Estimating the Regional Surface Fluxes of Carbon Dioxide Using the Kalman Filter

A Thesis

Presented to

The Academic Faculty

by

Danielle Elizabeth Haas-Laursen

In Partial Fulfillment

of the Requirements for the Degree
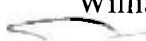
Doctor of Atmospheric Chemistry

Georgia Institute of Technology

September 1997

Estimating the Regional Surface Fluxes of Carbon Dioxide Using the Kalman Filter

Approved:

_____
Dana Hartley, Chairperson

_____
William Chameides

_____
Derek Cunnold

_____
C.S. Kiang

_____
John Mak

Date Approved September 22, 1997

DEDICATION

This is dedicated to my friends and family, who have surrounded me with love and

laughter - especially my parents, sister and husband.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

x

11

SUMMARY

Carbon dioxide is a radiatively important trace gas in the atmosphere, whose concentration has been rising since the industrial revolution. Understanding the sources and sinks that regulate carbon dioxide is crucial for scientists to assess future global change effects. Previous estimates of carbon dioxide reveal that a "missing" sink has been unaccounted for in the northern hemisphere. While various methods have attempted to assess the magnitude of fluxes, the location of the "missing" sink has alluded scientists since first published by Tans et al. (1989).

In an effort to answer this question, an inverse method for deducing time-varying fluxes of trace gases is first tested and developed using a simple model and then in a full three-dimensional model. This method is based on the Kalman filter, which has an added strength over similar techniques in its use of weighting error estimates in the optimization. To fully utilize this potential, studies are also presented here to account for model error and error that arises from taking flask samples at low frequency.

Having found satisfactory results in the testing stages, the method is then applied to invert for carbon dioxide fluxes based on 1992 measurements. The inverse solution indicated that the "missing sink" lies in the northern hemisphere biosphere, with two contributing components. The increased vegetative sink is due in part to an earlier growing season and in part to a decreased fall/winter source in America and Eastern

China.. Recent anecdotal evidence from satellites (Myneni et al., 1997) provides further support for the earlier growing season. With multiple sensitivity studies, this solution is found to be fairly robust.

Observations are sparse in the Southern Hemisphere, and, therefore, the inverse results in these areas are limited. With future growth in developing countries and continued land use changes, understanding the fluxes in these regions will become increasingly important. One of the major recommendations from this work is the need to improve the measurement network in the southern hemisphere.

This methodology can be applied to any number of gases on a variety of time and spatial scales. Future application to improve the budgets of other radiatively important trace gases, like carbon monoxide, methane, and nitrous oxide, is recommended.

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Carbon Dioxide: the Natural Component

Carbon dioxide has been in the earth's atmosphere since early geologic time, and its role as a greenhouse gas has made the earth a more habitable planet. Of the greenhouse gases other than water vapor, carbon dioxide contributes the majority of radiative forcing (64%) (IPCC, 1995). If the atmosphere was devoid of greenhouse gases, the surface of the earth would be approximately 253 Kelvin, well below the freezing point of water (Chameides and Perdue, 1996). At this temperature, surface water would be in a solid rather than liquid state, and life on earth would be extremely different, if it existed at all.

Carbon dioxide is cycled through the atmosphere, ocean, lithosphere, and biosphere. Carbon is stored in several reservoirs on Earth, with the largest storage area in the oceans (38,100 Gt). The interaction of the various reservoirs are depicted in the following cartoon (Figure 1), where the reservoir amount is expressed in Gt and the fluxes between reservoirs are denoted by arrows. Note that there is a large amount of uncertainty in these numbers, particularly the fluxes between reservoirs. The values expressed are taken from Schlessinger (1991) and IPCC (1995) and are meant to represent average fluxes of the present day carbon cycle (1980's). Each of the flux components will be considered in detail in this chapter.

1
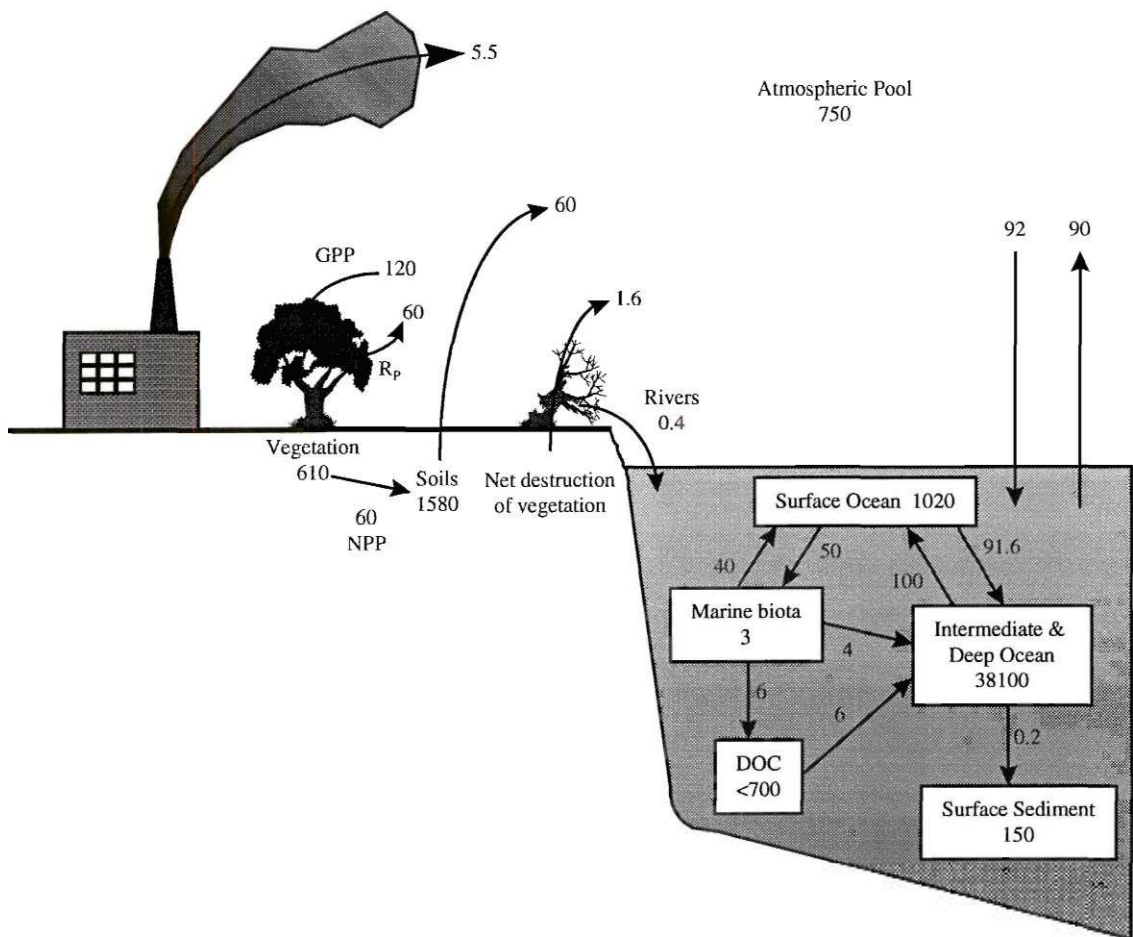
Figure 1: Carbon Cycle in Gt C (adapted from Schlesinger, 1991; IPCC, 1995)

1.1a Vegetative flux

Vegetation, which is defined here as trees, soils, plants, and the micro-organisms living in them, act as sources and sinks of CO_2 depending upon the season and amount of sunlight and rainfall. During the growing season, plants use sunlight as an energy source to drive photosynthesis, which converts CO_2 to oxygen (O_2) and hydrocarbons like

glucose ($C_6H_{12}O_6$). This process is what drives vegetative growth and results in a large sink of atmospheric CO_2 during the growing season. During other times of the year, there is a release of CO_2 due to respiration from organisms in the soils and the decay of dead vegetation. The net effect of vegetation on the CO_2 budget is uncertain, with estimates ranging from a net cancellation between the positive and negative fluxes to a net sink. This uncertainty will be described in more detail in section 1.3.

1.2a Ocean flux

In the latest Intergovernmental Panel on Climate Change (IPCC) report (1995) the net flux from oceans is estimated to be an uptake from the atmosphere of 2.0 ± 0.8 Gt C/yr. Despite the net effect being negative (a sink), there is actually a large amount of variability in ocean flux seasonally, geographically, and interannually. Ocean fluxes change between positive and negative fluxes depending on a number of important regulating physical, chemical and biological factors. These include temperature, wind speed, partial pressure of CO_2, salinity, alkalinity, and pH. These factors are highly non-linear, making the determination flux of CO_2 between the ocean and atmosphere complicated. In the simplest of cases, without additional sources, if the atmosphere and ocean were in complete equilibrium, the amount of CO_2 in each would remain constant. In general, if the atmosphere has greater amounts of CO_2 than the ocean, then CO_2 can be drawn into the surface water. However, the flux between mediums is actually regulated by the temperature, wind speed, pH, and salinity of the ocean, and the flux can move in both directions. Once CO_2 is drawn down into the surface waters, phytoplankton utilize

3

CO_2 (same as vegetation on land), converting it to hydrocarbons during photosynthesis. This process then changes the amount of CO_2 at the surface of the oceans, which in turn increases the difference in the partial pressure of CO_2 between the atmosphere and the ocean (ΔpCO_2). The greater this difference, the more CO_2 can cross between these two mediums. However, there is a different time scale between cycling in the surface waters and the deep ocean. Since the carbon can quickly be re-released via respiration and decay, it is only when it is converted to organic matter and sinks to the deep ocean reservoir that the carbon is a net sink. Despite this large amount of cycling between the atmosphere (~90 Gt C/yr) and the surface waters (~92 Gt C/yr), the annual net is the much smaller balance of these fluxes (~2 Gt C/y)r.

1.2 Carbon Dioxide: Anthropogenic Influence

Despite carbon dioxide's essential and natural abundance on earth, humans have been perturbing its distribution in the environment. Due to the burning of fossil fuels, deforestation, and land use changes, the natural budget of carbon dioxide has had an additional large anthropogenic source. In the 1950's, C. Keeling began taking measurements of CO_2 at Mauna Loa, Hawaii. These measurements show that there has been a significant increase in carbon dioxide (Figure 2). The increase from pre-industrial to present day concentrations is estimated to be 75 ppm or approximately a 20% increase (IPCC, 1995).
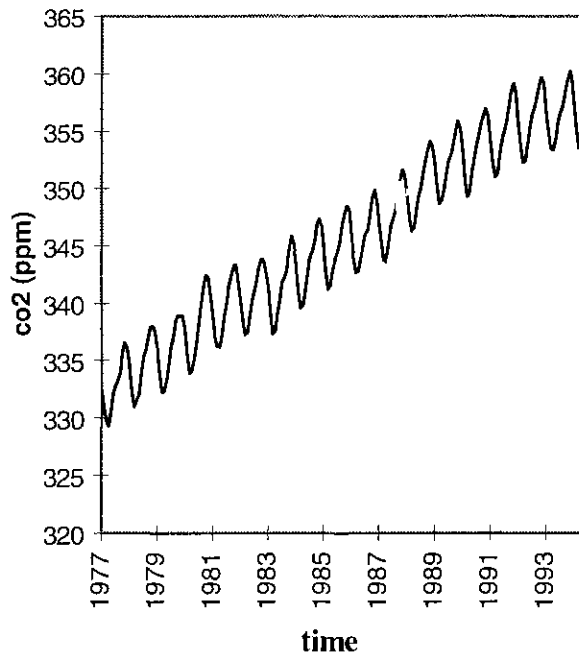
Figure 2: Mauna Loa Observations (adapted from NOAA/CMDL)

1.2a Fossil Fuel Source

The largest anthropogenic source of carbon dioxide to the atmosphere is the burning of fossil fuels, primarily in the form of gas, coal, and oil. This source has increased steadily since the industrial revolution and is currently estimated for the 1980's to be 5.5 ± 0.5 Gt Carbon per year and for 1994 to be 6.1 Gt (IPCC, 1995). This component is the best constrained of all of the sources and sinks of CO_2, partially due to the United Nations Framework Convention on Climate Change, which was drafted just prior to the U.N. Earth Summit in 1992 and went into force in March 1994. One provision of this treaty, which 167 nations have ratified as of May 1997, requires all

nations to estimate all sources and sinks of carbon dioxide within their country and to report these findings to the public.

