a$ . Canopy water vapor and CO<sub>2</sub> fluxes must traverse an additional resistance  $r_c$ . Since it is assumed that water vapor and CO<sub>2</sub> exchanges occur from only one side of the leaf, the boundary-layer resistance is doubled for water vapor ( $2r_b$ ) and is more than doubled for CO<sub>2</sub> ( $2.8r_b$ ).

The sensible heat flux in the canopy air space,  $H_c$ , is given by

$$H_c = \rho c_p \frac{(T_c - T_a)}{r_b} = \rho c_p g_b (T_c - T_a), \quad (4)$$

where  $T_c$  and  $T_a$  are the temperature at the leaf surface and in the canopy air space respectively,  $r_b$  is the resistance,  $g_b = 1/r_b$  is the leaf boundary layer conductance in units of  $\mu\text{mol}/\text{m}^2/\text{s}$ , and  $\rho$  and  $c_p$  are the density and specific heat of air. For water vapor, the latent heat flux is expressed as

$$\lambda E_c = \frac{\rho c_p}{\gamma} g_b (e_s - e_a) = \frac{\rho c_p}{\gamma} g_s (e_i - e_s), \quad (5)$$

where  $e_a$ ,  $e_s$ , and  $e_i$  are the water vapor partial pressure in the canopy air space, at the leaf surface, and inside the leaf (saturated), respectively,  $g_s = 1/r_s$  is the stomatal conductance, and  $\gamma$  is the psychrometric constant. The leaf interior is always assumed to be saturated with respect to water vapor, so that  $e_a/e_s = h_s$ , which is the relative humidity at the leaf surface. Similarly, the diffusive flow of CO<sub>2</sub> into the stomata (net assimilation,  $A_n$ ) is expressed as

$$A_n = \frac{g_b}{1.4} * \frac{(c_a - c_s)}{p_s} = \frac{g_s}{1.6} * \frac{(c_s - c_i)}{p_s}, \quad (6)$$

where  $A_n$  is the assimilation of a single plant leaf and is expressed in  $\mu\text{mol CO}_2/\text{m}^2/\text{s}$ ,  $p_s$  is the surface pressure in Pa, and the constant factors 1.4 and 1.6 account for the different molecular diffusivities of  $\text{CO}_2$  and water vapor in the leaf boundary layer and the stomatal pores, respectively (Denning, 1994). The concentration of  $\text{CO}_2$  in the leaf interior,  $c_i$ , is calculated from the leaf surface concentration and the assimilation rate using

$$c_i = c_s - \frac{1.6 A_n}{g_s} p_s. \quad (7)$$

The photosynthesis model of Farquhar *et al.* [1980] and the stomatal model of Ball [1988], which was expanded by Collatz *et al.* [1991,1992], are the basis for the photosynthesis-conductance model in SiB2. Photosynthesis and conductance are explicitly connected by the Ball-Berry relationship, which is written as

$$g_s = m \frac{A_n h_s}{c_s} p + b, \quad (8)$$

where  $m$  and  $b$  are empirical coefficients prescribed for each biome from observations. The slope  $m$  is unitless, and the intercept  $b$ , in units of  $\mu\text{mol}/\text{m}^2/\text{s}$ , represents the resistance when the stomata are completely closed. Equation (5) shows that leaf conductance for the influx of  $\text{CO}_2$  and the simultaneous efflux of water are directly linked through a simple dependence on assimilation, relative humidity,  $\text{CO}_2$  concentration, pressure, and two vegetation dependent constants.

In the model, the leaf assimilation rate is difference between the minimum of three limiting factors and the leaf maintenance respiration. This can be expressed as

$$A_n \leq \text{Min}(\omega_c, \omega_e, \omega_s) - R_d, \quad (9)$$

where  $\omega_c$  is the Rubisco (leaf enzyme) limited rate of assimilation,  $\omega_e$  is the light-limited rate of assimilation,  $\omega_s$  is the carbon compound export limitation, and  $R_d$  is the leaf respiration rate in  $\mu\text{mol}/\text{m}^2/\text{s}$ . The Rubisco-limited rate is primarily a function of the leaf's enzyme reserves, which

can be thought of as the biochemical processing of the leaf. The light-limited rate is a function of the amount of PAR captured by the chlorophyll. The export limitation,  $w_s$ , describes the capacity of the leaf to export or utilize the products of photosynthesis, and the capacity to utilize the immediate products of photosynthesis scales linearly with the Rubisco concentration in the leaf. In SiB2, the rate limits are smoothed using a quadratic to avoid abrupt transitions from one limitation to another.

Equations (4) through (9) are only relevant to a single leaf with known physiology, physical properties, and forcing conditions. To describe the canopy photosynthesis,  $A_c$ , and the canopy conductance,  $g_c = 1/r_c$ , the equations must be integrated over the depth of a vegetation canopy. Sellers *et al.* [1992] applied Beer's Law for interception and absorption of PAR to express the net assimilation and stomatal conductance for the entire canopy as closed-form integrals that are solved analytically. Assuming that plants allocate nitrogen to the leaves with the most light and that the canopy depth profile of leaf nitrogen and  $V_{\max}$  follow the time-mean radiation-weighted profile of PAR, the photosynthetic rate and conductance of an entire canopy can be estimated by multiplying a calculation of the performance of the uppermost leaves in the canopy, which are exposed to the maximum incident PAR flux and have the highest photosynthetic capacity, by a canopy PAR use parameter  $\Pi$  (Sellers *et al.*, 1997).  $\Pi$  is driven by satellite data and is calculated as

$$\Pi = \int_0^{L_T} e^{-kL} = \frac{1 - e^{-\bar{k}L_T}}{\bar{k}} = \frac{\overline{FPAR}}{\bar{k}}, \quad (10)$$

where  $L$  is the leaf area index (leaf area per unit ground surface area, LAI),  $L_T$  is the total LAI,  $k$  is the extinction coefficient for PAR flux in the canopy,  $FPAR$  is the fraction of PAR absorbed in the canopy, and the over-bars indicate time-mean values of the quantities. The quantities in this expression can be estimated from spectral vegetation indices, which are combinations of the radiances in the visible and near-infrared regions as observed over vegetated land surfaces by

satellite sensors (Sellers *et al.*, 1997). In this study, the quantities are calculated using the Normalized Difference Vegetation Index (NDVI). Using the canopy integration factor allows the assimilation in the entire canopy to be expressed as

$$A_c = A_{n0} * \Pi, \quad (11)$$

where  $A_{n0}$  is equal to  $A_n$  for top green leaves. The stomatal conductance for the canopy,  $g_c$ , is calculated using the expression

$$g_c = m \frac{A_c h_s}{c_s} p + b. \quad (12)$$

The soil respiration in SiB,  $R_g$ , is calculated as

$$R_g = R_f * S_s \quad (13)$$

where  $R_f$  is a respiration scaling factor and  $S_s$  is the soil scale, which is a function of both temperature and moisture. The scaling factor is an indication of how much carbon is in the soil to respire. Since SiB is a balanced model, the respiration over a year must equal the assimilation in that year, which can be written as

$$\sum_{1\text{year}} R_g = \sum_{1\text{year}} A_n. \quad (13)$$

The respiration factor can then be calculated by dividing the annual assimilation by the annual sum of the soil scale, or

$$R_f = \frac{\sum_{1\text{year}} A_n}{\sum_{1\text{year}} S_s}. \quad (14)$$

### 2.3.2 RAMS

The meteorological model coupled to SiB2 is the Brazilian version of the CSU Regional Atmospheric Modeling System (RAMS). RAMS is a comprehensive mesoscale meteorological modeling system designed to simulate atmospheric circulations spanning in scale from hemispheric scales down to large eddy simulations of the planetary boundary layer. RAMS has

been applied to the simulation of flows at scales as small as buildings and is aptly suited for studying the interactions between the atmosphere and terrestrial ecosystems (Nicholls *et al.*, 1993; Nicholls *et al.*, 1995; Denning *et al.*, 2003, Nicholls *et al.*, 2004). RAMS solves the equations of motion, radiative transfer, and thermodynamics for specified regions, and the equations set used is the quasi-Boussinesq non-hydrostatic equations described by Tripoli and Cotton [1982]. The model is discussed in detail by Pielke *et al.* [1992] and Cotton *et al.* [2002].

A significant feature of RAMS is the incorporation of a telescoping nested-grid scheme, which allows the model to solve the equations simultaneously on any number of interacting computational meshes of differing spatial resolution. Based on the two-way grid interactive procedures of Clark and Farley [1984], RAMS has the ability to represent a large-scale area and then to nest to progressively smaller scales. The highest resolution meshes model details of small-scale atmospheric systems, while coarse meshes provide boundary conditions for the fine mesh regions and also simulate large scale atmospheric systems that interact with the smaller scale systems resolved on the finer grids.

The turbulence closure option used in this study is the Mellor Yamada [1982] scheme for vertical diffusion and the Smagorinsky [1963] scheme for horizontal diffusion. The Mellor Yamada scheme employs a prognostic turbulent kinetic energy (TKE). The horizontal diffusion is based on the original Smagorinsky formulation, which uses a deformation-based mixing coefficient that is the product of the horizontal deformation rate (the horizontal gradient of horizontal velocity) and the square of a length scale.

A combination of the Grell parameterization scheme and the RAMS microphysics scheme was used in this study. The coarse domains used the Grell convective cumulus scheme, which is an entraining plume model based on the quasi-equilibrium assumption (Grell, 1993; Grell and Devenyi, 2002). The Grell scheme is a single cloud scheme with updraft and downdraft fluxes and compensating motion that determine the heating and moistening profiles. A study by Grell, Kuo, and Pasch [1991] found that the inclusion of downdrafts is crucial in predicting the

correct feedback in the mid-latitudes; and they also found that the Grell scheme is useful for smaller grid scales, as it tends to allow a balance between the resolved-scale rainfall and the convective rainfall. The Grell scheme makes use of a wide variety of assumptions that lead to a realistic simulation of mesoscale convective systems (MCS), including assuming a large cloud size, using a stability closure, parameterizing moist downdrafts, and not assuming unrealistically large lateral mixing.

The finer resolution domains with grid spacing of 5 km and 1 km are intended to represent cloud processes explicitly using the bulk microphysics parameterization in RAMS. Details of the microphysical parameterization are presented in Meyers *et al.* [1997] and Walko *et al.* [1995]. The scheme predicts the mixing ratio and number concentration of cloud water, rain, pristine ice, snow, graupel, aggregates, and hail from conservation equations which include advective, diffusive, and precipitation tendencies and source terms resulting from interactions between each category and the other forms of water substance (Walko and Tremback, 2002). The general gamma distribution is the basis function used for hydrometeor size in each category. Additional features include: use of stochastic collection for number concentration tendency; breakup of rain droplets formulated into the collection efficiency; diagnosis of ice crystal habit dependent on temperature and saturation; evaporation of each species assuming that the smallest particles completely disappear first; and more complex shedding formulations which take into account the amount of water mass on the coalesced hydrometeor (Meyers *et al.*, 1997).

This study uses the two-stream radiation scheme developed by Harrington, which treats the interaction of three solar and five infrared bands with the model gases and cloud hydrometeors (Harrington, 1997; Harrington *et al.*, 1999). The Harrington parameterization accounts for each form of condensate (cloud water, rain, pristine ice, snow, aggregates, graupel, and hail) as well as water vapor; and it also utilizes information on ice crystal habit (Walko and Tremback, 2002). Using a two-stream model for both solar and longwave radiation has the advantage that computational costs only go up linearly with the number of vertical levels, as

opposed to older schemes that used an emissivity approach to longwave radiation where the costs increased with the square number of the levels (Cotton *et al.*, 2002). In addition, the Harrington model responds to the detailed liquid and ice hydrometeor size spectra simulated with both the bulk and bin-resolving microphysics models. This response permits the simulation of changes in the radiative heating as droplet spectra broaden to precipitation sizes, and it also permits feedbacks of radiation on droplet and ice particle vapor deposition growth (Harrington *et al.*, 1999; Harrington and Olsson, 2001).

In this study, the radiation condition discussed by Klemp and Wilhelmson [1978] is used at the lateral boundaries, which assumes that disturbances reaching the boundaries move as linearly propagating gravity waves. This boundary condition is intended to allow most disturbances to propagate out of the model domain without strongly reflecting back to the interior (Walko and Tremback, 2002). Additionally, this simulation utilized analysis nudging by Newtonian relaxation that is implemented in RAMS, in which the model fields on the lateral boundaries can be nudged toward observational data as the simulation progresses. RAMS currently uses the technique that has been termed analysis nudging for its 4DDA scheme, where the observational data is first objectively analyzed to the model grid, then the model field is nudged to the gridded analysis (Walko and Tremback, 2002). The lateral boundary nudging in RAMS is an implementation of the Davies [1983] scheme, where a number of grid points in a boundary region of the coarsest resolution grid are nudged toward data. Time-series of gridded horizontal wind, potential temperature, and total water mixing ratio values are provided; and the atmospheric model solution is relaxed toward the analyzed data. Lateral boundary nudging introduces time-varying information in the model domain while damping information propagating from the model interior toward the lateral boundary.

## 2.4 Input Data for the Model

### 2.4.1 Vegetation Classes

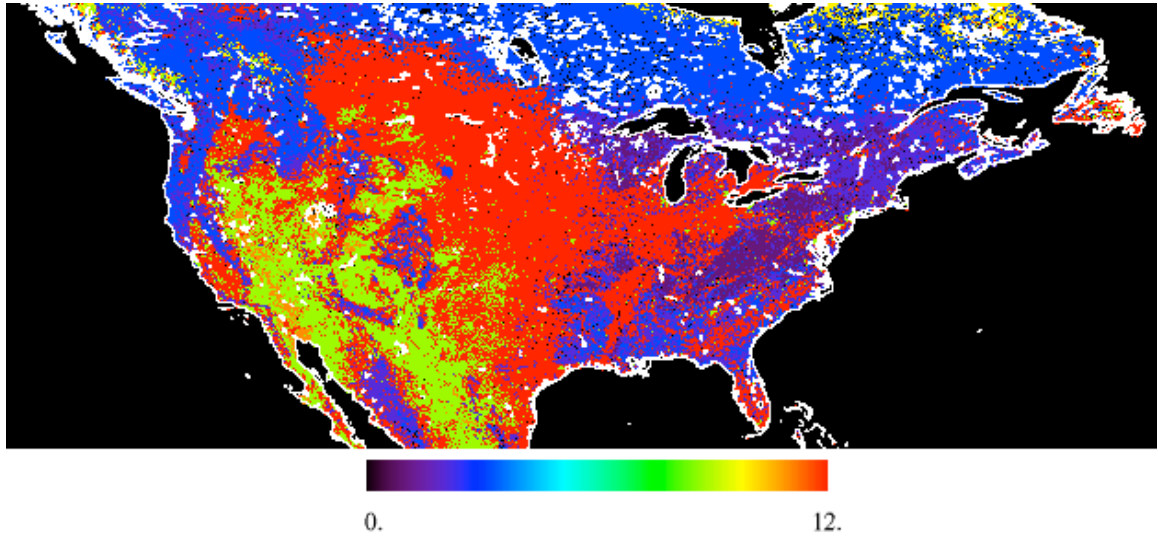
To determine the vegetation class for each grid cell in SiB2-RAMS, this study uses the 1-km land cover classification data derived from AVHRR at the University of Maryland, which is described in Hansen *et al.* [1998, 2001]. Hansen *et al.* originally classify the data using thirteen different classes. These classes have been converted to SiB classes to make the data compatible with SiB2-RAMS, and Table 2.3 depicts the scheme used to convert them.

Original Class by Hansen <i>et al.</i> [1998]	Corresponding SiB Class
0 Water	0 Water
1 Evergreen Needleleaf Forest	4 Tall Needleleaf Trees
2 Evergreen Broadleaf Forest	1 Tall Broadleaf Evergreen Trees
3 Deciduous Needleleaf Forest	5 Tall Needleleaf Deciduous Trees
4 Deciduous Broadleaf Forest	2 Tall Broadleaf Deciduous Trees
5 Mixed Forest	3 Tall Broadleaf and Needleleaf Trees
6 Woodland	4 Tall Needleleaf Trees above 48N 3 Broadleaf and Needleleaf Trees below 48N
7 Wooded Grassland	4 Tall Needleleaf Trees above 48N 12 Grasslands/Agriculture below 48N
8 Closed Shrubland	10 Ground Cover (Tundra) above 50N 9 Short Shrubs below 50N
9 Open Shrubland	10 Ground Cover (Tundra) above 50N 9 Short Shrubs below 50N
10 Grassland	12 Grasslands/Agriculture
11 Cropland	12 Grasslands/Agriculture
12 Bare Ground	13 Ice above 60N 9 Shrubs with Bare Soil from 60N-45N 11 No Vegetation (Desert) below 45N
13 Urban and Built-Up	9 Broadleaf Shrubs with Bare Soil

**Table 2.3:** Conversion table between the AVHRR vegetation classification and the SiB classes

A map of the vegetation classification for much of North America is shown in Figure 2.6. The figure shows that most of Canada is classified as tall needleleaf trees, while much of the United States is C3 grasslands and agriculture. The eastern portion of the United States is a combination of broadleaf deciduous trees and mixed forest, and the southwest is primarily shrubs and desert.

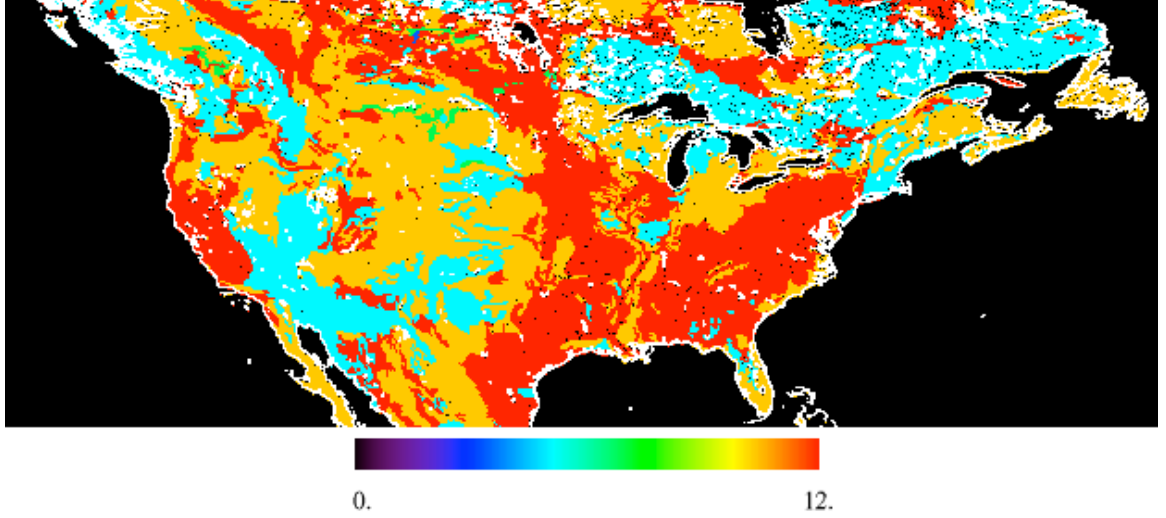




**Figure 2.6:** The 1-km biome classification over North America from Hansen *et al.*

### 2.4.2 Soil Type

The soil data used in this study was generated by the SoilData System, which was developed by the Global Soil Data Task Group of the International Geosphere-Biosphere Programme (IGBP) Data Information System, and it is available on CD-ROM from the Oak Ridge National Laboratory Distributed Active Archive Center in Oak Ridge, TN (Global Soil Data Task, 2000). The original data was a 5' x 5' (~10 x 10 km) resolution map containing values of % sand, % clay, and % silt. Using a triangulation program, the data was converted to SiB soil classes and then masked to the vegetation map to ensure the maps agree on the location of water pixels. If the vegetation map had water in a pixel that did not originally correspond to a water pixel in the soil map, then the soil pixel was set to water. In contrast, if the biome map had a vegetation class and the soil type was water, then that pixel was set to the nearest soil class. A map of SiB soil classes over the domain used in this study is displayed in Figure 2.7.



**Figure 2.7:** Map of the 10-km SiB soil classification from IGBP.

### 2.4.3 NDVI

This study uses global SPOT 10-day composited Normalized Difference Vegetation Index (NDVI) from the United States Department of Agriculture Foreign Agriculture Service (USDA/FAS) through collaboration with the Global Inventory Modeling and Mapping Studies (GIMMS) Group at NASA/GSFC. The SPOT data is from the VEGETATION instrument on board the SPOT-4 polar orbiting satellite. NDVI is computed from the equation

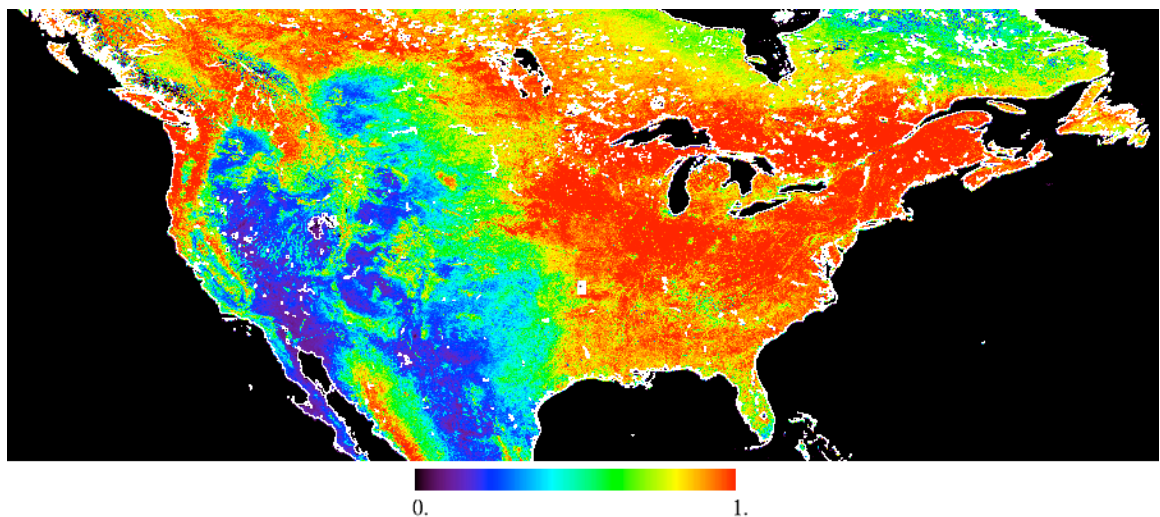
$$NDVI = \frac{IR - R}{IR + R}, \quad (13)$$

where IR represents the radiances in the infrared band and R is the radiances in the red band.

Ten-day composite image data were constructed by selecting pixels with the maximum NDVI during the period, which minimizes cloud cover and water vapor effects that strongly reduce NDVI.

NDVI values range from 0 to 1, with 0 representing water. In general, high values of NDVI represent high reflectance in the IR band and low reflectance in the red band, which means the foliage is healthy and abundant with green vegetation. In contrast, low values of NDVI indicate a lack of green leaves, higher pigmentation, or unhealthy foliage. The intensity of the reflected radiation depends on the pigmentation and the overall health of the vegetation, and it varies for different types of foliage.

To prepare the NDVI data for SiB2-RAMS, the data went through several processing steps. First, the data was filled using status maps also provided by USDA/FAS, which indicate whether the datum at each pixel is valid or contains errors. Since the northern portion of North America does not receive any solar radiation during the winter, SPOT is unable to retrieve NDVI values in the high latitudes during the winter months. To fix this, the missing NDVI pixels were filled by linearly interpolating the values at the same pixel between the last value SPOT retrieved in the fall and the first value it retrieved for that pixel in the spring. In addition, all points where the status map indicate that either the ground was covered with ice/snow or that the area was covered by clouds during the 10-day period were also filled by interpolating the closest values in time at that pixel. Then, since the NDVI maps were provided in Mercator projection, they were re-projected into latitude and longitude coordinates. Finally, the NDVI data were masked to the vegetation map from the University of Maryland. Any pixel that was water in the vegetation map was required to be water in the NDVI maps; and any pixel that had a vegetation class other than water but was considered water in the original NDVI map was assigned a non-zero NDVI by interpolating the NDVI values from the nearest pixels.



**Figure 2.8:** 1-km resolution NDVI values from SPOT on August 1, 2001.

Figure 2.8 displays the NDVI values on August 1, 2001. As expected, the figure shows there is considerable variability across North America. The east coast has high values of NDVI,

due to the higher availability of moisture and the more dense and greener vegetation. The NDVI values are lower over the western half of the United States, with values of 0.4-0.6 across the plains. Deserts experience very low NDVI values due to the lack of vegetation.

In addition to using the maps of NDVI, two morphological parameters used by SiB also depend on NDVI. These two parameters the 5<sup>th</sup> and 98<sup>th</sup> percentile of NDVI, written as NDVI<sub>5</sub> and NDVI<sub>98</sub>, respectively, are important parameters in SiB that are used to calculate fPAR and LAI. Using the NDVI data set, we calculated NDVI<sub>5</sub> and NDVI<sub>98</sub> for the vegetation classes. After this calculation, we then followed rules established by Los [1993] in assigning the NDVI<sub>5</sub> and NDVI<sub>98</sub> values. It was assumed that the lowest 5% values for bare soil and shrubs represented no vegetation conditions, and NDVI<sub>5</sub> for every vegetation class was set to this value. For NDVI<sub>98</sub>, the values assigned to each biome class are displayed in Table 1.4.

Vegetation Class	Class used for NDVI <sub>98</sub>
1	1
2	2
3	3
4	4
5	4
6	1
7	12
8	1
9	12
10	12
11	12
12	12

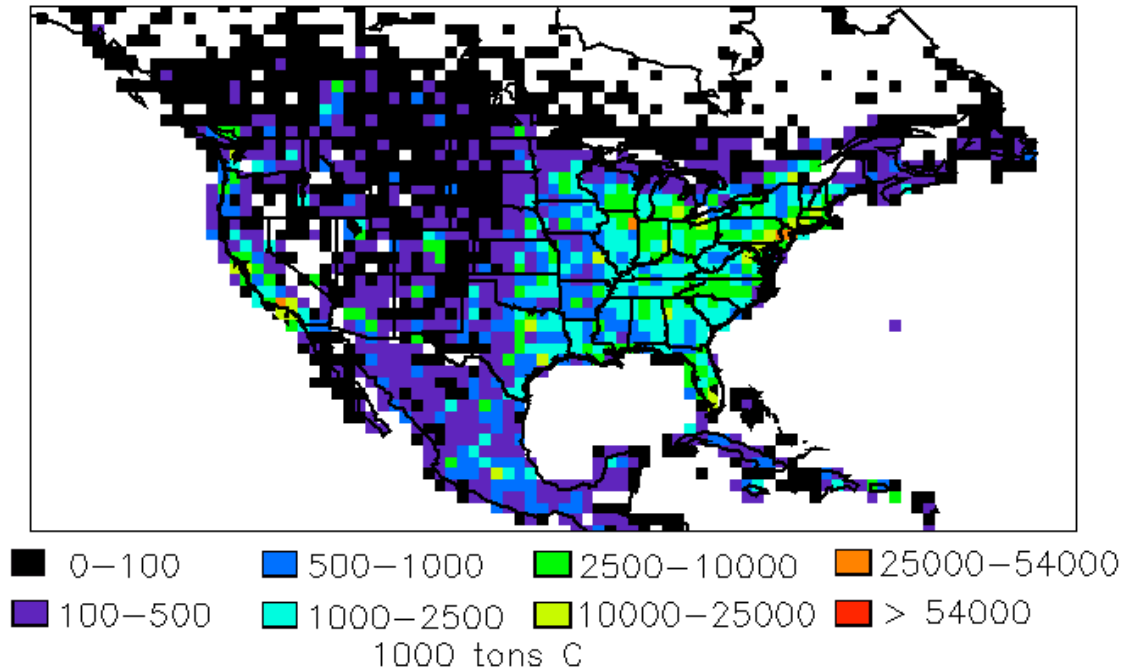
**Table 2.4:** NDVI<sub>98</sub> values assigned to each biome class in SiB using the rules from Los [1993].

#### 2.4.4 Fossil Fuel Combustion

Surface fluxes of carbon to the atmosphere due to fossil fuel combustion, cement production, and gas flaring were derived from the 1995 CO<sub>2</sub> emission estimates of Brenkert [1998]. The data was obtained from the Carbon Dioxide Information Analysis Center (CDIAC) at the Oak Ridge National Laboratory in Tennessee. Since the resolution of the data is one degree by one degree, the fossil fuel fluxes are interpolated to the finer resolutions used in this study.

The 1995 carbon emissions estimates were converted to carbon fluxes in units of  $\text{kgC}/\text{m}^2/\text{s}$ , and the surface flux was scaled from the annual 1995 estimates to estimates for August 2001. The surface fluxes were multiplied by a scaling factor of 1.124, which was calculated using monthly estimates of global anthropogenic emissions (A. Wang, personal communication).

The 1995 carbon emissions are shown in Figure 2.9. The emissions map shows that the main sources of carbon are the major cities, as Los Angeles, Denver, Chicago, and New York can be seen clearly in the figure. In general, the east coast contributes more anthropogenic emissions to the atmosphere, while areas in the west and Canada have very little fossil fuel emissions.

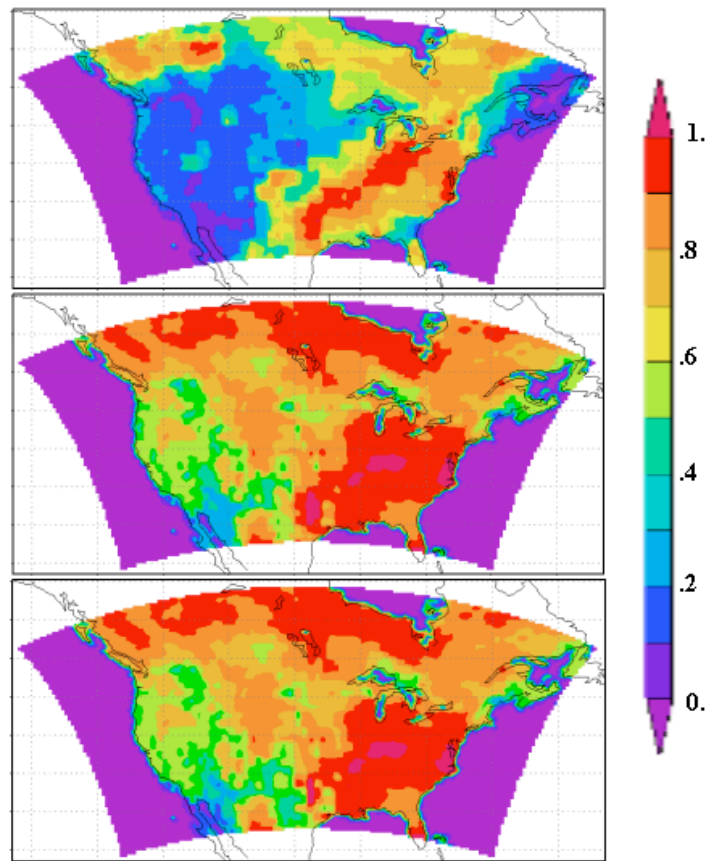


**Figure 2.9:** The 1995  $\text{CO}_2$  emission estimates from Brenkert [1998].

#### 2.4.5 Soil Moisture, Soil Stress, and Respiration Factor

To spin-up the soil moisture to reasonable values, SiB was run offline for ten years, from 1991 through 2001. The soil moisture in SiB2-RAMS is initialized from the soil moisture fields produced by offline SiB driven from NCEP re-analysis. The resolution of the offline model is one degree by one degree, and the soil moisture for all resolutions in this study is initialized with the one degree by one degree soil moisture values. Figure 2.10 displays the soil moisture

values for August 1, 2001. The top panel shows the soil moisture volumetric fraction of saturation in the top layer of soil, the middle panel depicts the soil moisture in the root zone, and the bottom panel shows the soil moisture in the recharge layer. The top layer of soil is much drier than the other two layers, but the moisture in this layer can change quite rapidly. In the root zone, the soil in Canada and the eastern United States is moist, while the western portion of the United States is slightly drier, although still having values from 0.3 to 0.6. The soil moisture in the third soil layer is very similar to the soil moisture distribution in the root zone.



**Figure 2.10:** The initial soil moisture volume fraction of saturation for SiB2-RAMS, produced from offline SiB. The top panel is the moisture field for the top layer, w1; the middle panel displays the soil moisture from the root zone, and the bottom panel displays the soil moisture from w3.

The respiration factor in SiB2-RAMS was also calculated from offline SiB, using the ten-year run of SiB. The respiration factor for 2001 was obtained by dividing the total assimilation for 2001 by the total soil scale for that year, as discussed in section 2.3.1. For grids

in SiB2-RAMS with a smaller resolution than the offline 1 by 1 degree model, if the grid cell in SiB2-RAMS has the same biome type as the offline SiB cell or any immediately adjacent pixels, then the grid cell of SiB2-RAMS has the same respiration factor as the offline SiB cell with the same biome class. If the SiB2-RAMS biome type does not match either the SiB cell it resides in or the adjacent cells, then the SiB2-RAMS respiration factor is set to the biome-based global average respiration factor (A. Wang, personal communication).

## **2.4 Lateral Boundary Forcing**

The meteorological fields in SiB2-RAMS are initialized using the National Center for Environmental Prediction (NCEP) mesoscale Eta-212 grid analysis (AWIPS 40-km). The Eta-212 analysis covers most of North America, has 40-km horizontal resolution, and three-dimensional output that is provided on about 25 constant pressure surfaces spanning from 1000 mb to 25 mb at intervals of 50 mb. In addition to using the Eta-212 data for initial conditions, the lateral boundaries are nudged towards the 3-hourly Eta analysis.

The CO<sub>2</sub> field in SiB2-RAMS is initialized to 370 ppm, and the lateral boundaries are also set to 370 ppm in this study.

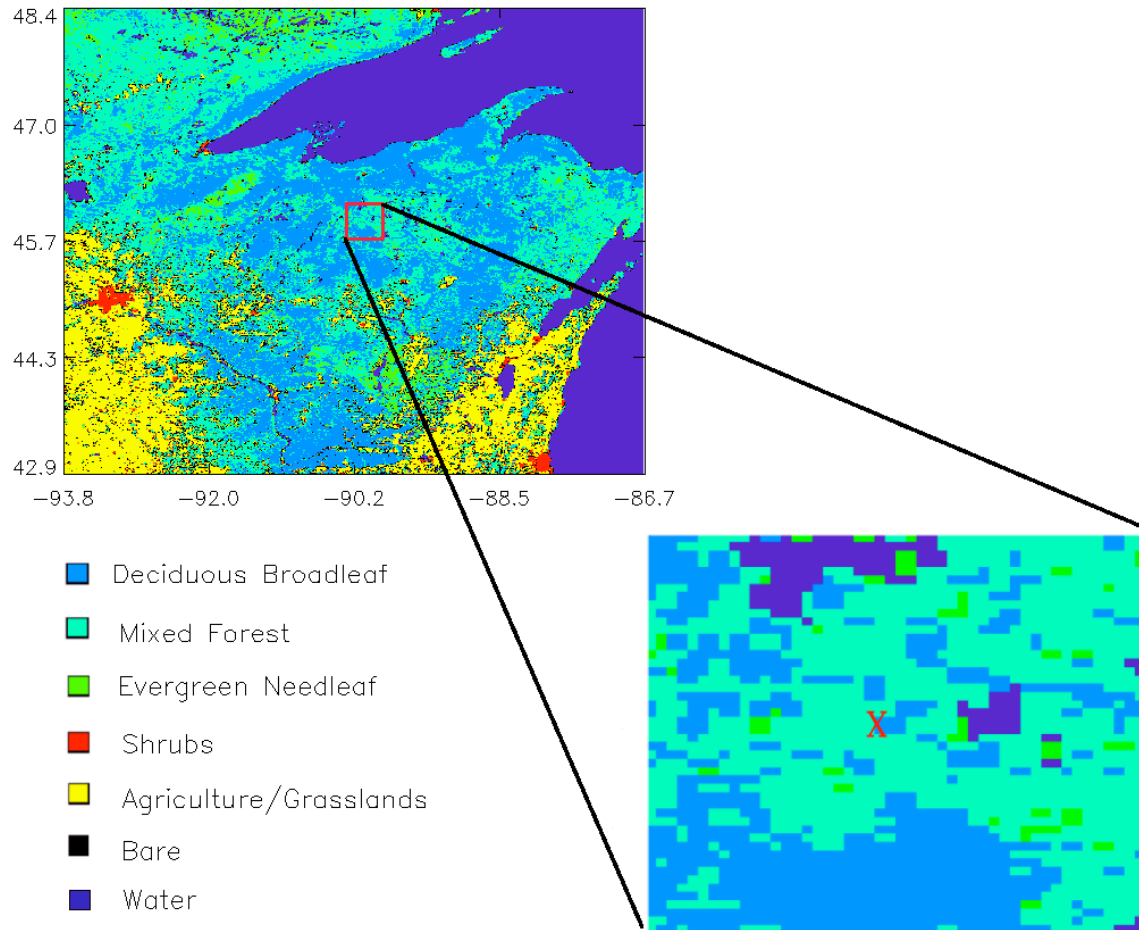
## **2.5 Case Descriptions**

### **2.5.1 Case 1: 5-Day Clear-Sky Simulation centered at WLEF**

The first case that will be analyzed in this study is a five-day, clear-sky simulation. For this case, Lara Prihodko provided the model output, which was used in a recent study by Nicholls *et al.* [2003]. This simulation uses a coarse domain and three nested grids all centered on the WLEF tall tower in Wisconsin, which was discussed in section 2.1.1. The horizontal grid increments are 16 km, 4 km, 1 km, and 333 m for grids 1, 2, 3, and 4 respectively. This study will analyze grids 1 and 3. Figure 2.11 shows the grid set-up for case 1. The backgrounds plotted in the figure display the vegetation classes. Grid 1, which is the top panel in Figure 2.11, has 40



grid cells in both the x and y direction, covers an area of 640 km by 640 km, and includes Lake Superior as well as part of Lake Michigan. The bottom panel in Figure 2.11 zooms in on grid 3, which is covered primarily by forest with a lake on the central northern boundary and a smaller lake straight east of the tower. Grid 3 has a horizontal grid increment of 1 km and is  $(38 \text{ km})^2$ .



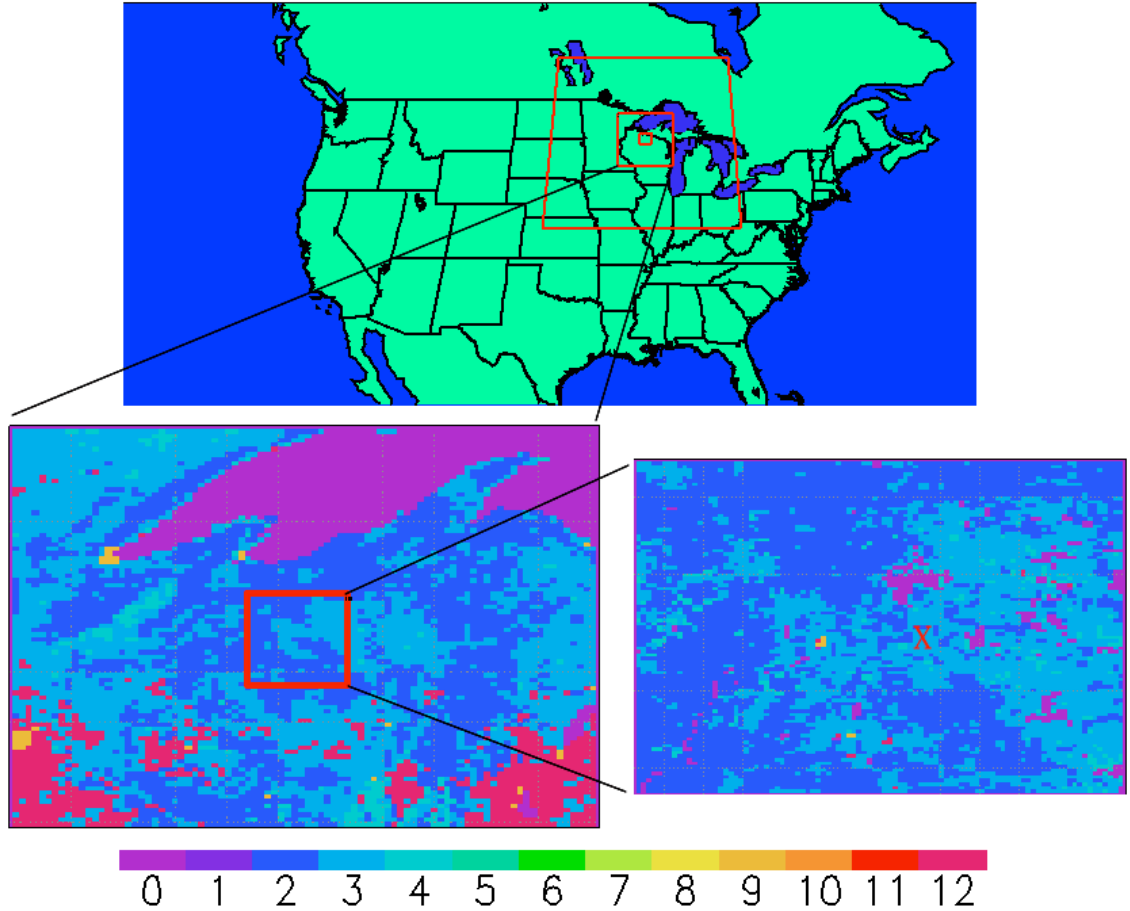
**Figure 2.11:** The grid location and vegetation classes for grids 1 (top) and 3 (bottom). The red x indicates the location of the WLEF tower.

The time frame of this case is five days, beginning 6 AM LST July 26, 1997 and ending 6 AM LST July 31, 1997. The day prior to the study was cloudy with some precipitation, but the dates simulated did not include any clouds. The first two days, July 26 and 27 were warm and humid with winds predominantly from the west. During the night of July 27, the winds strengthened and veered, which brought colder air from the north; however, the temperature



warmed gradually in the following days and the winds lessened, eventually becoming southwesterly.