1.2b Land Use Change Source

Another source of carbon dioxide is also due to vegetation; however, it is separated into its own category due to its anthropogenic origins. Typically called land use change, this source refers to the released carbon from deforestation and biomass burning. In the case of deforestation, carbon dioxide is released when the trees are clear-cut and when the debris left behind decays. Furthermore, this has a long lasting effect since the vegetation is no longer there to uptake CO_2. Biomass burning also releases carbon dioxide. The land use change source is estimated for the 1980's to be 1.6 ± 1.0 Gt C/yr. This value is interpreted from land clearing rates, biomass inventories, satellite data, and modeling (IPCC, 1995).

Due to this rapid increase in carbon dioxide and its climate forcing potential, knowing the regional carbon dioxide sources and sinks is essential to understanding and predicting global change. Without knowing exactly where the sources and sinks are, effective means to control atmospheric concentrations cannot be developed. However, as shown in Table 1, there is a large amount of uncertainty in the current best estimates of surface fluxes for CO_2. When these fluxes are summed together and atmospheric increase is taken into account, the net value should be approximately zero. Instead, there is an imbalance, or implied sink, of, 1.3 ± 1.5 Gt C /year. In order to determine more

information about this sink, Tans et al. (1989) ran an atmospheric model with these surface fluxes and determined that the additional sink is needed in the northern hemisphere to explain the discrepancy between model output and observed measurements (Tans et al., 1989).

Table 1: Estimates of net CO_2 Fluxes to the Amosphere for the 1980's (IPCC, 1995)

Sources	Gt of C per year
fossil fuels	5.5 ± 0.5
land use change	1.6 ± 1.0
Sinks	**Gt of C per year**
oceans	2.0 ± 0.8
northern hemisphere vegetation	0.5 ± 0.5
Atmospheric increase	3.3 ± 0.2
Implied "Missing" Sink	**Gt of C per year**
??	1.3 ± 1.5

Much debate exists in the scientific community trying to explain this so termed "missing sink" of carbon dioxide, with discussion focusing on the relative roles of terrestrial and oceanic sinks.

1.3 Methods to Determine CO_2 Fluxes

Various techniques have been used in an attempt to quantify the carbon dioxide fluxes and settle the debate over the location of the "missing" sink. One method is to make use of isotopic measurements of carbon to differentiate between vegetative and

oceanic uptake of anthropogenic carbon dioxide. Since fossil fuel and biomass burning have a low $^{13}C/^{12}C$ ratio compared to the background atmosphere, isotopic measurements can be used to partition the amount of anthropogenic carbon in the land, ocean, and atmosphere. Using this technique, Quay et al. (1992) has estimated that the oceans uptake 2.1 ± 0.8 Gt C /yr and the vegetative component is approximately a net zero.

Another method to estimate the ocean flux is to make use of measurements of the partial pressure of carbon dioxide between the atmosphere and ocean. These measurements have been collected by ship cruises throughout the world's oceans. The ocean flux (F) can be constrained by an exchange coefficient (k) and the average estimated magnitude of the difference between atmospheric and oceanic partial pressure of carbon dioxide (ΔpCO_2), using the following equation:

$$F = k \, \Delta pCO_2$$

Using this technique, various estimates of ocean uptake have been made ranging from -0.5 ± 0.3 Gt C/yr by Tans et al., (1990) to -2.0 ± 0.6 Gt C/year by Siegenthaler and Sarmiento (1993). There are several uncertainties that exist with this method of calculation. First, there are discrepancies between the exchange coefficient calculated from the bomb experiments and laboratory calculations (Takahashi et al., 1992; Siegenthaler and Sarmiento, 1993). There is even a range of estimates caused by different interpretations of the bomb inventory, with a recent study finding that the flux calculated from the radiocarbon measurements is overestimated by 25% (Hesshaimer et al., 1994). Second, the measurements of ΔpCO_2 are sparse (Siegenthaler and Sarmiento,

8

1993) and have been found to have extreme variations over relatively small horizontal distances (Watson et al., 1991). In a cruise of the North Atlantic, Watson et al. (1991) found variations of 10 μatm over 100 km distance. To put this into perspective, an error of 1 μatm in the annual ΔpCO_2 results in an uncertainty of 0.2 Gt C /yr (Watson et al., 1991); thus, a range of 10 μatm results in an uncertainty of 2.0 Gt C/yr. This large amount of variability has significant impact on the uncertainty associated with estimating fluxes using an average ΔpCO_2. Table 2 summarizes the range of estimates for ocean uptake from the various techniques.

Table 2: Ocean Flux Estimates Using Various Methods

Method	Source	Estimate (Gt C /yr)
^{13}C ratio	Quay et al.., 1992	-2.1 ± 0.8
flux equation	Tans et al., 1990	-0.5 ± 0.5
flux equation	Siegenthaler and Sarmiento, 1993	-2.0 ± 0.6

In addition to the previously described method using isotope measurements, there are also other methods to estimate the vegetative flux. The most common method relies on forest inventories and estimated flux rates, used in conjunction with a model (Houghton, 1993). In attempting to identify the northern hemispheric "missing" carbon dioxide sink, a wide range of estimates of the vegetative fluxes have been proposed. Two general camps can be identified: (1) those that believe there is a net zero flux (Melillo et

al., 1988; Houghton et al., 1987; 1993) and (2) those that estimate an uptake ranging from 1 - 2 Gt C /yr (Kauppi et al., 1992; Sedjo, 1992; Dai and Fung, 1993).

The general methodology is to estimate an accumulation of regrowth for an area of the northern hemisphere forest and then extrapolate to the larger regions based on forest inventories. These methods rely on biological models of net primary productivity, which have uncertainty associated with them. The types of models, the number of feedbacks, and the processes considered vary among the studies, accounting for the wide range of estimates (Taylor, 1993). Another technique employed by Tans et al. (1990) was to infer a vegetative sink by trying to match different scenarios with observations. Using various scenarios to try and improve agreement between model output to observations, Tans et al. (1990) estimate the vegetative uptake to be between 1.5-2.0 Gt C/yr. Table 3 summarizes the range of estimates for vegetative uptake in the Northern Hemisphere forests from the various techniques.

Table 3: Northern Hemisphere Vegetative Flux Estimates Using Various Methods

Method	Source	Estimate (Gt C /yr)
isotope ratio	Quay et al.., 1992	0.0
extrapolation / NPP	Houghton, 1987; 1993; Melillio et al., 1988	0.0
extrapolation / NPP	Kauppi et al., 1992	-1.0 ± 0.2
extrapolation / NPP	Sedjo, 1992	$-0.7 \pm ?$
infered sink	Tans et al., 1990	-1.7 ± 0.3

1.4 Inverse Methods to Estimate CO_2 Fluxes

Another approach to determine the poorly defined budgets of trace gases is to use an inverse method. This involves using a model of the atmosphere to predict atmospheric concentrations given an initial estimate of surface fluxes. By running a model with the "best" estimates of fluxes and then comparing the predicted model output to observed data, one can tell whether or not the initial flux values are good estimates by how well the resulting output matches observations. It is at this point that the inverse method is employed. By minimizing the difference between observations and model output, the inverse technique optimizes the solution to determine the best fit of fluxes. Under appropriate conditions, this inversion can produce an optimized and unique solution.

Previously, the Kalman filter has been applied to estimate fluxes for nitrous oxide and CFC-11 (Cunnold et al., 1983) using a 12 box model and CFC-11 (Hartley and Prinn, 1993) using a full three-dimensional general circulation model. The Kalman filter has been applied to various geophysical problems including wind assimilation (Daley and Menard, 1993; Gauthier et al., 1993) and ocean data assimilation (Ghil, 1989; Bennett and Budgell, 1987). However, until this thesis, the Kalman filter has not been used for inferring the fluxes of the carbon dioxide.

Studies using other inverse methods (Green's Function, Singular Value Decomposition) have been used to estimate the fluxes of carbon dioxide (Tans et al., 1989; Enting and Mansbridge, 1989;1991; Ciais et al., 1995) and methane (Brown, 1993). In all of these previous studies, only two-dimensional models have been used. While 2-D models are computationally less demanding, they average around a latitude circle and therefore, provide no insight into longitudinal variation. In the case of carbon dioxide, major sources are located in the industrialized northern middle latitudes, complicated by potential sinks (ocean and vegetative) all at the same latitude. A two dimensional model is not capable of distinguishing the emissions from different industrialized regions such as Europe, East Asia, and the United States and their surrounding oceans. Thus, there is no hope of resolving the issue of the "missing" sink using only two dimensions because it is impossible to differentiate between land and ocean. Therefore, a three dimensional model is essential to determine the net fluxes of these gases from such regions. This thesis will use a three-dimensional model.

The inverse with the 3-D model has been thoroughly tested with CFC-11 including pseudo data studies and using time series of actual observations (Hartley 1992; Hartley and Prinn, 1993). However, there is an important issue that must be addressed when moving from inverse studies with CFC-11 to gases like carbon dioxide and methane. CFC-11 has reasonably constant annual emissions, while both carbon dioxide and methane have seasonally varying sources and sinks (Thompson et al., 1986; Khalil et al., 1990). These adaptations will be addressed in detail in the methodology section (Chapter 3.1).

Another limitation that previous inverse studies (Enting and Mansbridge, 1989;1991; Brown, 1993) have found is an artifact termed "error amplification." Clearly, there is error associated with many components of the system (i.e. measurements, model, initial flux estimates). In order to account for error in the system, one method employed is to truncate the spatial and time domains, effectively smoothing the data (Enting an Mansbridge, 1989; 1991; Brown, 1993). Even with this step, these studies typically yield significant error amplification (Enting an Mansbridge, 1989; 1991; Brown, 1993).

Rather than using truncation to control error, the Kalman filter actually weights the various errors in determining the optimum solution; thus, it should be a more stable optimal estimator for the purpose of estimating the fluxes of carbon dioxide. This ability to adjust for various noises and uncertainties is a strength of the Kalman filter technique. The Kalman filter will be described in detail in Section 3.1 and 3.2, and error estimates for the filter are determined and described in Sections 3.3.

When utilizing an inverse method to deduce flux estimates, there are several factors that can complicate the solution. Among these is the lifetime of the gas, which must be longer than the transport time to a measurement site. Carbon dioxide is long lived; thus, one does not have to worry about transport times for capturing the fluxes. Another complicating factor is chemistry. If the gas reacts chemically in the atmosphere, then the gases that are involved in the reaction must also be well represented in the model. This also is not an issue for carbon dioxide since it is essentially inert in the atmosphere. Despite being broken down in the terrestrial biosphere and the oceans (Sections 2.1a and 2.1b), in the gaseous phase in the atmosphere, carbon dioxide is extremely stable. There is only a small $in\ situ$ sink for CO_2, which is photolysis to CO and $O(^3P)$ in the stratosphere and mesosphere (Warneck, 1988). Other than this small chemical sink, there are no $in\ situ$ sinks for CO_2 in the atmosphere. Therefore, carbon dioxide does not have chemistry complicating the relationship between sources and sinks within the domain of the atmosphere. While these issues do not affect the inverse for carbon dioxide, they should be kept in mind when applying the methodology to other gases in the future.

As an outline of the dissertation, each section is described briefly here. The methodology is divided into two main sections: Chapter 2 and Chapter 3. Chapter 2 explains the Kalman filter and addresses the issue of how to adapt the optimization technique to deduce time varying fluxes. This development is all done using a simplified test model. Chapter 3 uses a full three-dimensional model and characterizes various

errors to augment the Kalman filter. This includes assessing model error and error that arises from comparing flask samples to model output. Chapter 4 presents the inverse results. The first section includes tests of the developed methodology using pseudo data in the three-dimensional model. This section includes exploring how to best group regions together for the inverse, as well as determining the optimum year to invert for given the current measurement network. Having found satisfactory test results in this study, we extend the inverse methods to real observations in an attempt to deduce the regional CO_2 fluxes in the atmosphere. These results and discussion are presented in section 4.2. A series of sensitivity studies explores how robust the solution is in section 4.3. Finally, conclusions and recommendations appear in Chapter 5.

CHAPTER 2

METHODOLOGY I: KALMAN FILTER ADAPTATIONS

2.1 Inverse Method to Deduce Fluxes

Given the importance and current uncertainty in the various fluxes, this thesis is an attempt to improve our knowledge of these fluxes of CO_2 utilizing an optimal linear estimator, known as an inverse method. Inverse methods have been employed in numerous fields to solve a host of problems. As a working definition of optimal estimators, the following is from Gelb (1994):

> *An optimal estimator is a computational algorithm that processes measurements to deduce a minimum error (in accordance with some stated criterion of optimality) estimate of the state of a system by utilizing: knowledge of system and measurement dynamics, assumed statistics of system noises and measurement errors, and initial condition information.*

These methods are useful whenever one has noisy measurements of a system and a model of the system that can relate the state of the system to its measurements. The relationship between the system, its measurements, the model and the optimal filter can be visualized with the cartoon in Figure 3. In the figure, boxes represent the algorithms (or models)

16

and the circles are inputs and outputs of the algorithms (or models). The diagram is
labeled in general terms with a case specific example provided in parenthesis.
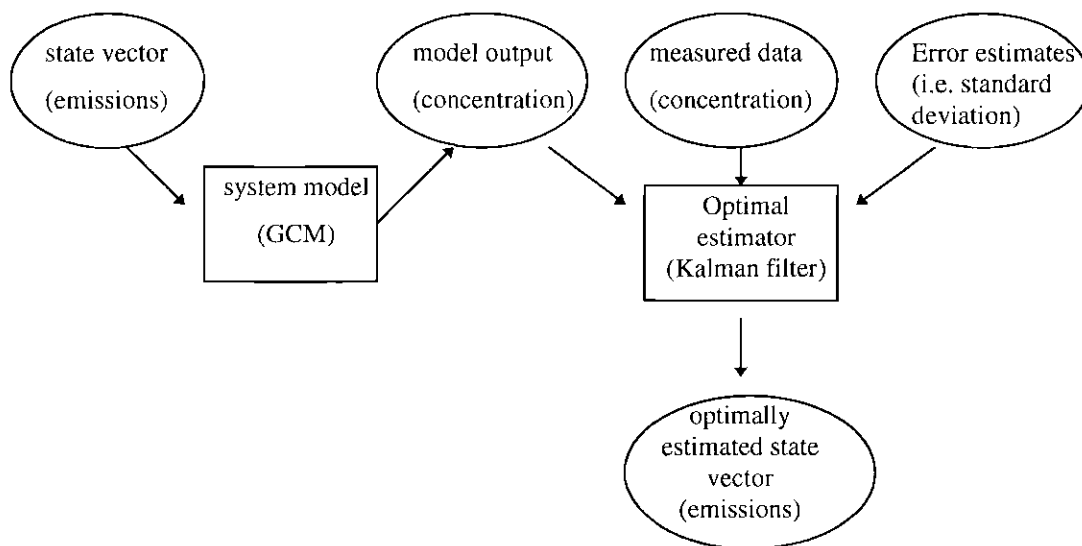
Figure 3: Diagram of Relationships to an Optimal Estimator

The optimal estimator used in this work was first developed in 1960 by R.E.
Kalman and is so called the Kalman filter. It is the most commonly used of the optimal
estimators (Gelb, 1994), having been used in fields as varied as missile guidance systems,
robotics, digital signal processing, and biological studies of yeast fermentation. One
advantage of the Kalman filter over other inverse methods is that it includes error
weighting functions within the filter. Thus, if a particular measurement has a large error,
it will be given less weight than a measurement with very little error. Likewise, if the

17

initial estimate of the state vector is well known to begin with, the amount of freedom to change the value in the solution will be limited; whereas, if the state vector is poorly constrained initially, then the filter will allow large changes to the estimate.

The Kalman filter can be derived as the minimization of a cost function, J. Several assumptions go into the posing of the cost function. First, a linear relationship, P, exists that relates emissions (E) and observations (χ) described as χ^{obs} = PE. Second, the estimate of emissions differs from the true emissions state, $\hat{E}$, by some value, ΔE, and this difference is defined as $\Delta E = E_{t+1}^{est} - E_t^{est}$. Finally, the amount of error in the emissions estimate can be described by an error estimate $\sigma_{E,t}$, where the relationship is

$$E_t^{est} = \hat{E}_t + \sigma_{E,t}.$$

With these assumptions in mind, the cost function is posed to consist of two components. One is the difference in observations and the optimum model predicted concentrations, $\chi^{obs} - P\hat{E}_t$. The other component is the size of the step change in emissions (ΔE). As described previously, the strength of the Kalman filter is that error estimates are contained within the actual filter. Thus, each of these components is weighted by associated error estimates. The covariance matrix (C) estimates the constraints on the initial guess, E, while the noise matrix (N) estimates the amount of error contained in the observations, χ^{obs}. The cost function is given in the following equation:

$$\mathbf{J} = (\chi_t^{obs} - \mathbf{P}_t \hat{\mathbf{E}}_t)^T \mathbf{N}_t^{-1}(\chi_t^{obs} - \mathbf{P}_t \hat{\mathbf{E}}_t) + (\Delta \mathbf{E})^T \mathbf{C}_t^{-1}(\Delta \mathbf{E})$$

<div align="right">Equation 1</div>

A derivation of the solution can be found in Appendix A; however, the solution itself is shown in the following equation:

$$\Delta \mathbf{E}_t = (\mathbf{C}_t \mathbf{P}_t^T [\mathbf{P}_t \mathbf{C}_t \mathbf{P}_t^T + \mathbf{N}_t]^{-1})(\chi_t^{obs} - \chi_t^{model})$$

<div align="right">Equation 2</div>

A description of each of the components in the solution and how the Kalman filter is applied in practice is included here. Initial guesses of emissions are taken to begin the estimation procedure. These emissions are used to calculate the model mixing ratio (χ^{model}) at the first time step. This χ^{model} is compared to observations in order to update the emissions using the following equation:

$$\mathbf{E}_2 = \mathbf{E}_1 + \mathbf{G}_1(\chi_1^{obs} - \chi_1^{model})$$

<div align="right">Equation 3</div>

Note that equation 3 is simplified but identical to equation 2. This sequence of steps is then repeated to calculate emissions at each subsequent time step.

G_t is called the gain matrix and it is defined as follows:

$$\mathbf{G}_t = \mathbf{C}_t \mathbf{P}_t^T [\mathbf{P}_t \mathbf{C}_t \mathbf{P}_t^T + \mathbf{N}_t]^{-1}$$

<div align="right">Equation 4</div>

P_t is the partial derivative matrix,

$$\mathbf{P}_t = \mathbf{P}_{ikt} = \frac{\partial \chi_{it}^{model}}{\partial \mathbf{E}_{kt}} \ .$$

<div align="right">Equation 5</div>

P_t^T is the transpose of the P_t matrix.

The covariance matrix, C_t, is initialized with diagonal elements that contain the initial estimate the variance in emissions,

$$\mathbf{C}_t = [\sigma_{k,t}^E \sigma_{k,t}^E] \ .$$

<div align="right">Equation 6</div>

The covariance matrix is updated throughout the inversion by the following equation:

$$C_{t+1} = C_t - G_t P_t C_t \qquad \text{Equation 7}$$

Finally, N_t is the noise matrix, which is also diagonal and defined as follows:

$$N_t = [N_{ijt}] = [\sigma_{it}^{obs} \sigma_{it}^{obs}]. \qquad \text{Equation 8}$$

2.2 Adapting the Kalman Filter for Time Varying Fluxes

The purpose of this initial test case is not an attempt to quantify emissions, but rather to assess how to apply the Kalman filter as in Hartley and Prinn (1993) to time varying sources. In this section, various adaptations that exist in Kalman filter theory will be explored in an attempt to identify the optimum method for deducing time varying regional surface sources/sinks. It is not practical to run numerous test cases in a complex model. In order to be able to clearly identify the components of the inverse problem and afford to run a wide variety of tests, a simplified analytical regional scale model is employed.

2.2a Simplified Model Description

For the transport model, we use the Effectively Sampled Region (ESR) model (Prinn, 1988), which provides a steady state solution to the 2-D coupled advection, diffusion and chemistry problem in cylindrical coordinates. The relative placement in the latitude/longitude plane of the assumed source regions and observation sites is shown in Figure 4. Note that the relative placement of the sources and observations can be

20

conceptualized as either a global scale or urban scale situation. The time scale would also change depending on the geographic scale. For example, in global problems, the model and observations are compared on monthly time scales; whereas, in urban studies, the time scale is hourly. The wind barbs drawn are for winter (easterly), summer (south-easterly), and all/spring (east-south easterly). The magnitudes are depicted by the relative size of the wind barb.
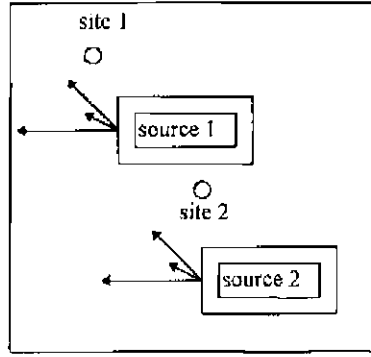
Figure 4: Map of fictitious sites.