### 2.5.2 Case 2: 10-Day Simulation centered at WLEF



**Figure 2.12:** The grid setup of case 2 (top), with the nested grids outlined in red. The vegetation coverage for grid 3 is shown in the bottom left and the vegetation for grid 4 is displayed in the bottom right panel. The red x indicates the location of the WLEF tower.

The second case we will analyze in this study is a ten-day simulation that is also centered on the WLEF tower. Similar to the first case, this case also has a coarse domain and three nested grids. The horizontal grid increments in this case are 40 km, 10 km, 5 km, and 1 km and the number of grid cells in each domain are 150 x 90, 150 x 150, 90 x 90, and 97 x 97 for grids 1, 2, 3, and 4, respectively. In this study, we will analyze both grid 4 and grid 3. Although grid 3 has a resolution of 5 km, which is large for the cloud-resolving microphysics, we will still

analyze this grid because it will help us understand the errors introduced by using satellite measurements to represent a larger 450 km domain. The grid setup is displayed in Figure 2.12. The top panel shows the coarse domain, with the size and placement of the three nested grids outlined in red. The bottom two panels show the vegetation classification for grids 3 (left) and 4 (right). Grid 3 includes part of Lake Superior. The upper and middle portions of the domain are dominated by mixed forest, while the southern third contains significant areas of agriculture and cropland. Zooming in to grid 4, the vegetation is primarily mixed forest and broadleaf deciduous trees with a few patches of evergreen trees and agriculture. Grid 4 contains several small lakes, with a relatively significantly sized lake just north of the WLEF tower.

This case simulates twenty days, from 00 GMT August 1 to 00 GMT August 21, 2001. The first ten days spinup the model to help make this run more realistic. The time period we will analyze from this simulation is from 00 GMT August 11 to 00 GMT August 21. During this ten-day time period, three fronts passed over the WLEF tower. A more in-depth analysis of the weather during the simulation will be discussed in the next chapter.

## **2.6 Methods for Simulating OCO Using SiB2-RAMS Output**

Since OCO will estimate total column  $\text{CO}_2$  concentrations, the  $\text{CO}_2$  concentrations from the model simulations are vertically integrated by pressure weighting, yielding total column mixing ratios. Since the vertical coordinate of the model output is the height in meters rather than pressure, we will use a standard atmosphere to find the pressure corresponding to the vertical heights. To more closely match the vertical height, the standard atmosphere was first linearly interpolated to a resolution of only 5 m near the surface. The pressure corresponding to the model height is then used to weight the  $\text{CO}_2$ : the  $\text{CO}_2$  concentration at each height is multiplied by the difference in pressure between the upper and lower boundaries surrounding the current level. The values are then summed and divided by the sum of the pressure changes. The final

result is the CO<sub>2</sub> concentration for the entire vertical column of air, which is comparable to the total column CO<sub>2</sub> measurements from OCO.

After computing the total column CO<sub>2</sub>, we next had to imitate a satellite track. For simplicity, we assume that the satellite travels due south. We also assume that all the footprints over the different latitudes sampled within the domain will be averaged together to yield only one concentration for the grid. To emulate this spatial averaging in the satellite measurements, we averaged all the latitudinal values together to produce one CO<sub>2</sub> concentration for each longitude. The final consideration is the width of the satellite track we are imitating. For case one, we assume that the width of the satellite track is the width of the horizontal grid increment: the width of a satellite track in grid 1 is 16 km and the width of a satellite track in grid 3 is 1 km. Using all these assumptions, case one has 40 different possible satellite tracks (i.e. one possible track for each grid in the x-direction) for grid 1 and 38 possible satellite tracks for grid 3. Each emulated satellite track produces one total column CO<sub>2</sub> concentration. For case two, the width of the emulated satellite swath is 10 km to match OCO. For grid 4, a single satellite measurement from one track is the mean of ten 1-km grid cells, or pixels, in the x-direction and all the pixels (97) in the y-direction. Grid 4 has 88 different possible satellite tracks at any one point in time. Similarly, for grid 3 a satellite track (which produces one concentration of total column CO<sub>2</sub>) is the mean of two 5-km pixels in the x-direction and all the pixels (90) in the y-direction; and grid 3 has 89 different possible satellite tracks.

## Chapter 3: Meteorological Evaluation of Case 2

In this chapter, we analyze the meteorological conditions during case 2 in the first section. The second section evaluates the performance of the model in case 2 by comparing both large-scale patterns and point measurements to observations.

### 3.1 Meteorological Analysis

Looking at the meteorological conditions for grids 3 and 4 of case 2 using NCEP weather charts archived at the National Climatic Data Center (NCDC) and measurements from the WLEF tower, the ten day time period began with mild weather: the temperature was moderately warm and the wind was light and predominantly from the southwest. The pressure began to drop the night of August 11, and the wind shifted to westerly moving to northwesterly early in the morning on the 12<sup>th</sup> as the wind speed strengthened. A front came through around noon local time on the 12<sup>th</sup> with a relatively strong northerly wind. Just prior to the front the CO<sub>2</sub> concentration jumped up. When the front actually passed through, the CO<sub>2</sub> concentration decreased again, only to jump up even higher the next hour. This increase in CO<sub>2</sub> concentration could be due to higher CO<sub>2</sub> air being advected from the southwest. Due to high temperatures causing high respiration rates, areas around Texas and Oklahoma may be large sources of CO<sub>2</sub> and the front could be incorporating these higher concentrations. Following the front, the wind speed decreased but the wind remained out of the north. The 13<sup>th</sup> is a sunny summer day: the temperature is moderate with a light wind while the pressure increased and the CO<sub>2</sub> concentration decreased. That evening, the wind shifted directions and again became southwesterly. During the day on the 14<sup>th</sup>, a cold front formed in Canada. That night at the WLEF tower, the surface pressure began to drop and the southwesterly wind strengthened. Around midnight the CO<sub>2</sub>

concentration dropped dramatically. The morning of the 15<sup>th</sup> the eastern portion of the cold front north of WLEF remained stationary over Canada, while the western portion moved south into northern Minnesota and the Dakotas. In addition, there is a low-pressure center over Kansas and Nebraska that is beginning to develop into a front. By late afternoon on the 15<sup>th</sup>, the cold front over northern Minnesota has continued to move south while the low-pressure center over Kansas has traveled northeast and developed into a warm front through southern Illinois and Indiana attached to a cold front in Kansas and Oklahoma. This second separate front in the midwest is causing cyclogenesis ahead of the northern-most front in Minnesota and Canada, and the next several hours are complicated by the merging of these two low-pressure systems. That evening the eastern portion of the cold front over Minnesota and Canada shifted south into northern Wisconsin where it remained stationary for the next twelve hours, while the western half of the front over the Dakotas continued to move south into Nebraska and Kansas. The front located in the Midwest moved north until the low-pressure center was in southern Wisconsin the morning of the 16<sup>th</sup>. By late morning on the 16<sup>th</sup> the northern front again began to move south over Wisconsin, passing over the tower around noon. The two fronts finally merged the afternoon of the 16<sup>th</sup> and continued to track southwest.

Looking at the measurements taken at the WLEF tower for August 15 and 16, the surface pressure actually reached its minimum value around noon on the 15<sup>th</sup>. The wind speed was about 13 m/s at 396 m, and the wind was still southwesterly. The pressure remained low until after the front passed over the tower and began to climb the afternoon of the 16<sup>th</sup>. The wind speed at 396 m on the tower dropped just after midnight on the 16<sup>th</sup>, and the wind remained relatively calm while the northern front remained stationary and the midwestern front moved north. While the wind speed is low, the wind direction began to shift to the north. Around noon when the front passes over the tower the wind speed jumped back up, bringing air down from the north. The air behind the front is actually rather warm, and the temperature at 30 m the afternoon of the 16<sup>th</sup> was still around average for this time period, despite the passage of the front.

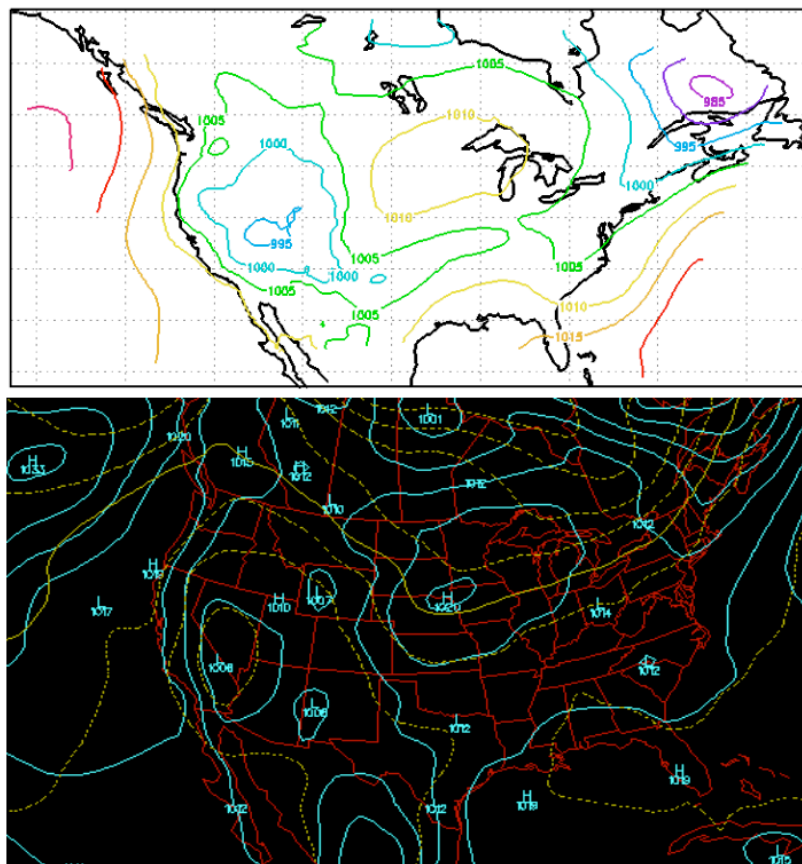
During these two days, the CO<sub>2</sub> concentration at the WLEF tower increased dramatically from its minimum on midnight the 15<sup>th</sup> until the morning of the 16<sup>th</sup>. This increase in CO<sub>2</sub> concentrations could have been due to the cyclogenesis seen in the region: the midwestern front could have drawn high CO<sub>2</sub> air up from Texas and Oklahoma, where high temperatures may have caused high respiration rates and high CO<sub>2</sub> concentrations. Since the northern front was stationary over northern Wisconsin, the increase in CO<sub>2</sub> may be more dramatic in this case than in the previous front on the 12<sup>th</sup> because the air could have been coming from higher source regions for a longer period of time due to the second front in the midwest. The CO<sub>2</sub> concentration began to decline a few hours before the front, possibly due to northerly winds bringing down lower CO<sub>2</sub> from Canada, and it continued to decrease after the front passes through.

Moving on to the end of the ten-day time frame, early morning on the 17<sup>th</sup> the wind began to relax and shift to the west while the pressure increased and the CO<sub>2</sub> concentration decreased. The afternoon of the 17<sup>th</sup> there was a cold front in North Dakota and Montana, and there was also a cold front in the east extending up into Canada that caused considerable cyclogenesis. At the tower, the wind direction continued to shift counterclockwise: the wind shifted from being westerly to southerly to easterly. The evening of the 17<sup>th</sup> the cold front in North Dakota moved south and east, and the strong cyclonic flow seen behind the cold front in the east generated a warm front in Wisconsin that connected with the cold front in Minnesota. Around midnight on the 18<sup>th</sup> the warm front passed over the WLEF tower, strengthening the wind, which had continued circling counter-clockwise from the east and was from the northeast. The warm front caused a slight increase in the CO<sub>2</sub> concentration at 396 m, but the increase was much less than the increase seen in the previous two fronts. Perhaps the CO<sub>2</sub> increase is much less significant because the front came from the northeast, and rather than bringing in high CO<sub>2</sub> air from the southwest, the cyclogenesis brought air with relatively low concentrations from further north and east. Since the front passed over at night, the slight increase in CO<sub>2</sub>

concentrations may be due to the slightly higher nighttime CO<sub>2</sub> concentrations coming over the great lakes. While the pressure increased on the 18<sup>th</sup>, a strong northeasterly wind continued and the temperatures were much warmer than the previous few days. On the 19<sup>th</sup>, the wind speed decreased and began its clockwise shift from the north through easterly to southwesterly as the temperature and pressure both increased. The 20<sup>th</sup> continued the moderate weather conditions seen on the 19<sup>th</sup>, with an increasing surface pressure, warm temperatures, and a light wind from the southwest.

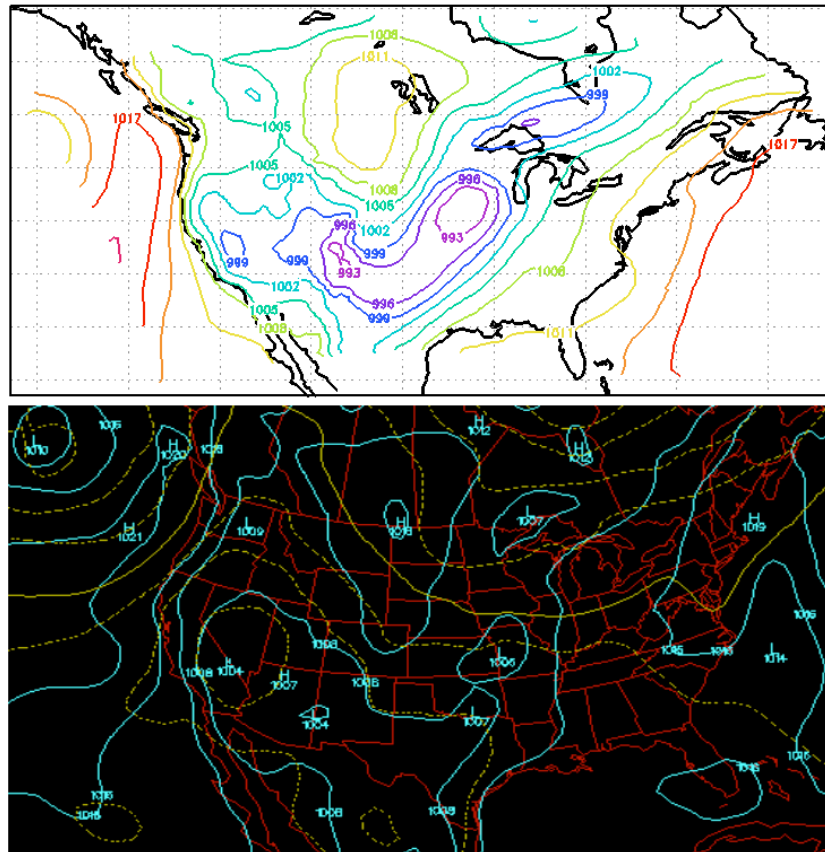
## 3.2 Model Evaluation

### 3.2.1 Large-Scale Patterns



**Figure 3.1:** Pressure contours at 0 GMT on August 11, 2001. The top panel is the modeled sea-level pressure from case 2, and the bottom panel is the sea-level pressure from ETA, provided by Unisys.

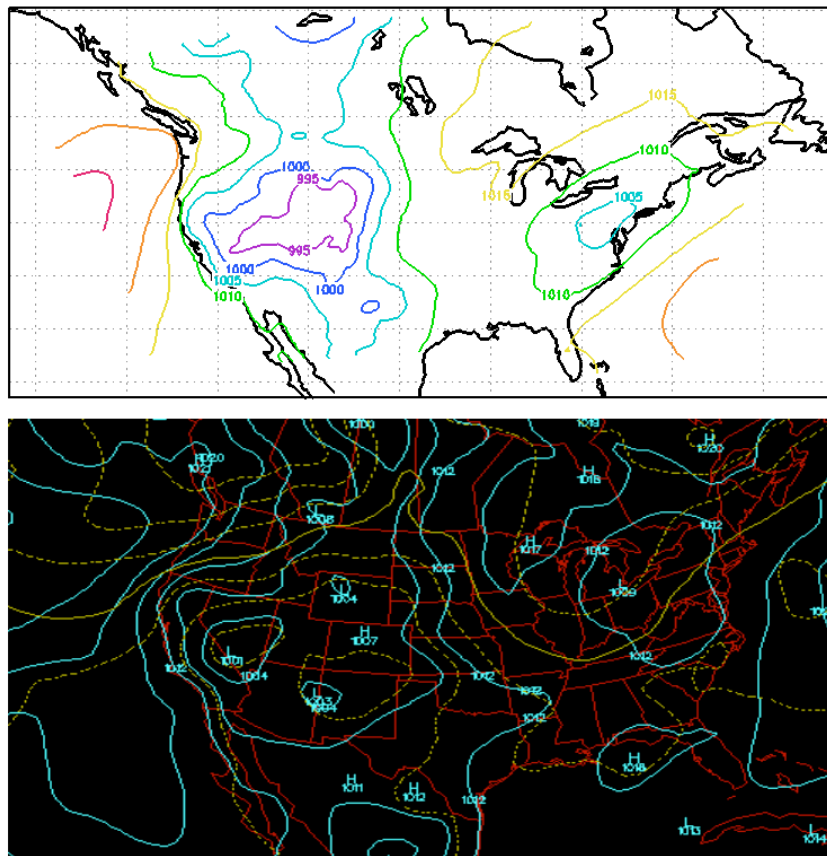
Taking a snapshot at the start of the ten-day time frame we will be analyzing, Figure 3.1 shows the sea-level pressure at 0 GMT on August 11, 2001. The top panel displays the sea-level pressure in the SiB2-RAMS simulation and the bottom panel shows ETA sea-level pressure analysis map from Unisys. At this point in time, the model captures the overall pressure patterns quite well: there is a high-pressure area east of the Great Lakes and a low-pressure region in the southwest. Up into Canada, low-pressure contours also exist in both the model and the analysis.





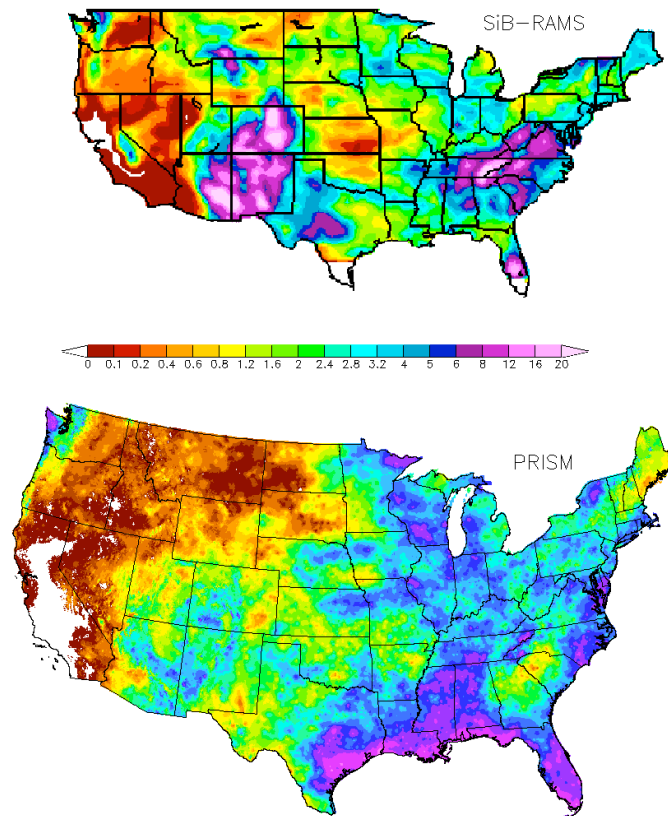
Canada. The model also captures the low pressure region over Lake Superior, down the Midwestern US, and extending west into Texas and New Mexico; however, the low pressure centers are shifted slightly and the region does not extend as far west as seen in the analysis.

Finally, to investigate the meteorological conditions at the end of the simulation, we compared the surface pressures at 0 GMT on August 20, 2001. Figure 3.3 displays both the modeled and the observed sea-level pressure. Yet again, the regional pattern in the model is similar to the surface analysis: there is a region of high-pressure north and east of Lake Superior, a low-pressure center over New England, and a low pressure region located in the western states. Overall, looking at the sea-level pressure indicates that SiB2-RAMS does capture the basic weather events during this time period.



**Figure 3.3:** Similar to Figures 3.1 and 3.2, but for 0 GMT on August 20, 2001.

Although SiB2-RAMS does a reasonable job at capturing the sea-level pressure, one problem in case 2 is the precipitation pattern. Figure 3.4 displays the accumulated precipitation in the model in the top panel, where the precipitation amount includes the precipitation from 00 GMT August 1 to 00 GMT on August 21, 2001. The bottom panel shows a map of the August 2001 precipitation from the Spatial Climate Analysis Service at Oregon State University. To get the spatial map of precipitation in August, Oregon State interpolated the point measurements from the monitoring stations to grid points using PRISM (Parameter-elevation Regressions on Independent Slopes Model). PRISM is an analytical model that uses point data and a digital elevation model (DEM) to generate gridded estimates of monthly precipitation (Daly *et al.*, 1994). Although Figure 3.4 compares twenty days of accumulated precipitation in the model to the precipitation for the entire month, it is useful to investigate the spatial patterns.



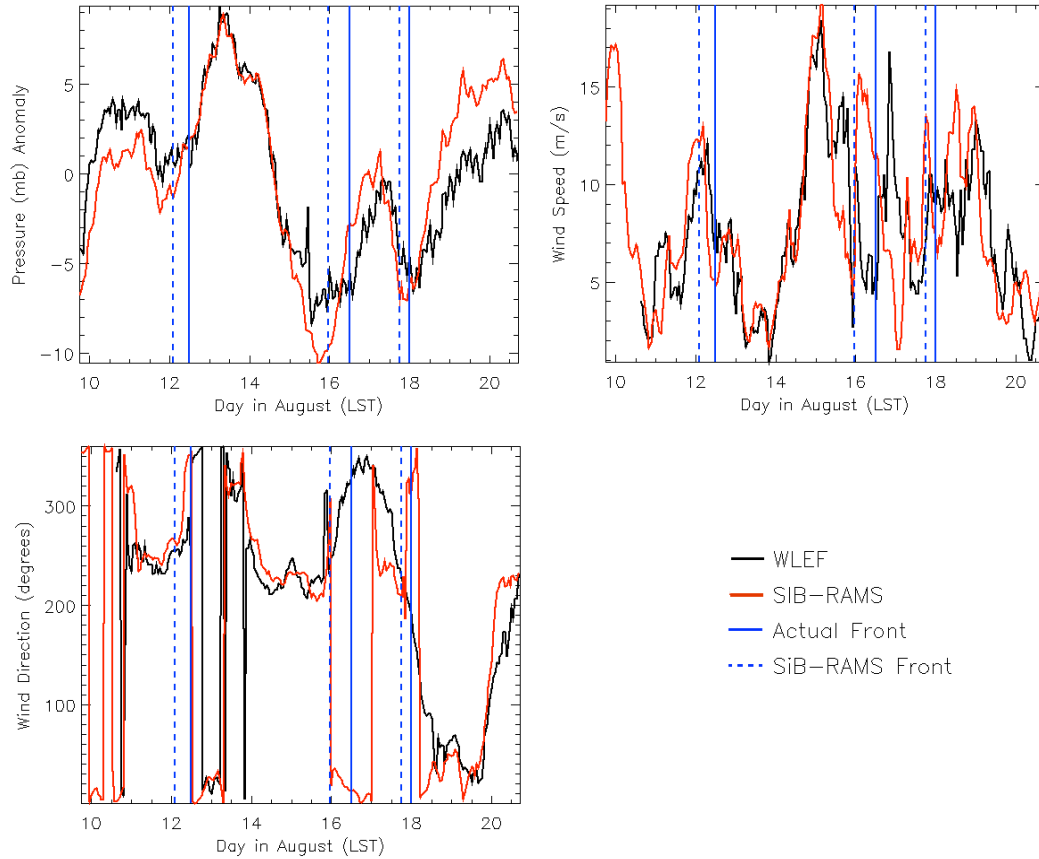
**Figure 3.4:** Accumulated precipitation maps in inches. The top panel is the precipitation from case 2 (00 GMT August 1 through 00 GMT August 21, 2001). The bottom panel is the August 2001 precipitation from Oregon State University using PRISM.

At first glance, the overall precipitation pattern is acceptable: the west is dry while the east is wet; however, upon further examination, SiB2-RAMS does a poor job in replicating the precipitation. The most obvious difference is the precipitation over Colorado, New Mexico and northern Mexico. Although the map generated from PRISM does show considerable precipitation in this region, the model grossly overestimates the amount and produces over 20 inches of rain in only twenty days! SiB2-RAMS seems to persistently underestimate the pressure in that region, indicating that the model has an overactive monsoon. Looking at other areas, we see that the precipitation is again overestimated in the southern Appalachian Mountains and underestimated on the Gulf coast. Over our area of interest the precipitation distribution appears to be realistic. Although the precipitation is underestimated in this region, it could be due to the fact that only twenty days of model output is displayed rather than the entire month.

Previous studies have investigated the capability regional climate models have in simulating precipitation, and these studies show that models have a difficult time simulating precipitation and that the precipitation is strongly sensitive to the initial soil moisture, soil stress, domain size, and areal coverage (e.g. Seth and Giorgi, 1998; Liang *et al.*, 2004). In an attempt to acquire a reasonable precipitation pattern, this study used a coarse domain that extended from off the east coast all the way west to the Pacific Ocean. The domain also extended south to include part of the Gulf of Mexico. In addition, as stated in section 2.3.5, the initial soil moisture was set to the pattern from offline SiB. A three-month spin-up run from May through July 2001 over the same first 2 grids indicates that the soil moisture may have been overly stressed in some regions throughout the domain [A. Wang, personal communication]. Having lower soil moisture stress was shown to increase the respiration over the fine grids, and it may also help to reduce the strength of the monsoon, although further investigation using adjusted soil moisture is required. Even though our simulation does not have a completely accurate precipitation pattern, we assume that the precipitation over the area we will be analyzing is reasonable and that the differences will not dramatically effect the CO<sub>2</sub> concentrations. In addition, my study does not require the

simulation to exactly replicate the observations as long as the model is self-consistent, which is the case in this simulation as the model is a perfect simulation of itself.

### 3.2.2 Point Comparisons



**Figure 3.5:** Meteorological variables for both the model (red line) and the observations (black line), where the x-axis is the local time. Surface pressure anomalies are displayed in the top left panel, wind speed at 396 m is shown in the top right panel, and wind direction at 396 m is shown in the bottom panel. The solid blue line indicates the frontal passages at the WLEF tower and the dashed blue line denotes the fronts in the model.

In addition to looking at the large-scale pattern, we also compared the simulated fields at WLEF in the model to the measurements taken at the tower. Figure 3.5 shows three meteorological variables at the WLEF tower and in the model. The top left panel displays the surface pressure anomaly, the top right panel shows the wind speed at 396 m, and the wind direction at 396 m is shown in the bottom panel. In all three panels the measurements from the tower are in red while the model results are shown in black. To help analyze the model-data comparison, the solid blue lines show the timing of the fronts from both the measurements and

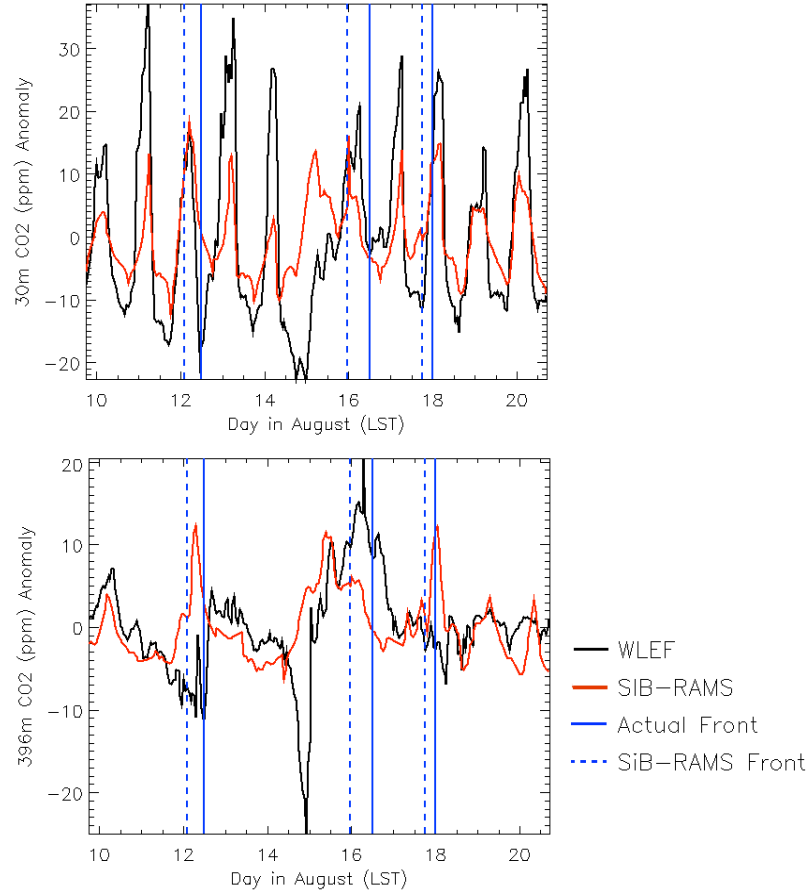
from an analysis of NCEP weather charts archived at the National Climatic Data Center. The dashed blue lines indicate the frontal passages seen in the model. The x-axis in the figure is the date in August in local time, and the times used in this discussion will all be local times rather than GMT. The pressure anomaly was calculated by subtracting the mean over the time shown from the time-series.

Overall, the model does a good job simulating these three variables over the ten-day time frame; however, the model does not capture the timing of the details around each of the three fronts. Looking at the 12<sup>th</sup>, both the model and the measurements have the minimum in pressure around midnight, but in the model the front passes over the tower around ten hours early: rather than the front passing over around noon on the 12<sup>th</sup> as seen in the observations, it passes over the tower closer to 2 AM in the model. Corresponding to an earlier front, the wind speed in the model increases earlier than observed and the wind direction shifts to northerly prior to the actual direction change. Between the afternoon on the 12<sup>th</sup> and the afternoon on the 14<sup>th</sup> all three of these variables in the model match the observations remarkably well. Beginning the morning of the 15<sup>th</sup> the pressure and wind in the model begins to drift slightly from the observations, and a combination of these variables indicates that the front in the model passes over the tower around 11 PM on the 15<sup>th</sup>, rather than near noon on the 16<sup>th</sup>. In the model, the midwestern front seen in the observations does not develop and the northern front over Wisconsin continues to move south the night of the 15<sup>th</sup> rather than remain stationary while the second front in the midwest travels north. Rather than the pressure reaching a minimum around noon on the 15<sup>th</sup> and remaining low until after the front passes around noon on the 16<sup>th</sup>, the pressure in the model continues to decrease until around 7 PM on the 15<sup>th</sup> and begins increasing that night. The wind speed at 396 m in SiB2-RAMS also continues its decrease until about 10 PM on the 15<sup>th</sup>, when it jumps up to a considerably higher speed and becomes northerly just before the front passes over around 11 PM. Looking at the measurements, the wind remains calm the night of the 15<sup>th</sup>, as the tower is between the two fronts discussed in section 2.4.2. Then, when the measured winds strengthen

and become northerly as the observed front passes over on the 16<sup>th</sup>, the wind speed in the model decreases. Starting around noon on the 16<sup>th</sup>, the model again does a reasonable job representing the observations until the afternoon of the 17<sup>th</sup>: the pressure increases the night of the 16<sup>th</sup> and decreases the morning of the 17<sup>th</sup>, the wind speed at 396 m is moderate and the wind direction is changing from northerly to southwesterly. The afternoon of the 17<sup>th</sup> the pressure in the model drops rapidly until around 6 PM when the variables indicate that a front passes over this area. At this time, the wind speed in the model increases and the wind direction shifts again to westerly and then becomes northerly. The front in the model is about six hours ahead of the observed front, and rather than the wind circling around from the east as is seen in the observations, this front behaves similar to the previous two fronts. From the afternoon of the 18<sup>th</sup> to the end of the simulation, the model simulates all three variables reasonably well, although the pressure is overestimated.

Looking at the CO<sub>2</sub> concentrations, Figure 3.6 displays the CO<sub>2</sub> concentration anomalies from both the model and the tower, with the red line indicating the model values and the black line depicting the measured values at the WLEF tower. The top panel shows the anomalous concentration at 30 m and the bottom panel displays the CO<sub>2</sub> concentration anomalies at 396 m. We can see from the top panel that the model underestimates the daily variability of CO<sub>2</sub>: during the day the model overestimates the concentration at WLEF while the model underestimates the concentrations at night. One possible explanation for the lack of variability in the model is that the model may be overestimating the depth of the mixed layer. During the day, SiB2-RAMS may be mixing low CO<sub>2</sub> concentrations over too deep a boundary layer, which would increase the concentrations. At night, the stable boundary layer may be too deep with too much mixing, causing the high CO<sub>2</sub> concentrations at the surface from respiration to be mixed with the low CO<sub>2</sub> air above. Another possibility for the underestimation of the CO<sub>2</sub> concentration at night in the model may be that the resistance between the canopy air space and the reference level may be too

low, which would also overestimate the mixing and underestimate the CO<sub>2</sub> build-up from respiration.



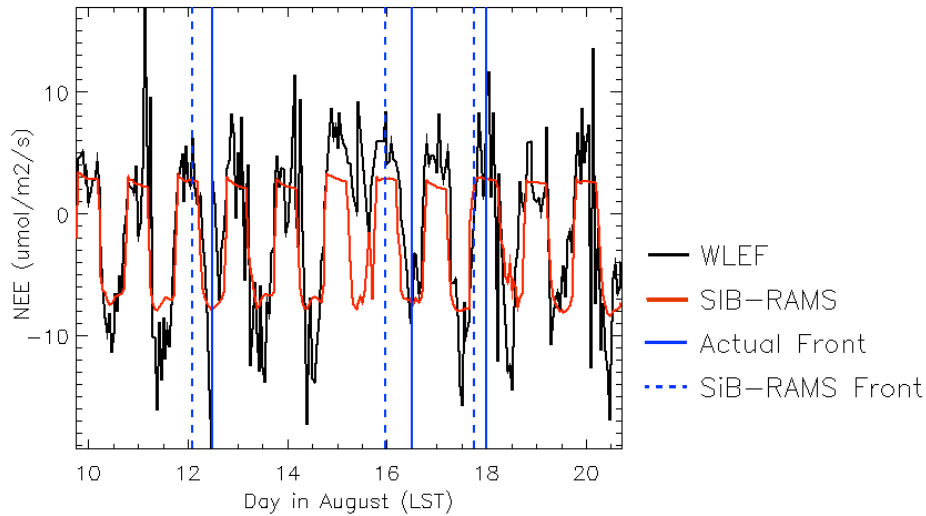
**Figure 3.6:** CO<sub>2</sub> concentration anomalies at 30 m (top) and at 396 m (bottom). The red line depicts the CO<sub>2</sub> from SiB2-RAMS and the black line shows the concentrations at WLEF.

Looking at the bottom panel, the modeled CO<sub>2</sub> concentration at 396 m has three distinct spikes corresponding to the three fronts that passed through. The observations show similar increases for two of the fronts; however, the CO<sub>2</sub> concentration anomalies in the model occurred before the actual CO<sub>2</sub> maximum primarily due to the earlier frontal passages in the model. The behavior of CO<sub>2</sub> at 396 m in the model surrounding the first front is similar to the observations: the CO<sub>2</sub> concentration increases initially prior to the front, decreases for about an hour near when the front passes through, then increases substantially again. Between the first front and the second front, the CO<sub>2</sub> concentrations at 396 m decrease in both the model and the observations.

The response of the CO<sub>2</sub> concentration to the second front is again similar to the observed concentration anomalies, as the large increase in CO<sub>2</sub> occurs entirely before the front in both the observations and the model and prior to the increase there is a decrease in CO<sub>2</sub>, although the decrease in the model is quite small. Both increases start approximately 18 hours before the front; however, in the model the increase is much quicker and the maximum concentration occurs several hours before the front instead of only being one to two hours before the front as seen in the observations. The underestimation of both the magnitude and the length of the CO<sub>2</sub> increase in the model could be due to the model not capturing the midwestern front. This second front south of the main front that was not captured in the model caused the northern front to be stationary for approximately twelve hours, and this delay lengthened the amount of time available for the CO<sub>2</sub> increase and could have drawn in higher CO<sub>2</sub> concentrations from source regions, which would amplify the magnitude of the concentration increase. Looking at the second half of the simulation, the CO<sub>2</sub> concentration decreases in both the model and the observations following the second front. The third front in the simulation again caused the CO<sub>2</sub> concentration at 396 m to increase; however, the observed increase was much smaller. In this front, the model behaves very similarly to the first front: the CO<sub>2</sub> concentration increases before the front, the concentration decreases slightly when the front passes over, and the concentration again increases even more after the front. In the model, the wind also behaves the same as in the first case: the wind increases slightly before the front and the wind direction shifts from the south through the west until it finally becomes northerly. However, in real life, the wind direction at the tower circles around the other way, becoming easterly first and then northerly. The model did not capture the correct wind direction, causing the CO<sub>2</sub> influence to be primarily from the south and the west, rather than from the east and the north. This difference in the behavior of the front could have caused the CO<sub>2</sub> concentration in the model to jump while in the observations it did not because the air was coming from an area of lower concentration.



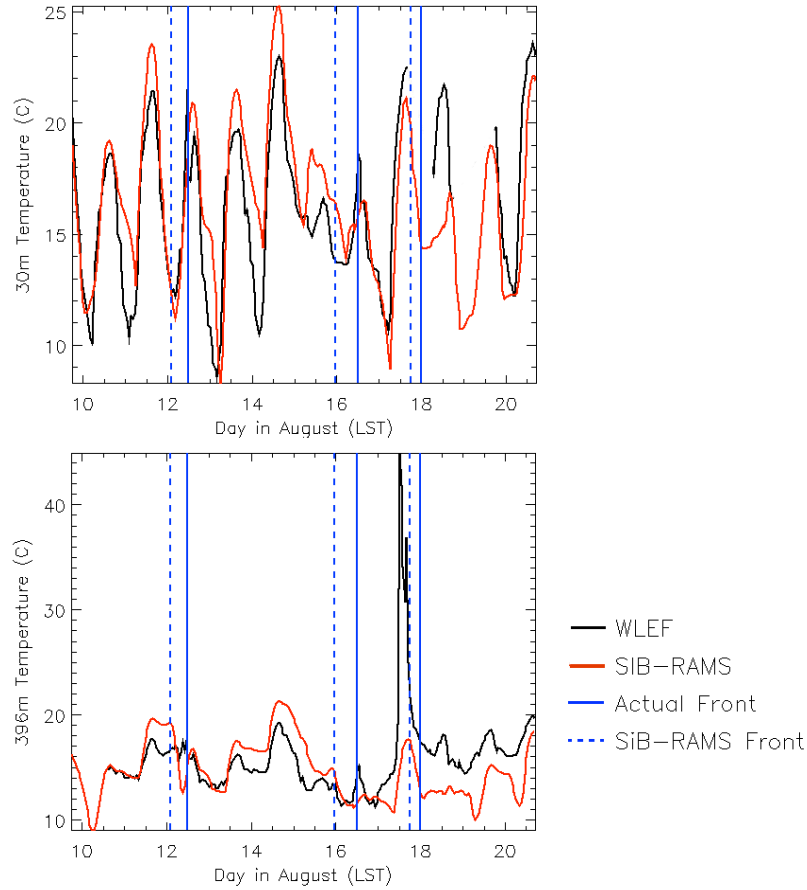
Another variable we evaluated was NEE, which was calculated in SiB2-RAMS by subtracting the assimilation from the respiration. Since the flux footprint of the tower is larger than  $1\text{ km}^2$ , we calculated the NEE in the model for a 4 by 4 km region centered on the tower. The NEE comparison is shown in Figure 3.7, again with the model values in red and the WLEF measurements in black. Overall, the model underestimates the assimilation during the day and also underestimates the respiration at night. The decreases in NEE during the day, as seen on the 15<sup>th</sup> and the 18<sup>th</sup>, are due to clouds decreasing the incoming radiation. Although the model does have some variability in NEE due to clouds, the model does not capture the day-to-day variability that is seen in the observations, as the behavior in SiB2-RAMS is nearly the same every day.



**Figure 3.7:** Observed NEE (black line) and modeled NEE (red line).

The next plot, Figure 3.8, shows the observed and modeled temperature at both 30 m and at 396 m. Looking at the temperatures at 30 m, the model does a good job overall at capturing the diurnal temperature variability until August 15. That afternoon, which is just before the cold front in the model, SiB2-RAMS overestimates the temperature, probably due to not capturing the second cold front in the midwest. Then, the model overestimates the strength of the cold front from Canada and underestimates the temperature on the 16<sup>th</sup> and also on the 17<sup>th</sup>. Looking after the third front during this time period, the model severely underestimates the temperature at 30 m

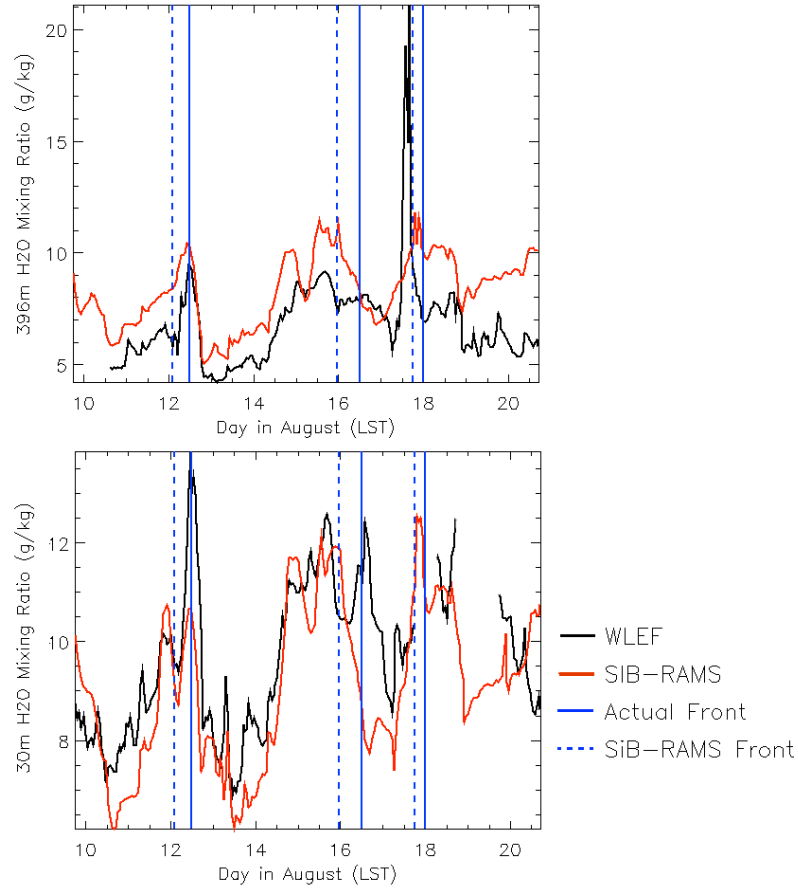
on the 18<sup>th</sup>, helping to support that while the model captured a front, it did not capture the correct behavior for the front and simulated a cold front from the north rather than a cyclonic warm front from the southeast.



**Figure 3.8:** Observed (black) and modeled temperature (red) at 30 m (top) and 396 m (bottom).

The dominant feature in the temperature at 396 m (bottom panel) is the temperature spike around noon on the 17<sup>th</sup>. Although the high temperature remains for more than one hour, the magnitude of the temperature is unrealistic for this area. The remaining temperatures at 396 m look reasonable and the model follows the observations in general, although at times there is a considerable offset between the measurements and the observations. In the middle of the simulation the model overestimates the temperatures for about three days prior to the second front. Following the second front, SiB2-RAMS underestimates the temperatures. The underestimation at 396 m agrees with the underestimation at 30 m, again indicating SiB2-RAMS

overestimates the strength of the second cold front and that it simulates a cold front on the 18<sup>th</sup> rather than a warm front. Despite the offset from the measurements, the model does a fair job at tracking the individual events, such as the temperature increase the night of the 11<sup>th</sup>, the temperature increase on the 13<sup>th</sup>, and the temperature decrease on the 15<sup>th</sup>.



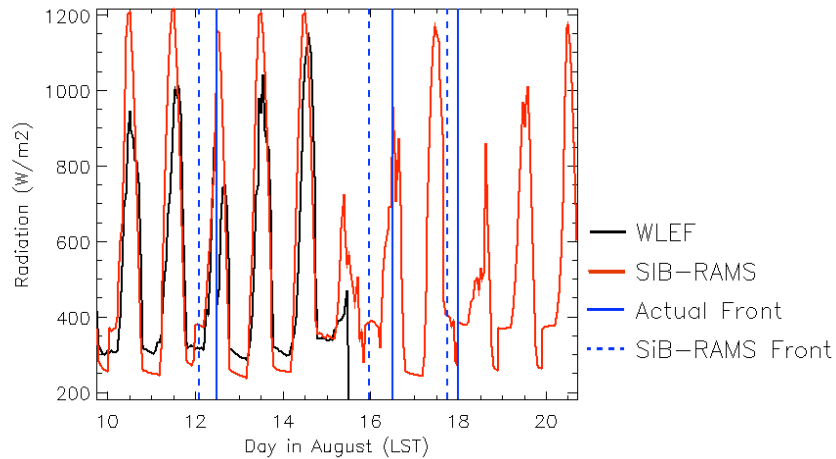
**Figure 3.9:** Water vapor mixing ratio at 30 m (top panel) and at 396 m (bottom panel). The black line is the observed mixing ratio, the red line is the modeled ratio, the blue line indicates the timing of the observed fronts, and the dashed blue line indicates the frontal passages in the model.

Figure 3.9 shows the water vapor mixing ratio at both 30 m and 396 m. The water vapor present at 30 m in the model is relatively close to the observed values, while being consistent with the timing of the fronts in the model. Prior to the first front, the water vapor mixing ratio increases. When the front passes through, the water vapor at 30 m decreases due to precipitation, and immediately following the front the water vapor increases again, similar to the observations. For the few days between frontal systems, the mixing ratio is quite low until it

starts increasing as the second front builds. Around noon on the 15<sup>th</sup> the water vapor content at 30 m decreases as considerable precipitation occurs in the model. Following the precipitation and prior to the front, the water vapor mixing ratio in the model again increases until after the frontal passage the night of the 15<sup>th</sup>, when it drops significantly. Prior to the third front, the water vapor mixing ratio again increases. When the front passes over the night of the 17<sup>th</sup>, there is again considerable precipitation in the model, which decreases the water vapor content. The night of the 18<sup>th</sup> the mixing ratio decreases sharply despite the lack of precipitation in the model; and the water vapor content steadily increases until the end of the simulation.

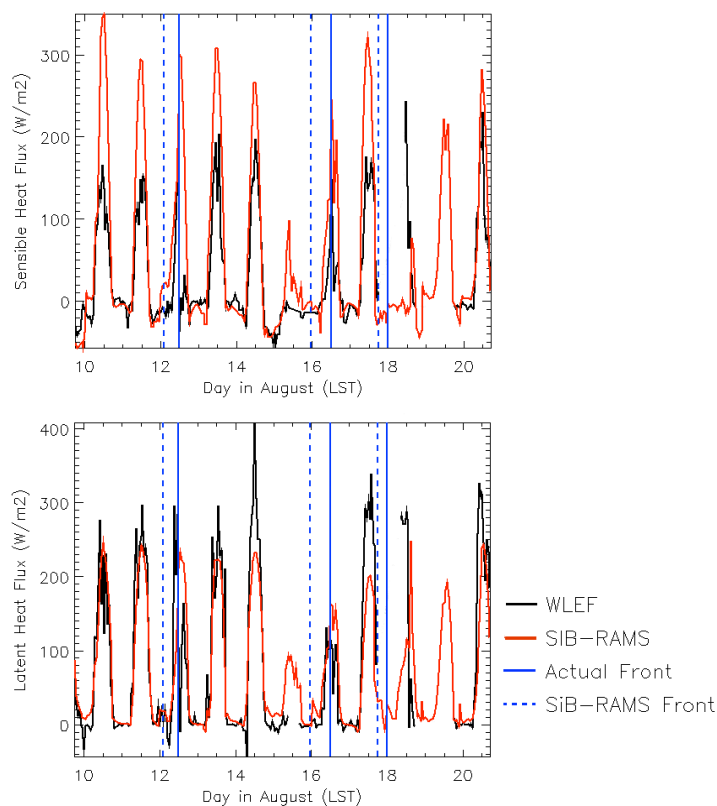
At 396 m, the model overestimates the water vapor mixing ratio, although it does capture much of the variability. Similar to the unrealistic temperature spike in the observations around noon on the 17<sup>th</sup>, the water vapor mixing ratio at 396 m also increases significantly for several hours, but this data is obviously unrealistic and should not be evaluated.

The WLEF tower also measures the incident downwelling radiation, which is shown in Figure 3.10. Clouds cause the sharp decreases seen in the model at specific hours. Even though the second half of the run does not have valid measurements, the figure displays that in general the model overestimates the solar radiation. The model also underestimates the longwave radiation, as the radiation present at night is lower than observed.



**Figure 3.10:** Incident downwelling radiation at the tower (black) and in the model (red).

Finally, the last point comparisons are the sensible and latent heat fluxes at 30 m, shown in Figure 3.11. Since the flux footprint of the tower is larger than  $1\text{km}^2$ , we compared the tower observations to the mean of a 4 km by 4 km region centered on the tower in the model. In general, the model overestimates the sensible heat flux, which is how it compensates for the overestimated radiation. Looking at the bottom panel in Figure 3.11, the model does quite a good job simulating the latent heat flux, particularly at the beginning of the times shown. The biggest differences in the latent heat flux occurs near the end of the simulation when the model underestimates the latent heat flux the afternoons of August 17 and 18, corresponding to the third front. Since the sensible heat flux is overestimated and the latent heat flux is underestimated, the simulated Bowen ratio is too high. This is probably due to the lack of wetlands in the model compared to the real landscape surrounding the WLEF tower, which is 30-40% wetlands (Baker *et al.*, 2003).



**Figure 3.11:** Sensible heat flux at 30 m (top panel) and latent heat flux at 30 m (bottom panel). As in previous plots, the black is measured and the red is modeled.

The point comparisons between the model and the observations indicate that the model captures the overall atmospheric conditions at the tower quite well. In addition, the large-scale maps indicate that the model was also capable of reproducing the dominant pressure fields, even though the specific behavior associated with each front may have been different in the model than observed. Overall, the meteorological and biological variables simulated using SiB2-RAMS are realistic and thus justifies using this simulation to analyze satellite biases.

# Chapter 4: Spatial Representativeness of Satellite CO<sub>2</sub> Measurements

Satellite measurements can be used in inverse models to improve the data constraint; however, the width of the satellite track is not the same as the size of the grid cell in an inverse model. If the inverse model optimizes the concentrations in an entire grid cell to satellite concentrations from only a small fraction of the domain, errors may be introduced into the inversion as the satellite mixing ratios may not be representative of the entire grid cell. To reduce errors in the resulting fluxes, the representativeness of the satellite measurements must be quantified. Using case 1 and case 2 in SiB2-RAMS, this chapter discusses the spatial variability of CO<sub>2</sub> and investigates the spatial representativeness of the satellite measurements by analyzing the concentration differences between a single satellite track and a sample inversion grid cell.

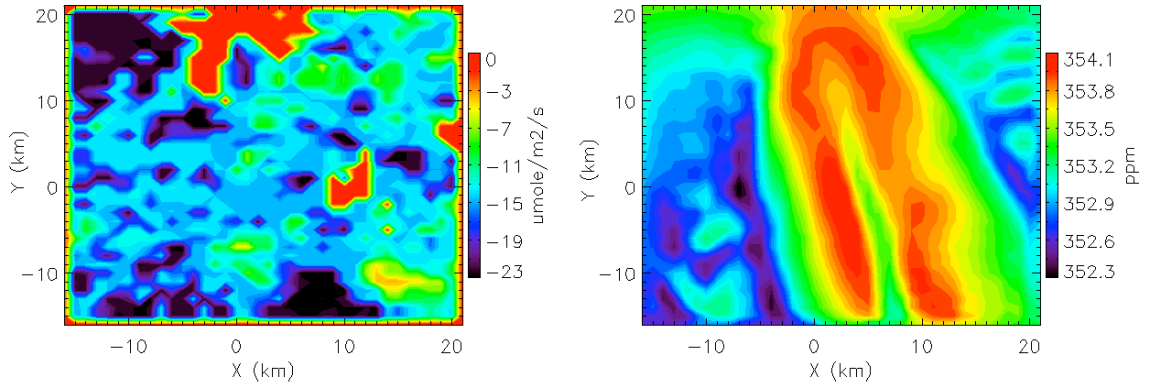
## 4.1 Case 1

Since CO<sub>2</sub> is highly variability at the surface, first we will look at the low-level variability in SiB2-RAMS for both grid 3 and grid 1. The spatial variability of both these domains was first discussed by Nicholls *et al.* [2004].

### 4.1.1 Overview of Surface CO<sub>2</sub> Spatial Variability

Looking at grid 3, which has a domain size of only 38 by 38 km, Figure 4.1 shows the CO<sub>2</sub> flux at 30 m in the left panel and the CO<sub>2</sub> concentration at 396 m in the left panel, both at 1 PM on July 28, 1997. A cold front came through on this day, with strengthened winds bringing cold air from the north. During the day there is a strong drawdown of CO<sub>2</sub> over the land, indicated by the negative flux values. The lakes in this domain are clearly evident because they have zero flux; and due to the lack of photosynthesis, the lakes typically have higher

concentrations during the day. The right panel of Figure 4.1 shows the relatively high concentrations of CO<sub>2</sub> over the larger northern lake being advected south. Nicholls *et al.* show that the temperature of the air above the lake is considerably colder than that over land, which has warmed during the day due to the sensible heat flux. The cold air advects southward and slowly sinks, bringing down higher values of CO<sub>2</sub> from aloft. Just outside the regions of subsidence, air is lifted causing upward motion that raises the CO<sub>2</sub> depleted air from the surface.

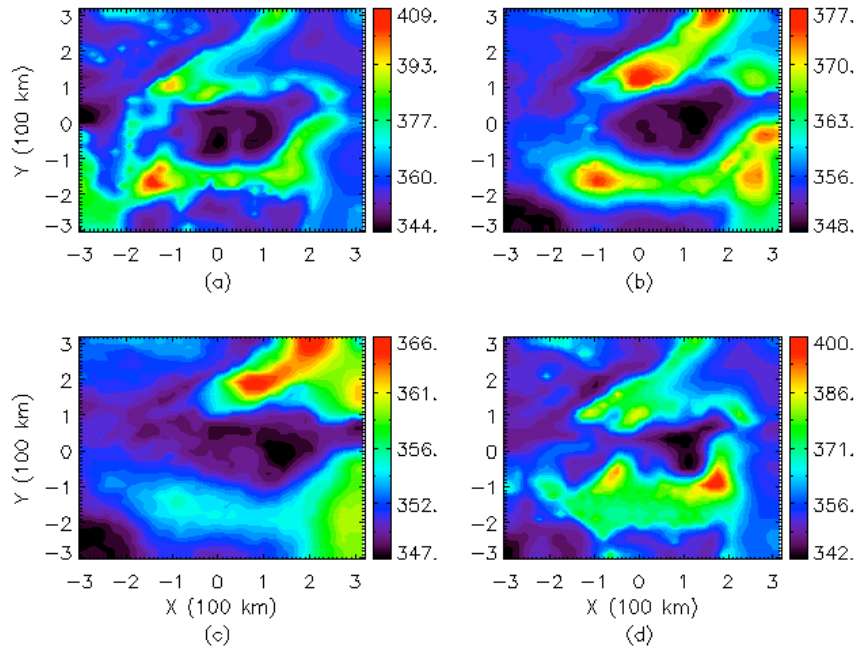


**Figure 4.1:** Simulated CO<sub>2</sub> flux at 30 m (left panel) and CO<sub>2</sub> concentration at 396 m (right panel) at 1 PM for grid 3 on July 28, 1997.

Moving to the larger domain, Figure 4.2 shows horizontal cross-sections of CO<sub>2</sub> concentrations 100 m above the surface at various times for grid 1, which is 640 km. At 4 am on July 29, which is panel (a) in Figure 4.2, there is a region of low CO<sub>2</sub> concentration near the center of the domain surrounded by a ring of higher concentrations. Excluding the lakes, the CO<sub>2</sub> flux is positive throughout the domain due to respiration, and the strongest positive CO<sub>2</sub> flux occurs on the southern border of the domain. The topography and geography cause the airflow to be divergent over the plateau near the center of the domain and convergent in the valleys and over the lakes. The upward motion associated with divergence causes the high surface CO<sub>2</sub> over the land to be lifted and advected, creating the ring of higher CO<sub>2</sub> concentrations. The high CO<sub>2</sub> air from respiration on land continues to advect over the lakes, where the highest concentration occurs just after daybreak.



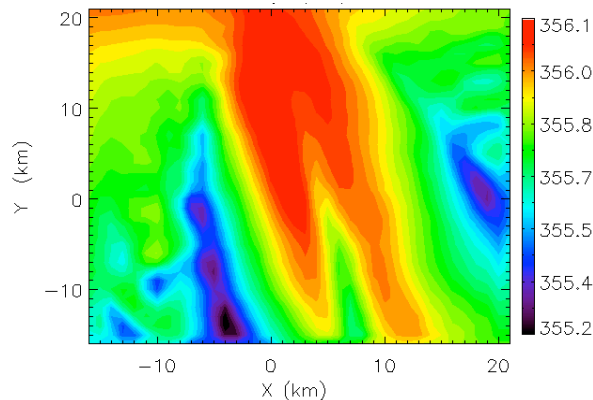
Panel (b) of Figure 4.2 shows the concentrations at 10 am on July 29. At this time, the ring of high  $\text{CO}_2$  has started to disappear as photosynthesis is underway. During the day, the strongest uptake of  $\text{CO}_2$  is near the center of the domain, and by 2 PM the  $\text{CO}_2$  concentrations over the land have dropped to lower values than the concentrations over the lakes. As the evening progresses, the concentration over the lakes decreases as low  $\text{CO}_2$  air is advected from the photosynthetically active land and subsides over the lakes. The concentration over the lakes reaches a minimum around midnight and then begins to increase again due to the advection now of high  $\text{CO}_2$  air from the land. By 4 am the next morning, shown in panel (d), the concentrations over the lakes have started to increase again and the high  $\text{CO}_2$  ring in the domain has returned. To ensure that these patterns were not a boundary effect in the model, Nicholls *et al.* [2003] ran the same simulation with a larger domain and found that these effects still occurred.



**Figure 4.2:** Grid 1  $\text{CO}_2$  concentrations at 100 m. Panel (a) displays the concentration at 4 am on July 29, panel (b) shows the  $\text{CO}_2$  variability at 10 am, panel (c) shows the concentrations at 2 PM, and panel (d) displays the  $\text{CO}_2$  concentrations at 4 am on July 30, 1997.

### 4.1.2 Total Column CO<sub>2</sub> Variability

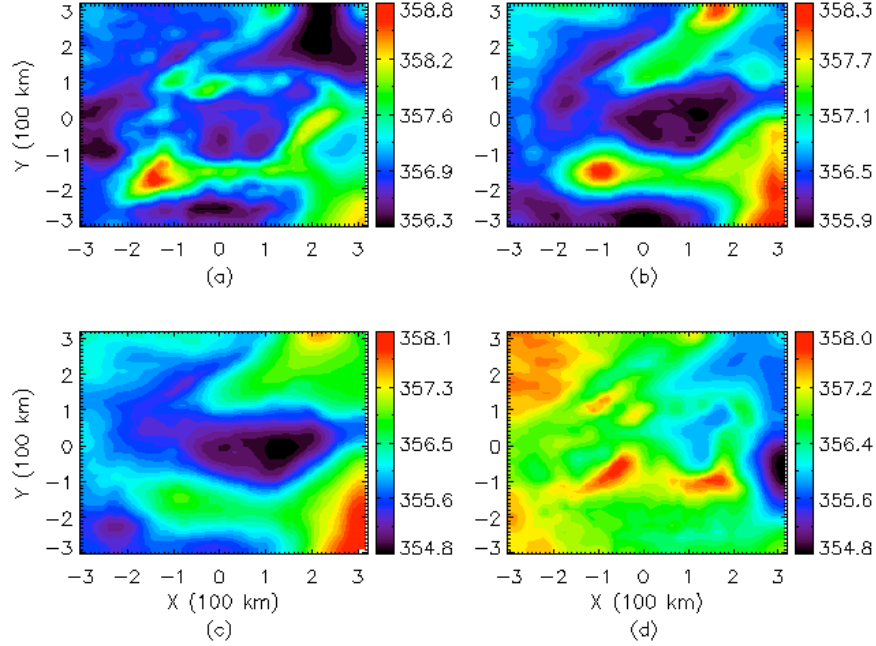
Since the satellite will measure total column mixing ratios, it is important to study the spatial variability of total column CO<sub>2</sub>. Figure 4.3 shows the total column CO<sub>2</sub> concentrations over grid 3 at 1 PM on July 28 and can be compared to the right panel in Figure 4.1. Although including the CO<sub>2</sub> concentrations above the boundary layer in total column measurements dampens the variability near the surface, the surface spatial patterns of CO<sub>2</sub> can still be seen: high CO<sub>2</sub> over the lakes is being advected south while the western portion of the grid has lower concentrations due to the photosynthetic draw-down. While the total column concentrations capture the surface spatial patterns, the primary difference between total column and near-surface measurements is the range of variability. Rather than varying by several ppm over a 38 km domain, the range of total column concentrations on the same domain is only several tenths of a ppm up to 1 ppm.



**Figure 4.3:** Grid 3 total column CO<sub>2</sub> concentrations at 1 PM on July 28, 1997.

Moving to the spatial variability on the large domain, Figure 4.4 shows the total column CO<sub>2</sub> concentrations for grid 1. Figure 4.4 can be compared to the near-surface variability shown in Figure 4.2. The patterns of variability for total column CO<sub>2</sub> are nearly identical to the variability at 100 m, except that the range of concentrations is smaller. Although the spatial variability is greater on grid 1 than on grid 3 due to the larger spatial area, the total column concentrations only varied by approximately 3 ppm on the large domain, rather than the 60 to 70

ppm range seen near the surface. Overall, investigating the spatial variability of total column CO<sub>2</sub> concentrations shows that total column concentrations reflect the surface CO<sub>2</sub> patterns but have a smaller concentration range.

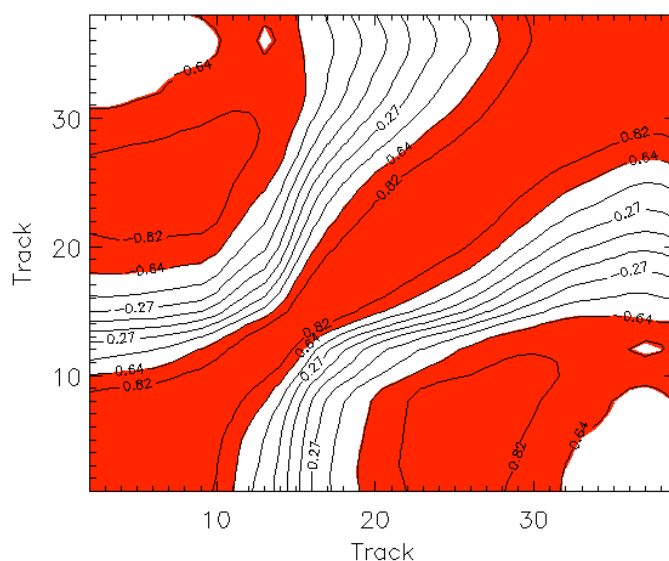


**Figure 4.4:** Grid 1 total column CO<sub>2</sub> concentrations. Similar to Figure 4.2, (a) is 4 am on July 29, (b) is 10 am, (c) is 2 PM, and (d) is 4 am on July 30, 1997.

#### 4.1.3 Satellite Track Spatial Variability

Before looking at the spatial representativeness of a single satellite track, this study will look at the spatial variability of the tracks over the five-day simulation using correlation maps. Correlation maps will help to investigate how independent the information from a single track is in relation to the entire domain and also how the track values vary throughout the domain. Prior to calculating the correlation coefficients, the CO<sub>2</sub> time-series was detrended and the diurnal cycle was removed. To get the correlation values, each possible satellite track was correlated in time to every other possible track in the domain. To determine the 95% significance level, we used the two-tailed t-statistic and ten degrees of freedom (two degrees of freedom per day). In the correlation maps shown, the contour lines denote the correlation coefficients and the red regions indicate areas that met the 95% significance requirements.

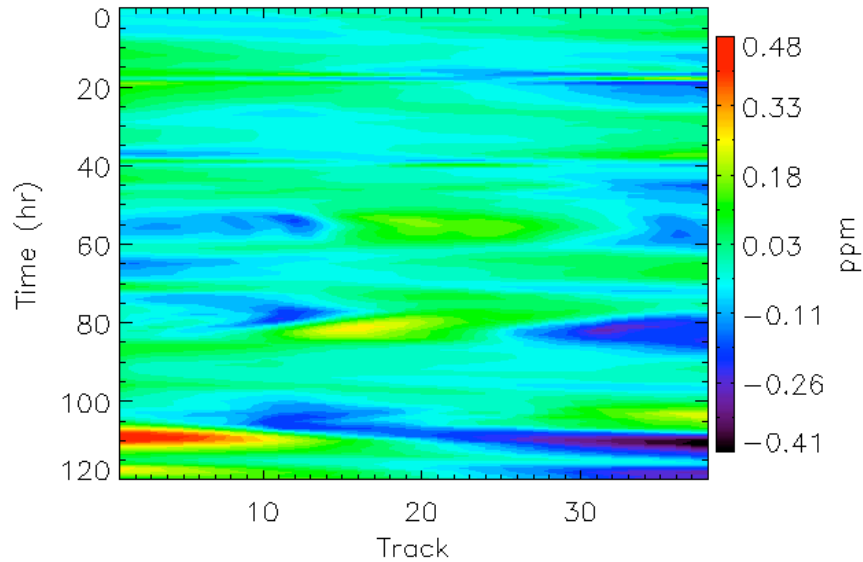
The correlation map for grid 3 is shown in Figure 4.5. Both the x and y axis represent the satellite track number, where track 1 represents the value a satellite would retrieve if it flew on the western edge of the domain and track 38 represents the CO<sub>2</sub> concentration retrieved from values on the far eastern edge of grid 3. Figure 4.5 reveals that the concentrations from each track are significantly correlated in time with themselves as well as five to ten adjacent tracks. If all the tracks had been significantly correlated, then the domain would not have any spatial variability and any single track would represent the domain; however, in this case not all swaths are significantly correlated as the correlation coefficient drops off rapidly beyond the neighboring tracks. Since the significant correlation for each swath extends for approximately five neighboring tracks in either direction, the satellite swath is accurately capturing the CO<sub>2</sub> concentration for a region nearly ten times its size but may not be capturing the concentration for the entire domain.



**Figure 4.5:** Correlation map of satellite tracks in grid 3. Red regions indicate 95% significance.

One interesting feature in Figure 4.5 is the significant negative correlations between the tracks slightly to the west of the tower (tracks 10-15) and the tracks on the eastern edge of the domain. These negative correlations indicate that high CO<sub>2</sub> concentrations on the western side of the domain correspond in time to low CO<sub>2</sub> concentrations on the eastern part of the domain. To

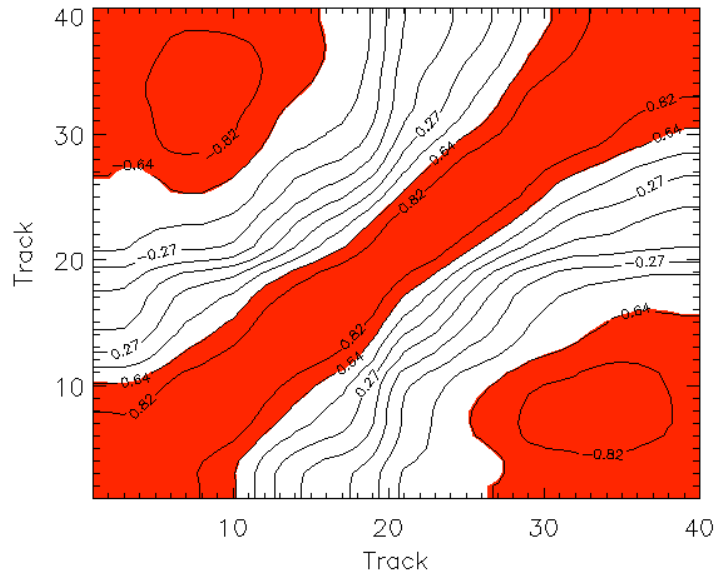
help determine the cause of this negative correlation, a Hovmuller diagram is shown in Figure 4.6. This figure shows the tracks on the x-axis and the time in hours on the y-axis, where time 0 is 6 am on July 26. The concentrations plotted in this figure are the hourly CO<sub>2</sub> concentrations from each emulated satellite track minus the hourly domain mean CO<sub>2</sub> concentration. Figure 4.6 shows that concentration anomalies typically begin on the western edge of the domain and are advected eastward and that at any specific moment in time the concentration differences near the western side of the domain are opposite in sign from the differences in the tracks on the eastern edge. This negative correlation is primarily due to the topography and advection.



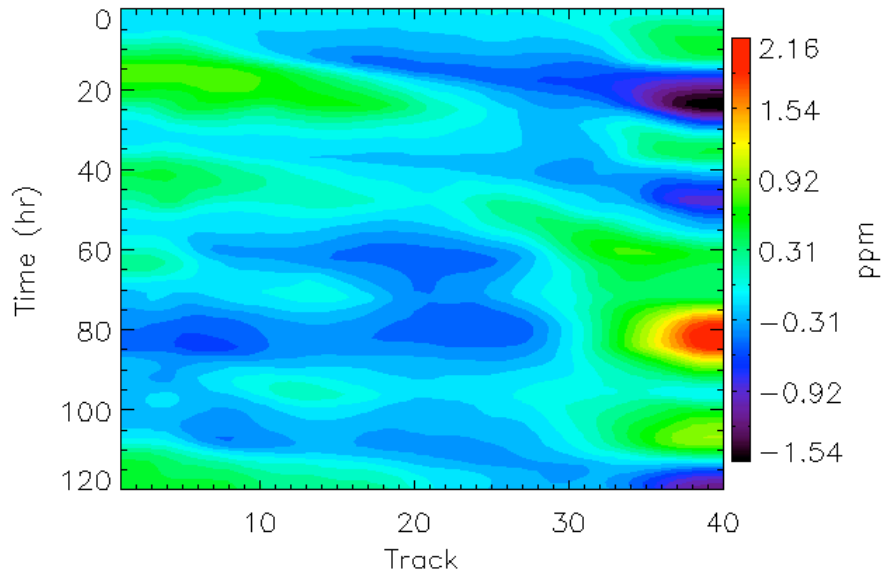
**Figure 4.6:** Hovmuller diagram for grid 3. Time 0 corresponds to 6 am on July 26 and time increases downwards. The CO<sub>2</sub> concentrations are the differences relative to the hourly domain average.

Figure 4.7 shows the correlation plot for grid 1. The correlation map is similar to Figure 4.5: the tracks are significantly correlated with themselves down the diagonal and tracks on opposite sides of the domain are negatively correlated. The width of the diagonal correlation band is slightly smaller than the width for grid 3, indicating that there is more variability throughout this domain. The size and location of the negative correlation is also different between grid 1 and grid 3. The tracks on the western side of the grid are negatively correlated with more tracks on the eastern half of the domain than in grid 3. The larger areal extent of the

negative correlation on grid 1 is due to the physical features in the domain: the western portion of the grid is predominantly forest and the eastern portion is covered by the great lakes, and the different CO<sub>2</sub> diurnal cycle over land and lakes can explain part of the negative correlation.



**Figure 4.7:** Correlation map for grid 1. Red shading indicates 95% significance.



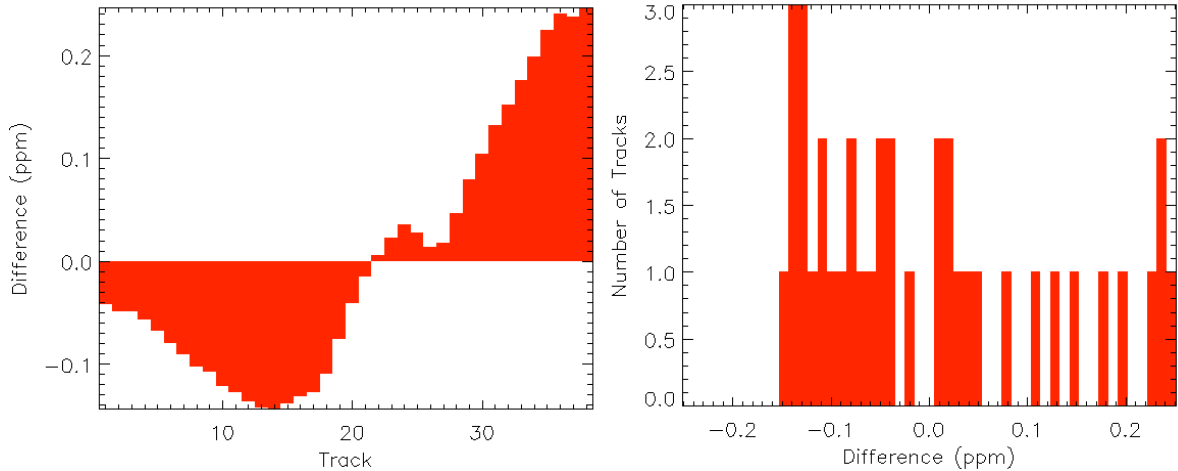
**Figure 4.8:** Hovmuller diagram similar to Figure 3.6 but for grid 1.

Another contributing factor to the negative correlation is large-scale synoptic patterns. Figure 4.8 shows the Hovmuller plot for grid 1, and again we see that concentration anomalies move from east to west and that concentrations tend to be anomalously low (high) over the

western portion of the domain while being high (low) on the eastern portion of the domain. The advection of CO<sub>2</sub> combined with the opposite behavior of CO<sub>2</sub> over land and over water could be causing the negative correlations seen in Figure 4.7. The correlation plots emphasize the spatial variability in the domains and indicate that since all the tracks are not significantly correlated a single track does contain some independent data and may or may not be representative of the entire domain.

#### **4.1.4 Spatial Representativeness Errors**

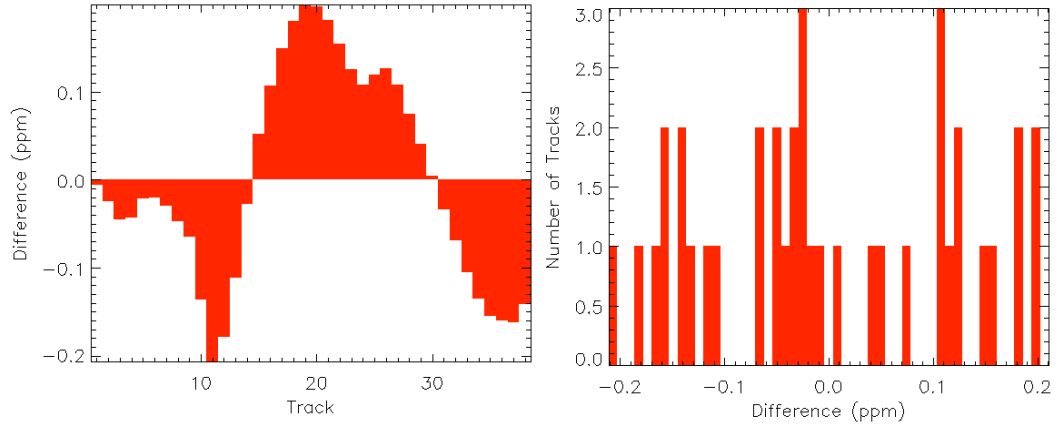
In investigating the ability of a single satellite track to represent the spatial variability over the entire domain at 1pm, we will analyze grid 3 first. As a first case, we will determine the difference between each of the 38 swaths on grid 3 and the domain mean on a relatively calm day, July 29. For this comparison, the domain mean at 1 PM is subtracted from the value of each track. The results are shown in Figure 4.9. The left panel of the figure shows the error for each track and the right panel compiles the differences into a sampling distribution, with the difference on the x-axis and the number of tracks corresponding to that difference value displayed on the y-axis. Figure 4.9 shows that at 1 PM on July 29 the tracks on the western half of the domain underestimate the true CO<sub>2</sub> concentration while the eastern half of the grid overestimates the concentrations, with the largest deviation from the true mean occurring on the eastern edge of the domain. At that time, the concentrations on the western half of the grid were low due to photosynthesis and the northeast corner of the grid had higher CO<sub>2</sub> concentrations from high CO<sub>2</sub> air being advected from the lake. While none of the swaths captured the mean concentration, all of the tracks captured the mean CO<sub>2</sub> within 0.26 ppm. Since the differences between the satellite tracks and the true mean are small, all possible satellite tracks on July 29 closely match the true CO<sub>2</sub> concentration.



**Figure 4.9:** Grid 3 spatial representativeness errors at 1 PM on July 29, which were calculated by subtracting the domain average at 1 PM from the simulated satellite concentrations at 1 PM. The left panel displays the track number on the x-axis and the corresponding difference on the y-axis. The right panel compiles the differences into a sampling distribution, with the x-axis showing the difference and the y-axis displaying the number of tracks corresponding to each value.

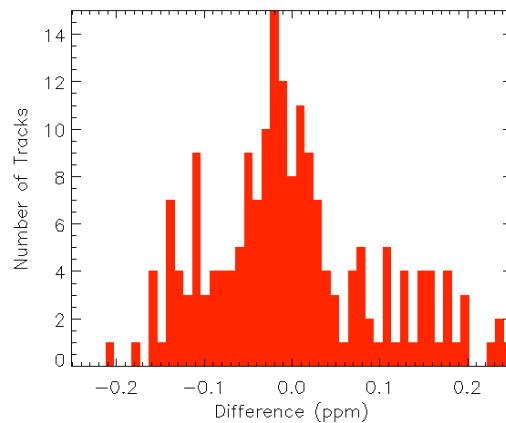
Since a single simulated satellite track has the capability to closely capture the true concentration of  $\text{CO}_2$  on a typical calm and sunny day, next we will look at the spatial representativeness errors on July 28 when the cold front passed through. Although the front was not accompanied by any clouds, the winds changed and strengthened, altering both the pattern of spatial variability and the source region of  $\text{CO}_2$ . The results, shown in Figure 4.10, display that for this case the tracks on the eastern and western boundaries underestimated the average  $\text{CO}_2$  concentration while the swaths in the middle of the domain overestimated the mean. Looking at the total column  $\text{CO}_2$  concentrations for this time shown in Figure 4.3, we can see that the tracks that overestimated the concentration were the tracks over the plume of high  $\text{CO}_2$  from the lake. The tracks on the western boundary and the tracks surrounding the high  $\text{CO}_2$  plume on the edges of the updraft were the most representative tracks in the domain. The range of errors on this day was actually less than the following day, and all the tracks captured the concentration within 0.21 ppm.





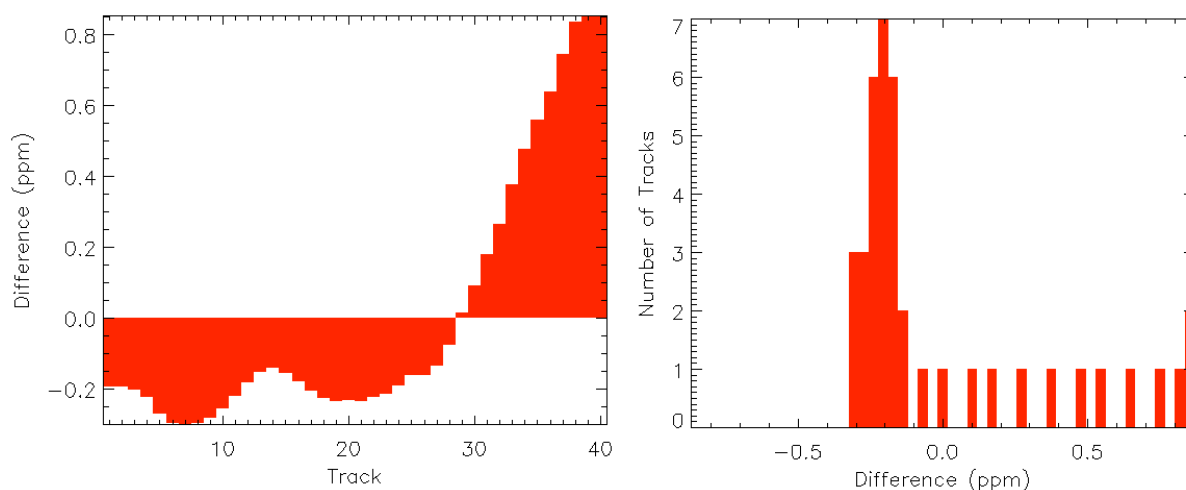
**Figure 4.10:** Grid 3 spatial representativeness errors at 1 PM on July 28, 1997.

Finally, we will investigate the representativeness of a single track that may be taken on any day at 1 PM over the five-day simulation. Following the same procedures described above, the sampling distribution in this case is a compilation of the differences between each track at 1 PM and the corresponding domain-averaged column mean mixing ratio at 1 PM for all five days and is shown in Figure 4.11. The spatial representativeness error is primarily centered on zero, with a maximum error of only 0.25 ppm and with 95% of the tracks capturing the mean concentration within 0.2 ppm. This simulation indicates that the representativeness error in satellite total column CO<sub>2</sub> measurements used on a small domain will be small and relatively evenly distributed between underestimating and overestimating the CO<sub>2</sub> concentration.



**Figure 4.11:** Spatial sampling distribution for all tracks at 1 PM. Similar to Figures 4.9 and 4.10, the difference on the y-axis is calculated by subtracting the domain average at 1 PM from the total column simulated satellite concentrations. This figure is a compilation of the errors from each day in the simulation.

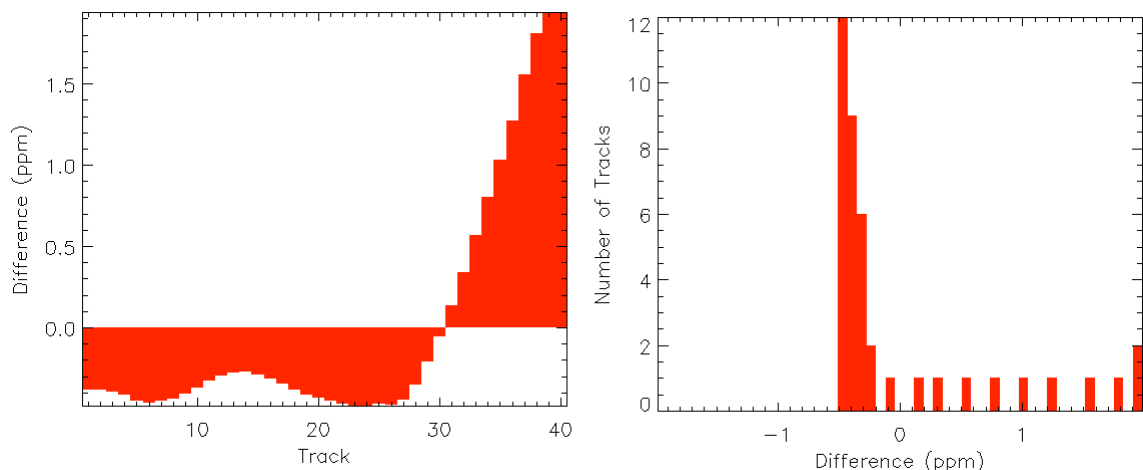
After examining the small domain, we will now analyze grid 1 using the same techniques. The first case will be to determine the accuracy of each of the 40 tracks over this 640 km domain on July 29. Figure 4.12 shows the mean  $\text{CO}_2$  concentration for grid 1 at 1 PM subtracted from the  $\text{CO}_2$  concentration from each track. Similar to grid 3, the figure clearly shows that the tracks over the western portion of the domain underestimate the average  $\text{CO}_2$  concentration while the tracks on the eastern side of the domain overestimated the true  $\text{CO}_2$  value; however, the sampling errors from each track are greater for grid 3 than they are for the smaller domain.