The equation for the trace gas concentration in the ESR is as follows:

$$\chi_t = \frac{E_t}{\pi h v_t l_t} \exp\left(\frac{r\cos\phi_t}{l_t}\right) k_0\left(\frac{\varepsilon_t \, r}{l_t}\right)$$

21

where $l_t = \dfrac{2\tau \sigma_t{}^2}{v_t}$ and $\varepsilon_t = \left[\dfrac{1 + 2l_t}{v_t t_{net}} \right]^{\frac{1}{2}}$

χ_t = concentration (mass/volume)

E_t = emissions (mass/time)

r = radial distance of observations from the source (length)

$\sigma_t{}^2$ = variance of horizontal wind speed (length^2/time^2)

v_t = average horizontal wind speed through region (length/time)

h = boundary layer height (length)

t_{net} = the chemical lifetime (time)

ϕ_t = wind direction relative from the source to the observation site (radians)

τ = time scale of mixing (time)

k_0 = Bessel function of the zero order

This transport model is run using a given set of emissions at each source, and χ_t is calculated for the time over which the emissions are valid. In lieu of real observations of concentrations, this ESR generated χ field is used as "pseudo" data for our region, as previously described by Hartley and Prinn (1993). This is a crutial testing step in any inverse, as one must first be able to invert for a known solution before any credence may be given to a previously unknown solution.

For the inversion process, we alter the winds in the model from those used to calculate the "pseudo" data. In this way, we are dealing with a more realistic modeling situation, where the model is imperfect. We can compare runs using the same emissions and these two different wind fields to calculate a transport residual, δ, defined by the equation, $\delta = \chi_{\text{perfect}}^{\text{model}} - \chi_{\text{imperfect}}^{\text{model}}$. This δ will be used in the Kalman filter equations, by adding it to the noise matrix.

The noise matrix contains an estimate of the error associated with observations. This error is a measure of both instrument precision and natural variability at an observation site (Hartley and Prinn, 1993; Chang et al., 1996). In this test study, we are working with pseudo-data based on a simplified model, so we define N_t to be greater than $P_t C_t P_t$, as is the case with previous 3D model studies (Hartley and Prinn, 1993). We then augment the noise matrix with the transport residual. Equation 7 is modified as follows (Hartley and Prinn, 1993):

$$N_t = [N_{ij}] = [\sigma_{it}^{\text{obs}} \sigma_{jt}^{\text{obs}}] + [\delta_{it}^{\text{obs}} \delta_{jt}^{\text{obs}}] \qquad \text{Equation 9}$$

Thus, the noise matrix now accounts for model error in the comparison of model calculated concentration and "pseudo" observations, as well as the typical error associated with the measurements.

Furthermore, we account for the propagation of incorrect emissions in global models by augmenting the ESR during the inversion. To do this, we carry over a

percentage of the difference between model and "pseudo" observations through time,

with the following equation: $\chi_{t,history}^{model} = \chi_t^{model} + 0.05\sum_{i=1}^{t}(\chi_i^{obs} - \chi_i^{model})$. Thus, a poor

emissions guess is propagated temporally. This error propagation is evident in the model

results, which show increasing error as time progresses.

2.2b Adaptations to Kalman Filter

When we apply the Kalman filter in the ESR model to solve for constant

emissions over time, the filter is consistently able to converge to the "correct" emissions

(Figure 5). Note that the constant emissions case s the situation found in Hartley and

Prinn (1993) when inverting for CFC-11.

24
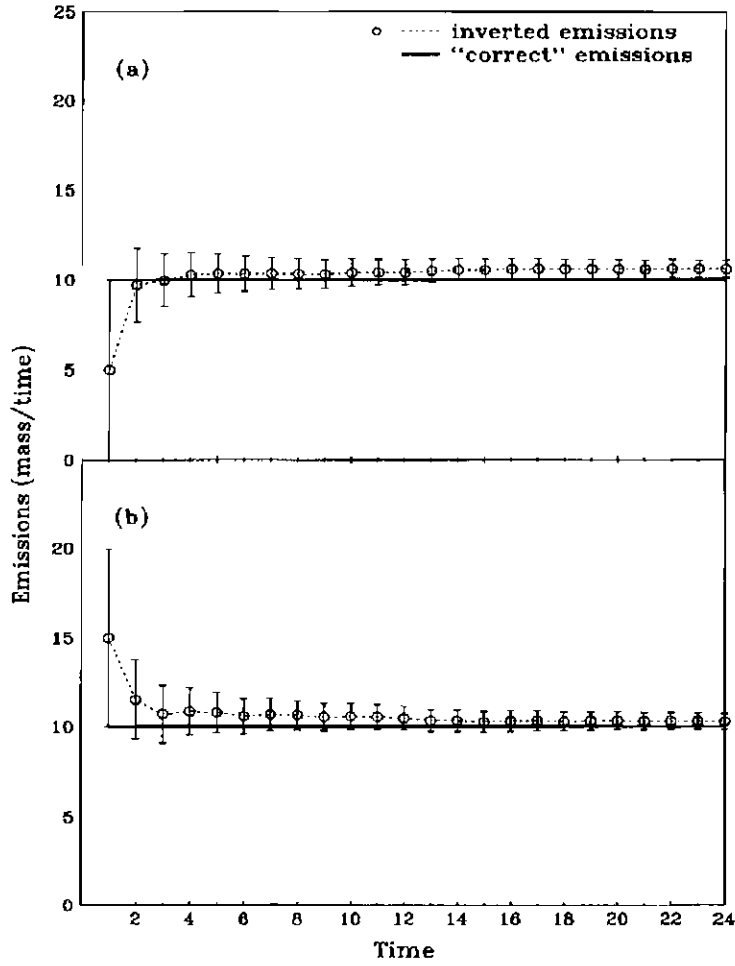
Figure 5: Kalman filter for constant fluxes, where (a) is source 1 and (b) is source 2.

However, when we apply the same Kalman filter in the ESR to invert for time varying emissions, the filter can no longer converge on the "correct" emissions (Figure 6).

25
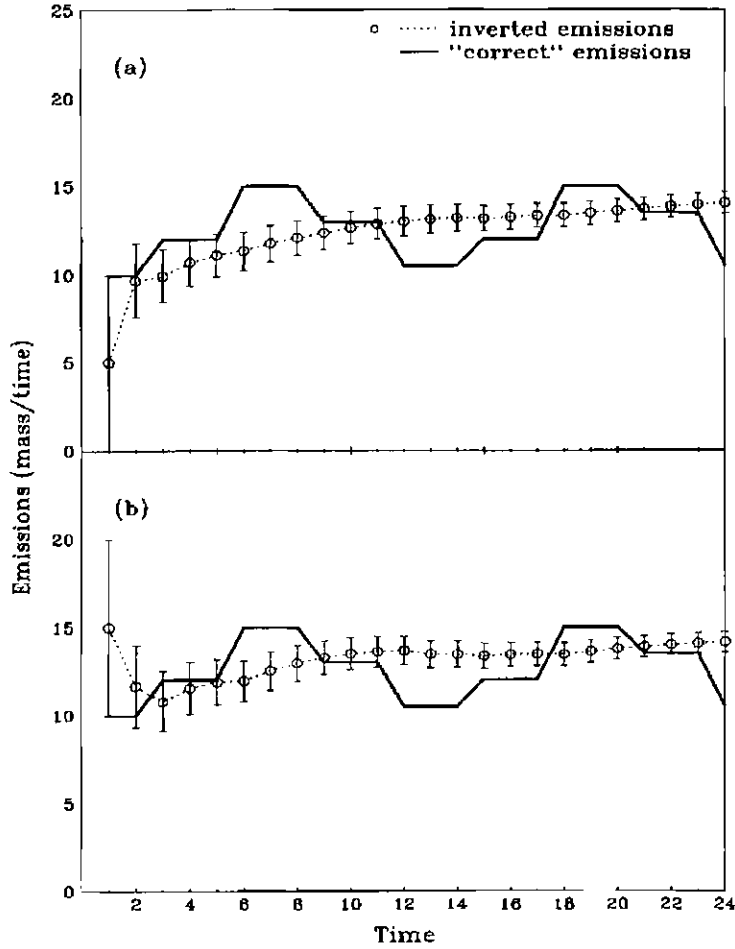
Figure 6: Kalman Filter for Time Varying Fluxes

One method used to invert for time varying emissions that has been applied in 2-D models solves for the coefficients of a function that includes simple harmonics to describe the seasonal cycle. Forms of this function include $E = (A\cos(\omega t) + B\sin(\omega t))$, where $2\pi/\omega$ is the period and A and B are unknown constants that represent the magnitude and phase of the seasonal cycle (Surendran and Mulholland, 1986). Other

representations for the seasonal cycle have also been used (Enting and Newsam, 1990; Brown, 1993) for gases like CO_2 and CH_4.

In doing this type of multi-variable inversion, two issues arise. The first is that the inversion must solve for two or more terms (e.g. A, B) for each source region. This large number of unknowns will generally require a large number of independent observation sites in order to arrive at a unique solution. Since the networks of observation sites for essentially all trace gases are very sparse spatially, this is not presently a useful choice in 3-D models, as it will result in an ill-posed problem. Furthermore, the period (i.e. ω) may also be a function of time, adding more unknowns and exacerbating the ill-posed problem. If the period is not solved for separately, the standard Kalman filter cannot deduce the change in period. This is further extended by the fact that the seasonal cycle may not be a simple harmonic. While not an issue for carbon dioxide, this would become a very important issue for a gas like methane. In the case for methane, each source has its own distinct annual cycle. For instance, in the Tae-Ahn Peninsula (TAP), Korea, the rice paddy emissions have a peak during the summer months, while wetlands and tundra emissions peak in October, and biomass burning peaks in May (Dlugokencky et al., 1993). Since these cycles of differing amplitudes and phases overlap throughout the year, it is unrealistic to fit with simple harmonics. Due to these complexities in the sources and in the interest of developing a robust method that can be extended beyond carbon dioxide, we chose to be more general and make no *a priori* time dependent assumptions about the seasonal cycle.