**Figure 4.12:** Grid 1 spatial representativeness errors at 1 PM on July 29, 1997. The errors are calculated by subtracting the domain average total column  $\text{CO}_2$  concentration from each of the simulated satellite mixing ratios (as in the figures for grid 3).

The larger differences between the track and the mean are due to the greater heterogeneity seen in this domain: the variation in  $\text{CO}_2$  concentration at this time is 3.28 ppm, which is much larger than the 1 ppm variability seen on the smaller domain. Knowing the diurnal cycle of  $\text{CO}_2$  over land differs from the diurnal cycle over water, it is no surprise that the tracks on the western half of the domain over land underestimate the  $\text{CO}_2$  concentration due to the photosynthetic uptake while the tracks over the water overestimate the total column concentration. Looking at the sampling distribution in the right panel of Figure 4.12, most of the tracks underestimated the concentration by 0.3 to 0.1 ppm, indicating that the total column  $\text{CO}_2$

concentration over the forest is relatively homogeneous. Looking at the tracks that overestimate the  $\text{CO}_2$  concentration, the more water that the satellite track covered, the larger the overestimation of  $\text{CO}_2$ . All of the tracks capture the true domain concentration at this time within 0.9 ppm.

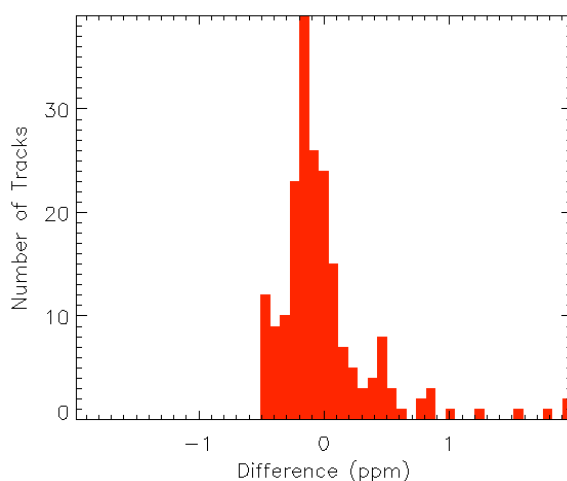


**Figure 4.13:** Grid 1 spatial representativeness errors at 1 PM on July 28, 1997.

The representativeness of a single swath during the cold front on July 28, shown in Figure 4.13, is very similar to the following day: the tracks over land underestimated the  $\text{CO}_2$  concentration while the tracks over the lakes overestimated the mean total column  $\text{CO}_2$ . Although the sampling distribution has a similar shape, the errors and the standard deviation for this case is much larger, indicating that a single track is less representative during the front. In this case, all of the tracks captured the mean concentration on the domain within 1.9 ppm, which is twice as large as the maximum error on July 29.

The last case on grid 1 will investigate the total column  $\text{CO}_2$  concentration of a random single track taken at 1 PM any day of the study. Similar to grid 3, this distribution was calculated by subtracting the mean value at 1 PM from the concentration of  $\text{CO}_2$  for each track on the corresponding day and compiling the results together into a single sampling distribution plot. The results for this calculation are shown in Figure 4.14. The figure shows that the distribution is shifted slightly negative, with approximately two-thirds of the tracks underestimating the true

CO<sub>2</sub> concentration and the other third overestimating the concentration. 95% of the tracks capture the true concentration within 0.74 ppm, and the mean variability at 1 PM over the five days on grid 1 is 3.4 ppm. The tracks that overestimate the concentration by more than 1 ppm are the tracks over the lakes on the day of the front, as we saw in the previous case. Overall, studying this case indicates that the representative errors are smaller than the 0.5% instrumental error and that the errors are much smaller for grid 3 than for the larger domain, primarily due to the smaller range of CO<sub>2</sub> variability in a smaller domain.



**Figure 4.14:** Grid 1 spatial representativeness errors at 1 PM for all five days of the simulation.

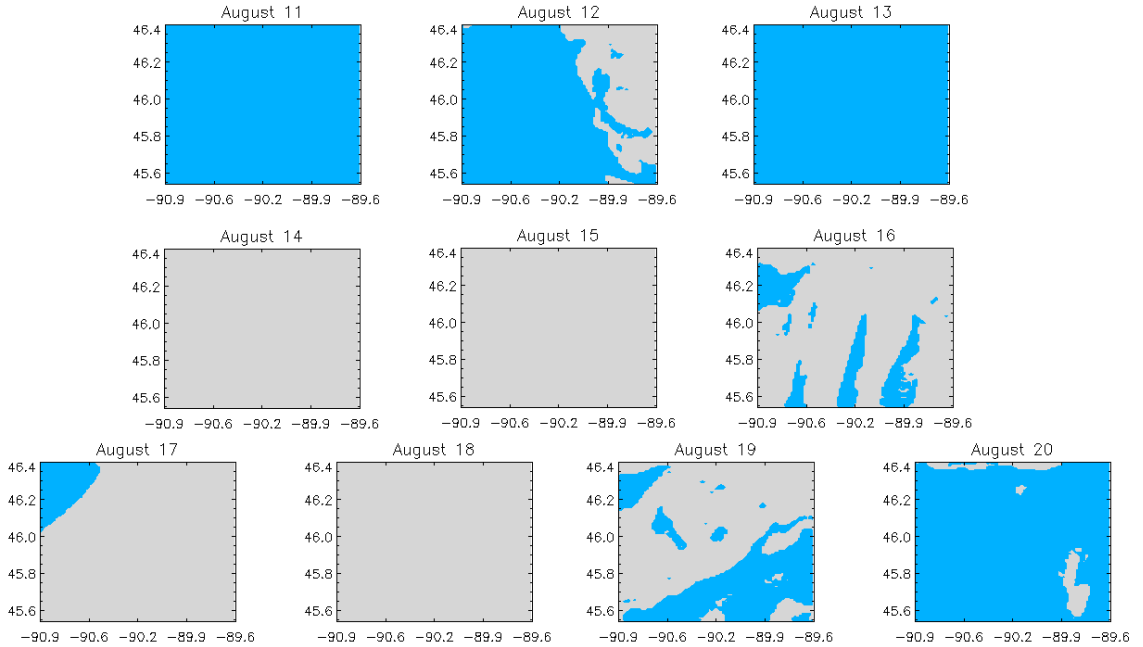
Similar to Figures 4.13 and 4.12, the difference on the y-axis is calculated by subtracting the domain average at 1 PM from the total column simulated satellite concentrations. This figure is a compilation of the errors from each day in the simulation.

## 4.2 Case 2

Similar to our analysis for case 1, we will investigate the ability of a single satellite track to represent the average concentration throughout the entire domain. Since this case includes clouds, we will first discuss the cloud cover patterns at 1 PM for each day. Next, we will investigate NEE patterns and the total column variability at 1 PM for both grids 4 and 3. Lastly, we will analyze the representativeness error from each clear-sky track at 1 PM for both grids.

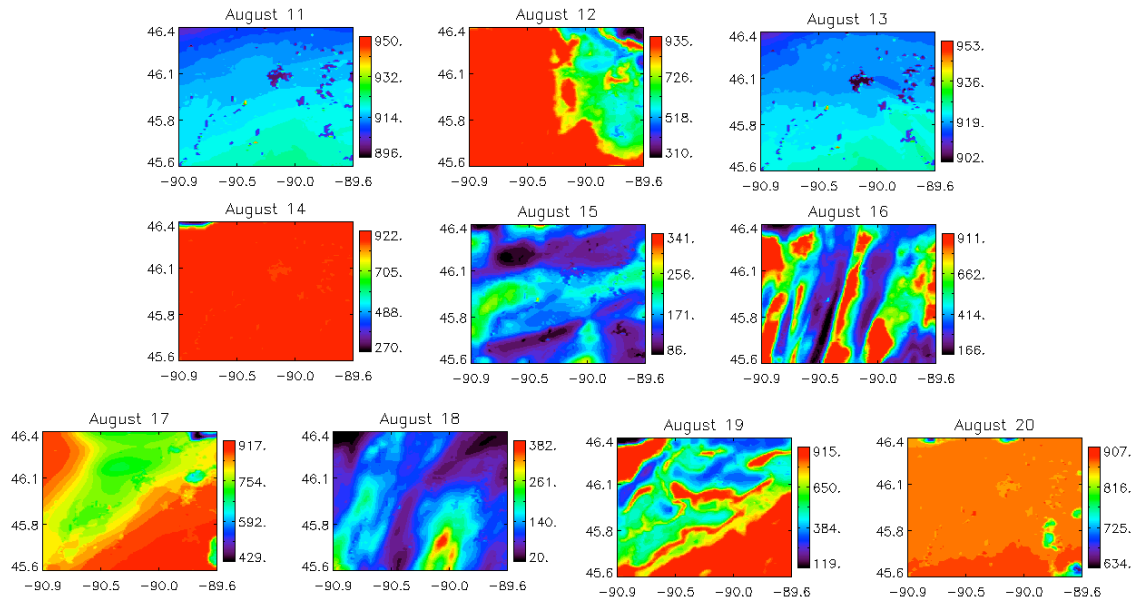
### 4.2.1 Cloud Cover

Since case 2 includes clouds, we need to identify the cloudy pixels in the simulation at 1 PM. For grid 4, a single 1 km by 1 km pixel in the domain was considered cloudy if any vertical level in the atmosphere had any liquid water or ice condensates. Maps of the cloud cover from each day at 1 PM for grid 4 are shown in Figure 4.15, where blue indicates the sky is clear and gray indicates the pixel is considered cloudy. During our ten-day simulation, two days are completely clear, four days are partly cloudy, one day is mostly cloudy, and three days are completely cloudy. The two clear days are at the beginning of the run before the first front and between the passage of the first front and the second front. Except for August 20, the partly cloudy days are the days just after a frontal passage: a front passed over the region in the early morning on the 12<sup>th</sup>, in the late afternoon on the 15<sup>th</sup>, and in the evening on the 17<sup>th</sup>. In this simulation, the days when the domain is completely cloud covered correspond to the afternoon before a front. For example, August 14 and 15 are completely cloudy prior to the passage of the second front and August 18, the day of the third front, is also cloudy.



**Figure 4.15:** Grid 4 cloud cover at 1 PM for each of the ten days in the simulation. Grey indicates clouds, blue is clear-sky.

Simulated radiation values are helpful to determine the thickness of the cloud cover as regions covered by optically thick clouds have significantly lower downwelling shortwave radiation at the surface. Figure 4.16 shows maps of the shortwave downwelling radiation incident at 1 PM on the surface of grid 3 for each day of the simulation. Looking at the partly cloudy days, on August 12 the eastern portion of the domain is on the edge of a cloud deck approximately 1 km off of the ground that extends further northeast. The radiation measurements help indicate that the cloud-deck continues beyond the edge of the domain as the radiation values, which are low on the far eastern portion of the domain, gradually increase toward the middle of the grid as the cloud cover thins.



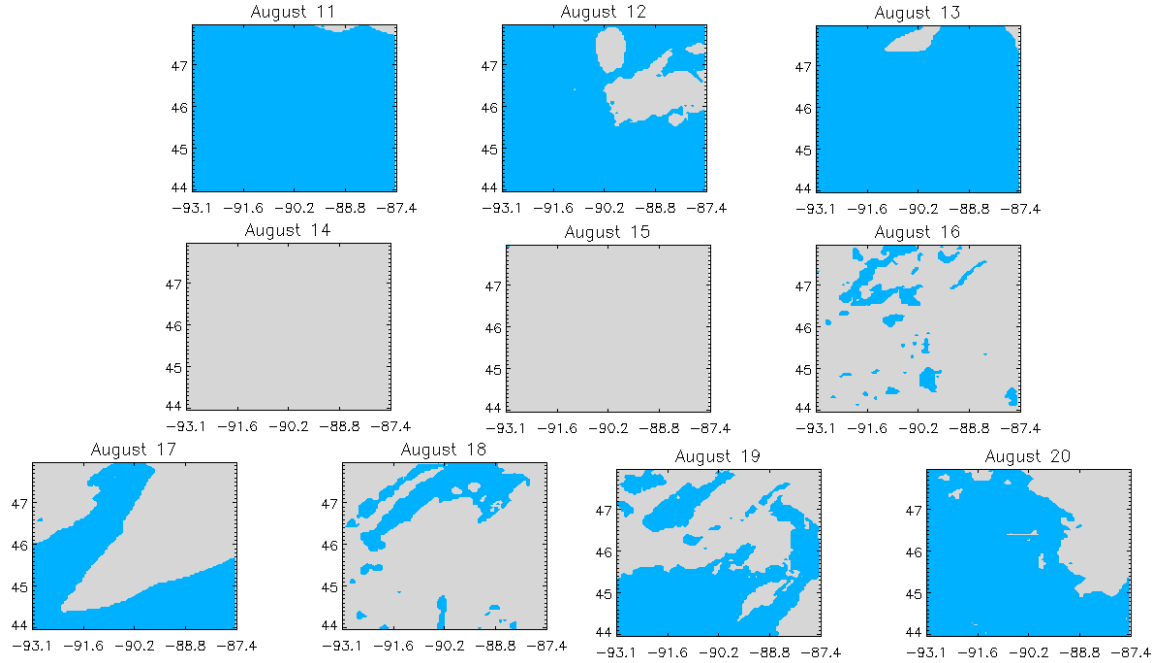
**Figure 4.16:** Daily 1 PM maps of shortwave downwelling radiation incident on the surface of grid 4. The x-axis is the longitude, the y-axis is the latitude, and the units contoured are  $\text{W/m}^2$ .

Moving on to the 16<sup>th</sup> of August, the clear patches in the domain are stripes in the relatively thick low cloud cover, which is about 1 km above the surface and which extends well beyond the boundary of grid 4. Corresponding stripes of high radiation also exist; however, they are shifted slightly to the west. Similar to August 12, the clear corner on August 17 is the edge of the cloud cover as the radiation values decrease from the northwest corner. The clouds in the

middle of the domain on the 17<sup>th</sup> are relatively low clouds situated about 1 km above the ground that block a considerable amount of shortwave radiation; however, the clouds on the southeastern corner of the domain are most likely thin cirrus clouds, as the radiation values are quite high and the cloud level is approximately 8 km above the surface. August 19 is partly cloudy with low, thicker clouds as well and is relatively evenly split between clear and cloudy pixels. Finally, on August 20 the sky is mostly clear and the clouds are relatively small in areal coverage. On this day, the clouds are puffy cumulous clouds sitting on top of the mixed layer approximately 2 km above the ground.

On August 14, which is completely cloudy, high cirrus clouds are the primary cloud cover as the incident downwelling radiation is quite large and the cloud cover is quite high with condensates present from 6 to 13 km. In contrast, the cloud cover on the 15<sup>th</sup> is a thick stratus layer that is effective at blocking the incoming radiation and cloud liquid and ice condensate is present in all layers of the model. Similarly, the cloud cover on August 18 is also thick since the radiation values are low, but on this day rather than the clouds extending further up in the atmosphere, the cloud top is around 5 km.

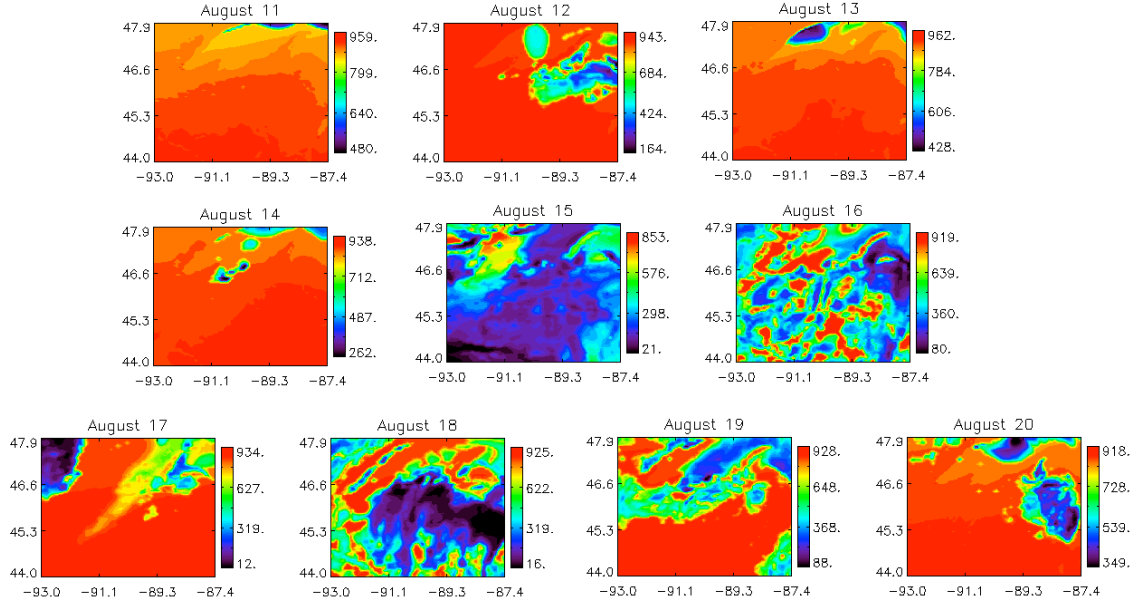
Moving on to the larger grid, Figure 4.17 shows cloud cover maps for grid 3. We used the same method to determine if the sky was clear or cloudy that we used for grid 4: a single 5 km by 5 km pixel in the grid was cloudy if there was liquid or ice condensates anywhere in the atmosphere. For grid 3, two days are mostly clear, five days are partly cloudy, and two days are completely cloudy.



**Figure 4.17:** Grid 3 cloud cover at 1 PM for August 11 through August 20, 2001. Grey indicates clouds while blue is clear-sky.

To help understand the types of cloud cover in the simulation, Figure 4.18 shows maps of the 1 PM downwelling shortwave radiation incident at the surface of grid 3. On August 11, 12, and 13, the small clouds on the northern portion of the grid are probably cumulous clouds, as a vertical profile of the liquid and ice condensates indicates that the clouds exist approximately 2 to 3 km above the ground. As seen in the small domain, August 14 is primarily covered by thin cirrus 6 to 14 km above the surface that allow the majority of the shortwave radiation to pass through. August 15 is completely cloudy, with low radiation values and condensates apparent throughout nearly the entire vertical profile. On August 16, the cloud cover appears to be breaking up, with many clear patches through the low-level, thick clouds. Looking at August 17, 18, and 19, the clouds are patchy and exist from levels near the surface up to approximately 15 km. Finally, on August 20 the sky is still only partly cloudy, but rather than having low and high level clouds, only cumulus clouds 1 to 2 km above the surface are present in the domain.



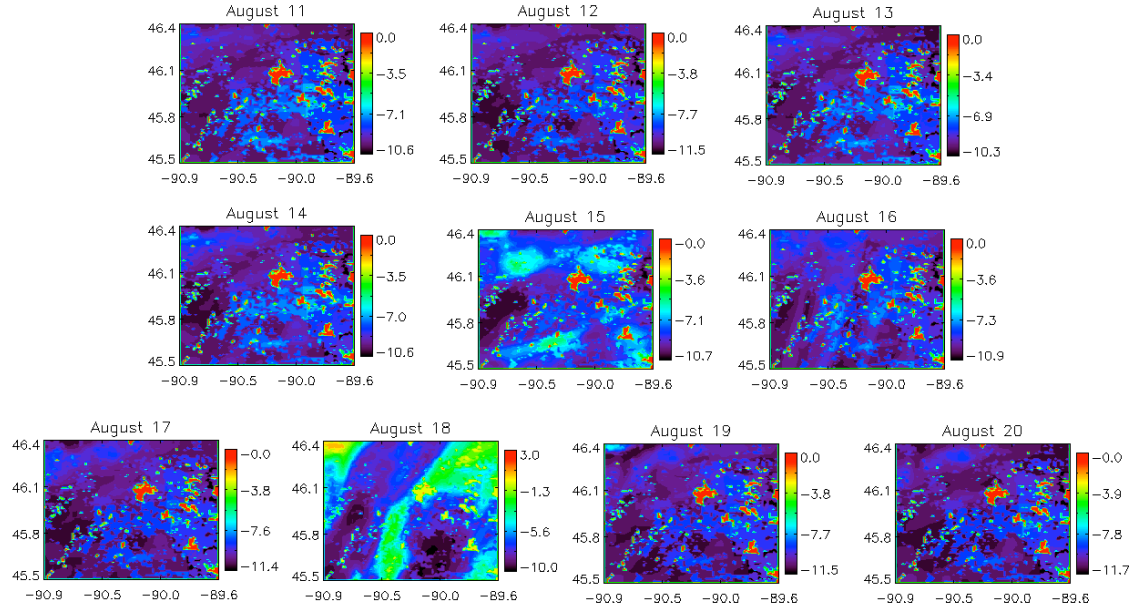


**Figure 4.18:** Daily 1 PM maps of shortwave downwelling radiation incident on the surface of grid 3. The x-axis is the longitude, the y-axis is the latitude, and the units contoured are  $\text{W/m}^2$ .

#### 4.2.2 NEE Variability

To help determine the cause of the  $\text{CO}_2$  concentration patterns and to determine the effects of clouds on the  $\text{CO}_2$  flux, we will investigate the net ecosystem exchange (NEE) on grid 4 and grid 3. The modeled NEE is calculated by subtracting the assimilation from the respiration.

Figure 4.19 shows 1 PM maps of NEE on grid 4 for each day in case 2. The lakes are clearly visible as they have no flux. Throughout the entire domain at 1 PM, the vegetation is photosynthesizing and is a sink of carbon. NEE is remarkably similar from day-to-day and in general the pattern looks like the biome map for this grid: areas covered with broadleaf deciduous trees have the greatest carbon uptake with an average NEE of  $\sim -10 \mu\text{mol/m}^2/\text{s}$ , mixed forests uptake less carbon and have an average NEE of  $\sim -7 \mu\text{mol/m}^2/\text{s}$ , and areas with short vegetation have the lowest  $\text{CO}_2$  flux of only  $\sim -4 \mu\text{mol/m}^2/\text{s}$ .



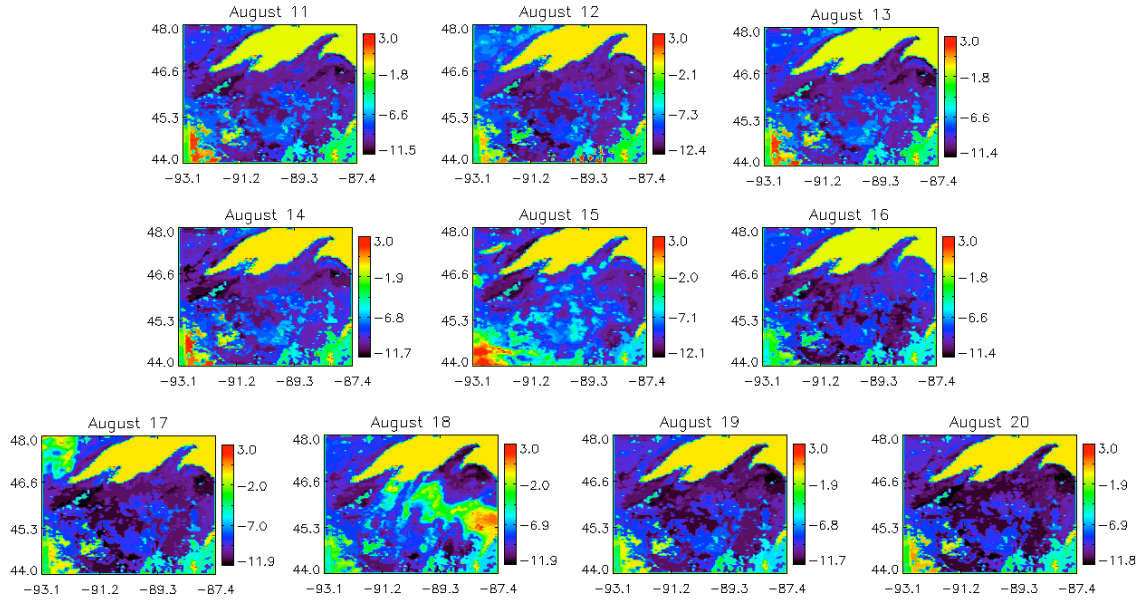
**Figure 4.19:** Daily NEE maps for grid 4, with units of  $\mu\text{mol}/\text{m}^2/\text{s}$ .

Looking at the patterns more in depth, the two main factors that appear to be influencing the NEE on this grid are clouds and temperature. In this simulation, the clouds have to be optically thick enough to reduce the incoming radiation to only  $\sim 300 \text{ W}/\text{m}^2$  to effect the vegetation. In this simulation, photosynthesis is not enhanced by slight cloud cover increasing the diffuse fraction of light; and this could be due to the model incorrectly treating the photosynthetic efficiencies for diffuse and direct radiation (Dai *et al.*, 2004). The majority of the days in the simulation do not have enough cloud cover to alter NEE, but the days that do have optically thick clouds which noticeably decrease the photosynthetic uptake are August 15 and 18. On the 15<sup>th</sup>, several areas of reduced photosynthesis can be seen in the northern and southern portions of the domain. These areas, which are contoured in green and have an NEE of only  $\sim -4 \mu\text{mol}/\text{m}^2/\text{s}$ , correspond to areas of very low incoming solar radiation, and the decrease in light is causing the plants to photosynthesize less. Similar circumstances exist on the 18<sup>th</sup>: the northeastern corner extending down the middle of the domain has less photosynthesis because the assimilation is restricted by the low incoming solar radiation. In the northwestern corner, the clouds on the 18<sup>th</sup> are blocking so much radiation that the assimilation is nearly zero and the area

is actually a small source of CO<sub>2</sub>, which can be seen because the color scale for this day extends up to 3  $\mu\text{mol}/\text{m}^2/\text{s}$ . A third day that shows a slight change in the CO<sub>2</sub> flux that corresponds to the cloud cover is August 16. Examining closely, stripes of alternating low and high NEE exist. The most evident stripes are in the western half of the domain, but stripes in the center and eastern portion exist with very slight changes in NEE. Although the stripes in NEE do not line up directly with the cloud cover, they do coincide with the stripes seen in the radiation, which are offset slightly from the cloud cover. Changes in the radiation appear to be causing small NEE changes. Interestingly, both the assimilation rate and the respiration rate are altered on the 16<sup>th</sup>. The assimilation is greater where the radiation is higher, but the respiration is also greater because the temperature is slightly higher. The two effects cancel each other, but the change in assimilation is slightly greater causing the weak striping pattern.

On the remaining days, the small changes in NEE are due primarily to temperature: as the temperature increases the respiration increases and the plants are less productive. For example, NEE is slightly higher on August 11 than on August 12 because the temperature is warmer causing more respiration, particularly in the western half of the domain. On this grid, plants are taking up carbon throughout the entire domain with small changes day-to-day primarily due to optically thick clouds and temperature.

For grid 3, daily NEE maps at 1 PM are displayed in Figure 4.20. Lake Superior and other smaller lakes are clearly evident because they do not have a CO<sub>2</sub> flux and are contoured in yellow. Orange and red contours indicate regions of CO<sub>2</sub> flux into the atmosphere and darker colors indicate uptake of carbon into the biosphere. Similar to grid 4, the grid 3 NEE patterns are quite similar for each day and resemble the biome map: deciduous broadleaf forests uptake the most CO<sub>2</sub>, followed by mixed forests, short vegetation, and finally agriculture, which has the least uptake and is commonly a source of CO<sub>2</sub> in this domain.



**Figure 4.20:** Daily NEE maps for grid 4, with units of  $\mu\text{mol}/\text{m}^2/\text{s}$ .

The changes in NEE in this grid are due to clouds and temperature, as well as soil moisture and soil classification. On August 11 in the southwestern corner there is a sharp gradient from an upward flux of  $\text{CO}_2$  to a downward flux. This gradient is actually caused by a combination of soil moisture and soil classification. The soil to the east of the gradient line is clay loam and the soil on the western side is loam that contains much less clay and more sand. Due to the soil properties, the loam will evaporate quicker causing the loam to dry out much sooner than the clay loam, which holds the moisture longer. Because the soil is dry in August, respiration increases with soil moisture and the corresponding changes in respiration cause the large gradient in NEE. No precipitation had occurred in the few days prior to the 12<sup>th</sup>, causing the gradient to be large because the eastern side still contained sufficient moisture to have significant respiration, while the western side of the gradient is much drier and thus does not respire as much. The night of the 11<sup>th</sup> some precipitation fell on this area causing the surface soil moisture to increase. Since both soils have more moisture, particularly the loam, the gradient is weaker. This gradient in the southwestern corner of the domain increases on the 13<sup>th</sup> and 14<sup>th</sup> due to the lack of precipitation and the drying out of the sandier loam. A significant precipitation event occurs the morning of the 15<sup>th</sup> providing sufficient soil moisture to increase the respiration in both

classes and to eliminate the gradient, which increases again the 16<sup>th</sup> and 17<sup>th</sup> as the soil moisture again evaporates. The final precipitation in the domain occurs the evening of the 17<sup>th</sup> and again weakens the gradient in the southwestern corner the following day. Due to a decrease in soil moisture that occurs more for the loam soil than the clay loam, the gradient again increases the final two days of the simulation.

Although different soil classes combined with soil moisture cause the gradient in the southwestern corner of grid 3, the majority of the remaining variability in NEE is due to the reduction of radiation by cloud cover and changes in temperature. Since the first three days are primarily clear, temperature causes the small variability seen between the days. The warmest day, August 11, has slightly less uptake than August 12, which is much cooler. This slight increase in productivity occurs because respiration increases exponentially with soil temperature, which is rather quickly affected by the air temperature in the top layers. On August 14 the pattern is similar to the previous three days despite significant cloud cover because the clouds are optically thin and do not decrease the radiation enough to effect the vegetation.

The days that clouds reduce the incoming radiation to less than  $\sim 300 \text{ W/m}^2$  and alter photosynthesis are August 15, 17, and 20. On the 15<sup>th</sup>, the majority of the domain receives less radiation because of the clouds. The northwestern corner receives the most radiation, which corresponds to the areas of greatest uptake on this day. The southwestern corner and up into the middle of the domain receives very little solar radiation, and as a result the assimilation is suppressed. The southwestern corner is a larger source of  $\text{CO}_2$  due to reduced assimilation combined with increased respiration due to increased soil moisture as discussed above. On August 17, the northwestern corner is covered by optically thick clouds that decrease the photosynthetic uptake considerably; and finally, on August 18 clouds on the eastern side of the domain decrease the photosynthetic uptake by  $\sim 10 \mu\text{mol/m}^2/\text{s}$ .

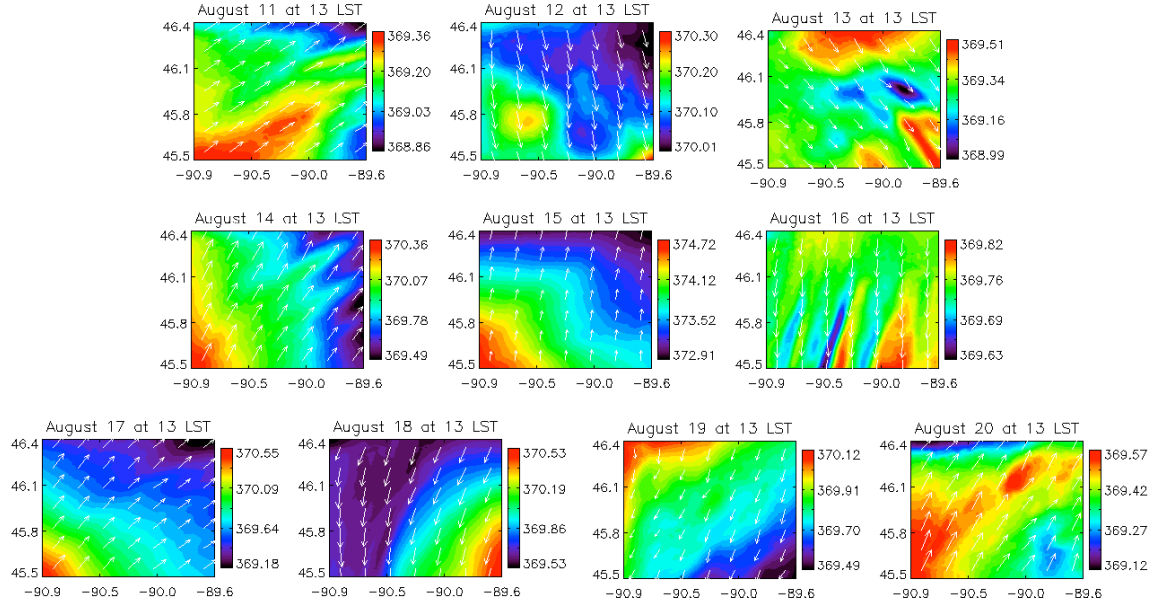
### 4.2.3 Total Column CO<sub>2</sub> Spatial Variability

Spatial maps of total column CO<sub>2</sub> concentrations in grid 4 at 1 PM for every day in case 2 are shown in Figure 4.21. Overlaid on the CO<sub>2</sub> concentrations are the wind vectors displaying the direction and speed of the near-surface wind. The maps display the spatial variability on a 97 by km domain that a single satellite track must represent. At 1 PM, the daily variability of total column CO<sub>2</sub> concentrations in the domain ranges from 0.2 to 1.8 ppm, with an average variability of 0.8 ppm on this domain. In general, the spatial patterns do not resemble the NEE maps shown in section 3.2.2.1. Since the spatial patterns of total column CO<sub>2</sub> concentration are not consistent with the carbon flux, the main driver to the hourly variability in this domain is advection. In addition to showing the 1 PM daily variability in total column CO<sub>2</sub>, Figure 4.21 also shows the considerable day-to-day variability that is seen in the total column concentrations. The CO<sub>2</sub> flux actually varies very little from day to day, indicating that synoptic weather systems and advection are also causing the changes in the magnitudes of the CO<sub>2</sub> concentrations.

The first day of the simulation, August 11, is a clear day with low CO<sub>2</sub> concentrations. The wind is southwesterly and the CO<sub>2</sub> concentration has a gradient from high concentrations in the southwest corner to low concentrations in the opposite northeast corner. Since the NEE map does not indicate that the southern corner is a source of CO<sub>2</sub>, the higher CO<sub>2</sub> concentrations are being advected from outside the domain. On August 12, the magnitude of the concentrations increases slightly due to the first front advecting high CO<sub>2</sub> from source regions to the south and the west. By 1 PM on the 12<sup>th</sup>, the front has already passed over and the northerly wind is advecting lower CO<sub>2</sub> concentrations from the north. Although clouds cover half of the domain, at this time the concentration gradient does not match the cloud contour, suggesting that on this scale advection is dominating over the biological effects that may be caused by clouds. Following the frontal passage, the CO<sub>2</sub> concentrations decrease as the wind lessens and returns to being southwesterly.

Just a few hours before the second front, 1 PM on August 15 has the highest mixing ratios as the pressure is dropping, the wind is increasing, and high concentrations are being advected from the southwest. After the second front, the concentrations again decrease dramatically, dropping over 4 ppm in total column concentrations in only one day. Looking at the pattern of CO<sub>2</sub> concentrations after the front, the southern half of the grid on the 16<sup>th</sup> contains alternating stripes of high and low CO<sub>2</sub> that strongly resemble cloud cover and NEE, although the variability is very small (< 0.2 ppm). On this day, the clouds are influencing the biological activity, decreasing photosynthesis and thus increasing the concentrations. Another possibility is that higher CO<sub>2</sub> from the lakes due to the lack of photosynthesis in the north-central and northeastern portions of the domain is being advected south.

Looking at the last four days in the simulation, the total column concentrations remain relatively low. The concentrations increase slightly again with the third front on August 17; however, the main increase in CO<sub>2</sub> is at night. On the 18<sup>th</sup>, the CO<sub>2</sub> decreases despite significant cloud cover. On this day, low CO<sub>2</sub> is being advected from the north with a relatively strong gradient in the southeastern corner of the domain. The NEE map shows that the cloud cover altered the rate of photosynthesis with regions of CO<sub>2</sub> flux to the atmosphere, which could also be contributing to the spatial pattern of CO<sub>2</sub>. On the 19<sup>th</sup>, although the higher CO<sub>2</sub> region corresponds to a cloudy region and the low CO<sub>2</sub> in the southeast corner corresponds to clear sky, NEE does not have the same pattern and suggests that the gradient is again caused by advection rather than the biology. Figure 4.21 reveals that at 1 PM the spatial variability is small on a 97 km domain with minimal biological influence, while the day-to-day variability is much greater due to synoptic weather systems and advection.



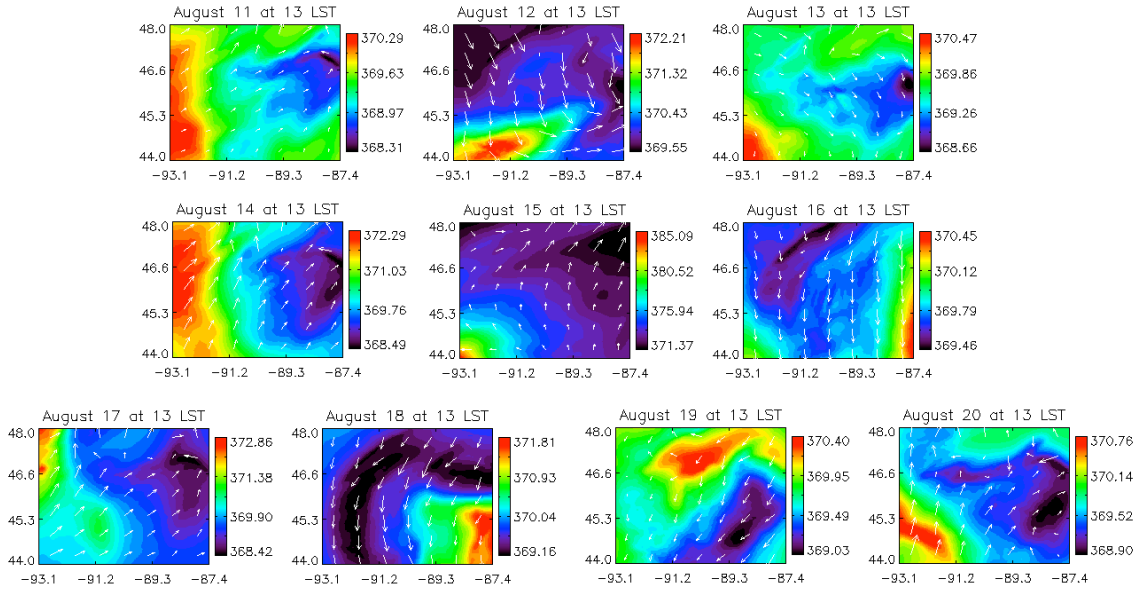
**Figure 4.21:** Total column CO<sub>2</sub> spatial variability at 1 PM for grid 4 overlaid by the wind vectors at 30 m. The x-axis is longitude, the y-axis is latitude, and the units are in ppm. The 1 PM total-column variability is shown for each day of the ten-day simulation, beginning with August 11 and ending with August 20, 2001.

Moving on to the larger domain, the spatial maps of total column CO<sub>2</sub> for grid 3 are displayed in Figure 4.22. The daily total column CO<sub>2</sub> variability at 1 PM in grid 3 is considerably larger than grid 4, ranging from 1 ppm to 13.7 ppm. These patterns in CO<sub>2</sub> are influenced both by advection and photosynthesis, as increases in the concentration due to the effects of cloud cover on the photosynthetic rate can be seen in the concentration maps. Looking at the day-to-day variability, the daily magnitudes of the concentrations are very similar to grid 3, indicating that advection and synoptic scale weather conditions cause the greatest day-to-day changes in CO<sub>2</sub>.

Starting at the beginning of the simulation, on August 11 the magnitude of the total column CO<sub>2</sub> concentrations is very similar to grid 3, with mixing ratios around 369 ppm. In the presence of a southwesterly wind, a CO<sub>2</sub> concentration gradient of a few ppm exists from high CO<sub>2</sub> on the western side of the grid to low concentrations on the eastern portion of the domain. The high CO<sub>2</sub> on the western third of the grid is most likely due to a combination of NEE and advection. Looking at the flux of CO<sub>2</sub>, the southwestern corner of the domain is actually a small source of CO<sub>2</sub>, primarily from the pixels that are classified as agriculture and grasslands. In



addition to the source of CO<sub>2</sub> in the domain, the winds could be advecting higher CO<sub>2</sub> concentrations from just outside the domain. On the eastern side of the domain, the lower CO<sub>2</sub> concentrations coincide with a region of strong uptake and lighter winds.



**Figure 4.22:** Total column CO<sub>2</sub> spatial variability at 1 PM for grid 3 with the wind vectors at 30 m displayed in white. The x-axis is longitude, the y-axis is latitude, and the units are in ppm. The 1 PM total column variability is shown for each day of the ten-day simulation beginning on August 11, 2001 and ending on August 20, 2001.

On August 12, the magnitude of the concentrations increases. The increase in the overall concentration is primarily due to the first front advecting higher CO<sub>2</sub>, since the photosynthetic uptake is actually slightly greater overall than on the previous day. However, the spatial pattern resembles the photosynthetic activity: the northwestern corner has low CO<sub>2</sub>, the southwest corner has high CO<sub>2</sub> due to an upward flux of CO<sub>2</sub>, and the area in the middle of the eastern edge has lower CO<sub>2</sub> corresponding to a region of uptake. Advection by winds is also contributing to this pattern, as moderately high CO<sub>2</sub> concentrations are being advected from Lake Superior and also from the southwestern corner.

After the first frontal passage, the magnitude of the concentrations again decreases on August 13 and begins to increase again on the 14<sup>th</sup> before the second front. Prior to the front the night of the 15<sup>th</sup>, the concentrations at 1 PM are considerably higher than the surrounding two

days. In addition, August 15 has the greatest daily total column  $\text{CO}_2$  variability as the concentration ranges over 10 ppm in the 450 km domain. The higher concentrations indicate that both biology and advection influence the  $\text{CO}_2$  concentrations. The  $\text{CO}_2$  spatial pattern resembles the NEE map, with the highest concentrations in a source region. In addition to naturally being a source of  $\text{CO}_2$  at this time of year on clear days, the southwestern corner of the domain is covered by clouds that severely reduce the radiation and cause an even greater upward flux of  $\text{CO}_2$ . However, the source of  $\text{CO}_2$  from this area is probably not enough alone to cause the  $\text{CO}_2$  increase, as the majority of the domain except the lakes is still a sink of  $\text{CO}_2$ ; and it is likely that high  $\text{CO}_2$  air is being advected from outside the domain by the southwesterly winds.

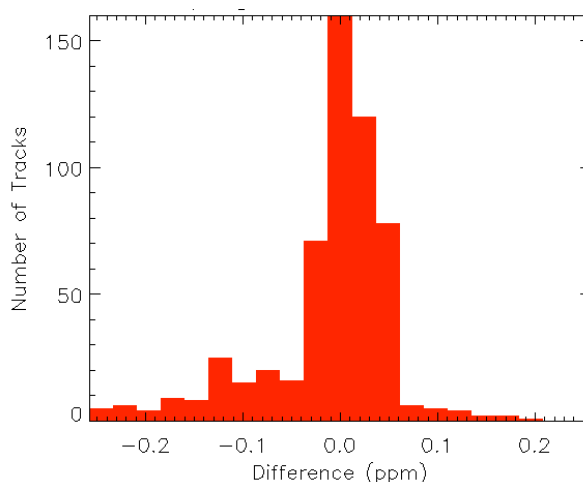
Following the front, the concentrations throughout this larger domain drop again as they did in grid 4. On August 17 before the third front, the concentrations are still low, with the highest concentrations corresponding to an upward flux of  $\text{CO}_2$  in the northwestern corner due to the cloud cover and reduced radiation. The next day the overall magnitude of the concentrations is similar to the 17<sup>th</sup>, but the winds are southeasterly and the mixing ratio is low in the northern portion of the domain and high in the southwestern corner. The low concentrations over the lake, where there is no photosynthesis, are probably due to the advection of low  $\text{CO}_2$  from Canada in combination with the strong uptake just south of Lake Superior. The region of high  $\text{CO}_2$  in the southeast corner is due to the flux of  $\text{CO}_2$  from reduced photosynthesis due to the optically thick clouds.

Finally, the last two days of the simulation have relatively low  $\text{CO}_2$  concentrations with stable meteorological conditions, as the pressure is increasing, the wind speeds are low, and the skies are relatively clear. On August 20 the wind circulations are primarily local breezes, and the  $\text{CO}_2$  pattern strongly resembles the flux of  $\text{CO}_2$ . Over Lake Superior the concentrations are slightly higher due to no flux of  $\text{CO}_2$ , and small sea breezes are advecting the higher concentrations over the lands. Beginning in the northwestern corner and extending south and east to the opposite corner, the concentrations are low above photosynthetically active areas. Last, the

bottom left corner of the domain has high concentrations coinciding with a slight source of CO<sub>2</sub> in this area that is being advected north. Analyzing the total column concentrations on grid 3 reveals that over a larger area the CO<sub>2</sub> concentrations are considerably more variable and that both biology and advection contribute to the spatial pattern of CO<sub>2</sub> on this scale.

#### 4.2.4 Spatial Representativeness Errors

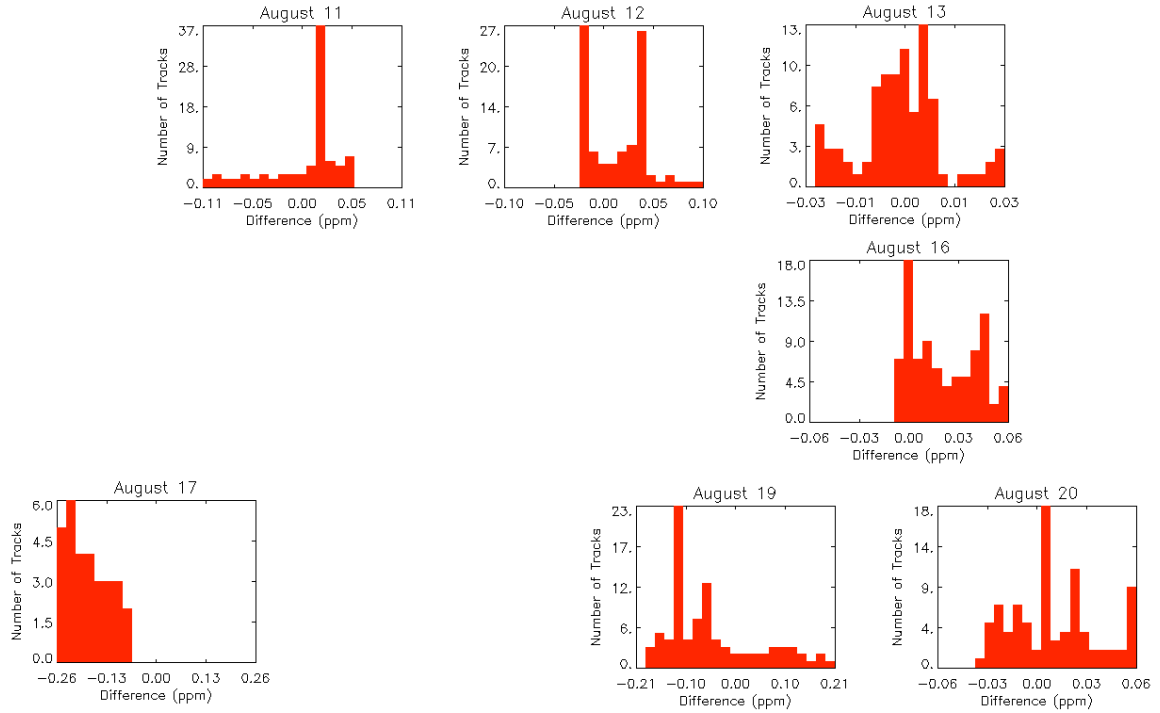
In this section we will investigate the spatial representativeness errors introduced by using 10-km wide satellite tracks to represent a 97 by 97 km domain and a 450 by 450 km domain. To calculate the spatial representativeness errors, we subtracted the domain-averaged concentration at 1 PM from each simulated satellite concentration using only clear-sky pixels.



**Figure 4.23:** Grid 4 sampling distribution of the spatial representativeness errors at 1 PM compiled from all ten days of the simulation. The x-axis is the difference between the emulated satellite concentration and the domain mean concentration and the y-axis is the number of satellite tracks that correspond to each difference. Negative values indicate an underestimation by the emulated measurements and positive values indicate an overestimation.

Figure 4.23 displays the compiled representativeness errors for grid 4 at 1 PM from all ten days of the simulation. All of the tracks capture the spatial variability in grid 4 quite well. The majority of the tracks actually capture the domain average within 0.05 ppm. All of the emulated satellite values are within 0.26 ppm of the domain mean and 95% of the satellite tracks have a spatial error of less than 0.2 ppm. The small magnitude of the error is not too surprising,

since the variability at 1 PM was on average only 0.8 ppm. The distribution of this error is relatively symmetrical and centered on zero, indicating that the small spatial errors will be random and are not biased.

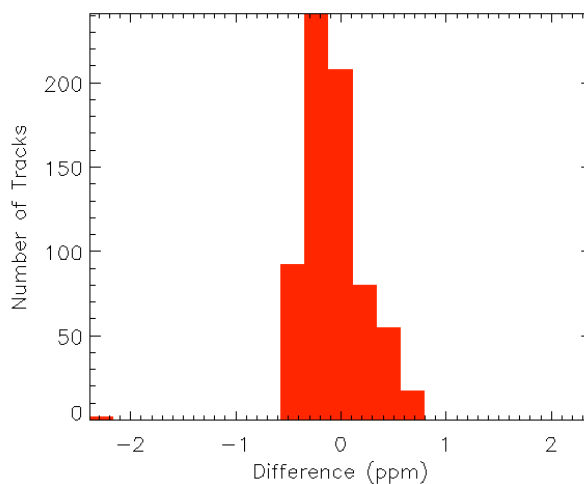


**Figure 4.24:** Grid 4 daily sampling distributions of the spatial representativeness errors at 1 PM. The days with no corresponding distribution were completely covered with clouds.

Breaking down the representativeness errors into the contributions from each day, Figure 4.24 displays the daily sampling distributions. Each day has a different distribution and a different range of spatial errors. As seen in the compiled distribution, most of the days capture the domain average within 0.06 ppm. The larger errors in the distribution are due to August 17 and 19. On the 17<sup>th</sup>, high CO<sub>2</sub> was being advected into the southwestern corner of the domain prior to the third front, and the overall variability was quite large, with concentration differences of 1.4 ppm across the domain. The large spatial gradient and the cloud cover pattern caused the emulated satellite concentrations to underestimate the domain averaged concentration, and the largest errors at 1 PM for the entire simulation are the satellite tracks on the 17<sup>th</sup> that only see a few clear pixels on the edge of the cloud cover near the north-central part of the domain. On

August 19, the errors are relatively large, but in this case there are tracks that both overestimate and underestimate the mean 1 PM concentration. The tracks that overestimate the concentration are the tracks on the western side of the domain that see the high CO<sub>2</sub> in the northwest corner, and the tracks that underestimate the CO<sub>2</sub> are the tracks that see the lower concentrations in the southern portion of the domain. More tracks underestimate the concentration at this time because more of the clear area is in the portion of the domain with the lowest concentrations.

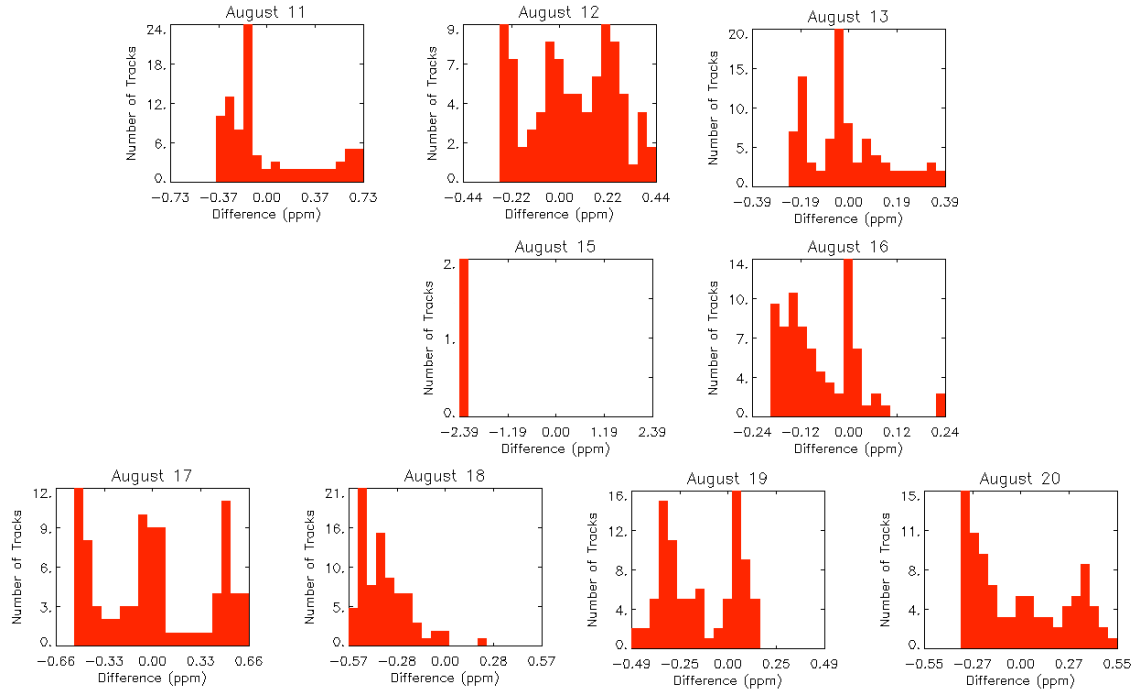
The sampling distribution of the compiled representativeness errors at 1 PM for grid 3 is shown in Figure 4.25. The range of errors on this grid is considerably larger than for grid 3, primarily because the daily CO<sub>2</sub> variability over the domain was larger, averaging 3.5 ppm. The largest number of tracks underestimate the mean concentration by ~0.25 ppm. Over 99% of the tracks have a spatial error less than 0.8 ppm, but two tracks underestimate the domain-average concentration by more than 2 ppm. Similar to grid 3, the distribution of the errors is relatively symmetrical, again demonstrating that the spatial representativeness errors will be random.



**Figure 4.25:** Grid 3 sampling distributions of spatial representativeness errors at 1 PM compiled from all ten days of the simulation.

To examine the errors more closely and to determine what caused the large underestimation, Figure 4.26 breaks the compiled distribution into sampling distributions from each day. The days sampled before a front (August 11, 15, and 17) had larger spatial errors due

primarily to the higher CO<sub>2</sub> variability on those days. In contrast, days after the fronts (August 12, 16, and 19) all had spatial representation errors less than 0.5 ppm. On the 11<sup>th</sup>, the tracks on the western edge of the domain, in the high CO<sub>2</sub> region, overestimated the concentration by ~0.7 ppm. The largest errors came from the two tracks that had clear pixels on August 15, where the emulated satellite concentrations underestimated the mean concentration by more than 2 ppm. These two concentrations are from a few pixels of clear sky in the far northwestern corner. The large errors occur because the northwestern corner had the lowest concentrations in a gradient of more than 2 ppm in total column CO<sub>2</sub>. On August 17, the tracks on the western edge of grid 3 overestimated the concentrations by ~0.6 ppm and the tracks on the eastern edge underestimated the domain-averaged total column mixing ratio by ~0.5 ppm. Looking at the 18<sup>th</sup>, since the high concentrations are only in the southeastern corner that is cloud-covered and not along the entire edge, nearly all the satellite tracks to underestimate the mean because the satellite tracks cannot see the higher concentrations under the clouds present throughout much of the domain. Analyzing grid 3 illustrates that the magnitude of the errors increases with an increase in the size of the region the satellite measurements are representing. On this larger domain, over 99% of the emulated satellite concentrations had spatial representativeness errors of 0.8 ppm or less. This case also demonstrates that large errors of over 2 ppm are possible when looking through small holes in the clouds and using the measurements to represent a large area, but the number of tracks with an error more than 1 ppm is less than 1%.



**Figure 4.26:** Daily sampling distributions of spatial sampling errors at 1 PM for grid 3.

# Chapter 5: Temporal Representativeness of Satellite CO<sub>2</sub> Measurements

OCO will only measure total column CO<sub>2</sub> mixing ratios at 1:15 PM. Diurnal and temporal sampling errors can occur if an inverse model optimizes diurnal average CO<sub>2</sub> concentrations rather than mid-afternoon concentrations, and these errors may result in incorrect flux estimates. By comparing simulated satellite concentrations at 1 PM to the domain-averaged diurnal mean CO<sub>2</sub> concentration, we can analyze the temporal sampling errors introduced into an inverse model that uses diurnally averaged concentrations. This chapter investigates the diurnal variability in SiB2-RAMS and also discusses the diurnal sampling errors in case 1 and case 2.

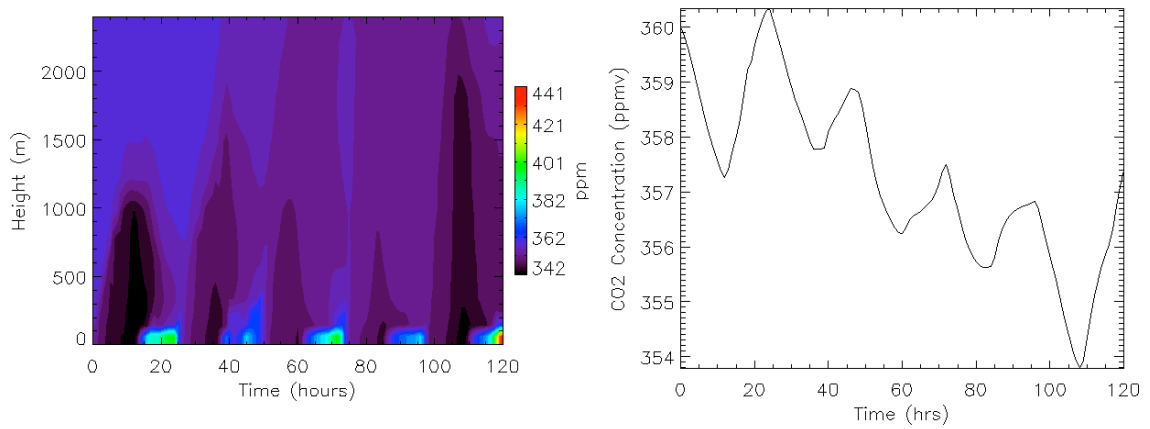
## 5.1 Case 1

### 5.1.1 Diurnal Cycle

Figure 5.1 displays the diurnal cycle for grid 3. The left panel shows vertical profiles of the domain-averaged CO<sub>2</sub> concentration for all five days from case 1, illustrating the key points of the diurnal cycle discussed in chapter 1. At the beginning of the run, the figure shows the growth of the boundary layer to ~1 km, which contains well-mixed low CO<sub>2</sub> concentrations due to photosynthesis. At night the stable and shallow boundary layer is clearly visible with high CO<sub>2</sub> concentrations due to respiration. The surface concentrations are decoupled from the residual boundary layer, which continues to have lower mixing ratios. In the morning the figure shows the buildup of another deep boundary layer as the high CO<sub>2</sub> is convected upwards and mixed with the residual layer. The amplitude of the variability at the surface is ~60 ppm on average; however, the strength and depth of the stable nocturnal boundary layer varies from night to night.



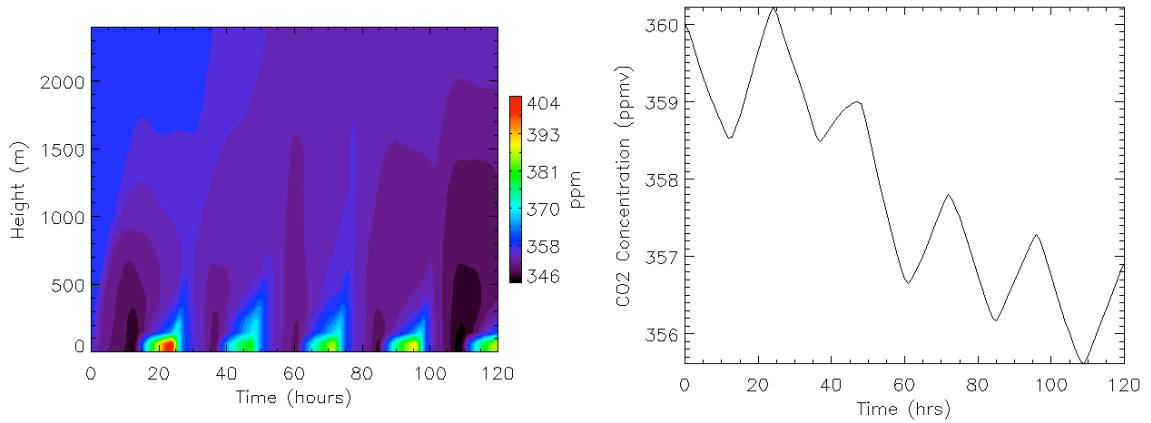
The night of July 27, beginning hour 40, the boundary layer is much less stable and the buildup of high CO<sub>2</sub> is less. That night strengthened winds brought cooler air from the north and mixed the nocturnal concentrations reducing the surface buildup. The following day the CO<sub>2</sub> concentrations remained slightly higher. The nocturnal boundary layer on the third night, July 28, is once again more stable with high CO<sub>2</sub> in the lowest 200 m of the atmosphere. The most daytime vertical mixing occurs on July 30, which had higher temperatures and weaker winds helping the boundary layer to grow to almost 2 km. That night the stable boundary layer has the highest CO<sub>2</sub> concentrations in this domain.



**Figure 5.1:** The diurnal cycle of CO<sub>2</sub> for grid 3. The left panel shows vertical profiles of the domain-averaged concentrations for the entire simulation, beginning 6 AM July 26 at 0 hours and ending 6 AM July 31 at 120 hours. The right panel shows the diurnal cycle of the domain-averaged total column CO<sub>2</sub>.

The right panel of Figure 5.1 shows the domain averaged total column CO<sub>2</sub> diurnal cycle. A downward trend of decreasing concentrations is clearly evident, indicating that for these five days in July the biosphere sink is larger than the sources from respiration and advection. These five days were clear with only a minor cold front, so the uptake by the plants and the minimal advection is not surprising. The figure shows that the amplitude of the diurnal cycle ranges from ~0.8 ppm to nearly 3 ppm peak to peak. The highest concentrations in the total column are around 6 AM due to the buildup of high CO<sub>2</sub> from respiration, and the lowest mixing ratios occur ~6 PM, when the solar zenith angle becomes large.

The diurnal cycle for grid 1 is shown in Figure 5.2. Similar to the small domain, grid 1 has a deep mixed layer of low  $\text{CO}_2$  during the day and high concentrations in a shallow layer at night. The amplitude of the diurnal cycle at the surface is slightly less than on grid 3, ranging from concentrations of  $\sim 350$  ppm to  $\sim 390$  ppm. The growth of the boundary layer into the residual layer and the corresponding mixing and diluting of the high near-surface  $\text{CO}_2$  is more obvious on grid 1, as the higher  $\text{CO}_2$  concentrations contoured in green and blue are clearly visible; however, the day-to-day variability in the depth of the nocturnal boundary layer is less, as the depth of the boundary layer is nearly identical every day in the simulation.



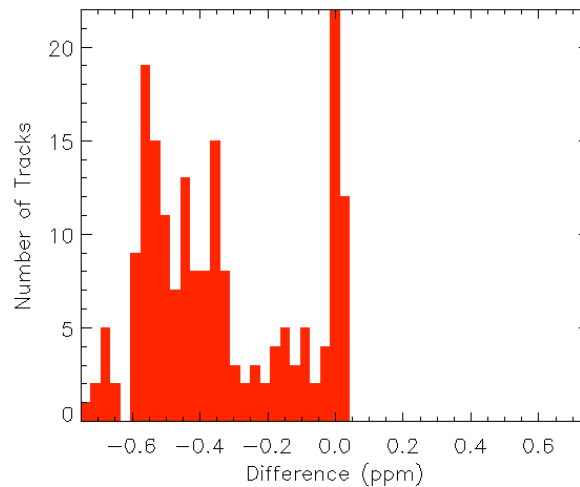
**Figure 5.2:** The domain-averaged diurnal cycle of  $\text{CO}_2$  for grid 1. The left panel shows vertical cross-sections for the entire simulation, and the right panel shows the cycle of total column concentrations.

Looking at the right-hand panel of Figure 5.2, the diurnal variability of total-column  $\text{CO}_2$  on grid 1 is quite similar to grid 3: the most notable feature is the decreasing concentrations during the simulation due to the strong photosynthetic uptake and the timing of the maximum and minimum  $\text{CO}_2$  concentrations occur around 6 PM and 6 AM, respectively. The amplitude is slightly less on grid 1, with differences of  $\sim 0.5$  ppm to  $\sim 2$  ppm; however, both grid 3 and grid 1 have considerable diurnal variability in total column concentrations.

### 5.1.2 Diurnal Representativeness at 1 PM

To calculate the diurnal errors from sampling only at 1 PM, we subtracted the domain-averaged diurnal mean from 6 AM LST to 6 AM the following morning from each emulated 1

PM satellite concentration for the corresponding day; and we compiled the daily results into a single sampling distribution. Figure 5.3 shows the sampling distribution for grid 3. Rather than being evenly distributed, the errors are almost always negative, indicating that the concentrations at 1 PM on all five days are less than the diurnal average. The errors in the emulated satellite concentrations range from slightly overestimating the concentration by  $<0.05$  ppm to underestimating the total column concentration by as much as 0.75 ppm. July 27 had the smallest errors compared to the diurnal mean, when a few tracks slightly overestimated the mean concentration. The large negative errors occurred on July 26 and 29. Looking at the right panel of Figure 5.1, the nighttime buildup of  $\text{CO}_2$  on the 27<sup>th</sup> was less because of the small cold front. In addition, the concentrations at 1 PM on the 27<sup>th</sup> were higher than on the other days as the minimum concentration occurred slightly later. The fact that the emulated satellite tracks underestimate the time-averaged total column concentration suggests that using satellite concentrations at 1 PM will introduce a negative bias compared to the diurnal average, and the mean bias over the five day simulation on grid 3 is  $-0.34$  ppm.

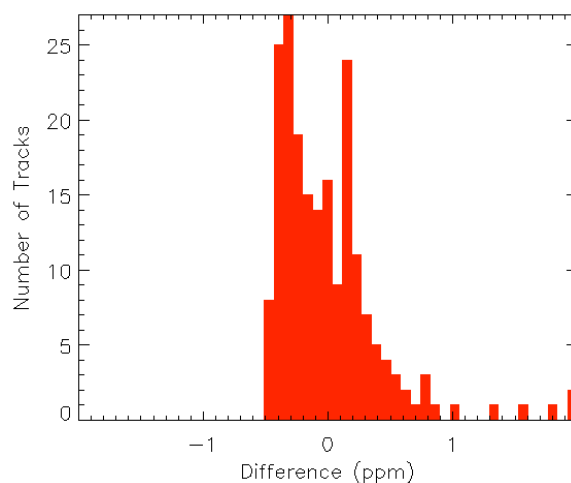


**Figure 5.3:** Sampling distribution of the 1 PM errors in the emulated satellite concentrations compared to the diurnal mean for grid 3. This distribution is a compilation of the errors from all five days in the simulation.

In addition to comparing the emulated satellite tracks from each day with the corresponding diurnal average, we compared the domain-averaged concentration over the entire

simulation to the domain-averaged concentration at 1 PM. The domain average time mean is 357.24 ppm, and the domain average at 1 PM is 357.9 ppm. This calculation also suggests that for a small relatively homogeneous domain the concentrations at 1 PM will be negatively biased.

The grid 1 diurnal error sampling distribution at 1 PM is shown in Figure 5.4. Unlike grid 3, the distribution for this larger domain is centered near zero. Nearly 2/3 of the tracks underestimate the concentration by as much as 0.5 ppm, and the remaining 1/3 overestimate the concentration, with a few tracks overestimating the diurnal average by more than 1 ppm. These tracks that significantly overestimate the diurnal mean are the tracks on the eastern edge of the domain on July 29 that only see the lakes.

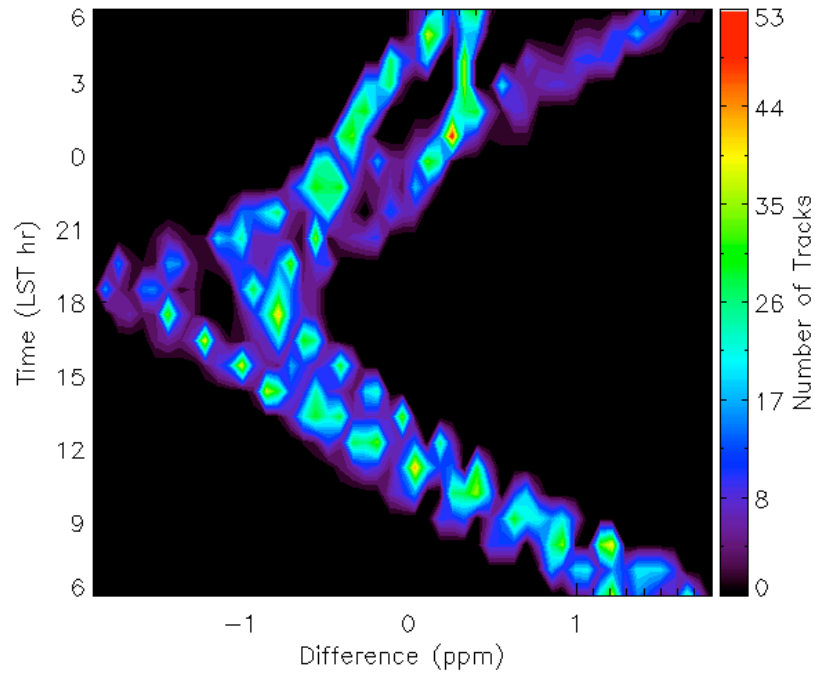


**Figure 5.4:** Sampling distribution of diurnal errors at 1 PM for grid 3.

The relative symmetry of the sampling distribution suggests that the diurnal errors on a larger domain are random due to the inclusion of various land-cover types. In addition, the mean over the entire simulation for grid 1 is 357.815 ppm and the mean at 1 PM is 357.820 ppm. Since the temporal mean over 24 hours is nearly identical to the temporal mean at 1 PM, the diurnal errors are likely to be small and unbiased. In this case, the inclusion of portions of Lake Superior and Lake Michigan makes the error random since these large bodies of water do not have the same diurnal cycle as land and have higher concentrations during the day. The complexity of the land cover in this domain illustrates that the diurnal cycle error will depend on the region the

satellite concentrations are representing. The diurnal errors over regions that are homogeneously covered by vegetation with strong photosynthetic uptake will likely have a large negative diurnal error and introduce a bias, as seen in grid 3. In contrast, regions that are not homogeneously covered by land could have large diurnal errors for any single track but overall the errors may be random and unbiased, although this result may vary for different domains.

### 5.1.3 Hourly Diurnal Representativeness



**Figure 5.5:** 24-hour sampling distribution of the diurnal errors from all five days for grid 3. The x-axis is the difference between the emulated satellite concentrations and the diurnal domain-averaged mean, and the y-axis is the hour, with 6 AM on the bottom and 6 AM the following morning on the top of the figure. The number of tracks corresponding to each error bin is contoured.

Since the emulated satellite concentrations on grid 3 underestimated the diurnal average, the next question we can investigate using SiB2-RAMS is what hour is ideal to capture the diurnal average. To answer this, we analyzed a contour plot displaying the 5-day compiled sampling distributions for every hour. This plot for grid 3 is shown in Figure 5.5. Similar to the sampling distribution at 1 PM, we calculated the errors by subtracting the domain-averaged 6 AM to 6 AM LST diurnal mean from the emulated satellite concentrations at every hour. These

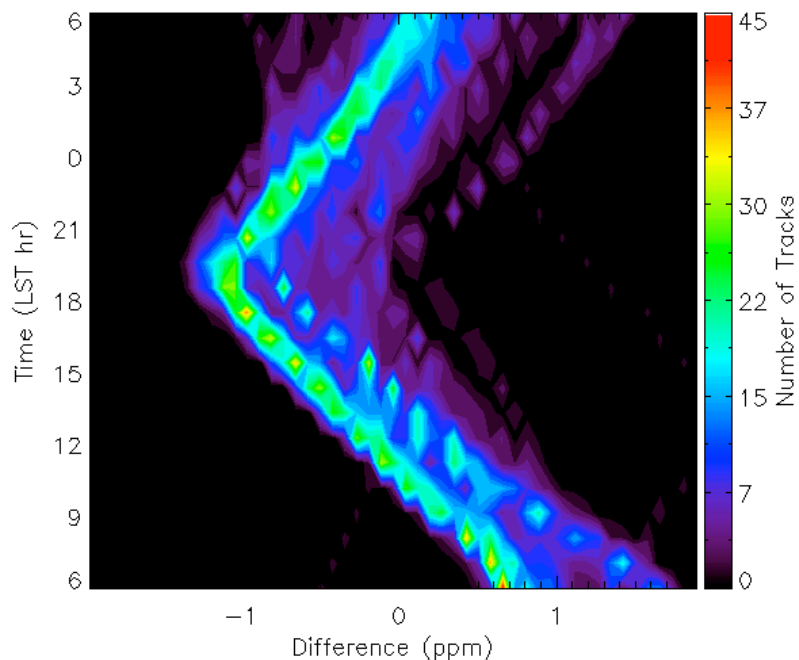
hourly distributions have been compiled into a single contour plot that shows the errors on the x-axis and the corresponding number of tracks that have errors in each error bin contoured in colors, with black representing zero tracks and red representing the most tracks, which in this case is 53.

The diurnal cycle of total column CO<sub>2</sub> is clearly evident, as photosynthesis causes an underestimation of the diurnal average while respiration at night causes an overestimation. The sample distributions are centered near zero once in the morning and once at night. For this domain, the best time to capture the diurnal mean is 11 AM, when all the concentrations are within 0.55 ppm of the diurnal mean. The time is earlier than 1 PM for this domain because grid 3 is predominantly covered with forest that actively photosynthesizes during the summer and lowers the concentration over the entire domain earlier in the morning.

Although OCO requires sunlight, we can also investigate the time at night when the emulated concentrations are closest to the diurnal mean. At night the sampling distributions are bimodal, having a large number of tracks that both underestimate and overestimate the diurnal mean. The bimodal distribution is due to the different behavior of the stable boundary layer for each night. July 26, 28, and 30 all had stable nocturnal boundary layers that had high near-surface CO<sub>2</sub> concentrations increasing the nighttime total column concentrations early in the evening (see Figure 5.1). On these days the best time to sample the concentrations to represent the diurnal mean is ~11 PM. On July 27 and 29, the nocturnal boundary layer did not develop as strongly and the high near-surface CO<sub>2</sub> was diluted and advected throughout the domain, causing the concentrations to be closest to the diurnal mean later at night ~3 AM. The different nighttime behavior caused the bimodal distribution of errors seen in Figure 5.5, with the time at night that comes closest to the diurnal concentration depending on the night that is sampled.

The grid 1 sampling distributions for all 24 hours are shown in Figure 5.6. Similar to grid 3, the diurnal cycle is evident; however, the range of errors from each hour is much greater than for grid 3. This larger spread of errors is due to the heterogeneous surface and the large

range of total column concentrations. The best time to sample the concentrations to capture the diurnal mean on this larger domain is 1 PM, when we saw in the previous section that the maximum error for 95% of the tracks was 0.75 ppm. As seen in grid 3, the nighttime tracks have a greater range of errors, and the best time to capture the diurnal mean at night is 4 AM.



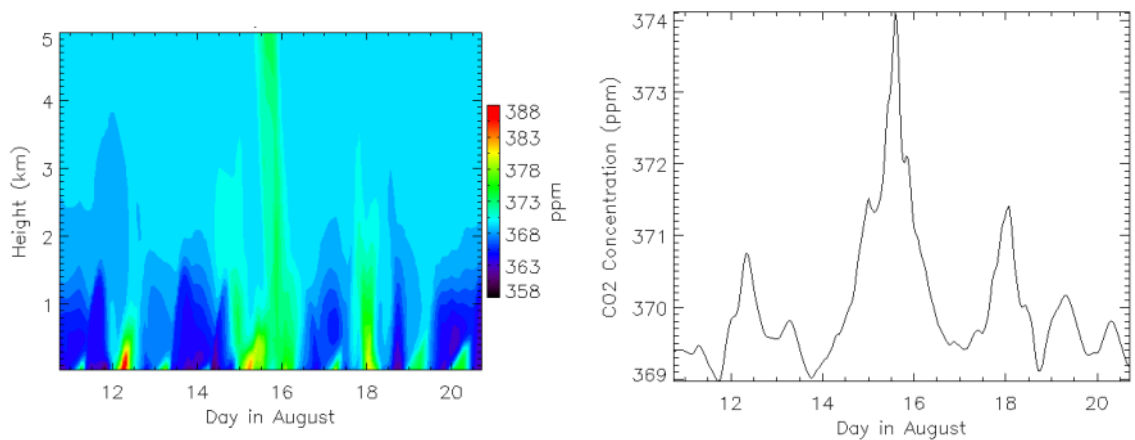
**Figure 5.6:** Compiled 24-hour sampling distribution of the diurnal errors for grid 1.

Investigating the sampling distributions for every hour in this case reveals that the best time to measure the total column concentrations in order to have the concentrations closely match the diurnal mean depends on the size and heterogeneity of the domain. If the domain is covered predominantly by productive vegetation, then the best time to measure the concentrations is late morning; however, this simulation indicates that if the region is more heterogeneous and contains a combination of land and water, then the best time to measure  $\text{CO}_2$  is slightly later in the afternoon. Since OCO will provide global coverage, this case suggests that in general the concentrations over much of the Earth will have random diurnal area; however, regions that have large homogeneous land cover that is actively photosynthesizing may have a negative bias.

## 5.2 Case 2

### 5.2.1 Diurnal Cycle

In case 2, Figure 5.7 shows the diurnal cycle of  $\text{CO}_2$  for grid 4. Similar to the figures from case 1, the left panel shows vertical profiles of the domain averaged  $\text{CO}_2$  concentrations through the 10-day simulation and the right panel shows the domain-averaged total column concentrations. Once again, the figure displays the daytime well-mixed and low concentrations in the boundary layer up to  $\sim 2$  km and the high concentrations at night in the shallow nocturnal boundary layer from respiration, which are decoupled from the residual boundary layer from the previous day; however, the depth of the boundary layer in this case is much more variable from day to day. A stable boundary layer does not develop on nights when a front passes through. Instead, high concentrations of  $\text{CO}_2$  exist throughout the atmospheric column. The high  $\text{CO}_2$  is advected from southwest of the domain, and due to both advection and vertical transport associated with the front, high  $\text{CO}_2$  extends above the residual boundary layer, with the highest concentrations at the surface due to both advection and respiration. The days following a frontal passage generally develop a slightly shallower boundary layer that grows over time until the next front.

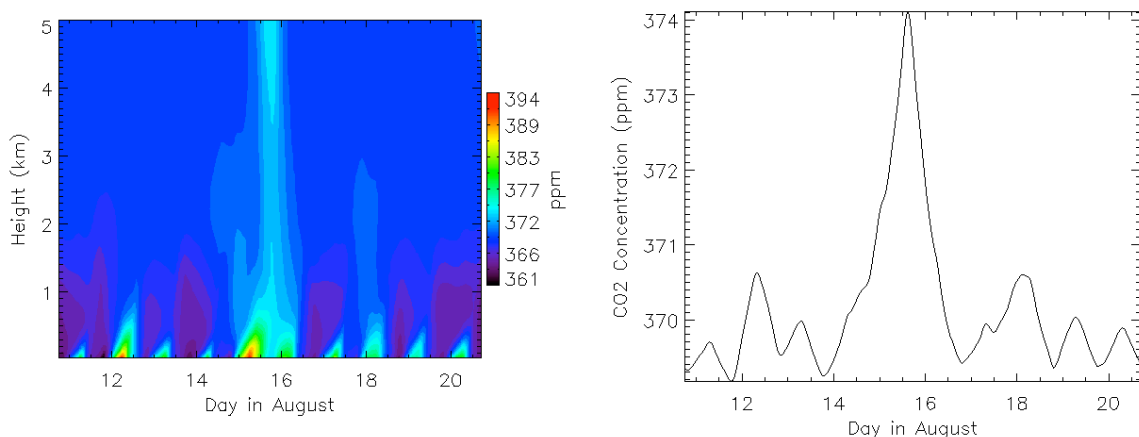


**Figure 5.7:** Domain average  $\text{CO}_2$  for grid 4. The left panel shows vertical profiles of the domain-averaged concentrations for the entire simulation, beginning 6 PM August 10 and ending 6 PM August 20. The right panel is the domain-averaged total column  $\text{CO}_2$ .



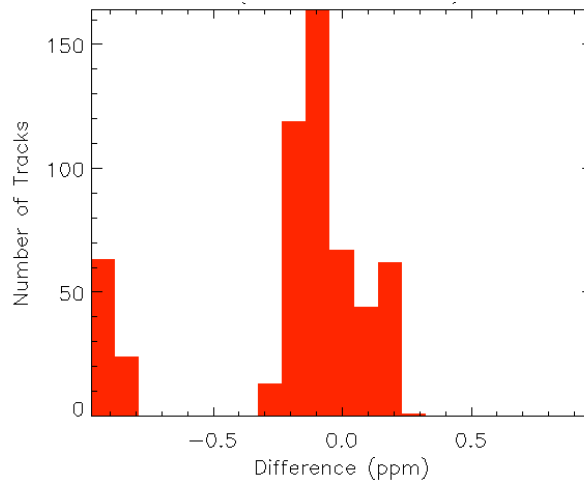
The right panel of Figure 5.7 shows that much of the variability in total column  $\text{CO}_2$  is synoptic variability rather than diurnal variability. The amplitude of the diurnal cycle in the total column concentrations is  $\sim 0.6$  ppm; however, synoptic weather conditions alter the total column concentrations by up to 5 ppm in a single day. Since the diurnal variability in this simulation is small, differences between the simulated satellite concentrations and the domain averaged diurnal mean due to the biology are expected to be small. In contrast, differences from the diurnal mean due to the timing of the fronts may cause large errors in the emulated satellite concentrations. This simulation is not representative of fronts at WLEF, as all three frontal passages occur at night; and the errors due to the frontal passages should be random over a representative period.

Figure 5.8 shows the diurnal cycle for grid 3. This figure is remarkably similar to grid 4: the nights with fronts passing through have high concentrations throughout the vertical profile, the main driver of the variability is advection and synoptic weather rather than the biological diurnal cycle, and the amplitude of the biological diurnal cycle for grid 3 is  $\sim 0.5$  ppm. Similar to case 1, the daytime boundary layer is not as deep in the larger domain.