Another way to make the estimated emissions approach the "correct" solution is to use a transition matrix, Φ, that contains the seasonal cycle. Differing from the previous method that inverts for the coefficients in the seasonal cycle, a transition matrix simply prescribes a time varying function, that updates emissions as $E_{t+1} = \Phi_t E_t$ (Gelb, 1994). The transition matrix is also added to the covariance matrix, as $C_{t+1}^- = \Phi_t C_t \Phi_t^T$, where the minus sign indicates that the step is performed prior to the update in equation 6. Utilizing a transition matrix would be ideal if he seasonal cycle of the gases were known; however, the reality is that we do not positively know the seasonal cycle, as the goal of this work is to invert for gases with unknown emissions. When the seasonal cycle is unknown, utilizing a transition matrix will bias the outcome. For instance, if we prescribed an incorrect seasonal cycle, $\Phi = \cos(\omega t)$, in our filter equations, the inverted emissions are shown in Figure 7.
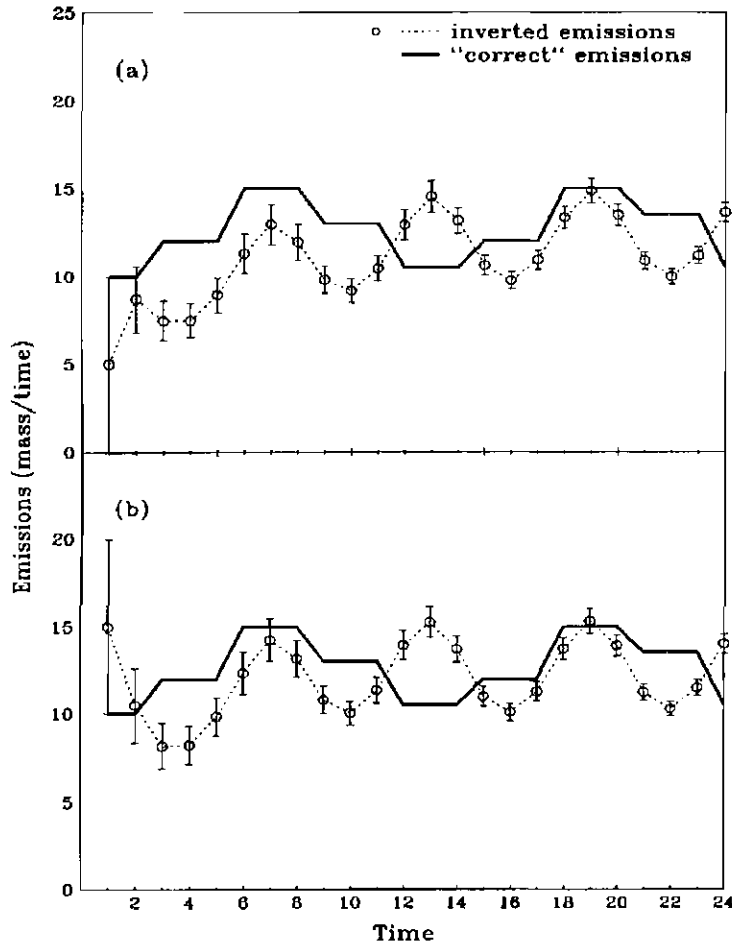
Figure 7: As in Figure 6, with transition matrix (ϕ) defined

We intentionally chose a poor, but reasonable, choice for the cycle, in order to show the type of error that can occur when the seasonal cycle is not prescribed correctly. Thus, we have only succeeded in adding an incorrect bias to the inverted emissions. In Kalman filter theory, the distinction in approach is that we will be using the Kalman filter to estimate emissions across measurements, rather than between measurements, as is done when using a transition matrix (Gelb, 1994).

However, we can take a closer look at the reason for the lack of convergence in Figure 6 by examining the Kalman filter equations. In equation (2), emissions are adjusted based on the product of the gain matrix and the residual. Initially, the covariance matrix is large, which results in a large gain matrix (equation 3). In time, the covariance matrix continually decreases in magnitude (equation 6), and the gain matrix decreases accordingly. Therefore, once the model predictions are in good agreement with the pseudo-data, both the residual and the gain matrix have greatly decreased. For time varying emissions, however, the residual will increase in magnitude when the emissions change, while the gain matrix will remain small (according to equation 3), and the overall desired adjustment in emissions cannot be attained.

One solution to the problems caused by the diminishing gain matrix is to utilize an adaptive filter (Sastri, 1985). The adaptive filter works by resetting the covariance matrix when the model calculated mixing ratio begins to diverge from the observations. Upon detection of an increase in the magnitude of the residual ($\chi^{obs}-\chi^{model}$) from the previous time step, C_t is reset to its initial value. This in turn allows G to be larger and the needed adjustment in emissions can be made. Applied to the same pseudo data used in Figure 6, the adaptive filter improves the emissions estimate (Figure 8).
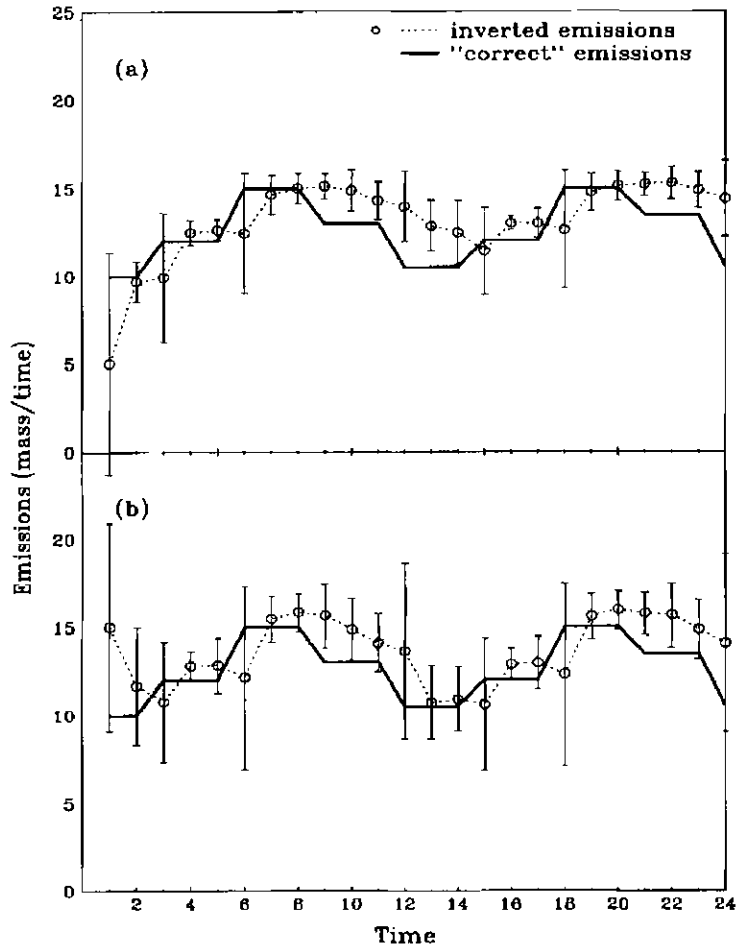
Figure 8: As in Figure 5, with an adaptive kalman filter.

However, there is an ensuing time lag. This time lag is more evident at the end, as the effect of "history" creates a deteriorating emissions estimate in time.

Note that in the adaptive filter the C_t matrix is enlarged every time the residual diverges. Therefore, the standard deviations for the estimated emissions are not truly the square roots of the diagonal elements of the covariance matrix. Thus, the quoted error we

31

use for the next set of runs in this section is calculated as $\delta E_t = P_t^{-1} \delta \chi_t$. Finally, the definition of "time lag" used in this thesis differs from the traditional usage in atmospheric science and as arises here. In atmospheric sciences, a time lag is thought of as a lag between emissions and detection at an observation site (i.e. a northern hemispheric source effecting a southern hemispheric observation site). This issue has previously been explored in 3D inverse studies with a tracer (Hartley, 1992; Hartley and Prinn, 1993) and found to be of second order importance with the current observational network. In this thesis, a time lag results from an estimation time lag in the filter (i.e. when there is divergence in the residual and the filter cannot adjust until after this divergence is detected one time step later).

One method used in Kalman filter analysis to deal with this time lag is an optimal smoother (Gelb, 1994). The Rauch-Tung-Striebel fixed-interval optimal smoother (Gelb, 1994) adjusts the emissions at the previous time step, where: $\tilde{E}_{t-1} = E_t + (C_{t-1} \Phi_{t-1}^T C_{t-1}^{-1})(E_t - E_{t-1})$. The tilde indicates a new value for emissions. In our calculations, Φ, the "transition" matrix, is an identity matrix as discussed previously. Thus, our equation simplifies to $\tilde{E}_{t-1} = 2E_t - E_{t-1}$. However, the smoother does not correct for the effect of this time lag on the computed χ^{model} in past time steps. Although an improvement (Figure 9) the smoother is not an ideal time lag correction because it does not correct errors in the model predictions. Thus, it still misses the correct solution as a result of "history" being propagated through the model concentrations.
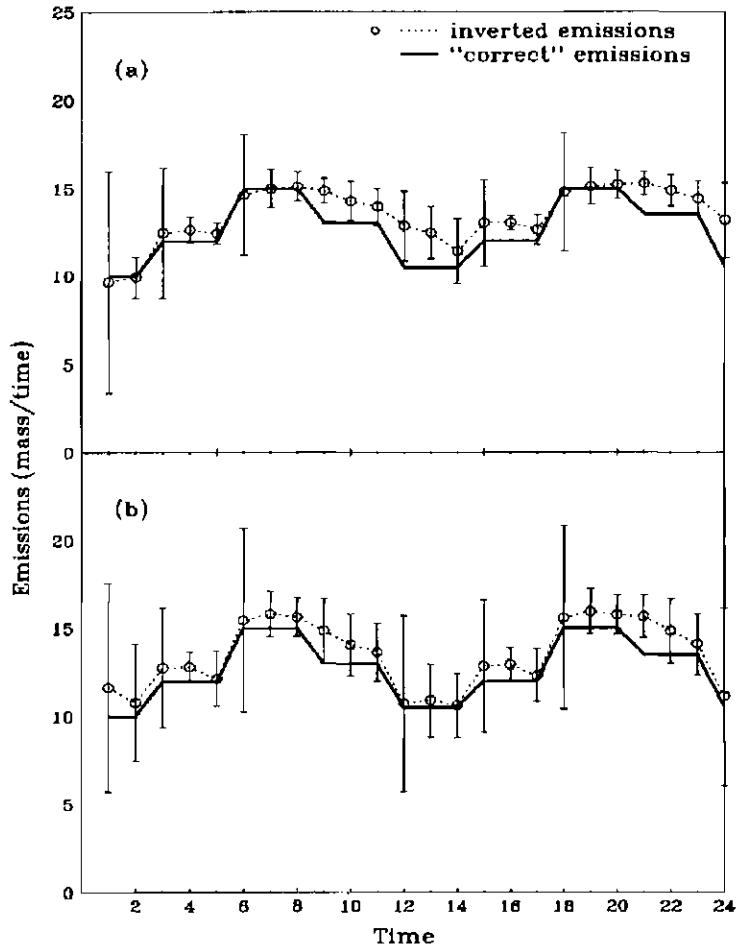
Figure 9: As in Figure 6, with a Smoother added to the Kalman filter

Applications of other time lag corrections also proved untenable. One such method, the forward-backward filter, required that all emission values be known prior to applying the correction (Gelb, 1994). This would not work in the case of determining unknown emission strengths. Another method, which uses a short period estimate, requires using a small inversion time step (Gutman and Velger, 1988). This is not always a practical option when, due to inadequate simulation of short-term transport variations,

model values need to be averaged in order to compare well with observed values.

Finally, we investigate applying an adaptive-iterative method. This is the recursive-iterative method (Young, 1984) used with the adaptive filtering method (Sastri, 1985). We iterate until the change in emissions from one iteration to another is sufficiently small. This approach has previously been used in another application of the Kalman filter for yeast fermentation (Bellgardt et al., 1986) and is considered to be a robust method for using inverse methods to solve for time varying state vectors (Young, 1984). The results are shown in Figure 10.
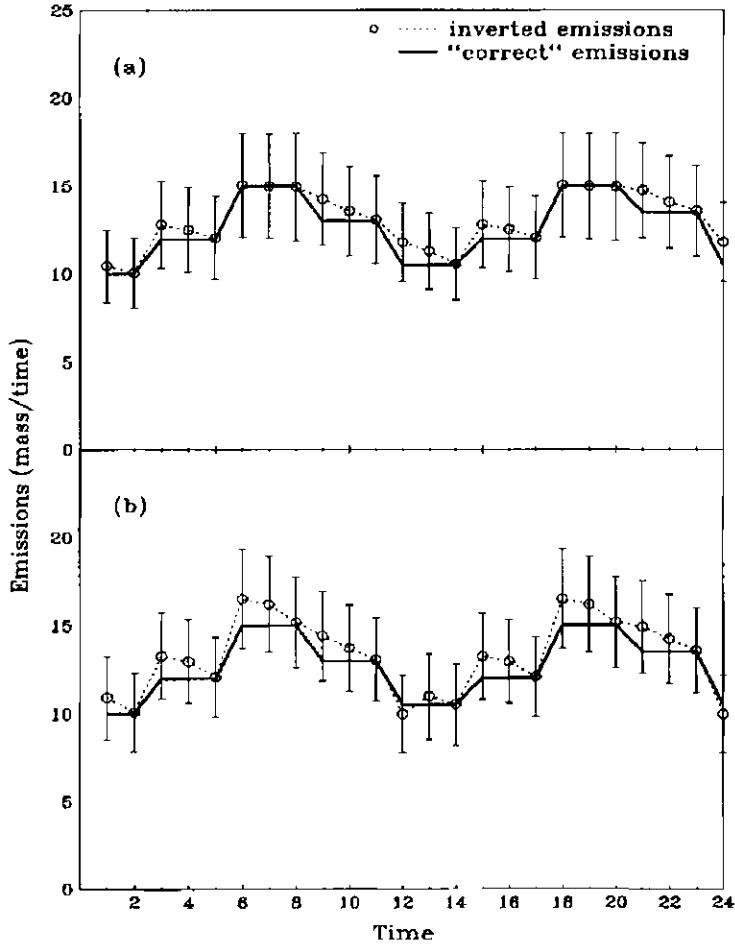
Figure 10: As in Figure 6, with an adaptive-iterative Kalman filter.

We note that in an iterative run, the covariance matrix is again allowed to update according to equation (6), both within each iteration and continuing to other times steps. In iterative filtering, the covariance matrix at the end of the first iteration at each time step is saved (Young, 1984). This value represents the statistical estimate of error as previously discussed. By saving the value after the first iteration, the error estimate is not artificially decreased based on the number of iterations at each time step. Therefore,

unlike the previous runs with the adaptive filter alone, the covariance matrix again represents the error in emissions and the error bars shown are as described previously. We note that even when the filter cannot converge on the "correct" solution, the solution is always within the error bars. Our results are summarized in Table 4, with the final iterative method clearly reducing the overall error in emissions prediction.

Table 4: Average percent difference between inverted emissions using various adaptations and the "correct" emissions.

Methodology	Figure Reference	avg. difference: region 1 (%)	avg. difference: region 2 (%)
non-adaptive filter	Figure 4	14.2	14.1
prescribed cycle	Figure 7	22.3	19.6
adaptive filter	Figure 8	13.6	13.6
smoother	Figure 9	7.6	6.5
adaptive-iterative	Figure 10	3.8	5.6

To summarize the development of the Kalman filter for time varying fluxes, we have attempted to simulate, using the ESR model, the behavior seen in three dimensional models in several ways, including adding "history" to the model, constraining the noise matrix to be greater than $P_t C_t P_t$, and using the "imperfect" model with slightly different winds than those which generated the "pseudo" observations. This has all been done to add complexity to the ESR model so that our results extend to global transport models. We tested various modifications to the standard Kalman filter including adaptive filtering

(Sastri, 1985), smoothing (Gelb, 1994), and iteration (Young, 1984; Bellgardt et al., 1986). We found that adaptive-iteration by far gave the best final solution. We note that this adaptive-iterative method is identical to the recursive-iterative method in Young (1984) with the inclusion of adaptive filtering as in Sastri (1985). A flow chart that summarizes the methodology is shown in Figure 11.

START

Time Step

emissions
estimation
$E_{i,t}$

Atmospheric Model
(e.g. GCM)

$\chi_{i,t}^{model}$

proceed
to next time
step
n=i
t = t+1
i = 1

Iterative
Loop:
i = i +1

No   does
i=1?

Yes

Is
$\Delta\chi_{i,t} > \Delta\chi_{n,t-1}$
?

Yes

adaptive filter:
reinitialize
Covariance matrix

No

Kalman Filter Equations:
(1) Calculate updated Emissions, $E_{i+1,t}$
(2) Update Covariance Matrix
(3) If i=1, save first update of Covariance

No   Is
$E_{i+1}- E_i$
>
threshold?   Yes
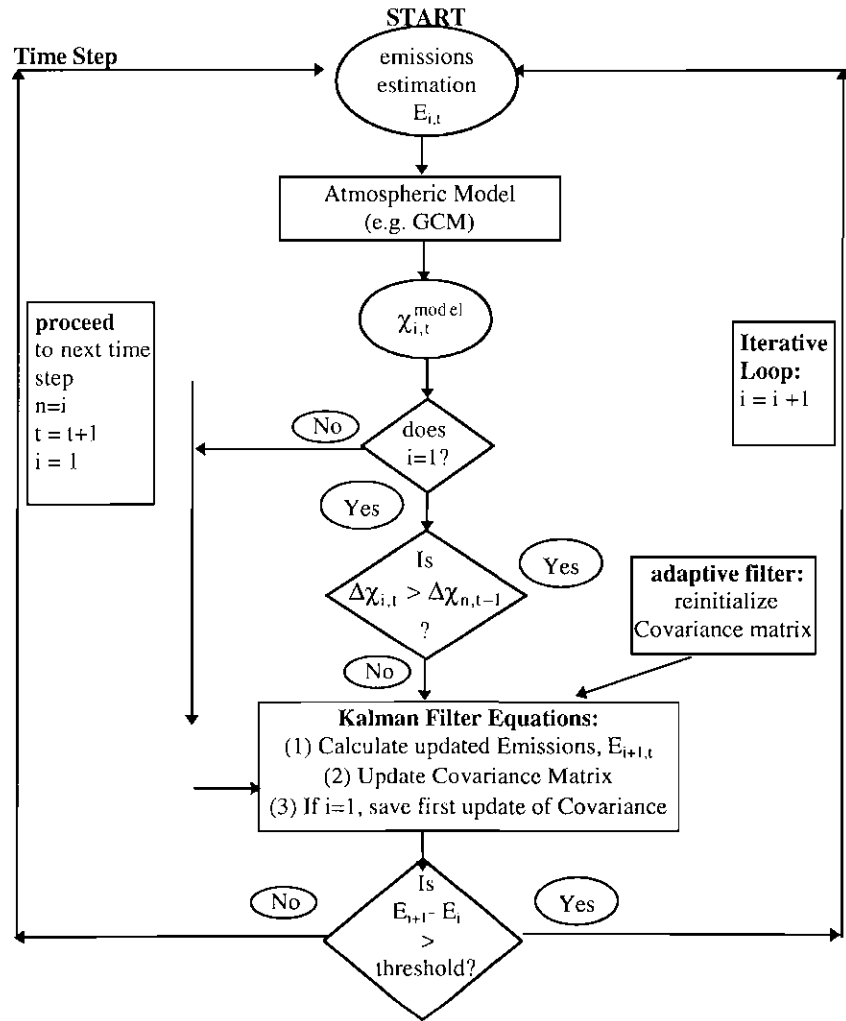
Figure 11: Flow Chart of Adaptive-Iterative Kalman Filter

CHAPTER 3

METHODOLOGY II: ERROR ESTIMATES

As previously described, one of the main advantages of using the Kalman filter is its use of error estimates in the filter. This provides a more stable solution since it does not force the model to agree with observations beyond the given error limitations. To take full advantage of this feature of the Kalman filter, this section will explore various error estimates that are added to the Kalman filter. All of these studies will be based on a full three dimensional general circulation model.

3.1 Description of GCM

The model used for all subsequent studies in this thesis is the offline version the NCAR Community Climate Model version 2 (CCM2) [Rasch et al., 1994; Hack et al., 1993]. The model has a T42 spectral resolution, which is approximately a 2.8° x 2.8° transform grid, and 18 vertical levels reaching 2.917 mb. Some of the main features of the dynamics in the model include a semi-Lagrangian transport scheme for water vapor and other chemicals, and a hybrid terrain following vertical coordinate [Rasch et al., 1994; Hack et al, 1993]. The CCM2 has been thoroughly tested with tracers and found to have a realistic interhemispheric transport rate (Hartley et al., 1994). This is extremely important because if the concentrations build up artificially in the northern hemisphere

due to model transport deficiencies, the inverse could falsely identify the sources and sinks. A detailed description of the CCM2 can be found in Hack et al. [1993].

The offline version of the model is essentially identical to the standard CCM2 with several exceptions. Instead of calculating the wind and temperature as the model moves forward in time (as in the standard CCM2), the offline version reads archived climatological winds from a previous CCM2 run. The winds and temperature are read in every six hours. This is an advantage over the standard CCM2 in that it saves a considerable amount of computer time. A one year tracer simulation using the offline version was compared to a previous run using the standard CCM2 to ensure that the offline version performed similarly. Further details can be found in Rasch et al. [1994] and Mahowald [1996].

3.2 Assessing Model Error in the GCM

In section 2.2, a transport residual, δ, was added to the noise matrix to account for model error. In the test model, this amount was easily derived by comparing the model output from the perfect to imperfect model as $\delta = \chi_{perfect}^{model} - \chi_{imperfect}^{model}$. For application in a GCM, we must derive a feasible, applicable methodology for assessing the contribution of imperfect winds to the differences between real observations and the model generated observations. We specifically propose using well known trace gases, such as CFCs, (Hartley et al., 1994) to test how well the model predicts the observed concentrations and

40

based on this comparison arrive at the transport residual. Once this transport residual is determined for the well understood trace gas, we can then appropriately adapt this residual for use with unknown gases like CO_2 and CH_4. One major assumption in this analysis is that CFC-11 emissions are perfectly constrained and therefore any difference in model output and observed data can be attributed to model error. This is, of course, an assumption; however it is not too far fetched. Particularly in the years that we simulate and compare (1987-1988), the emissions are reasonably well known.

Following analysis from Hartley et al. (1994), we compared CFC-11 concentrations at the 5 ALE/GAGE sites and 4 NOAA/CMDL sites with model output from the Community Climate Model 2 (CCM2), which is a full three-dimensional atmospheric general circulation model. We found there to be insignificant latitudinal variation, with an average percent difference between χ^{model} and χ^{obs} of $\pm1.5\%$. This is not a bias or systematic error, as the error fluctuates between positive and negative values. It is a random error that is on average 1.5% of the total concentration. In the case of CFC-11, this says that the model will agree on average with observations to approximately $\pm$ 5ppt.

We can extend this error to be valid for other gases by making use of the empirical relationship between measurement variation (σ) and the lifetime of the gas (Junge, 1974). This relationship determines the standard deviation by the following equation,

$$\sigma = \frac{b\overline{C}}{\tau}$$ Equation 10

where , $\overline{C}$ is the time mean concentration, τ is lifetime, and b is an empirically derived constant (0.16). As observed in the atmosphere, the longer the lifetime of a gas, the smaller the standard deviation in measurements.