**Figure 5.8:** The variation of simulated  $\text{CO}_2$  for grid 3. The left panel shows the domain averaged time-height cross-section, and the right panel shows the time varying domain average column mean mixing ratio.

### 5.2.2 Diurnal Representativeness at 1 PM

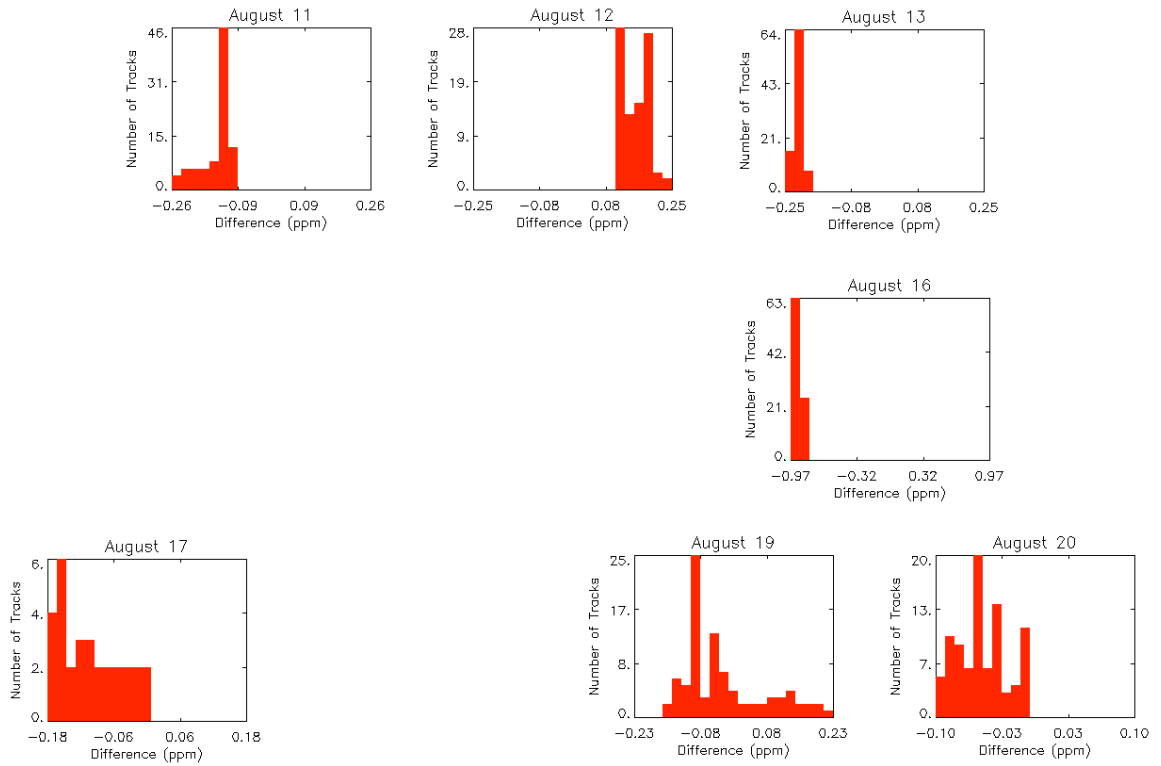


**Figure 5.9:** Grid 4 sampling distribution of the 1 PM errors in the emulated satellite concentrations compared to the diurnal mean from 6-6 PM LST. This distribution is a compilation of the errors from all ten days in the simulation.

To calculate the representativeness of emulated satellite measurements at 1 PM for case 2, we used the same procedure as in case 1: we subtracted the domain-averaged total column diurnal mean from 6 PM to 6 PM LST from the emulated clear-sky satellite concentrations at 1 PM and then compiled the results from all ten days into a single sampling distribution. The sampling distribution for grid 4 is shown in Figure 5.9. The error bin with the most simulated satellite tracks is negatively shifted from zero and is centered on  $\sim -0.1$  ppm. For this grid,  $\sim 1/3$  of the tracks overestimated or equaled the diurnal mean and  $\sim 2/3$  underestimated the mean. Nearly 20% of the tracks underestimated the diurnal mean by  $\sim 0.9$  ppm, which is a staggering error in the total column.

To determine what caused the large underestimation, Figure 5.10 displays daily sampling distributions of the differences between the emulated satellite tracks and the diurnal mean. The sign and magnitude of the differences depend primarily on the timing of the fronts. On August 11, the errors are all slightly negative with a maximum error of only  $-0.25$  ppm. Looking back at the column mean mixing ratio shown in Figure 5.7, the first day did not have any increases due to synoptic variability and slightly underestimated the concentrations because of lower

concentrations due to photosynthesis. The errors on the 12<sup>th</sup> are greater than zero because the diurnal mean included the second front and the concentrations at 1 PM are still increased slightly due to the advection of high CO<sub>2</sub>. The 13<sup>th</sup> is similar to the 11<sup>th</sup>, as the day is not affected by a front and the emulated satellite concentrations underestimate the diurnal mean by ~0.2 ppm. The large underestimations of the diurnal mean occur on August 16<sup>th</sup>, when all of the tracks underestimate the mean by at least 0.8 ppm. Since the second front passed over ~11 PM on the 15<sup>th</sup> and had high concentrations, the increased CO<sub>2</sub> is included in the diurnal mean for the 16<sup>th</sup>. Although the concentrations dropped rapidly early morning on the 16<sup>th</sup>, the inclusion of the high concentrations the night of the 15<sup>th</sup> raised the diurnal mean and made the concentrations at 1 PM appear to badly underestimate the diurnal mean.



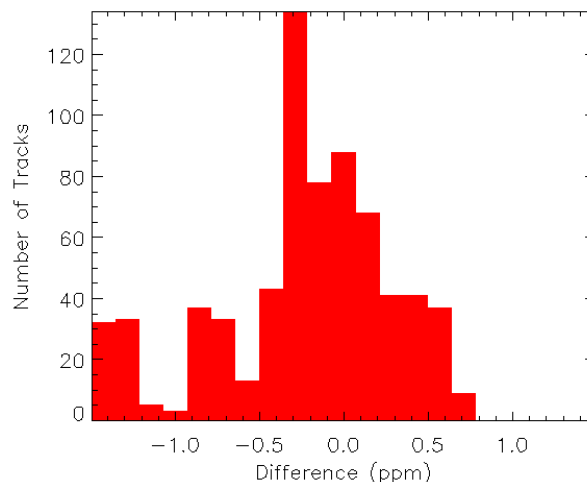
**Figure 5.10:** Daily sampling distributions of the difference between the emulated satellite tracks at 1 PM and the corresponding total column domain-averaged diurnal mean from 6-6 PM LST. The missing days are completely cloudy and did not have any possible satellite tracks.

After the second front, the satellite tracks have relatively small errors compared to the diurnal mean. On the 17<sup>th</sup> the tracks again underestimate the concentrations from the biological drawdown during the day. The 18<sup>th</sup>, which includes the third front in the diurnal mean, is cloudy and hence has no possible satellite tracks. If the day had not been cloudy, the 1 PM observations would have underestimated the mean by a substantial amount since the front advected high CO<sub>2</sub>. Tracks on the 19<sup>th</sup> both underestimate and overestimate the diurnal mean and have relatively small errors. This distribution includes zero because the low concentrations from the previous day decrease the diurnal mean to make it approximately the same as the concentrations at 1 PM. Finally, the 20<sup>th</sup> again slightly underestimates the diurnal mean.

The main determination of the temporal representation errors for grid 4 is the timing of the fronts. If the definition of the day is changed to begin at a different time, the distribution of errors would change for the days surrounding the fronts. This suggests that this case does not capture enough frontal samples and the temporal errors appear to be negatively biased because of the timing of the specific fronts in this case. Since fronts can pass over the area at any time, the temporal errors due to the synoptic events are expected to be random and not introduce a bias. Although the satellite measurements may not be biased compared to the diurnal average, since the main driver of the variability is synoptic events, large errors will remain when comparing satellite measurements to diurnal averaged concentrations. To avoid incorporating these large errors into inverse studies, inverse models must accurately model the synoptic-scale transport and not compare diurnal averaged concentrations to satellite measurements taken at one point in time.

The sampling distribution of the differences between the 1 PM emulated satellite concentrations and the domain-averaged diurnal mean for grid 3 is displayed in Figure 5.11. The shape of the distribution is similar to grid 4, as the largest number of tracks slightly underestimate the diurnal mean by ~0.3 ppm, ~1/3 of the tracks overestimate the mean, and ~1/5 of the tracks grossly underestimate the diurnal mean; however, the magnitude of the errors is greater for grid 3. The tracks that overestimate the concentration have errors up to 0.65 ppm and the emulated

satellite concentrations that underestimate the diurnal mean have errors as large as 1.5 ppm. As we saw in the spatial representativeness errors, the larger domain has greater errors due to the enhanced total column variability in CO<sub>2</sub>.

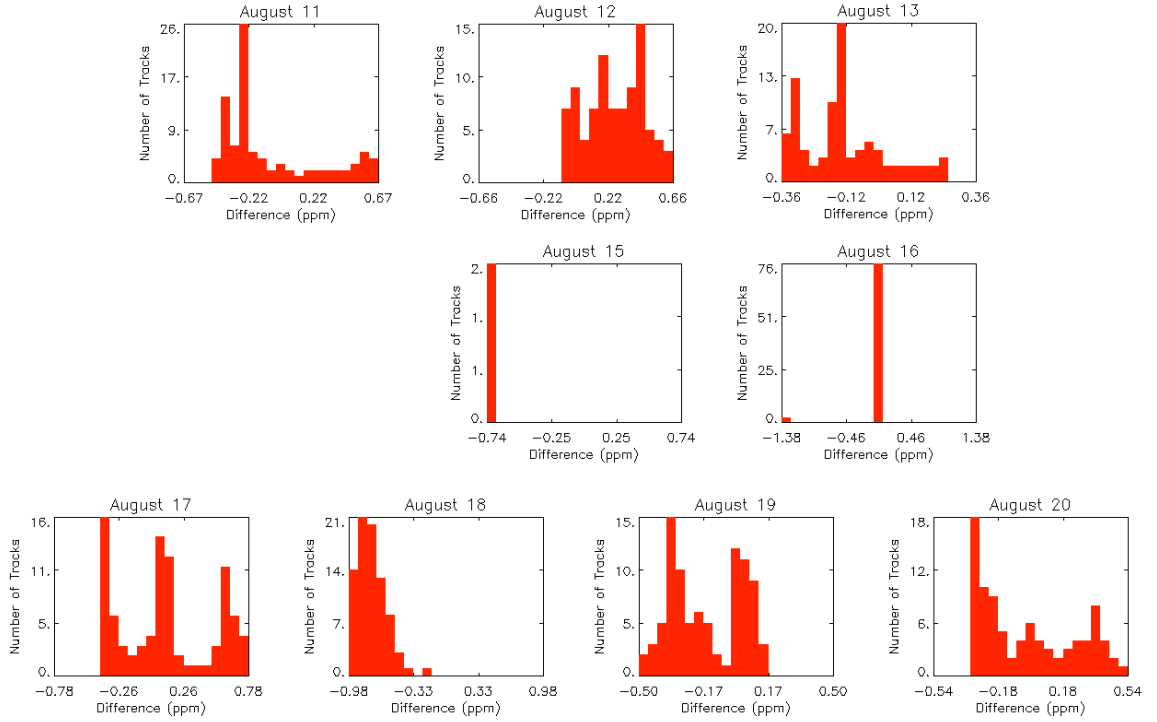


**Figure 5.11:** Grid 3 sampling distribution of the 1 PM errors in the emulated satellite concentrations compared to the diurnal mean from 6-6 PM LST.

The decomposition of the compiled ten-day distribution into daily sampling distributions is shown in Figure 5.12. Similar to grid 4, the magnitude and sign of the diurnal errors depends on the timing of the fronts more than on the biology and photosynthesis. On the mostly clear days, August 11, 13, and 20, the errors are small and include both underestimates and overestimates of the diurnal mean. Since overall the concentrations at 1 PM are near the diurnal mean, the tracks on the western side of the domain overestimate the total column diurnal mean because of the high concentrations from the agriculture and grasslands and the tracks on the eastern side of the domain underestimate the concentrations.

On the days surrounding the fronts, the magnitude and sign of the bias depends primarily on the timing of the front. The 12<sup>th</sup> overestimates the diurnal mean because the concentrations at 1 PM are still slightly elevated due to the front. The two tracks on the 15<sup>th</sup> underestimate the diurnal mean CO<sub>2</sub> concentration because they only see the very low CO<sub>2</sub> at 1 PM and because the front increases the CO<sub>2</sub> significantly that afternoon. Similarly, since the diurnal mean on the 16<sup>th</sup>

includes high concentrations from the front, measurements at 1 PM underestimate the diurnal mean because at that time the concentrations are low again from photosynthesis.



**Figure 5.12:** Daily sampling distributions of the difference between the emulated satellite tracks at 1 PM and the corresponding total column domain-averaged diurnal mean from 6-6 PM LST.

For case 2, the diurnal cycle in the total column  $\text{CO}_2$  concentration is weak. Instead of the diurnal cycle driving the  $\text{CO}_2$  temporal variability, we found that synoptic variability due primarily to advection of  $\text{CO}_2$  is the main influence to the mixing ratio. This analysis indicates that the temporal variability is not well sampled with a single measurement per day and may introduce large errors compared to the temporal average. To avoid large errors, inverse models need to accurately model synoptic-scale transport.

# **Chapter 6: Clear-Sky Errors in Satellite CO<sub>2</sub> Measurements**

## **6.1 Investigating the Clear-Sky Bias Using Observations**

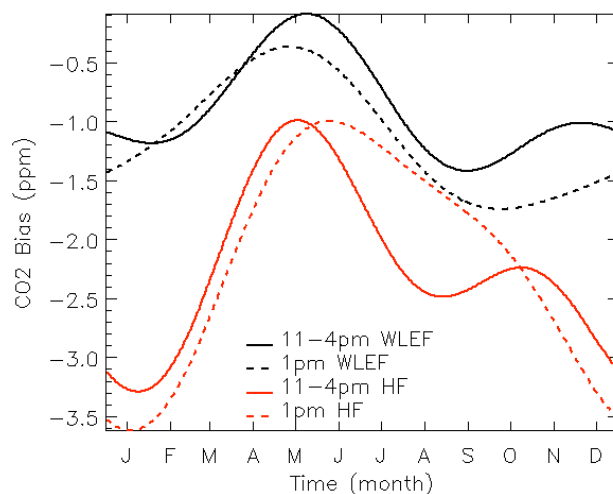
The first section in this chapter investigates the clear-sky errors using continuous observations from two towers. The errors found in this section include both local clear-sky errors due to being unable to see under the clouds at any one point in time and also temporal sampling errors from measuring the concentrations only on clear-sky days. These errors in the concentrations will appear in inversions that compare temporally averaged concentrations (weekly, bi-monthly, etc.) to satellite measurements that represent the concentration at only one snapshot in time.

The first subsection explains the results from the continuous observations, displaying the clear-sky bias in CO<sub>2</sub>, NEE, and CO. The next subsection explains a hypothesis for the cause of the shape and magnitude of the near-surface CO<sub>2</sub> bias. Finally, since the towers only take near-surface measurements, the errors in the total column concentrations will be different than the errors seen at the surface, and the last subsection discusses the expected errors in the total column satellite measurements from the analysis of near-surface concentrations.

### **6.1.1 Results**

Using the continuous measurements at WLEF and Harvard Forest (described in section 2.1), we investigated the clear-sky CO<sub>2</sub> bias, which was obtained by subtracting the two harmonic function fit to the complete data from the function fit to the clear-sky subset (discussed in section 2.2). The difference between the two fits is a sampling bias and not a random error because it

persists through all the years of data. Figure 6.1 shows the CO<sub>2</sub> bias at WLEF in the black lines and the bias at Harvard Forest in the red lines. The solid lines in the figure are the biases for the time period from 11 AM to 4 PM, while the dashed lines show the clear-sky bias at 1 PM.

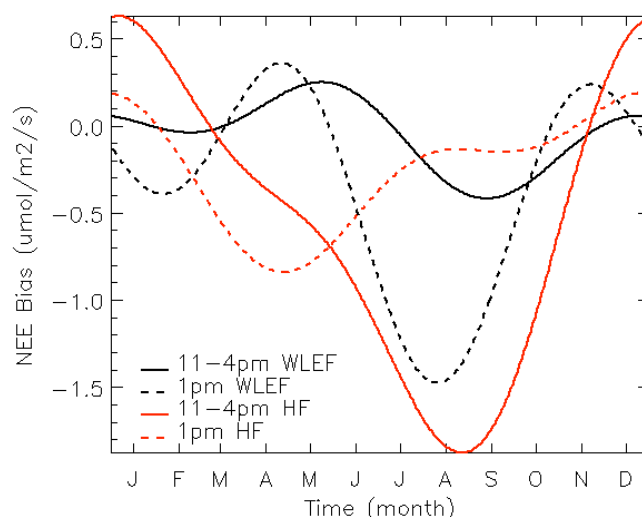


**Figure 6.1:** The clear-sky CO<sub>2</sub> sampling bias at both towers. The solid black line is the bias from 11 AM to 4 PM at WLEF and the dashed black line is the 1 PM bias. The red lines are the biases at Harvard Forest, with the solid line denoting 11-4 PM and the dashed line denoting the bias at 1 PM.

Measuring the CO<sub>2</sub> concentration only on clear days will underestimate the average mixing ratio. At both towers and at both times the shape of the bias is similar, with a greater near surface bias in the winter and a smaller bias during the summer months. This shape contradicts a priori expectations that the summertime bias will be positive due to the inability to see the enhanced NEE on cloudy days and that the winter bias will be small. Instead, the bias is always negative and is greatest during the winter, indicating that only measuring the concentrations on clear days will underestimate the CO<sub>2</sub>. At both locations the winter bias at 1 PM is slightly greater than the bias from requiring the entire afternoon to be clear, and at Harvard Forest the 1 PM bias during the summer is slightly less than the bias from 11-4 PM. The bias at the WLEF tower in Wisconsin is less severe than the bias at Harvard Forest, which could reflect the differences in vegetation, synoptic weather, or vertical mixing. The mean bias at 396 m on the WLEF tower from 11 AM to 4 PM is -1.0 ppm, and at Harvard Forest the mean bias is -2.1 ppm for 11 AM to



4 PM. Since the shape of the CO<sub>2</sub> bias is somewhat unexpected, we investigated the clear-sky bias in NEE and CO concentrations at the towers.



**Figure 6.2:** The NEE bias at both WLEF (black lines) and Harvard Forest (red lines). The solid lines depict the biases from 11 to 4 PM, while the dashed lines show the biases requiring clear conditions at 1 PM only.

Looking first at photosynthetic uptake, the NEE from both WLEF and Harvard Forest is shown in Figure 6.2. Both towers again exhibit a similar shape, with a large negative bias in the summer due to increased photosynthesis on clear days. The NEE bias is greater at Harvard Forest than at WLEF, which corresponds to the larger CO<sub>2</sub> bias at Harvard Forest. The largest NEE bias is at Harvard Forest from 11-4 PM, with a maximum bias of  $-2.1 \mu\text{mol}/\text{m}^2/\text{s}$  in August. Interestingly, the 1 PM bias at Harvard Forest has a very small negative bias in August. Instead, the maximum bias is earlier with a smaller magnitude of only  $-0.8 \mu\text{mol}/\text{m}^2/\text{s}$ . The cause of the shift in the 1 PM NEE bias at Harvard Forest is unclear, but perhaps could be due to meteorological factors such as water stress or increased temperature. At 1 PM, if the day is clear the water vapor mixing ratio may be low causing some stress on the plants and making them photosynthesize less, whereas looking from 11-4 PM the plants could be photosynthesizing more in the late morning and later afternoon. We performed an identical analysis of temperature; and

the temperature bias at Harvard Forest is positive during the summer, indicating that clear days are warmer. In addition, the temperature bias at 1 PM is  $\sim 0.4^{\circ}\text{C}$  greater than the temperature bias from 11-4 PM in August. The increased temperature at 1 PM could be increasing the respiration at the site, which would increase NEE.

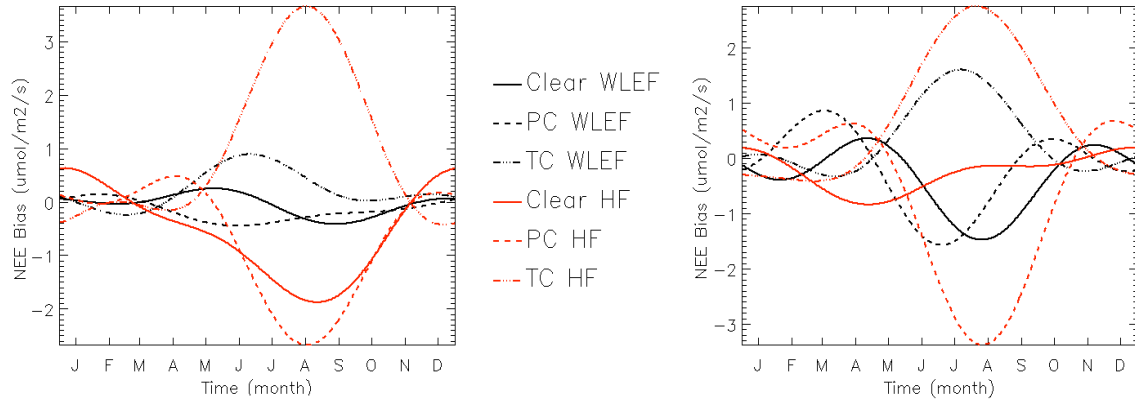
In contrast, the summer bias at WLEF is greater at 1 PM. The change of the largest bias from 11-4 PM at Harvard Forest to 1 PM at WLEF is partially due to the different years being analyzed at the two towers and the interannual variability in those years. As discussed in section 2.1, the data for Harvard Forest covers ten years while the NEE data at WLEF are only available for six years. We calculated the bias at Harvard Forest using the same years as at the WLEF tower and found the results were closer to the WLEF biases: from 11-4 PM the maximum bias was reduced to  $-1.7\ \mu\text{mol}/\text{m}^2/\text{s}$  and the bias at 1 PM shifted later to a maximum of  $-0.9\ \mu\text{mol}/\text{m}^2/\text{s}$  in mid June. Although the difference between the 1 PM bias and the 11-4 PM bias in August at Harvard Forest is reduced, the bias at 1 PM is still smaller and shifted earlier by a month. Another contributor to the 1 PM bias being larger at WLEF and smaller at Harvard Forest could be the different behavior of the weather and the different vegetation at the towers.

During the winter, both locations have a small NEE bias. The errors at WLEF are negligible, being slightly positive from 11-4 PM and negative at 1 PM. The winter errors at Harvard Forest are slightly positive, with a maximum bias of  $0.6\ \mu\text{mol}/\text{m}^2/\text{s}$ .

Even though the NEE differences are slightly different at the two towers and for the two time periods at Harvard Forest, we can still gain information from Figure 6.2. First of all, the summertime bias is negative at both locations while the winter differences are negligible or slightly positive. At WLEF, the summer NEE has a bias of  $\sim -0.6\ \mu\text{mol}/\text{m}^2/\text{s}$  on clear days and is slightly more for snapshots in the clouds than for completely clear days. At Harvard Forest, on clear summer days the NEE is negatively biased; however, the magnitude and timing of the bias is uncertain. For days that are completely clear and OCO can average adjacent measurements,

the errors are larger, and on the days that are only required to be clear for one hour the differences are less, particularly in August.

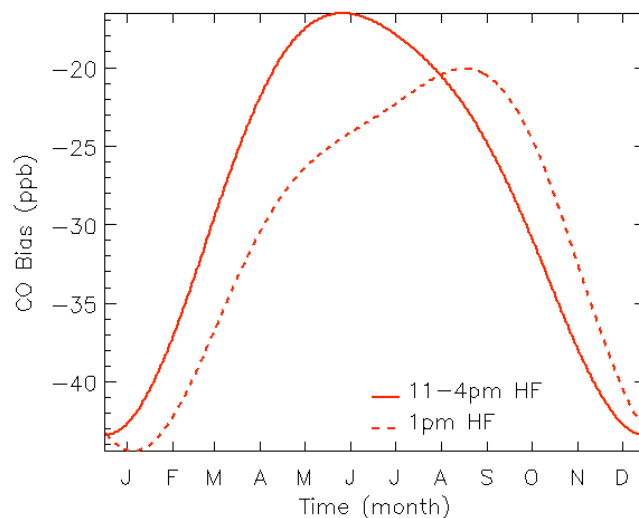
Recent literature shows that NEE is greatest on slightly cloudy days (e.g. Gu *et al.*, 1999; Freedman *et al.*, 2002; Law *et al.*, 2002), making the large negative summer bias surprising. This negative bias does not contradict with the literature, but rather is comparing different things. The published studies discussed in chapter 1 all investigated various levels of cloud cover to determine the optimal conditions for the greatest NEE, and all studies found NEE to be very low for overcast days. In this study, we are comparing clear days to all days, which includes clear, partly cloudy, and completely overcast conditions. Although clear days may not have the greatest NEE, these results indicate that photosynthesis is more active on clear days than it is on average for all days. This result is due to the average NEE being lowered by the cloudy days. Since the negative NEE bias during the summer is surprising, we analyzed the NEE bias for different sky conditions. At both locations and for both times we calculated the NEE bias for partly cloudy days and for completely overcast days. We considered 25% of the days to be partly cloudy, using the days with the next highest radiation measurements below the clear-sky days, and we considered 40% of the days to be overcast, using the days with the lowest radiation values. We compared each of these subsets to the total time-series to determine the bias from the associated sky conditions, and the results are displayed in Figure 6.3. The figure shows results that are similar to the literature. This figure shows that on partly cloudy days the NEE is indeed enhanced at both towers and for both times, having a greater bias in the summertime than the clear-sky subset. On cloudy days during the summer photosynthesis is reduced due to the light limitation, causing a large positive bias. The agreement of the NEE bias seen in this study with the literature for clear, partly cloudy and overcast days helps confirm that in the summer clear-sky days have slightly enhanced uptake compared to average.



**Figure 6.3:** NEE bias for clear sky (solid), partly cloudy (dashed), and totally cloudy or overcast days (dot-dash). The left-hand panel is the bias for 11-4 PM and the right panel is the bias at 1 PM. Similar to the previous plots, the black lines are the biases at WLEF and the red lines are the biases at Harvard Forest.

In addition to investigating NEE to help determine the shape of the CO<sub>2</sub> bias, we also calculated the bias in CO concentrations at Harvard Forest to determine the contributions of anthropogenic sources of CO<sub>2</sub>. Since CO is a ubiquitous by-product of the same combustion processes as CO<sub>2</sub> and has an average lifetime of only 3 months, CO measurements can provide information on the intensity of various anthropogenic activities, and previous studies have used CO to identify combustion and biosphere sources of CO<sub>2</sub> [Suntharalingam *et al.*, 2004; Palmer *et al.*, 2003; Lagenfelds *et al.*, 2002; Bergamaschi *et al.*, 2001; and Potosnak *et al.*, 1999]. In analyzing the CO bias, we will assume that the primary source of CO is fossil fuel combustion and that on average the combustion efficiency is 95%. Using those assumptions, a bias of 20 ppb of CO corresponds to a 0.5 ppm CO<sub>2</sub> bias, and a CO bias of 40 ppb yields a 1-ppm bias in CO<sub>2</sub> concentration.

Figure 6.4 displays the clear-sky CO bias. At both time periods, the CO concentration is lower on clear days, indicating that the contribution of fossil fuel combustion is less. The bias has a similar shape as the CO<sub>2</sub> bias, with a larger bias in the winter and a smaller bias in the summer. The CO bias indicates that part of the CO<sub>2</sub> bias is due to less anthropogenic concentrations on clear days, with an associated CO<sub>2</sub> bias of  $\sim -0.5$  ppm in the summer and  $\sim -1.2$  ppm in the winter.



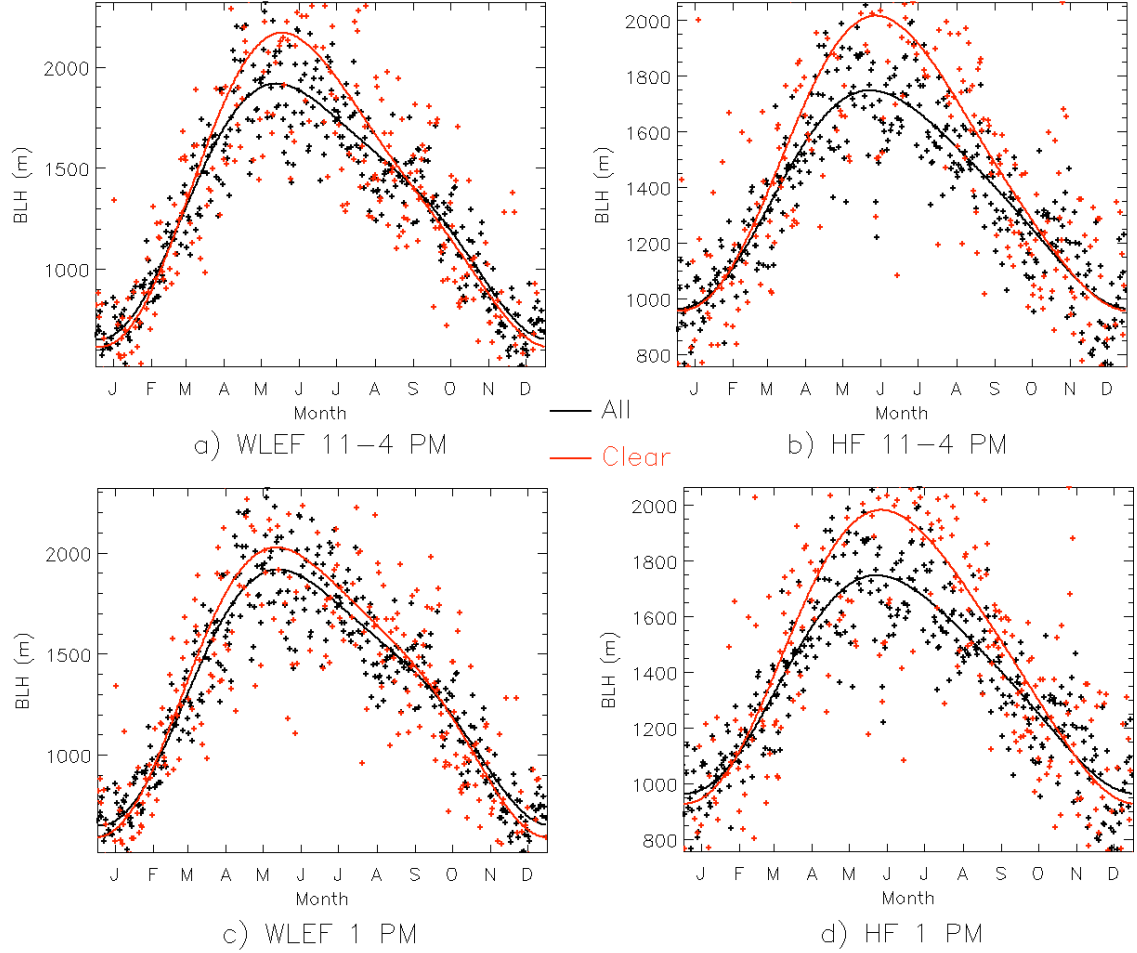
**Figure 6.4:** The clear-sky CO bias at Harvard Forest, with the bias from requiring clear conditions from 11 AM until 4 PM shown in the solid line and the bias at 1 PM displayed in the dashed line.

### 6.1.2 Explanatory Hypothesis: What Could Be Going On?

We propose that a combination of vertical mixing and advection are contributing to the clear-sky bias seen in the observations. The year-round negative bias seen in both the CO and CO<sub>2</sub> concentrations suggest that the boundary layer is deeper on clear days diluting the surface concentrations of gases with surface sources, or that clear days have less advection of air that has experienced high anthropogenic emissions. In addition, advection of air high in CO<sub>2</sub> from natural sources such as respiration could be more vigorous on cloudy days contributing to the negative CO<sub>2</sub> bias in clear conditions.

The first step to explaining the relative contributions from vertical mixing and from advection is to determine the bias in boundary layer heights on clear days. Since long-term boundary layer data is not available at either tower, we analyzed boundary layer heights from ERA-40 (ECMWF), which was discussed in section 2.1.5. We used a similar procedure in calculating the boundary layer bias as we did for the other variables; however, since ERA-40 only calculates the depth every six hours, the depths at 18 UTC are used for the total time-series at both 11-4 PM and 1 PM, making the clear-sky subset the only difference between these two

times. We used the same clear days that we used for the clear-sky  $\text{CO}_2$  subset. For example, to calculate the WLEF bias at 11-4 PM, the clear subset in the boundary layer depth time-series consisted of the same days that were sunny from 11-4 PM at WLEF when we calculated the  $\text{CO}_2$  bias. We fit two harmonics to both the complete time-series and to the clear subsets for both towers at both times.



**Figure 6.5:** Boundary layer depths at both towers, with the black + symbols representing the boundary layer heights from the complete dataset, the red symbols depicting the boundary layer heights on clear days, the black line depicting the fit to the complete time-series and the red line showing the fit to the clear-sky subset. Panel a) shows the data and the fits at WLEF for 11-4 PM, b) shows the boundary layer heights at Harvard Forest for 11-4 PM, c) shows the depths at WLEF at 1 PM, and d) shows the depths at Harvard Forest at 1 PM.

The data and the fits at both towers and for both times are displayed in Figure 6.5.

Although these data are produced from a model and have a large spread at any given time of year,

analysis of the ERA-40 boundary layer heights provide the best estimate of the actual depths at this point in time and give an estimate of the boundary layer depths on clear days that we can use to make more realistic calculations of the bias from enhanced vertical mixing. The primary feature in the data is the seasonal cycle of deeper boundary layers during the summer and shallower boundary layers in the winter. Looking at the fits, all four panels concur that during the summer the boundary layer is deeper on clear days, which agrees with our hypothesis; however, during the winter the boundary layer is actually slightly shallower on clear days.

Now that we have boundary layer depths, we can calculate an estimate of the bias from enhanced vertical mixing using a simple box model. In this calculation, we start with a background concentration,  $[CO_2]_B$ , of 360 ppm. From the background concentration, first we can calculate the  $\mu\text{mol}$  of  $CO_2$  in a box of air covering  $1 \text{ m}^2$  with a depth of  $Z_i$  m. Since units of ppm are a ratio of  $\mu\text{mol } CO_2$  per mol of air, the amount of  $\mu\text{mol}$  of  $CO_2$  in the box,  $\hat{C}$  is

$$\hat{C} = \frac{[CO_2]_B * \rho_{air} * Z_i}{M_d}, \quad (1)$$

where  $\rho_{air} = 1.275 \text{ kg/m}^3$  is the density of air and  $M_d = 28.97 \text{ mol/g}$  is the apparent molecular weight of air. Next, we add the local sources or sinks of carbon into the box. Since we do not have anthropogenic fluxed at the tower and since we can calculate the total contribution of anthropogenic emissions using  $CO$ , we only need to determine the flux of  $CO_2$  from plants. We will use an average daytime value of NEE, which is in units of  $\mu\text{mol/m}^2/\text{s}$ . To get the total  $\mu\text{mol}$  of  $CO_2$  being contributed into the box, we will multiply the NEE by a specified length of time and by the size of the box. In this study we assume that the NEE flux lasts for 10 hours, which is approximately the length of time plants are actively photosynthesizing for one day and is a realistic estimate of the length of time vegetation is contributing to an air mass. This step can be written as

$$\hat{C}_N = \hat{C} + (NEE * A * t), \quad (2)$$

where  $\hat{C}_N$  is the new concentration,  $A$  is the area of the box ( $1 \text{ m}^2$ ), and  $t$  is the time the NEE flux is contributing to the box, which in this case is 10 hours. Finally, the new concentration can be converted back into units of ppm to determine the change caused by the local flux of carbon. So the final concentration in the box,  $[CO_2]_F$ , is

$$[CO_2]_F = \frac{\hat{C}_N}{\text{mol air}} = \frac{\hat{C}_N * M_d}{\rho_{air} * Z_i}.$$

We performed this calculation for both the clear-sky days and the total days at both towers. The values for NEE and  $Z_i$  are summarized in Table 6.1. The total NEE values are the average daytime values for July in the summer and for January in the winter. The clear-sky NEE flux estimates are the total NEE plus the mean bias at each tower for July and for January. Similarly, the boundary layer heights are the mean of the heights from 11-4 PM and 1 PM at both towers in July and in January. To calculate the bias, we subtracted the concentration in the box for all sky conditions from the box concentration on clear days.

		January		July	
		NEE ( $\mu\text{mol/m}^2/\text{s}$ )	$Z_i$ (m)	NEE	$Z_i$
WLEF	Clear	1.	650	-8.	2150
	Total	1.	700	-7.	1900
Harvard Forest	Clear	1.	975	-13.5	2050
	Total	1.3	1000	-12.	1800

**Table 6.1:** Quantities used to determine the bias from NEE and vertical mixing.

The results are displayed in Table 6.2. In the summertime the bias appears small, with a slight overestimation of the  $CO_2$  concentration. This is because the deeper boundary layer is canceling the effect of the larger flux on clear days. The negative NEE bias seen in the observations depletes the  $CO_2$  concentration while the deeper boundary layer increases low  $CO_2$ , as it has to be mixed into a larger volume. The box model indicates that the summertime  $CO_2$  bias appears small at the tower because the lower concentrations from enhanced photosynthesis are mixed over a deeper boundary layer.



	WLEF	HF
Summer	-0.2	0.07
Winter	0.01	0.2

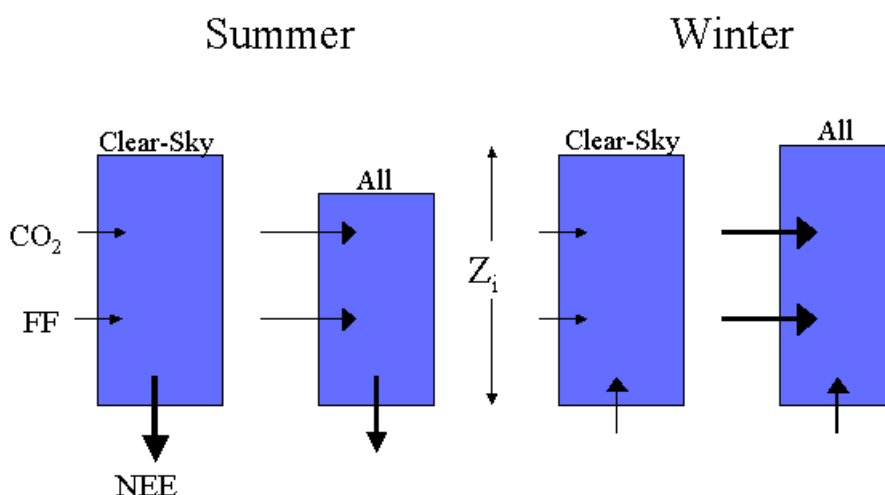
**Table 6.2:** The CO<sub>2</sub> bias in summer and winter at both towers from vertical mixing. The units are ppm.

Looking at the winter, the biases are also small, which is not surprising since the boundary layer depth and the NEE are nearly the same on clear days as they are on average. In this case, the box model illustrates that the large winter bias is probably not due to the vegetation or vertical mixing, but instead is likely due to advection.

Since the bias from vertical mixing of the CO<sub>2</sub> vegetation flux is small, especially in the winter, we can assume that the negative CO<sub>2</sub> bias must be due to less advection of high CO<sub>2</sub> concentrations on clear days. From the CO bias we know that advection of fossil fuels is contributing substantially to the negative CO<sub>2</sub> bias; however, this bias alone is not sufficient to create the large biases seen in the observations during the winter. The last factor that is contributing to the negative CO<sub>2</sub> bias is less advection of high CO<sub>2</sub> from natural sources or more advection of low CO<sub>2</sub> from natural sinks or across regional CO<sub>2</sub> gradients. Although we cannot calculate this contribution from point data, it is certainly possible that this factor could cause the remaining unexplained biases. On clear days, the air being advected over the tower could have lower CO<sub>2</sub> than average due to the accumulation of low CO<sub>2</sub> from enhanced photosynthesis or due to a different source region of air. Fronts and synoptic scale disturbances, which are associated with clouds, frequently contain air with higher CO<sub>2</sub> concentrations from a different region than the source region for clear days, and the negative bias could be due to under-sampling these events.

A cartoon summarizing our proposed mechanism to explain the CO<sub>2</sub> bias is shown in Figure 6.6, which we will now summarize. In the summer, the boundary layer is deeper on clear days canceling the effects of enhanced NEE. The small bias that is seen in the CO<sub>2</sub>

concentrations is due primarily to fossil fuel combustion, as the CO bias suggests that CO<sub>2</sub> has a bias of ~0.5 ppm from anthropogenic sources. The additional bias is due to the advection of lower than average CO<sub>2</sub> air. In the winter, the boundary layer is approximately the same depth on clear and on cloudy days, which does not contribute to the large negative bias. The CO concentrations suggest that air with lower CO<sub>2</sub> from fossil fuel than average is being advected on clear days, creating a negative bias of ~1.2 ppm. The additional bias is caused by cloudy days having higher CO<sub>2</sub> advected from natural sources such as enhanced respiration or suppressed photosynthesis. Undersampling events that advect high CO<sub>2</sub> could be causing a negative bias on clear days and could conceivably be contributing the ~2 ppm high surface CO<sub>2</sub> concentrations necessary to complete the large winter CO<sub>2</sub> bias at Harvard Forest.



**Figure 6.6:** Proposed mechanism to explain the CO<sub>2</sub> bias.

### 6.1.3 What Does this Mean for Satellites?

This study indicates that satellites will underestimate the CO<sub>2</sub> concentration due to a near surface bias; however, the magnitude of the bias in satellite CO<sub>2</sub> will require further investigation. Although the surface bias may be up to a few ppm of CO<sub>2</sub>, the satellite bias will be much less since OCO will measure total column CO<sub>2</sub>, diluting the surface by incorporating the well-mixed

CO<sub>2</sub> concentrations above the boundary layer. Since part of the bias at both towers appears to be dependent on the depth of the boundary layer, a satellite will not see the same bias as the observations at the towers. In the summer, the effects of a deeper boundary layer diluting the CO<sub>2</sub> bias from increased uptake will not appear in the total-column concentration and the satellite may experience a bias up to  $-0.2$  ppm due to increased photosynthesis. During the winter, the large bias at the surface may be primarily due to advection and under-sampling of cloudy days rather than due to the biology. This bias, which is present at both the towers during the summer and the winter, likely varies for different locations and will not bias the individual satellite measurements; however, the satellite measurements likely will not represent a temporal average and using the satellite measurements in inverse models will require accurate synoptic scale transport.

## **6.2 Investigating Clear-Sky Errors using SiB2-RAMS**

Since case 2 included clouds, in this section we will analyze the clear-sky errors, investigating both the local errors at 1 PM and the temporal sampling errors from under-sampling synoptic variations. The first subsection isolates the local clear-sky errors by comparing the concentration in a satellite track using only clear-sky pixels to the concentration in the satellite track using all pixels. An inversion that optimizes the concentrations from a domain that includes clear and cloudy areas to the satellite concentrations from only clear pixels will be introducing these local clear-sky errors.

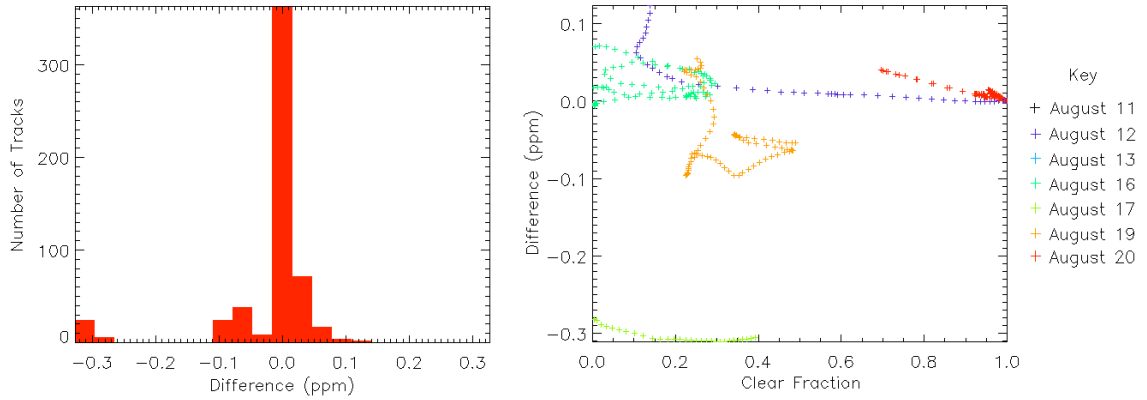
The second subsection investigates temporal sampling errors. An inversion will be subject to these errors if it compares temporally averaged concentrations in a grid cell to satellite total column concentrations taken only in clear columns at 1:15 PM. For example, if an inversion optimizes bi-monthly averaged concentrations to one satellite concentration, the inversion will be introducing temporal sampling errors in the concentrations that may alter the resulting fluxes. To

quantify these errors, we compared the emulated clear-sky total column satellite concentrations to the domain-averaged temporal mean concentration from the entire simulation.

### 6.2.1 Local Clear-Sky Errors

We calculated the local clear-sky total column CO<sub>2</sub> errors for both grid 4 and grid 3, which are the errors in each emulated satellite concentration from only seeing the clear pixels in the track at 1 PM compared to seeing all the pixels in the track the same day at the same time. For this calculation, we subtracted the emulated satellite total column CO<sub>2</sub> concentration at 1 PM using all pixels from the mixing ratios using only clear-sky pixels in the track at 1 PM. This calculation indicates how being able to only see clear pixels changes the concentration in comparison to the concentration if the satellite could measure the entire track.

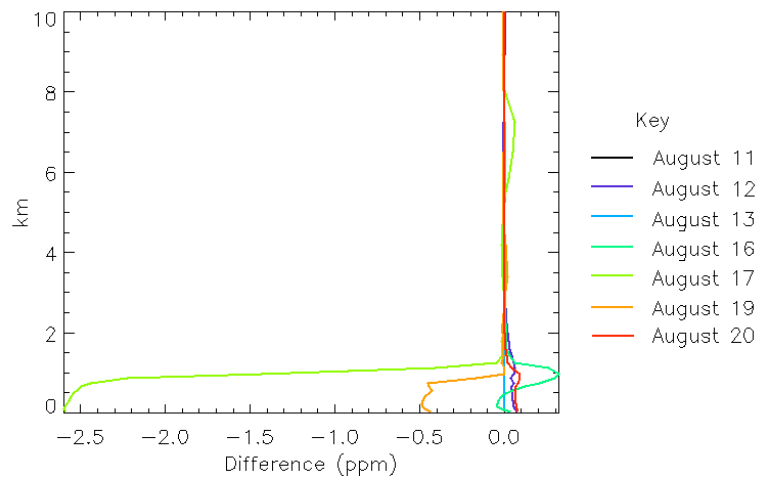
Figure 6.7 shows the local clear-sky errors for grid 4. The left panel displays a sampling distribution of the errors while the right panel displays the errors of each individual track. The sampling distribution in the left panel shows the differences on the x-axis plotted against the number of tracks with that error. The right panel breaks out the errors into the differences from each individual day. The x-axis of the right panel is the clear fraction of each emulated satellite track, with 1.0 being completely clear and 0 meaning the track is completely cloudy. Each individual + symbol represents one track and the colors denote the day corresponding to the error. August 14, 15, and 18 were completely cloudy and thus were unable to be measured. The local clear-sky errors are quite small, with the largest error being  $-0.33$  ppm. Nearly 60% of the satellite tracks do not have any clear-sky error, and as many tracks overestimate the concentration as underestimate the mean concentration. The right panel shows that the range of errors becomes larger as the track samples fewer pixels due to clouds. This result is expected, as the satellite sees fewer pixels it becomes more difficult to capture the mean concentration.



**Figure 6.7:** Local clear-sky total column CO<sub>2</sub> errors for grid 4, which are the differences between the emulated satellite concentration at 1 PM using only clear-sky pixels and the emulated satellite concentration at 1 PM the same day using all the pixels in the satellite track. The left panel is a sampling distribution of a compilation of the errors from all ten days, and the right panel shows the error from each individual track from all the different days of the simulation as a function of the clear fraction in the track, with 0 representing a completely cloudy track and 1. representing a clear track.

Looking at the individual days, the tracks on August 11 and 13 have no clear-sky errors because the entire domain is clear. The majority of the tracks on August 12 have very little error, but the error increases as the cloud cover in the tracks increases. The clouds are on the western side of the domain and are blocking the lowest concentrations and causing the tracks to have a positive clear-sky error of  $\sim 0.1$  ppm. The error is small because the spatial variability of CO<sub>2</sub> at 1 PM on the 12<sup>th</sup> is less than 0.3 ppm. On the 16<sup>th</sup>, all the tracks are able to see less than 40% of the possible pixels in the tracks because of the clouds. In contrast to expectations, the associated errors are small and positive, with the emulated clear-sky concentrations overestimating the mean concentration in the track by  $<0.1$  ppm. Similar to the 12<sup>th</sup>, the weak total column CO<sub>2</sub> variability in the domain helped keep the errors to a minimum. On the 17<sup>th</sup>, all the tracks underestimate the mean track concentration by the largest local clear-sky errors in this case. The figure illustrates that the majority of the tracks were covered by cloud, and the clear portion of the domain, which was the northwest corner, had relatively low concentrations with a strong gradient to high CO<sub>2</sub> in the southwestern corner. Since the tracks could not see the high CO<sub>2</sub> being advected into the

southwest corner, they underestimated the concentration. The 19<sup>th</sup> also had larger errors compared to the other days, and the tracks both underestimate and overestimate the concentration. Finally, the 20<sup>th</sup> has very small errors as the majority of the pixels in the tracks are clear. The tracks with the most cloud cover have positive errors because the areas under the cloud have lower concentrations, likely due to enhanced photosynthesis. Figure 6.7 indicates that the clear-sky errors on a 97 km by 97 km domain are small and relatively symmetrically distributed between underestimations and overestimations.

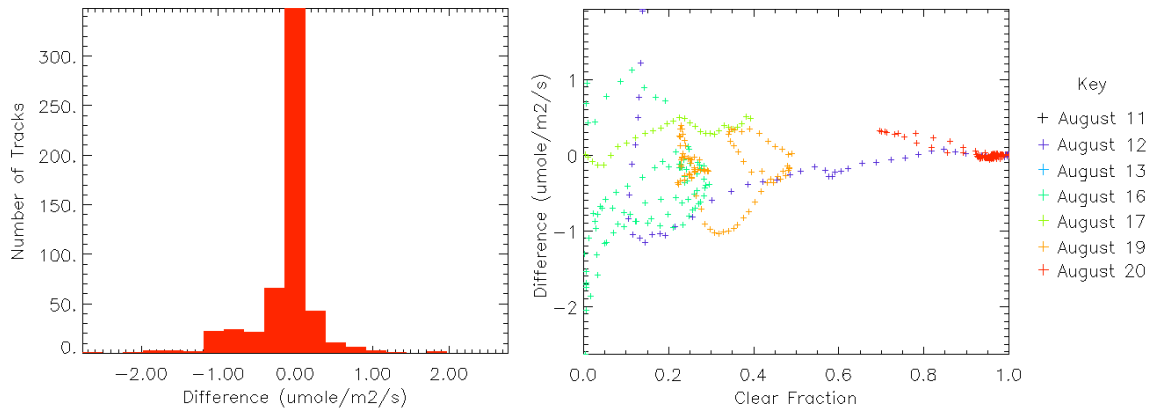


**Figure 6.8:** Vertical profiles of the clear-sky total-column CO<sub>2</sub> errors for grid 4, calculated by taking the domain average at each vertical level of the difference between the track CO<sub>2</sub> concentration using clear-sky pixels and the track mixing ratio using all the pixels in the swath for each vertical level. Each day is represented by a different color. August 14, 15, and 18 are completely cloudy.

Vertical profiles of the local clear-sky CO<sub>2</sub> errors from each day for grid 4 are displayed in Figure 6.8. These profiles are the domain average of the differences between the track CO<sub>2</sub> concentration using only clear-sky pixels and the emulated CO<sub>2</sub> concentration in the track using all pixels at each different vertical level. The majority of the difference in CO<sub>2</sub> comes from below 2 km, which is not surprising because most of the variability is close to the surface. August 12 and 20 have small positive errors, which come from overestimating the concentration primarily in the lowest km. August 16 also has positive total column CO<sub>2</sub> errors, which

surprisingly does not come from the surface but from  $\sim 1$  km above the surface. The 17<sup>th</sup> and 19<sup>th</sup> both have large negative errors that extend from the surface to  $\sim 1$  km up, and the 17<sup>th</sup> actually has a slight overestimation  $\sim 6$  km above the surface as well.

Since most of the total column  $\text{CO}_2$  variability is at the surface, to determine how NEE may be contributing to the clear-sky  $\text{CO}_2$  errors, we calculated the local clear-sky NEE errors. As in calculating the total column  $\text{CO}_2$  errors, the errors were calculated by subtracting the mean track NEE value using all pixels from the mean track NEE using only clear pixels. The clear-sky NEE errors for grid 4 are shown in Figure 6.9. The sampling distribution shows that the NEE errors from looking at clear-sky pixels only compared to the mean NEE including clouds are very small. Similar to the local clear-sky total column  $\text{CO}_2$  sampling distribution, the majority of the tracks do not have a clear-sky error. Surprisingly, the NEE errors are not one-sided but instead are rather symmetrical, with the clear-sky NEE under certain tracks underestimating the mean and in other areas overestimating the NEE flux.



**Figure 6.9:** Local clear-sky NEE errors for grid 4, which are the differences between the emulated satellite NEE value at 1 PM using only clear-sky pixels and the emulated satellite NEE value at 1 PM the same day using all the pixels in the satellite track. Similar to Figure 6.7, the left panel is a sampling distribution of the errors and the right panel shows the error from each individual track.

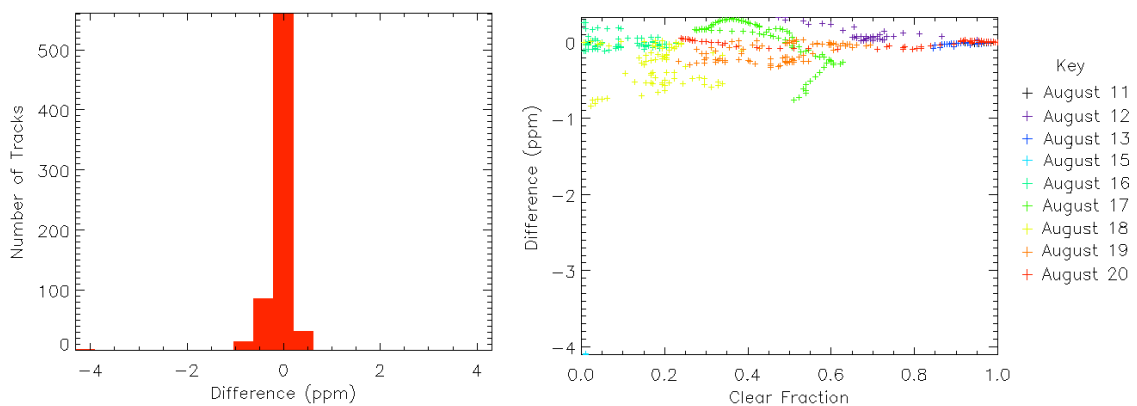
To get an idea of the correlation between NEE and the total column concentration, we can compare the right panel of Figure 6.9 to the errors in total column  $\text{CO}_2$  in Figure 6.7. The clear-sky NEE on August 12 has enhanced uptake by as much as  $-1.2 \mu\text{mol}/\text{m}^2/\text{s}$  in certain tracks;

however, in the total column clear-sky  $\text{CO}_2$ , the tracks all overestimate the concentration indicating that on this day the spatial pattern due to advection affected the  $\text{CO}_2$  concentration more than NEE. A few tracks do have less NEE in the clear pixels, and this overestimate could be contributing some to the  $\text{CO}_2$  errors, but advection changing the spatial pattern of  $\text{CO}_2$  appears to be the primary driver of the  $\text{CO}_2$  clear-sky errors. Similarly, on August 16<sup>th</sup> NEE is primarily enhanced in clear pixels, but in contrast the  $\text{CO}_2$  in the clear-sky tracks overestimates the mean concentration. It appears that the enhanced NEE is not altering the  $\text{CO}_2$  spatial pattern, which is confirmed by the vertical profile of the  $\text{CO}_2$  errors on the 16<sup>th</sup> indicating that the errors are not coming from the surface but rather from  $\sim 1$  km above the surface; however, recall that NEE had alternating stripes of enhanced photosynthesis in high radiation areas and that the  $\text{CO}_2$  concentration was slightly offset by the clouds. This offset could be due to advection and convection, which could be lifting and advecting higher concentrations of  $\text{CO}_2$  from the north and advecting the lower  $\text{CO}_2$  at the surface from enhanced photosynthesis slightly south into cloudy regions. On the 17<sup>th</sup> the NEE errors are very small, indicating that the  $\text{CO}_2$  errors are again due to the combination of advection and cloud cover, as high  $\text{CO}_2$  is being advected into the southwest corner but the clear patch over the northwest corner still has low  $\text{CO}_2$ . August 19 does have tracks that have corresponding negative errors in both NEE and  $\text{CO}_2$ , and on the 20<sup>th</sup> reduced uptake corresponds with higher concentrations. The areas under the clouds on August 20 actually had enhanced photosynthesis that lowered the concentration of  $\text{CO}_2$  in a region the satellite could not see, causing the positive errors. Comparing the NEE errors to the total column clear-sky errors indicates that overall NEE does not alter the  $\text{CO}_2$  concentration much on this domain, which is similar to our discussion in chapter 3 when we found that the spatial patterns of  $\text{CO}_2$  did not correspond to the NEE maps except on a few days of the simulation. On the days that do have altered  $\text{CO}_2$  correlating with altered NEE, the clear-sky errors are not one-sided but rather are symmetrical: on the 16<sup>th</sup> tracks with enhanced photosynthesis actually had higher  $\text{CO}_2$



concentrations with the errors coming from aloft, on the 19<sup>th</sup> clear-sky regions had enhanced photosynthesis and lowered CO<sub>2</sub> concentrations, and on August 20 clear-sky pixels actually had less uptake than the mean corresponding to slightly higher CO<sub>2</sub> concentrations. This grid indicates that the local clear-sky errors are not driven primarily by NEE differences and will be small and unbiased.

Now we will investigate the clear-sky errors for grid 3, which are shown in Figure 6.10. The majority of the clear-sky errors on the larger grid are also very small, with the magnitude of the error in over 80% of the tracks <0.3 ppm. The local clear-sky CO<sub>2</sub> errors are not biased, with some of the tracks overestimating the concentration, although a larger percentage underestimates the mean CO<sub>2</sub> concentration. Two tracks grossly underestimate the concentration by over 4 ppm in the total column.

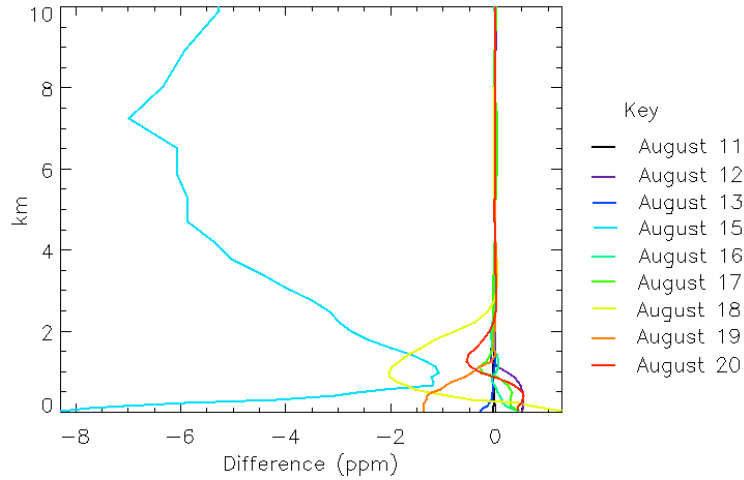


**Figure 6.10:** Local clear-sky total column CO<sub>2</sub> errors for grid 3, which are the differences between the emulated satellite concentration at 1 PM using only clear-sky pixels and the emulated satellite concentration at 1 PM the same day using all the pixels in the satellite track. The left panel is a sampling distribution of the errors and the right panel shows the error from each individual track with the different days having different colors. August 14 was completely cloudy.

Looking at the errors from each track, similar to grid 4 we see that the spread of errors increases as the cloud fraction decreases. The errors on August 11 and 13 are very small, as the majority of the domain is clear and the clouds do not alter the CO<sub>2</sub> concentration. On the 12<sup>th</sup> using only clear-sky pixels overestimates the mean CO<sub>2</sub> concentration in the tracks using all

pixels because the lowest concentrations in the domain are cloud covered. The large clear-sky errors come from August 15<sup>th</sup>, the day that also had large spatial and temporal errors. The two tracks with very few clear pixels drastically underestimate the mean concentration in the track because high CO<sub>2</sub>, with a large gradient, is being advected into the domain prior to the front but has not yet reached the clear pixels in the northwest corner. On August 17<sup>th</sup>, tracks both underestimate and overestimate the mean concentration, and on the 19<sup>th</sup> clear pixels have lower CO<sub>2</sub> causing relatively large negative errors. Finally, the 20<sup>th</sup>, which is mostly clear, has very small local clear-sky errors.

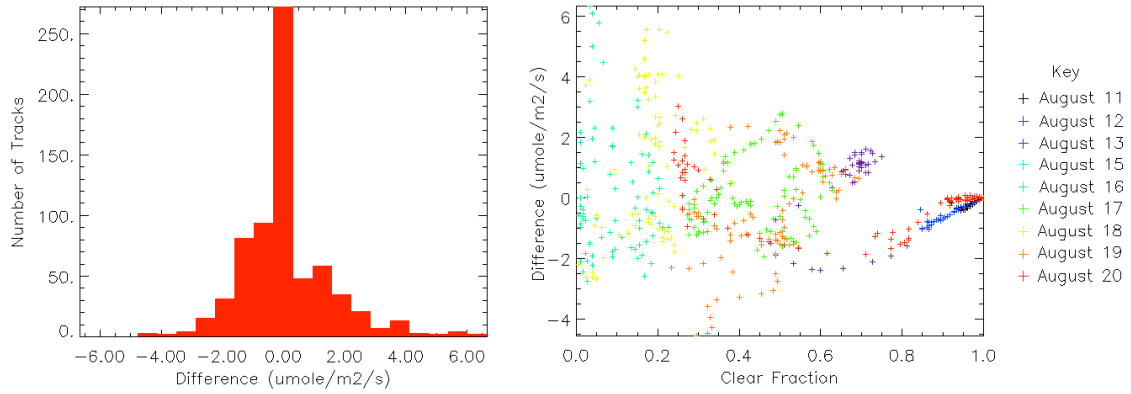
Vertical profiles of the clear-sky CO<sub>2</sub> errors for grid 3 are displayed in Figure 6.11. The vertical profiles are calculated by calculating the difference between the emulated track value using clear-sky pixels and the emulated track values using all the pixels in the track at each vertical level, and then taking the domain average of these errors for each level in the atmosphere. Similar to grid 4, the majority of the clear-sky error comes from the surface; however, the clear-sky error on the 15<sup>th</sup> extends all the way up to 10 km, indicating that the large error is not only due to surface variability but also to differences in the entire column. We saw in chapter 4 that at times close to frontal passages the entire column had increased CO<sub>2</sub> from advection, which is why the clear-sky error is also present in the entire column, making the clear-sky error large. The other days the clear-sky error comes from below 3 km. The vertical profiles show that on some of the days (August 12, 13, 16, and 19) the errors are close to the surface, while on other days (August 18 and 20) the error is higher up at ~1.5 km. Interestingly, the errors on the 18<sup>th</sup> and 20<sup>th</sup> switch signs in the profile and are positive close to the surface and negative at ~1.5 km. The vertical profiles for grid 3 also show that most of the variability is near the surface; however, with synoptic events the clear-sky errors exist throughout the entire column.



**Figure 6.11:** Vertical profiles of the grid 3 local clear-sky errors. The profiles from each day are the domain average at each vertical level of the differences between the clear-sky concentrations and the total concentrations for each track.

To determine the effects of NEE on the CO<sub>2</sub> errors, the local clear-sky NEE errors are displayed in Figure 6.12. The sampling distribution is relatively symmetrical and is centered on zero. Slightly more of the tracks have negative NEE clear-sky errors, but the largest errors are actually positive errors associated with less photosynthesis or more respiration in the clear regions.

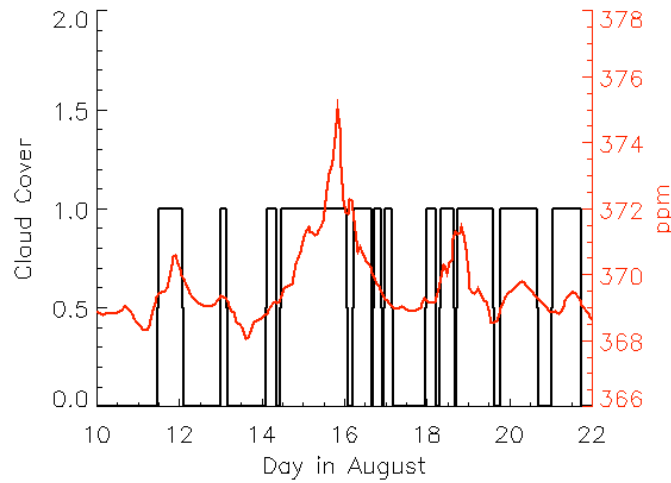
Looking at the daily distribution of errors as a function of the clear fraction in the track, August 13 has enhanced NEE in the clear regions, but the increased photosynthesis does not affect the CO<sub>2</sub> concentration. August 15<sup>th</sup>, the day with large CO<sub>2</sub> errors, actually has a positive NEE bias, which supports our earlier statement that advection is driving the gradient of CO<sub>2</sub> on this day. The 16<sup>th</sup> has both positive and negative errors in NEE and also in CO<sub>2</sub>, which indicates that changes in NEE are altering the total column CO<sub>2</sub> concentrations. In contrast, August 18<sup>th</sup> and 19<sup>th</sup> also have positive and negative NEE errors, but the clear-sky CO<sub>2</sub> concentration always underestimates the mean, which indicates that again advection is influencing the CO<sub>2</sub> spatial pattern more than NEE on these days. Lastly, August 20 is mostly clear and thus has small NEE errors. Similar to grid 3, the local clear-sky errors are affected more by advection than by local fluxes and are generally small; however, the errors can be large due to synoptic events.



**Figure 6.12:** Similar to Figure 6.9, local NEE clear-sky errors now for grid 3.

## 6.2.2 Temporal Sampling Errors

This subsection investigates the errors from using clear-sky measurements to represent the mean concentration over the entire simulation, which is analogous to comparing the clear-sky total column OCO concentrations to weekly or bimonthly CO<sub>2</sub> averages in inverse models. We are calling these errors temporal sampling errors because OCO will never be able to sample the concentrations on cloudy days.

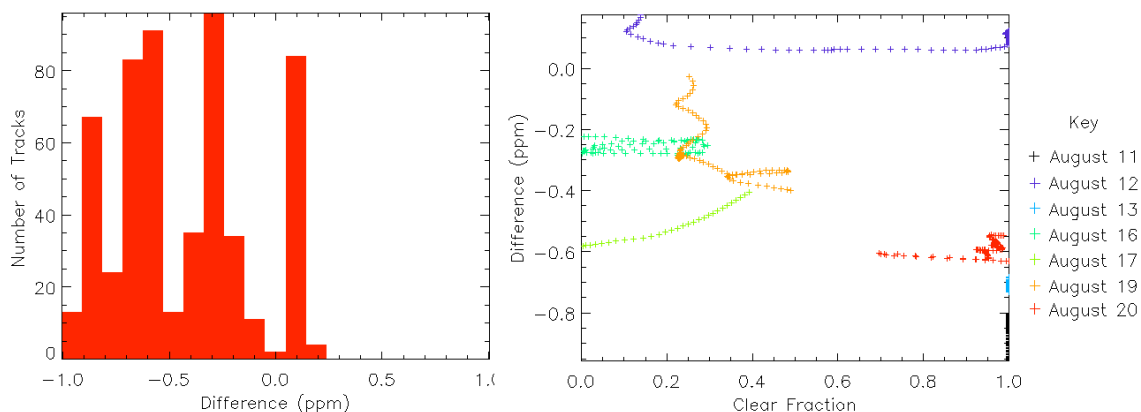


**Figure 6.13:** Total column CO<sub>2</sub> concentrations in case 2 sampled at the WLEF tower (red line) compared to the modeled sky conditions (black line), where a value of 0 indicates the sky was clear and 1 indicates the tower was cloud covered.

In this simulation we have seen that the fronts advect higher CO<sub>2</sub> from the southwest and that the concentrations on cloudy days are greater than the concentrations on clear or partly

cloudy days. This is illustrated in Figure 6.13, which shows both the modeled total column concentration at the WLEF tower and the modeled cloud cover. The red line shows the CO<sub>2</sub> concentrations, which are in ppm. The cloud cover is displayed in the black line, with 0 representing clear skies and 1 representing clouds. The figure shows that the three spikes in CO<sub>2</sub> concentrations associated with the three fronts all occur during cloudy periods. In contrast, the lowest concentrations seen in the simulation occur when the sky is clear.

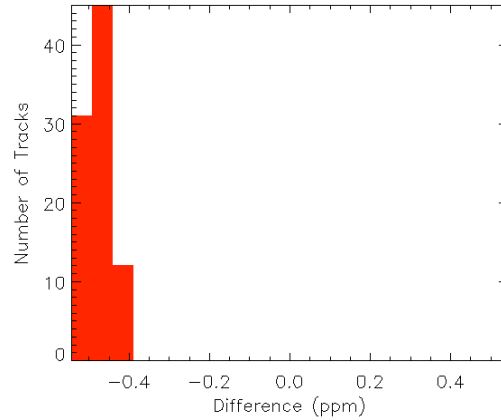
To calculate the clear-sky errors from sampling only clear days compared to the temporal average, we subtracted the mean domain concentration over the entire 10-day simulation from each clear-sky emulated satellite track at 1 PM. The compiled sampling distribution of these temporal sampling errors for grid 4 is shown in Figure 6.14, as is the error from each individual track. The majority of the satellite tracks underestimate the 10-day domain mean, illustrating that temporal averaging will introduce large errors and may have a negative bias in this domain. Each peak in the sampling distribution is a different day, indicating that the day-to-day variability of CO<sub>2</sub> is much greater than the spatial variability.



**Figure 6.14:** Grid 4 clear-sky CO<sub>2</sub> errors over the entire 10-day simulation, which are the differences between the emulated satellite concentration from each track using only clear-sky pixels and the 10-day domain average. The left panel is the compiled sampling distribution of the errors from all ten days. The right panel shows the errors from each track on the y-axis as a function of the clear fraction of the track, which is 1.0 if the track is completely clear and 0 if the track is overcast.

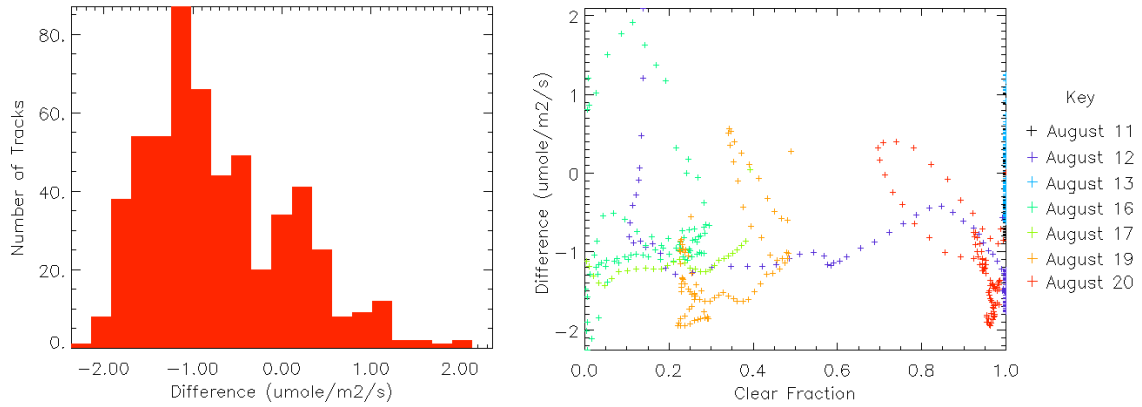
In the right panel of Figure 6.14 we can see the errors from each different emulated satellite track. In general, the higher the clear fraction the larger the underestimation of the mean. Also, the days following fronts typically have smaller errors because the concentration has not been depleted by photosynthesis yet. The two clear days, August 11 and August 13, have the largest errors and underestimate the 10-day mean by more than 0.6 ppm. August 20, which is also mostly clear, also significantly underestimates the 10-day mean. August 12, which is the afternoon after the front, has the smallest errors compared to the domain mean, with all of the tracks either capturing the domain mean or actually overestimating the 10-day average by a small margin. August 16, which also still has higher concentrations from the second front, also has small errors, although on this day the satellite tracks underestimate the mean concentration. The errors on the 17<sup>th</sup>, when most of the pixels in each emulated satellite track are cloudy, have the same magnitude as the errors on the 20<sup>th</sup>, which is mostly clear. Finally, August 19 has the largest range of errors, with the tracks in the clear southeast portion of the domain underestimating the 10-day mean by  $\sim 0.4$  ppm and the tracks on the western side of grid 4 only slightly underestimating the mean concentration.

Next we are going to look at the average clear-sky errors from sampling only clear days, which we calculated by subtracting the emulated track mean at 1 PM from all 10 days of the simulation using all pixels from the 10-day 1 PM track mean using only clear pixels. The results from this calculation are displayed in Figure 6.15. Compared to the 10-day track mean at 1 PM, the average clear-sky values from all the emulated satellite tracks have much lower concentrations and underestimate the mean track value by  $\sim 0.45$  ppm. Since the average of the clear-sky tracks underestimates the mean by at least 0.38 ppm, sampling only on clear days and using these measurements to represent a temporal average likely introduces a substantial bias because on average the satellite tracks on cloudy days have higher concentrations and because the majority of the tracks will capture this underestimation, at least for this simulation in Wisconsin.



**Figure 6.15:** Grid 4 sampling distribution of the average clear-sky total column CO<sub>2</sub> errors due to undersampling synoptic events. These errors are the differences between each emulated track mean at 1 PM from all days of the simulation using only clear pixels in the track and each emulated track mean at 1 PM from all days of the simulation using all pixels in the track.

To see how the NEE fluxes are contributing to the CO<sub>2</sub> errors, we will look at the clear-sky NEE errors compared to the 10-day 1 PM mean. Similar to the CO<sub>2</sub> errors over the entire 10-day simulation, we calculated the NEE errors by subtracting the 10-day mean at 1 PM from each clear-sky emulated satellite track. We used the 1 PM concentrations to avoid the diurnal cycle of photosynthesis. This calculation is still comparable to the CO<sub>2</sub> errors because the 1 PM mean CO<sub>2</sub> concentration was nearly identical to the domain average. Figure 6.16 shows the NEE errors for grid 4. The sampling distribution in the right panel shows that the NEE is typically greater in clear-sky conditions than on average, but some clear tracks do have less photosynthesis. The mode of the distribution is  $-1.2 \mu\text{mol}/\text{m}^2/\text{s}$ . Since not all of the tracks had increased uptake compared to the average, this simulation suggests that there is not a NEE bias on clear-days.



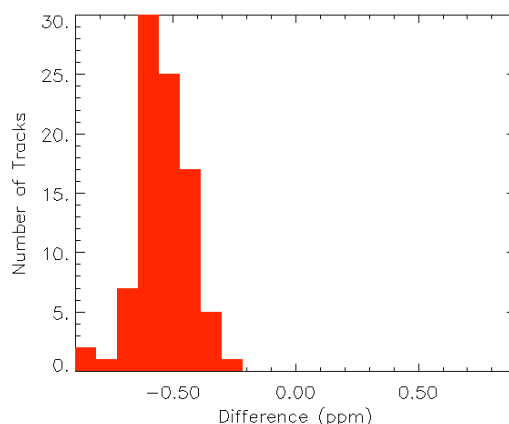
**Figure 6.16:** Grid 4 clear-sky NEE errors over the entire 10-day simulation, which are the differences between the emulated satellite NEE value from each track using only clear-sky pixels and the 10-day domain average. The left panel is the sampling distribution of the errors from the emulated satellite concentrations compared to the mean domain concentration at 1 PM. The right panel shows the error from each track.

Looking at the errors for the individual tracks, they appear to be relatively random, although definitely skewed towards enhanced photosynthesis compared to average. The totally clear days, August 11 and 13, had relatively small errors compared to the 10-day 1 PM mean, with some of the errors on the 13<sup>th</sup> being positive. August 12<sup>th</sup> had enhanced uptake of  $\sim 1 \mu\text{mol/m}^2/\text{s}$  on average, but did not cause an underestimation of the  $\text{CO}_2$  concentrations. Most of the clear-sky pixels on August 16<sup>th</sup> had enhanced photosynthesis, although a few tracks had reduced uptake compared to the 10-day average. August 17<sup>th</sup>, 19<sup>th</sup>, and 20<sup>th</sup> all had enhanced photosynthesis compared to the 10-day mean, and these days all had underestimates of the 10-day mean  $\text{CO}_2$  concentration. This correlation at the end of the simulation suggests that the photosynthesis is acting to lower the concentrations on clear days in the absence of a synoptic system.

Figure 6.17 shows the average clear-sky NEE errors from sampling only clear days, which we calculated by subtracting the emulated track mean at 1 PM from all 10 days of the simulation using all pixels from the 10-day 1 PM track mean using only clear pixels. Similar to the average  $\text{CO}_2$  errors, on average the tracks underestimate the 10-day mean NEE flux by  $\sim 0.5 \mu\text{mol/m}^2/\text{s}$ . This mean underestimation suggests that compared to the average, clear days do

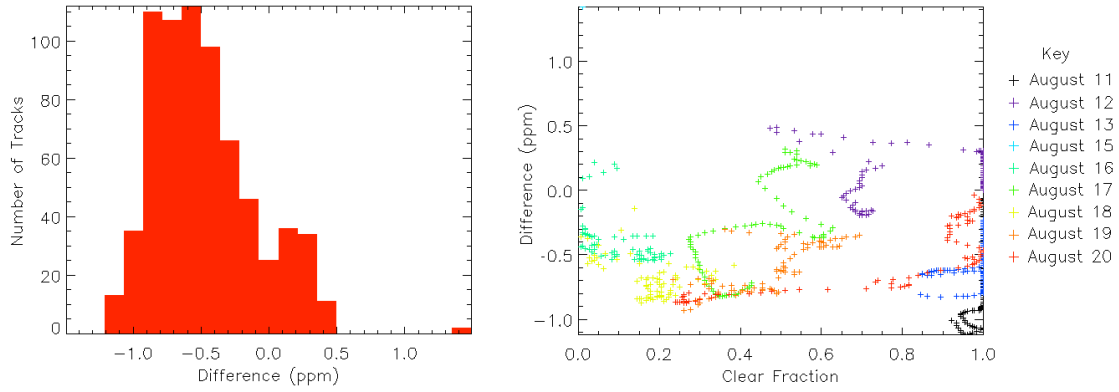


indeed have enhanced photosynthesis and that enhanced photosynthesis on clear days is contributing to the CO<sub>2</sub> errors. Since the NEE flux is greater on clear days on average, this suggests that the errors will likely be negative on average. Although it is possible for the NEE not to be biased in certain tracks, the majority of the tracks will have negative errors causing a bias.



**Figure 6.17:** Grid 4 sampling distribution of the average NEE errors due to undersampling synoptic events. These errors are the differences between each emulated track NEE mean at 1 PM from all days of the simulation using only clear pixels in the track and each emulated track mean NEE at 1 PM from all days of the simulation using all pixels in the track.

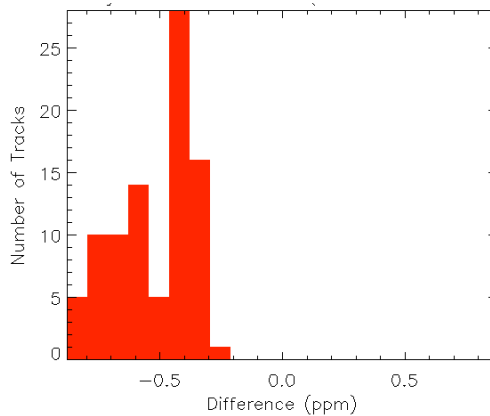
After analyzing the temporal sampling errors for grid 4, now we will analyze the errors for grid 3. The errors from a comparison between the emulated clear-sky satellite tracks and the 10-day domain average is shown in Figure 6.18. The 10-day domain averaged concentration was also nearly identical to the 1 PM 10-day domain average, so we only compared the emulated satellite concentrations to the 10-day domain average. As expected, the results show that most of the satellite tracks underestimate the 10-day domain average, although a few tracks do overestimate the concentration and 2 tracks grossly overestimate the mean. Although not all of the tracks underestimate the 10-day mean, since the majority of them do and the mean of the tracks is also definitely negatively biased, then using clear-sky concentration to represent a temporal average will introduce a negative error in the majority of the tracks, making the concentration more likely to be biased.



**Figure 6.18:** Grid 3 clear-sky CO<sub>2</sub> errors over the entire 10-day simulation, calculated using the same method used in Figure 6.14. The left panel is the sampling distribution of the errors from the emulated satellite concentrations compared to the 10-day domain mean concentration. The right panel shows the errors from each track as a function of the clear fraction of the track, which is 1.0 if the track is completely clear and 0 if the track is overcast.

Analyzing the errors on individual days, the right panel of Figure 6.18 shows the errors from each track for grid 3. Unlike grid 4, the errors do not have the tendency to be larger with a larger clear fraction. The tracks at 1 PM on the first day of case 2 underestimate the 10-day mean by the greatest magnitude, followed closely by the other predominantly clear day on the 13<sup>th</sup>. In general the days after the front in this case have smaller errors, although this is likely due to the timing of the fronts in this simulation. The largest error is the 1.4 ppm overestimation on the 15<sup>th</sup> by the two tracks with very few clear pixels. Due to the front, the concentrations are greater on the 15<sup>th</sup> in general, which is why these tracks overestimate the 10-day mean by such a large amount. Looking at the end of the simulation, most of the tracks underestimate the 10-day mean on these partly cloudy days by  $\sim -0.5$  ppm.

To look at the average of the errors from the emulated satellite, a sampling distribution of the errors from the difference between the 10-day track mean using clear-pixels only and the 10-day track mean using all the pixels is shown in Figure 6.19. Similar to grid 4, the average clear-sky concentration in all of the tracks is lower than the mean using all pixels, suggesting again that using clear-sky concentrations to represent a temporal average will introduce a negative bias of  $\sim -0.5$  ppm. The errors in this domain are larger, with a maximum error of  $-0.9$  ppm.