We can adapt this relationship to account for changes in model error with lifetime, based on the fact that the model error is largely a function of different variations in dynamics between the model and the real world. Thus, the model error, δ, can be defined as follows:

$$\delta = \frac{b'\overline{C}}{\tau}$$ Equation 11

where b′ is now 0.66, based on a CFC-11 lifetime of 44 years (Cunnold et al., 1994). Figure 12 shows the percent model error $\left(\frac{\delta}{\overline{C}}\right)$ for various lifetimes. Given an approximate lifetime of 50, 10, and 100 years, the percent model error is calculated as 1.32, 6.6, and 0.66% respectively for CO_2, CH_4, and N_2O.

A brief note on the lifetime approximation is included here. There is a distinction between cycling time from a reservoir (turnover time) and the adjustment or response time (often called lifetime), which is the characteristic decay of an instantaneous source in the reservoir. In the case of CFC-11, the turnover time and the response time are equivalent, since the only source and sink for CFC-11 is atmospheric. However, as we have already discussed in detail in the introduction, carbon dioxide moves between many

42

reservoirs (ocean, vegetation, atmosphere). The turnover time for CO_2 between the atmosphere, ocean, and biosphere is estimated at about 4 years (IPCC, 1995). However, much of the CO_2 that leaves the atmosphere each year and enters either the ocean or the biosphere is returned to the atmosphere within a few years. The adjustment (or response) time is instead esimtated at 50-200 years, which is the time it takes for carbon from the surface ocean to be deposited into the deep ocean. Since the purposes here is to approximate model error for the sole purpose of accounting for this error in the noise matrix, we will use the conservative side of this range (50 years) to ensure that we do not underestimate the model error.
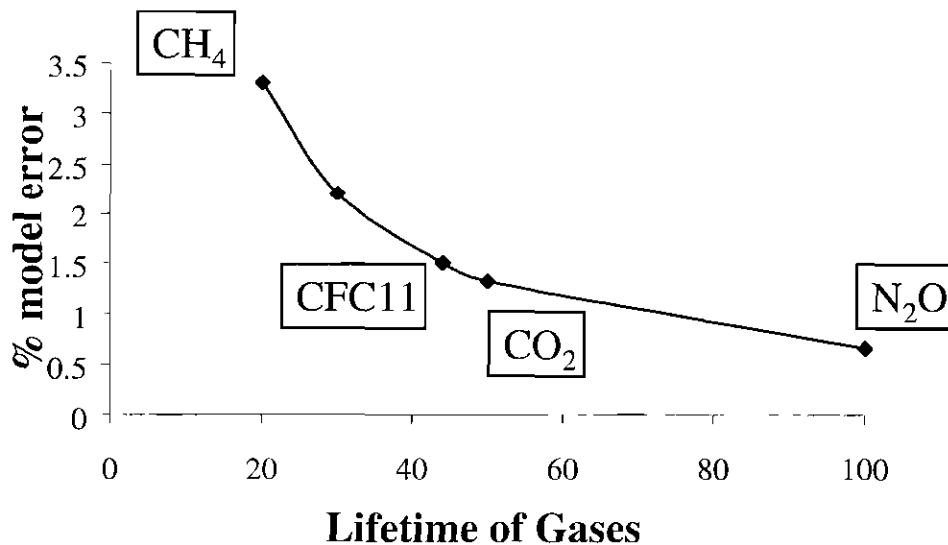
Figure 12: Modified Junge's Equation for the Relationship between Lifetime and Measurement Error.

3.3 Comparison of model output to CO_2 measurements

Observations of atmospheric trace gases are often compared with chemical transport models (CTMs) to evaluate the models (Hartley et al., 1994; Prather et al., 1987), to assess source-sink estimates (Erickson et al., 1996), and to constrain inverse problems (Hartley and Prinn, 1993; Ciais et al., 1995). However, the method used for comparison may depend on whether the observations are collected at high or low sampling frequency. The carbon dioxide *in situ* measurements are typically made at a high frequency (approximately every hour) and therefore are directly comparable to CTMs. The discrete air samples are often collected once or twice a week, suggesting that model output may need to be treated similarly. As the focus of the thesis is determine the fluxes of carbon dioxide, we will be examining carbon dioxide data from the NOAA/CMDL air sampling network, which is by far the largest sampling network of its kind.

It is important to recognize that the goal of the NOAA/CMDL air sampling network is to measure well-mixed air, not influenced by local CO_2 sources or sinks (Komhyr et al., 1985). To that end, many sites are at remote marine locations (Conway et al., 1988; 1994). To further reduce the effect of local sources and sinks, samples are not collected when the wind is from directions likely to be contaminated. Finally, a statistical method is used to reject outliers before calculating monthly or annual means (Conway et al., 1988; 1994).

44

Previous studies that have compared the NOAA/CMDL CO_2 flask data to model output have used several different approaches to make the two data sets comparable. One study filtered both the model and flask data with a lowpass digital filter (Erickson et al., 1996). Another study temporally smoothed the flask data to obtain curve fitted concentrations at 14 day intervals and then sampled the model every 14 days (Ciais et al., 1995). A recent study by Ramonet and Monfray (1996) specifically addressed this issue of how to compare model output to flask sample data at four of the NOAA/CMDL sites. They recommend obtaining model clean air by using back trajectories to determine if the model wind ever moved over a continental or typically "polluting" area. Whenever the back trajectory encounters an undesired region, the sample is eliminated. This is an extremely thorough approach and useful for ensuring that the model air comes only from a clean air sector. However, the same precautions are not taken in the field. When the flask samples are collected in the field, the only criteria for collecting a sample are wind speed and direction at the site at that moment. The sample is collected without any knowledge of whether the air has passed over a distant, undesired source. Thus, this back trajectory approach will result in "cleaner" air in the model than seen by the flask samples.

Our original approach was to use output from the model that meets the same criteria as used in the field and laboratory. Upon investigating how to implement this approach, we found that a discussion of the "clean air sectors" for the NOAA/CMDL network has only been published for the CMDL baseline observatories (BRW, MLO,

SMO, and SPO) (Ferguson and Rossen, 1992). However, the wind direction at the time of sampling has been recorded digitally for all measurements starting in the middle of 1992. Here, we present a three year statistical analysis of the wind direction during sampling for all NOAA/CMDL sites operational from 1993-1995. Comparing this wind distribution to ECMWF assimilated winds for the same three year period, we then determine which sites are actually screening for (or against) a particular direction(s). We then compare results of sampling the model using the field consistent method versus simply using the model predicted data with no pre-selection. Finally, we investigate the issue of sampling frequency.

3.3a Analysis of Wind Data.

The instructions for collecting samples at the CMDL cooperative global air sampling network sites include general guidelines concerning wind speed and direction. These guidelines are primarily intended to avoid local sources and sinks of the species to be measured. At most of the network sites, it is not practical to specify a "clean air sector" from which samples are always collected. Rather, the sites have been carefully selected so that there is a high probability of sampling well-mixed air unaffected by local sources or sinks. However, at some sites the general guidelines together with the local meteorology result in a de facto "clean air sector" from which the majority of the samples is collected.

When a flask sample is collected, the wind direction is estimated and manually recorded on a record sheet. At NOAA/CMDL, this information is entered into a data base. In most cases the wind direction is an estimate and not an instrument reading. For example, the person in the field may estimate that the wind is coming from the north east and record "NE" on the data sheet. This is then translated to degrees at NOAA/CMDL and digitally recorded as a wind coming from 45°. Thus, 45° should not be interpreted as an exact measurement, but merely representative of the approximate wind sector.

We have analyzed three complete years from January 1993 through December 1995. All of the available wind data at each of the 35 operational sites (shown in Figure 13) during this time period were divided into bins of 10° in wind direction. From this, we calculated basic statistical information about the frequency of wind at each site.

47
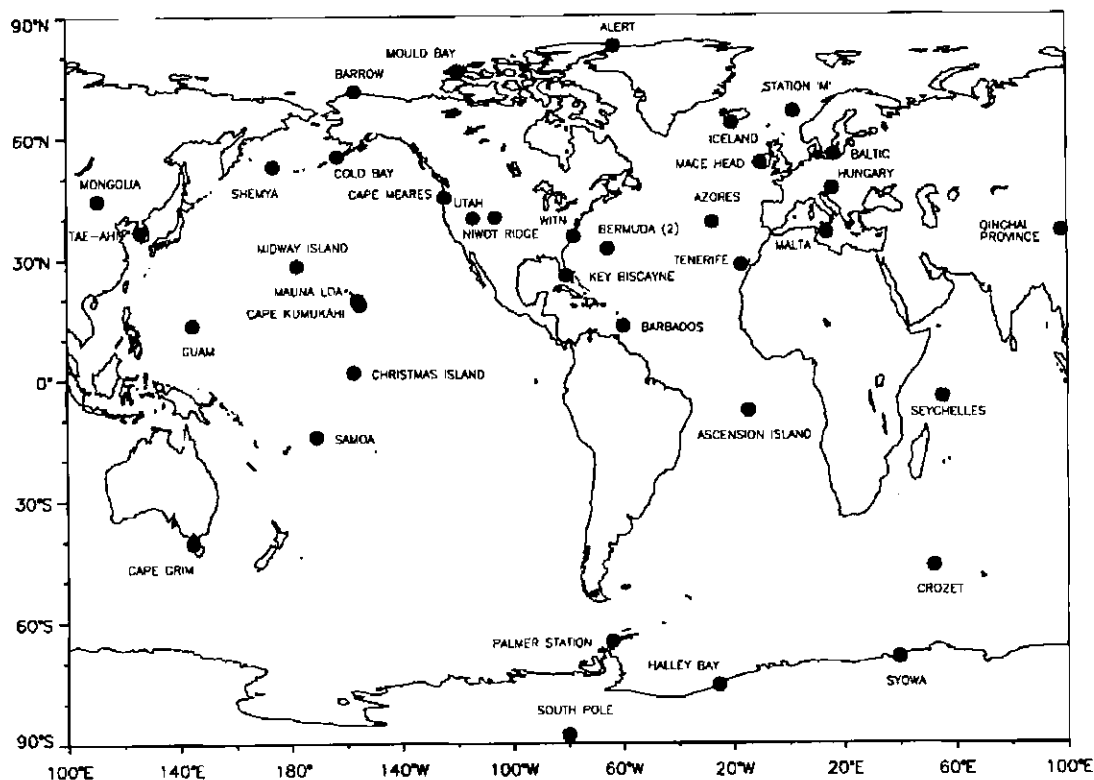
Figure 13: NOAA/CMDL sites included in statistical analysis.

While this analysis shows the typical wind directions during sampling at the sites,

it does not reveal any information about whether or not the wind was actually pre-selected

for a particular direction or if the predominant winds happen to be confined to that

particular direction. For example, as shown in Figure 14, the wind direction when

samples were collected at both Ascension Island (ASC) and Cape Grim (CGO) is

confined to a particular wind sector. Examining these distributions alone, there is no way

to determine if the wind was intentionally pre-selected or if the wind only blows in that

sector. Thus, we need to compare the NOAA/CMDL wind direction data with a

climatology of the wind at each site. Since actual climatologies are not available for each

48

site, we used the analyzed winds from the ECMWF. The ECMWF wind fields were

obtained from history tapes at National Center for Atmospheric Research (NCAR) at T42

grid resolution.
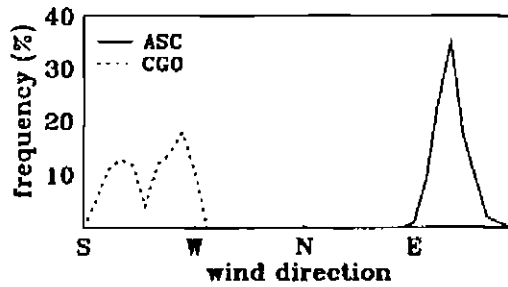
Figure 14: Frequency of wind occurrence in 10° bins at ASC and CGO wind direction for
3 year period from 1992-1995. Note: direction indicates origin of wind.