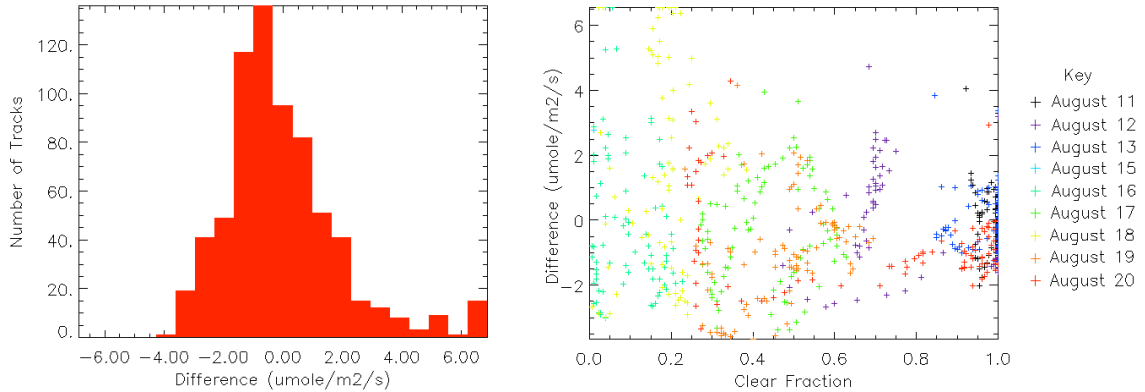


**Figure 6.19:** Grid 3 sampling distribution of clear-sky total column CO<sub>2</sub> errors due to undersampling synoptic events, similar to Figure 6.15.

To help determine if vegetation or advection caused the CO<sub>2</sub> errors, Figure 6.20 shows the temporal sampling clear-sky NEE errors for grid 3, which are the NEE differences that each satellite track sees. These errors are calculated the same way they were for grid 4. The sampling distribution in Figure 6.20 shows that the errors for grid 3 are centered more towards zero rather than being shifted towards negative errors, although the mode, which is  $\sim -0.5 \mu\text{mol}/\text{m}^2/\text{s}$ , still indicates more areas have enhanced uptake in clear conditions compared to average. The largest errors in the domain are positive errors, indicating that NEE was suppressed on some clear days by over  $6.0 \mu\text{mol}/\text{m}^2/\text{s}$ . The larger errors and the shift of errors towards more positive values are due to the complex vegetation on this domain compared to the relatively homogenous vegetation on grid 4.

The first impression from looking at the right panel of Figure 6.20 is that the NEE errors appear to be random rather than having a bias. The range of errors in the distribution appears to increase with the increase in cloud cover, unlike the CO<sub>2</sub> distribution, which had a large range of errors despite the sky conditions. The days that are primarily clear, August 11, 13, and 20, in general have relatively small errors centered on 0. August 12, which was partly cloudy, had suppressed NEE compared to the 10-day 1 PM mean. The areas under the clouds on the 12<sup>th</sup> have enhanced NEE compared to the previous day because the clouds were optically thin and provided

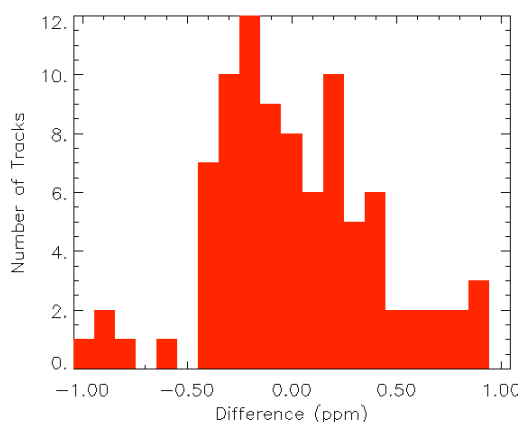
more optimal conditions for photosynthesis. The 12<sup>th</sup> also has positive CO<sub>2</sub> errors, which could be due to the suppressed NEE. August 15<sup>th</sup>, which had a large positive CO<sub>2</sub> error, has a positive NEE error of  $\sim 3 \mu\text{mol}/\text{m}^2/\text{s}$ , which is contributing to the CO<sub>2</sub> overestimation. August 16 and 17 have a large range of errors that are both positive and negative in sign. On the 16<sup>th</sup> the enhanced NEE is in the southwestern corner, where recent precipitation has moistened the soil.



**Figure 6.20:** Similar to Figure 6.16, grid 3 clear-sky NEE errors over the entire 10-day simulation. The left panel is the sampling distribution of the errors from the emulated satellite concentrations compared to the mean domain concentration at 1 PM. The right panel shows errors from each track.

Corresponding to the positive and negative NEE errors, the CO<sub>2</sub> errors are also distributed around 0 on the 16<sup>th</sup> and 17<sup>th</sup>. The emulated satellite tracks on the 18<sup>th</sup> have enhanced uptake in 1/3 of the tracks compared to the 10-day 1 PM mean concentration, while the rest of the tracks have suppressed NEE, with the largest errors occurring in the tracks on this day. The cloud cover on the 18<sup>th</sup> is covering much of the vegetation but not covering the lakes, and the large errors are due to the satellite track being unable to see the majority of the vegetation in the domain. As seen in Figure 6.18, the CO<sub>2</sub> errors are all negative, indicating that the errors are primarily due to advection. August 19 and 20 also have clear-sky tracks with enhanced and suppressed NEE compared to the 10-day mean, but the emulated tracks always underestimate the CO<sub>2</sub> concentration. While enhanced NEE is contributing to some of the negative CO<sub>2</sub> errors in grid 3, since some of the tracks have positive NEE errors yet still have less CO<sub>2</sub> than the 10-day mean, most of the CO<sub>2</sub> bias is due to advection.

Figure 6.21 shows the average clear-sky errors for each track and is analogous to Figure 6.17, which shows the sampling distribution for grid 4. For grid 3, the tracks do not have enhanced NEE in clear conditions on average, but instead the errors are relatively evenly distributed. This suggests that for this larger and more heterogeneous domain, NEE is not biased on clear days, and the NEE errors are not the primary drivers of the CO<sub>2</sub> negative errors.



**Figure 6.21:** Grid 3 sampling distribution of the average NEE errors due to undersampling synoptic events, calculated the same as in Figure 6.17.

Investigating the temporal sampling errors for both grid 4 and 3 show that undersampling of fronts and synoptic events leads to the underestimation of a temporal average using only clear-sky concentrations. Since nearly all the tracks have negative errors, this likely introduces a bias in the measurements. While enhanced NEE in clear conditions compared to the 10-day mean concentration may be causing some of the negative bias, particularly for grid 4, the dominant factor appears to be advection of high CO<sub>2</sub> during fronts that are associated with clouds. The large magnitude of errors and the potential bias at this location suggests that the satellite concentrations cannot be used to represent a temporal average and will introduce large errors in inversions that optimize temporally averaged concentrations to satellite measurements. To avoid these errors, inverse models will have to accurately model the transport and optimize concentrations taken at the same time and in the same meteorological conditions that the measurements were sampled.

# Chapter 7: Conclusions and Recommendations

## 7.1 Conclusions

The objectives of this research were to investigate three errors introduced when using satellite measurements in inverse studies:

- 1) **Spatial Representation Errors:** To what degree can one satellite track from a heterogeneous domain actually represent the average CO<sub>2</sub> concentration at the inversion resolution?
- 2) **Diurnal Average Errors:** Will measurements at 1:15 PM accurately capture the CO<sub>2</sub> diurnal average?
- 3) **Clear-Sky Errors:** What is the sign and magnitude of the local clear-sky errors? Will the measurements have temporal sampling errors from undersampling synoptic events?

To analyze these errors we used both continuous observations and a coupled ecosystem-atmosphere model, looking primarily at northern Wisconsin. The model simulations we analyzed included realistic clear-sky and cloudy conditions in the summer. To investigate these errors, we emulated the OCO measuring strategy and compared the results of simulated satellite concentrations to the domain average.

Spatial representativeness errors may be introduced into inversions that optimize CO<sub>2</sub> concentrations from a grid cell to satellite concentrations sampled over only a fraction of the

domain. Since current methods require inversions to use resolutions larger than the OCO sampling strategy, the magnitude of these errors must be quantified to reduce errors in the resulting fluxes. We found that the spatial representativeness errors are considerably smaller than the 0.5% instrumental error; however, the precision of the measurements will depend on the size and the heterogeneity of the domain the measurements are representing. In a relatively homogeneous domain consisting of similar vegetation types, 95% of the errors are  $<0.2$  ppm and are normally distributed. The spatial errors are small for a 100 km by 100 km domain because the variability at 1 PM on that scale is limited, typically  $<1.0$  ppm. As the domain size the satellite measurements are representing increases, the spatial variability in the domain increases due to the inclusion of more heterogeneous features. Looking at a relatively heterogeneous area in Wisconsin including portions of the Great Lakes, the spatial errors on a 450 km domain have increased to a standard deviation of  $\sim 0.4$  ppm and a chance exists of the satellite tracks having an error  $>1.0$  ppm in the total column compared to the domain average. To avoid introducing spatial representation errors, inverse models should obtain fluxes by optimizing concentrations on the finest possible resolution. If a coarse resolution is required, inverse models will have to de-weight the satellite measurements from heterogeneous domains to avoid introducing errors.

Inverse models can compare diurnally averaged concentrations to satellite measurements sampled at 1:15 PM. Optimizing the diurnal mean concentration can cause temporal errors in the inversion. We found that the diurnal cycle of total column  $\text{CO}_2$  from biology is small and that the main driver of the temporal variability in total column concentrations is synoptic events. This synoptic variability due to  $\text{CO}_2$  advection is not well sampled with one measurement time per day, and using satellite measurements from 1:15 PM to optimize diurnal average concentrations may introduce large errors in the inversion. These sampling errors can be avoided by simulating the transport in inversions and by comparing the satellite measurements to the modeled concentrations at the same time the satellite mixing ratios are sampled.

In this study we investigated two types of clear-sky errors: local clear-sky errors and temporal sampling errors. Local clear-sky errors are errors that may be introduced in inverse models that compare concentrations in a grid cell that may be partially cloudy to satellite mixing ratios sampled at the same time but only over clear regions. Our study suggests that local clear-sky errors at 1 PM are small and randomly distributed, with magnitudes  $<0.2$  ppm on a 100 km by 100 km domain. The SiB2-RAMS simulations suggest that the errors due to altered photosynthesis on cloudy vs. clear days are very small and that the local errors are primarily due to advection. To avoid larger clear-sky errors from  $\text{CO}_2$  variability caused by advection combined with biology, inverse models should use the smallest possible resolution; and our study indicates that inverse models using a 100 km resolution will have minimal local clear-sky errors.

The second type of clear-sky error we investigated is temporal sampling errors, which are errors that can be introduced into an inversion if the model uses satellite concentrations only in clear columns at 1:15 PM to optimize temporally-averaged concentrations. We found that these errors are large and introduce a significant bias. Analyzing a SiB2-RAMS simulation in Wisconsin illustrates that undersampling fronts associated with clouds and high  $\text{CO}_2$  can cause a bias of  $\sim -0.5$  ppm compared to the domain-averaged ten-day mean. The continuous measurements at both the WLEF tower and at Harvard Forest also support this conclusion, showing a negative year-round clear-sky bias in the observations, which may also be due to advection of high  $\text{CO}_2$  by synoptic systems that are associated with clouds. This negative temporal sampling bias is likely site-specific and may not be the same for all global locations; however, regardless of the region, due to the large variability in  $\text{CO}_2$  concentrations associated with synoptic events, satellite concentrations will not be able to represent a temporal average. To avoid these large errors, inverse models will have to accurately model the synoptic-scale atmospheric transport, and inversions will have to use the measured satellite mixing ratios to optimize modeled concentrations sampled at the same time.



## 7.2 Recommendations for Future Work

This study has investigated the spatial, diurnal, and clear-sky errors extensively in Wisconsin during the summer. Although we investigated a relatively heterogeneous domain, the errors could be different over various regions. While it is likely that the spatial and local clear-sky errors will remain small, they should be investigated at other locations, including regions that contain point sources. Since not all regions actively photosynthesize during the summer and since regions may be closer to anthropogenic sources, the temporal sampling errors will be different around the world and require further investigation.

In addition to sampling alternate regions, different times of year also need to be investigated. The results from the continuous observations suggest that at WLEF and Harvard Forest the temporal sampling errors are even larger during the winter than in the summer. A study during the winter using SiB2-RAMS could provide valuable information about this bias, as well as additional estimates about the spatial errors. Also, winter biases at other locations require further quantification.

Although the simulations in this study matched the observations at the WLEF tower relatively closely, advances in the model also may provide more realistic simulations and thus provide better error estimates. Currently the model fails to represent C4 vegetation in the United States, and a future project would be to update the model vegetation classes using a C3/C4 map. Changes in the vegetation classes could alter the photosynthesis and respiration, particularly in the plains regions. Another useful model advancement for this study would be a two-leaf model. Currently SiB2-RAMS is a single leaf model, which may not have the same benefits to cloud cover and diffuse radiation as actual ecosystems. Additionally, model simulations in different years could be compared to aircraft data, which would provide further evaluation of the model.

# References

- Bacastow, R.B., 1976: Modulation of Atmospheric Carbon Dioxide by the Southern Oscillation. *Nature*, **261**, 116-118.
- Baker, I., A.S. Denning, N. Hanan, L. Prihodko, M. Uliasz, P.L. Vidale, K. Davis, and P. Bakwin, 2003: Simulated and Observed Fluxes of Sensible and Latent Heat at the WLEF-TV Tower using SiB 2.5. *Global Change Biology*, **9** (9), 1262, doi:10.1046/j.13652486.2003.00671.
- Bakwin, P.S., P. Tans, D. Hurst, and C. Zhao, 1998: Measurements of Carbon Dioxide on Very Tall Towers: Results of the NOAA/CMDL Program. *Tellus*, **50B**, 401-415.
- Baldocchi, D. 1997: Measuring and Modelling Carbon Dioxide and Water Vapour Exchange Over a Temperate Broad-Leaved Forest During the 1995 Summer Drought. *Plant, Cell and Environment*, **20**, 1108-1122.
- Baldocchi, D., E. Falge, L.H. Gu, R. Olson, et al., 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, **82**, 2415-2434.
- Ball, J.T., 1988: An Analysis of Stomatal Conductance. Ph.D. Diss., Stanford University, 89 pp.
- Barnston, A.G. and R.G. Livezey, 1987: Classification, Seasonality, and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, **115** (6), 1083-1126.
- Battle, M., M.L. Bender, P. Tans, J.W.C. White, J.T. Ellis, T. Conway, and R.J. Francey, 2000: Global Carbon Sinks and their Variability Inferred from Atmospheric O<sub>2</sub> and  $\delta^{13}\text{C}$ . *Science*, **287**, 2467-2470.
- Bender, M., P.P. Tans, J.T. Ellis, J. Orchardo, and K. Habfast, 1994: A High Precision Isotope Ratio Mass Spectrometry Method for Measuring the O<sub>2</sub>/N<sub>2</sub> Ratio of Air. *Geochimica et Cosmochimica Acta*, **58**, 4751-4758.
- Bergamaschi, P., et al., 2000: Inverse Modeling of the Global CO Cycle. 1. Inversion of CO Mixing Ratios. *Journal of Geophysical Research*, **105**, 1909-1927.
- Bolin, B. and C.D. Keeling, 1963: Large-Scale Atmospheric Mixing as Deduced from the Seasonal and Meridional Variations of Carbon Dioxide. *Journal of Geophysical Research*, **68**, 3899-3920.

- Bousquet, P., P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein, and P.P. Tans, 2000: Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, **290** (5495), 1342-1346
- Brenkert, A., 1998: CO<sub>2</sub> Emission Estimates from Fossil-Fuel Burning, Hydraulic Cement Production and Gas Flaring for 1995 on a One Degree Grid Cell Basis. CDIAC, Oak Ridge, TN, NDP-058A.
- Buermann, W., B. Anderson, C.J. Tucker, R.E. Dickinson, W. Lucht, C.S. Potter, R.B. Myneni, 2003: Interannual Covariability in Northern Hemisphere Air Temperatures and Greenness Associated with El Nino-Southern Oscillation and the Arctic Oscillation. *Journal of Geophysical Research*, **108** (D13), 4396, doi:10.1029/2002JD002630.
- Chou, W.W., S.C. Wofsy, R.C. Harriss, J.C. Lin, *et al.*, 2002: Net Fluxes of CO<sub>2</sub> in Amazonia Derived from Aircraft Observations. *Journal of Geophysical Research*, **107**, 4614, 10.1029/2001JD001295.
- Clark, T.L. and R.D. Farley, 1984: Severe Downslope Windstorm Calculations in Two and Three Spatial Dimensions Using Anelastic Interactive Grid Nesting: A Possible Mechanism for Gustiness. *Journal of Atmospheric Science*, **41**, 329-350.
- Coakley, J.A. Jr., and P. Chylek, 1975: The Two-Stream Approximation in Radiative Transfer: Including the Angle of Incident Radiation. *Journal of Atmospheric Science*, **32**, 409-418.
- Collatz, G.J., J.T. Ball, C. Grivet, and J.A. Berry, 1991: Physiological and Environmental Regulation of Stomatal Conductance, Photosynthesis and Transpiration: A Model that Includes a Laminar Boundary Layer. *Agriculture and Forest Meteorology*, **54**, 107-136.
- Collatz, G.J., M. Ribas-Carbo, and J.A. Berry, 1992: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants. *Australian Journal of Plant Physiology*, **19**, 519-538.
- Conway, T.J., P. Tans, L.S. Waterman, K.W. Thoning, K.A. Masarie, and R.H. Gammon, 1988: Atmospheric Carbon Dioxide Measurements in the Remote Global Troposphere, 1981-84. *Tellus*, **40B**, 81-115.
- Conway, T.J., P.P. Tans, L.S. Waterman, and K.W. Thoning, 1994: Evidence for Interannual Variability of the Carbon Cycle from the National Oceanic and Atmospheric Administration Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network. *Journal of Geophysical Research – Atmospheres*, **99**, 22831.
- Cotton, W.R., R.A. Pielke Sr., R.L. Walko, G.E. Liston, C.J. Tremback, H. Jiang, R.L. McAnelly, J.Y. Harrington, M.E. Nicholls, G.G. Carrio, and J.P. McFadden, 2002: RAMS 2001: Current Status and Future Directions. *Meteorology and Atmospheric Physics*, doi:10.1007/s00703-001-0584-9.

- Crisp, D., R.M. Atlas, F.M. Breon, L.R. Brown, J.P. Burrows, P. Ciais, B.J. Connor, S.C. Doney, I.Y. Fung, D.J. Jacob, C.E. Miller, D.O'Brien, S. Pawson, J.T. Randerson, P. Rayner, P.J. Salawitch, S.P. Sander, B. Sen, G.L. Stephens, P.P. Tans, G.C. Toon, P.O. Wennberg, S.C. Wofsy, Y.L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, S. Schroll, 2003: The Orbiting Carbon Observatory (OCO) Mission. *Advances in Space Research*, **34**, 700-709.
- Dai, Y., R.I. Dickinson, and Y.P. Wong, 2004: A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance. *Journal of Climate*, **19** (12), 2281-2299.
- Daly, C., R.P. Neilson, and D.L. Philips, 1994: A Statistical-Topographical Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology*, **33**, 140-158.
- Davies, H.C., 1983: Limitations of Some Common Lateral Boundary Schemes Used in Regional NWP Models. *Monthly Weather Review*, **111**, 1002-1012.
- Davis, K. J., D.H. Lenschow, S.P. Oncley, C. Kiemle, G. Ehret, A. Giez, and J. Mann, 1997: Role of Entrainment in Surface–Atmosphere Interactions over the Boreal Forest. *Journal of Geophysical Research*, **102**, 29,219–29,230.
- Davis, K. J., P.S. Bakwin, C. Yi, B.W. Berger, C. Zhao, R.M. Teclaw, and J.G. Isebrands, 2003: The Annual Cycle of CO<sub>2</sub> and H<sub>2</sub>O Exchange Over a Northern Mixed Forest as Observed from a Very Tall Tower. *Global Change Biology*, **9**, 1278–1293.
- Denning, A.S., 1994: Investigations of the Transport, Sources, and Sinks of Atmospheric CO<sub>2</sub> Using a General Circulation Model. Ph.D. Dissertation, Colorado State University.
- Denning, A.S., M. Nicholls, L. Prihodko, I. Baker, P.L. Vidale, K. Davis, and P. Bakwin, 2003: Simulated Variations in Atmospheric CO<sub>2</sub> Over a Wisconsin Forest Using a Coupled Ecosystem-Atmosphere Model. *Global Change Biology*, **9**, 1-10.
- Dickinson, R.E., 1983: Land Surface Processes and Climate-Surface Albedos and Energy Balance. *Advances in Geophysics*, Academic Press, 48 pp.
- Engelen, R.J., A.S. Denning, K.R. Gurney, and G.L. Stephens, 2001: Global Observations of the Carbon Budget 1. Expected Satellite Capabilities for Emission Spectroscopy in the EOS and NPOESS Eras. *Journal of Geophysical Research*, **106** (D17), 20,055-20,068.
- Engelen, R.J., A.S. Denning, and K.R. Gurney, and TransCom3 Modelers, 2002: On Error Estimation in Atmospheric CO<sub>2</sub> Inversions. *Journal of Geophysical Research*, **107** (D22), 4635, doi:10.1029/2002JD002195.
- Farquhar, G.D., S. von Caemmerer, and J.A. Berry, 1980: A Biochemical Model of Photosynthetic CO<sub>2</sub> Fixation in Leaves of C<sub>3</sub> Species. *Planta*, **149**, 78-90.
- Farquhar, G.D. and M.L. Roderick, 2003: Pinatubo, Diffuse Light, and the Carbon Cycle. *Science*, **299**, 1997-1998.

- Fischer, H., M. Whalen, J. Smith, D. Mastroianni, and B. Deck, 1999: Ice Core Records of Atmospheric CO<sub>2</sub> Around the Last Three Glacial Terminations. *Science*, **283**, 1712-1714.
- Freedman, J.M., D.R. Fitzjarrald, K.E. Moore, R.K. Sakai, 2001: Boundary Layer Clouds and Vegetation-Atmosphere Feedbacks. *Journal of Climate*, **14**, 180-197.
- Global Soil Data Task. 2000. Global Soil Data Products CD-ROM (IGBP-DIS). CD-ROM. International Geosphere-Biosphere Programme, Data and Information System, Potsdam, Germany. Available from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. [<http://www.daac.ornl.gov>].
- Goulden, M.L., B.C. Daube, S.M. Fan, D.J. Sutton, A. Bazzaz, J.W. Munger, S.C. Wofsy, 1997: Physiological Responses of a Black Spruce Forest to Weather. *Journal of Geophysical Research*, **102** (D24), 28,987-28,996.
- Grell, G.A., Y.H. Kuo, and R.J. Pasch, 1991: Semiprognostic Tests of Cumulus Parameterization Schemes in the Middle Latitudes. *Monthly Weather Review*, **119**, 5-31.
- Grell, G.A., 1993: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations. *Monthly Weather Review*, **121**, 764-787.
- Grell, G.A., and D. Devenyi, 2002: A Generalized Approach to Parameterizing Convection Combining Ensemble and Data Assimilation Techniques. *Geophysical Research Letters*, **29** (14), 1693.
- Gu, L., J.D. Fuentes, H.H. Shugart, R.M. Staebler, T.A. Black, 1999: Responses of Net Ecosystem Exchanges of Carbon Dioxide to Changes in Cloudiness: Results from Two North American Deciduous Forests. *Journal of Geophysical Research*, **104** (D24), 31,421-31,434.
- Gu, L., D. Baldocchi, S.B. Verma, T.A. Black, T. Vesala, E.M. Falge, and P.R. Dowty, 2002: Advantages of Diffuse Radiation for Terrestrial Ecosystem Productivity. *Journal of Geophysical Research*, **107** (D6).
- Gu, L., D.D. Baldocchi, S.C. Wofsy, J.W. Munger, J.J. Michalsky, S.P. Urbanski, and T.A. Boden, 2003: Response of a Deciduous Forest to the Mount Pinatubo Eruption: Enhanced Photosynthesis. *Science*, **299**, 2035-2038.
- Gurney, K.R., R.M. Law, A.S. Denning, P.J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y.H. Chen, P. Ciais, S. Fan, I.Y. Fung, M. Gloor, M. Heimann, K. Higuchi, J. John, T. Maki, S. Maksyutov, K. Masarie, P. Peylin, M. Prather, B.C. Pak, J. Randerson, J. Sarmiento, S. Taguchi, T. Takahashi, C.W. Yuen, 2002: Towards Robust Regional Estimates of CO<sub>2</sub> Sources and Sinks Using Atmospheric Transport Models. *Nature*, **415**, 626-630.
- Hansen, J.W., J.W. Jones, C.F. Kiker, and A.W. Hodges, 1999: El Niño-Southern Oscillation Impacts on Winter Vegetable Production in Florida. *Journal of Climate*, **12**, 92-102.
- Hansen, M., R. DeFries, J.R.G. Townshend, and R. Sohlberg, 1998: 1Km Land Cover Classification Derived from AVHRR. College Park, Maryland: The Global Land Cover Facility.

- Hansen, M., R. DeFries, J.R.G. Townshend, and R. Sohlberg, 2000: Global Land Cover Classification at 1km Resolution Using a Decision Tree Classifier. *International Journal of Remote Sensing*, **21**, 1331-1365.
- Harrington, J.Y., 1997: The Effects of Radiative and Microphysical Processes on Simulated Warm and Transition Season Arctic Stratus. PhD Diss., Atmospheric Science Paper No 637, Colorado State University, Department of Atmospheric Science, Fort Collins, CO 80523, 289 pp.
- Harrington, J.Y., T. Reisin, W.R. Cotton, S.M. Kreidenweis, 1999: Cloud Resolving Simulations of Arctic Stratus. Part II: Transition-Season Clouds. *Atmospheric Research*, **55**, 45-75.
- Harrington, J.Y. and P.Q. Olsson, 2001: A Method for the Parameterization of Cloud Radiative Properties Within Numerical Models. Implications for Arctic Cloudy Boundary Layers. *Atmospheric Research*, **57**, 51-80.
- Hollinger, D.Y., F.M. Kelliher, J.N. Byers, J.E. Hunt, T.M. McSeveny, P.L. Weir, 1994: Carbon Dioxide Exchange between an Undisturbed Old-Growth Temperate Forest and the Atmosphere. *Ecology*, **75** (1), 134-150.
- Houghton, R.A., 1999: The Annual Net Flux of Carbon to the Atmosphere from Changes in Land Use 1850-1990. *Tellus*, **51B**, 298-313.
- Houweling, S., F.M. Breon, I. Aben, C. Rödenbeck, M. Gloor, M. Heimann, and P. Ciais, 2004: Inverse Modeling of CO<sub>2</sub> Sources and Sinks Using Satellite Data: A Synthetic Inter-Comparison of Measurement Techniques and their Performance as a Function of Space and Time. *Atmospheric Chemistry and Physics*, **4**, 523-538.
- Indermühle, A., T.F. Stocker, F. Joss, H. Fischer, H.J. Smith, M. Wahlen, B. Deck, D. Mastroianni, J. Tschumi, T. Blunier, R. Meyer, and B. Stauffer, 1999: Holocene Carbon-Cycle Dynamics Based on CO<sub>2</sub> Trapped in Ice at Taylor Dome, Antarctica. *Nature*, **398**, 121-126.
- Intergovernmental Panel on Climate Change (IPCC), 2001: Climate Change 2001: The Scientific Basis. Contribution of the Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (Houghton, J.T. et al., eds.), Cambridge and New York.
- Johnson, C. and D. Crisp, 2003: The Orbiting Carbon Observatory Mission.
- Jones, C.D. and P.M. Cox, 2001: Modeling the Volcanic Signal in the Atmospheric CO<sub>2</sub> Record. *Global Biogeochemical Cycles*, **15** (20), 453-465.
- Keeling, C.D., R.B. Bacastow, A.E. Bainbridge, C.A. Ekdahl, P.R. Guenther, L.S. Waterman, and J.S. Chin, 1976: Atmospheric Carbon Dioxide Variations at Mauna Loa Observatory, Hawaii. *Tellus*, **28**, 538-551.

- Keeling, C.D., R.B. Bacastow, A.F. Carter, S.C. Piper, T.P. Whorf, M. Heimann, W.G. Mook, and H. Roeloffzen, 1989: A Three-Dimensional Model of Atmospheric CO<sub>2</sub> Transport Based on Observed Winds: I. Analysis of Observational Data. *Aspects of Climate Variability in the Pacific and the Western Americas*, edited by D. Peterson, *Geophys. Monogr. Ser.*, **55**, chap. 15, AGU, Washington, D.C., 165-236.
- Klemp, J.B. and R.B. Wilhelmson, 1978: The Simulation of Three-Dimensional Convective Storm Dynamics. *Journal of the Atmospheric Sciences*, **35**, 1070-1096.
- Keeling, R.F. and S.R. Shertz, 1992: Seasonal and Interannual Variations in Atmospheric Oxygen and Implications for the Global Carbon Cycle. *Nature*, 358, 723.
- Krakauer, N.Y. and J.T. Randerson, 2003: Do Volcanic Eruptions Enhance or Diminish Net Primary Production? Evidence from Tree Rings. *Global Biogeochemical Cycles*, **17** (4), 1118, doi:10.1029/2003GN002076.
- Lagenfelds, R.L., R.J. Francey, B.C. Pak, L.P. Steele, J. Lloyd, C.M. Trudinger, and C.E. Allison, 2002: Interannual Growth Rate Variations of Atmospheric CO<sub>2</sub> and its d<sup>13</sup>C, H<sub>2</sub>, CH<sub>4</sub>, and CO between 1992 and 1999 Linked to Biomass Burning. *Global Biogeochemical Cycles*, **16** (3), 1048, doi:10.1029/2001GB001466.
- Law, B.E. *et al.*, 2002: Environmental Controls over Carbon Dioxide and Water Vapor Exchange of Terrestrial Vegetation. *Agriculture and Forest Meteorology*, **113**, 97-120.
- Leith, H., 1963: The Role of Vegetation in the Carbon Dioxide Content of the Atmosphere. *Journal of Geophysical Research*, **68**, 3887-3898.
- Liang, X.Z., L. Li, K.E. Kunkel, M. Ting, J.X.L. Wang, 2004: Regional Climate Model Simulation of Precipitation During 1982-2002. Part I: Annual Cycle. *Journal of Climate*, **17**, 3510-3529.
- Los, S.O., 1993: Calibration Adjustment of the NOAA AVHRR Normalized Difference Vegetation Index Without Recourse to Component Channel 1 and 2 Data. *International Journal of Remote Sensing*, **14**, 1907-1917.
- Lucht, W., I.C. Prentice, R.B. Myneni, S. Sitch, P. Friedlingstein, W. Cramer, P. Bousquet, W. Buermann, and B. Smith, 2002: Climatic Control of the Greening Trend and Pinatubo Effect. *Science*, **296**, 1687-1689.
- McCormick, M. P., L.W. Thomason, and C.R. Trepte, 1995: Atmospheric Effects of the Mt. Pinatubo Eruption. *Nature*, **373**, 399-404.
- Mellor, G.L. and T. Yamada, 1982: Development of a Turbulence Closure Model for Geophysical Fluid Problems. *Reviews of Geophysics and Space Physics*, **20**, 851-875.
- Meyers, M.P., R.L. Walko, J.Y. Harrington, and W.R. Cotton, 1997: New RAMS Cloud Microphysics Parameterization. Part II: The Two-Moment Scheme. *Atmospheric Research*, **45**, 3-39.

- Molineaux, B. and P. Ineichen, 1996: Impact of Pinatubo Aerosols on the Seasonal Trends of Global, Direct and diffuse Irradiance in Two Northern Mid-Latitude Sites. *Solar Energy*, **58** (1-3), 91-101.
- Nicholls, M.E., R.A. Pielke, and R.N. Meroney, 1993: Large Eddy Simulation of Microburst Winds Flowing Around a Building. *Journal of Wind Engineering and Industrial Aerodynamics*, **46**, 229-237.
- Nicholls, M.E., R.A. Pielke, J.L. Eastman, C.A. Finley, W.A. Lyons, C.J. Tremback, R.L. Walko, and W.R. Cotton, 1995: Applications of the RAMS Numerical Model to Dispersion over Urban Areas. *Wind Climate in Cities*, J.E. Cermak *et al.* Eds., 703-732.
- Nicholls, M.E., A.S. Denning, L. Prihodko, P.I. Vidale, I. Baker, K. Davis, and P. Bakwin, 2003: A Multiple-Scale Simulation of Variations in Atmospheric Carbon Dioxide Using a Coupled Biosphere-Atmospheric Model. *Journal of Geophysical Research Atmospheres*, **109** (D18117), doi:10.1029/2003JD004482.
- Nicholls, M.E., A.S. Denning, P.I. Vidale, I. Baker, K. Davis, and P. Bakwin, 2004: A Multiple Scale Simulation of Atmospheric Carbon Dioxide Using a Coupled Biosphere-Atmospheric Model. *Journal of Geophysical Research*, **109**, D18117, doi:10.1029/2003JD004482.
- O'Brien, D.M. and P.J. Rayner, 2002: Global Observations of the Carbon Budget 2. CO<sub>2</sub> Column from Differential Absorption of Reflected Sunlight in the 1.61  $\mu$ m Band of CO<sub>2</sub>. *Journal of Geophysical Research*, **107** (D18), 4354, doi:10.1029/2001JD000617.
- Pak, B.C. and Prather, M.J., 2001: CO<sub>2</sub> Source Inversions Using Satellite Observations of the Upper Troposphere. *Geophysical Research Letters*, **28**, 4571-4574.
- Palmer, P.I., *et al.*, 2003: Inverting for Emissions of Carbon Monoxide from Asia Using Aircraft Observations over the Western Pacific. *Journal of Geophysical Research*, **108**, 8828-8839.
- Petit, J.R., J. Jouzel, D. Raynaud, N.I. Barkov, J.M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V.M. Kotlyzkov, M. Legrand, V.Y. Lipenkov, C. Lorius, L. Pepin, C. Ritz, E. Saltzman, and M. Stievenard, 1999: Climate and Atmospheric History of the Past 420,000 Years from the Vostok Ice Core, Antarctica. *Nature*, **399**, 429-436.
- Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee and J.H. Copeland, 1992: A Comprehensive Meteorological Modeling System RAMS. *Meteorology and Atmospheric Physics*, **49**, 69-91.
- Potosnak, M.J., S.C. Wofsy, A.S. Denning, T.J. Conway, and D.H. Barnes, 1999: Influence of Biotic Exchange and Combustion Sources on Atmospheric CO<sub>2</sub> Concentrations in New England from Observations at a Forest Flux Tower. *Journal of Geophysical Research*, **104**, 9561-9569.



- Price, D.T. and T.A. Black, 1990: Effects of Short-Term Variation in Weather on Diurnal Canopy CO<sub>2</sub> Flux and Evapotranspiration of a Juvenile Douglas-Fir Stand. *Agriculture and Forest Meteorology*, **50**, 139-158.
- Randall, D.A., D.A. Dazlich, C. Zhang, A.S. Denning, P.J. Sellers, C.J. Tucker, L. Bounoua, J.A. Berry, G.J. Collatz, C.B. Field, S.O. Los, C.O. Justice, and I. Fung, 1996: A Revised Land Surface Parameterization (SiB2) for GCMs. Part III: The Greening of the Colorado State University General Circulation Model. *Journal of Climate*, **9**, 738-763.
- Rayner, P.J. and D.M. O'Brien, 2001: The Utility of Remotely Sensed CO<sub>2</sub> Concentration Data in Surface Source Inversions. *Geophysical Research Letters*, **28** (1), 175-178.
- Rayner, P.J., R.M. Law, D.M. O'Brien, T.M. Butler, and M. Dilley, 2002: Global Observations of the Carbon Budget: III. Initial Assessment of the Impact of Satellite Orbit, Scan Geometry and Cloud on Measuring CO<sub>2</sub> from Space. *Journal of Geophysical Research*, **107**, Article 4557, doi 10.1029/2001JD000618.
- Reichenau, T.G. and G. Esser, 2003: Is Interannual Fluctuation of Atmospheric CO<sub>2</sub> Dominated by Combined Effects of ENSO and Volcanic Aerosols? *Global Biogeochemical Cycles*, **17** (4), 1094-1102.
- Roderick, M.L., G.D. Farquhar, S.L. Berry, and I.R. Noble, 2001: On the Direct Effect of Clouds and Atmospheric Particles on the Productivity and Structure of Vegetation. *Oecologia*, **129**, 21-30.
- Russell, J.L. and J.M. Wallace, 2003: Annual Carbon Dioxide Drawdown and the Northern Annular Mode. *Global Biogeochemical Cycles*, **18**, GB1012, doi:10.1029/2003GB002044.
- Sarmiento, J.L. and N. Gruber, 2002: Sinks for Anthropogenic Carbon. *Physics Today*, **8**, 30-36.
- Sarmiento, J.L., 1993: Atmospheric CO<sub>2</sub> Stalled. *Nature*, **365**, 697-698.
- Schaefer, A.S. Denning, N. Suits, J. Kaduk, I. Baker, S. Los, L. Prihodko, 2002: Effect of Climate on Interannual Variability of Terrestrial CO<sub>2</sub> Fluxes. *Global Biogeochemical Cycles*, **16** (4), 1102, doi:10.1029/2002GB001928.
- Schaefer, K., A.S. Denning, O. Leonard, 2004: The Winter Arctic Oscillation and the Timing of Snowmelt in Europe. *Geophysical Research Letters*, **31**, L22205, doi:10.1029/2004GL021035.
- Schaefer, K., A.S. Denning, O. Leonard, 2005: The Winter Arctic Oscillation, the Timing of Spring, and Carbon Fluxes in the Northern Hemisphere. *Global Biogeochemical Cycles*, in press.
- Sellers, P.J., 1985: Canopy Reflectance, Photosynthesis, and Transpiration. *International Journal of Remote Sensing*, **8**, 1335-1372.
- Sellers, P.J. and Y. Mintz, 1986: A Simple Biosphere Model (SiB) for Use Within General Circulation Models. *Journal of the Atmospheric Sciences*, **43** (6), 505-531.

- Sellers, P.J., J.A. Berry, G.J. Collatz, C.B. Field, and F.G. Hall, 1992: Canopy Reflectance, Photosynthesis and Transpiration. Part III: A Reanalysis Using Enzyme Kinetics – Electron Transport Models of Leaf Physiology. *Remote Sensing of Environment*, **42**, 187-216.
- Sellers, P.J., D.A. Randall, G.J. Collatz, J.A. Berry, C.B. Field, D.A. Dazlich, C. Zhang, G.D. Collelo, and L. L. Bounoua, 1996a: A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part I: Model Formulation. *Journal of Climate*, **9**, 676-705.
- Sellers, P.J., S.O. Los, C.J. Tucker, C.O. Justice, D.A. Dazlich, G.J. Collatz, and D.A. Randall, 1996b: A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part II: The Generation of Global Fields of Terrestrial Biophysical Parameters from Satellite Data. *Journal of Climate*, **9**, 706-737.
- Sellers, P.J., R.E. Dickinson, D.A. Randall, A.K. Betts, F.G. Hall, J.A. Berry, G.J. Collatz, A.S. Denning, H.A. Mooney, C.A. Nobre, N. Sato, C.B. Field, A. Henderson-Sellers, 1997: Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere. *Science*, **275**, 502-509.
- Seth, A. and F. Giorgi, 1998: The Effects of Domain Choice on Summer Precipitation Simulation and Sensitivity in a Regional Climate Model. *Journal of Climate*, **11**, 2698-2712.
- Smagorinsky, J., 1963: General Circulation Experiments with the Primitive Equations. Part I: The Basic Experiment. *Monthly Weather Review*, **91**, 99-164.
- Smith, H.J., H. Fischer, M. Wahlen, D. Mastroianni, and B. Deck, 1999: Dual Modes of the Carbon Cycle Since the Last Glacial Maximum. *Nature*, **400**, 248-250.
- Suntharalingam, P. D.J. Jacob, P.I. Palmer, J.A. Logan, R.M. Yantosca, Y. Xiao, M.J. Evans, D.G. Streets, S.L. Vay, and G.W. Sachse, 2004: Improved Quantification of Chinese Carbon Fluxes using CO<sub>2</sub>/CO Correlations in Asian Outflow. *Journal of Geophysical Research*, **109**, D18S18, doi:10.1029/2003JD004362.
- Tans, P.P., I.Y. Fung, and T. Takahashi, 1990: Observational Constraints on the Global Atmospheric CO<sub>2</sub> Budget. *Science*, **247**, 1431-1438.
- Tennekes, H., 1973: A Model for the Dynamics of the Inversion Above a Convective Boundary Layer. *Journal of the Atmospheric Sciences*, **30**, 558-567.
- Tripoli, G.J. and W.R. Cotton, 1982: The Colorado State University Three-Dimensional Cloud/Mesoscale Model – 1982. Part I: General Theoretical Framework and Sensitivity Experiments. *Journal de Recherches Atmosphériques*, **16**, 185-220.
- Vukicevic, T., B.H. Braswell, and D. Schimel, 2001: A Diagnostic Study of Temperature Controls on Global Terrestrial Carbon Exchange. *Tellus, Ser. B*, **53**, 150-170.
- Walko, R.L., W.R. Cotton, M.P. Meyers, and J.Y. Harrington, 1995: New RAMS Cloud Microphysics Parameterization: Part I. The Single Moment Scheme. *Atmospheric Research*, **38**, 29-62.

Walko, R.L. and Tremback, C.J., 2002: RAMS Version 5.0 Model Input Namelist Parameters.

Wallace, J.M. and D.W.J. Thompson, 2002: Annular Modes and Climate Prediction. *Physics Today*, February, 28-33.

Washenfelter, R., G.C. Toon, J.F. Blavier, Z. Yang, P.O. Wennberg, S.A. Vay, D.M. Matross, 2005: Greenhouse Gas Column Abundances Retrieved from Ground-Based Near-Infrared Solar Spectra. Presented at the 2<sup>nd</sup> International Workshop on Greenhouse Gas Measurements from Space (IWGGMS), Pasadena, CA, March 24-25.

Zhang, Ni, 2002: Observations and Simulations of the Planetary Boundary Layer at a Tall Tower in Northern Wisconsin. Thesis, Colorado State University.