We performed the same statistical analysis on the ECMWF wind fields during the

same three year time period. The resulting frequency distributions for both data sets

(NOAA/CMDL and ECMWF) are plotted in Figure 15.
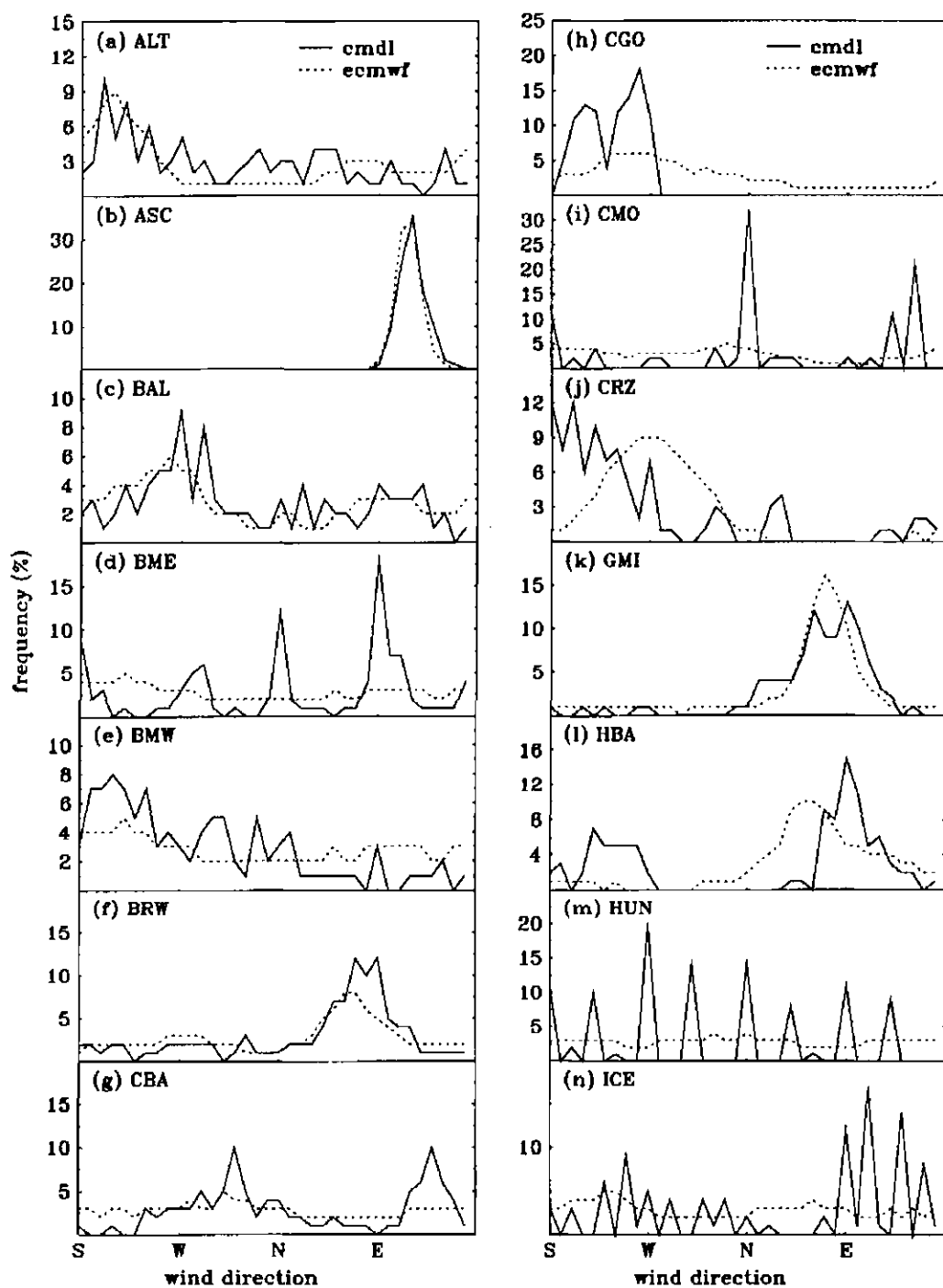
Figure 15: As in Figure 14, except includes NOAA/CMDL and ECMWF analyzed wind fields.
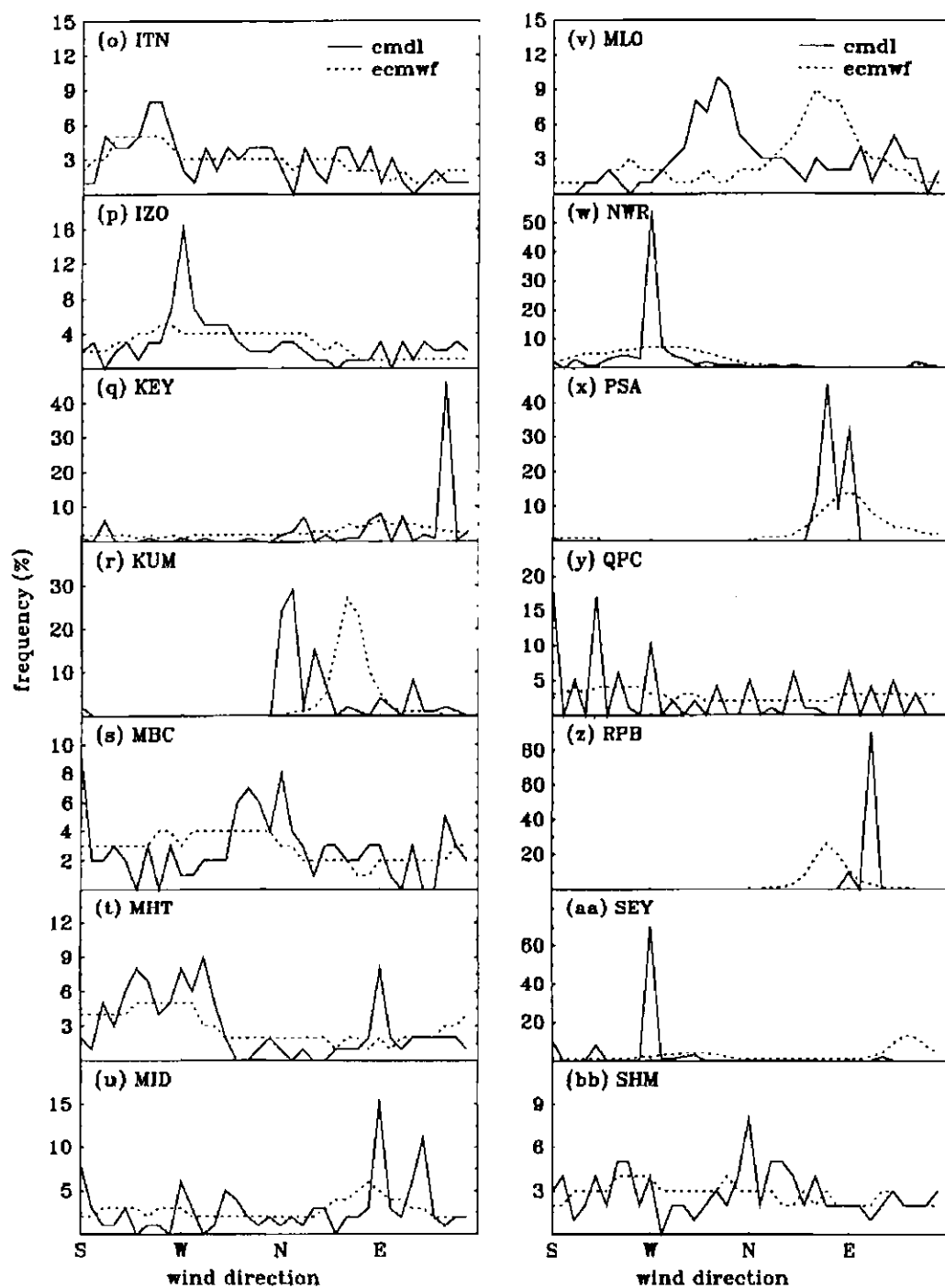
Figure 15(cont.).
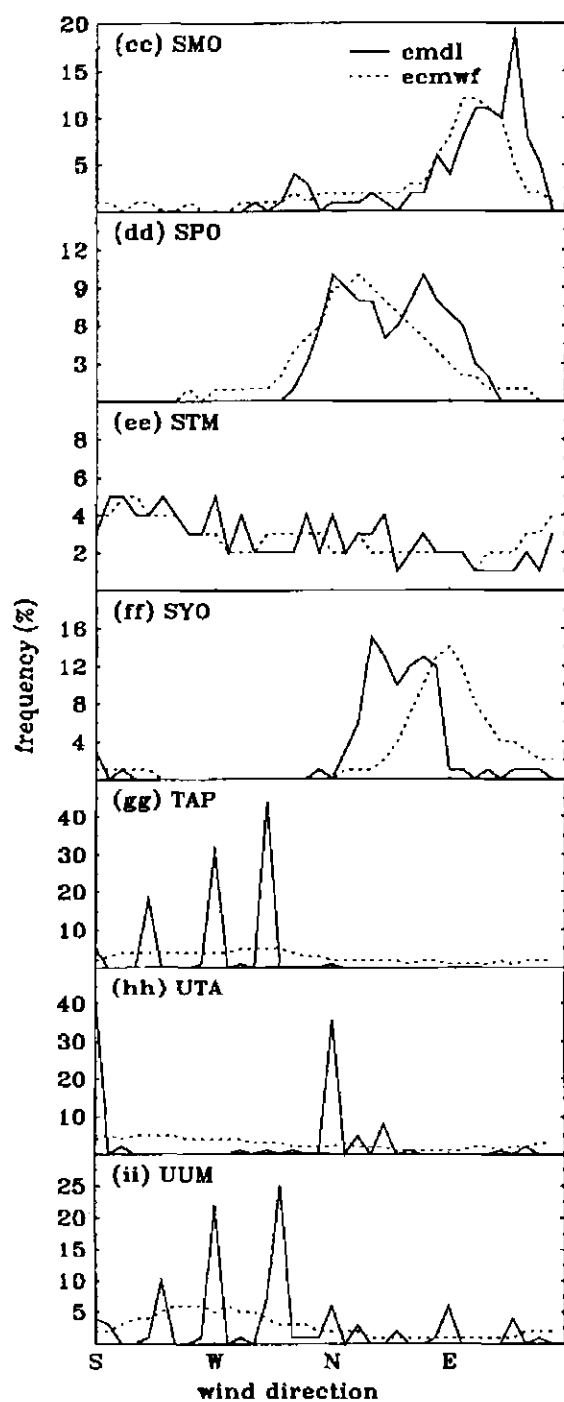
Figure 15 (continued)

As discussed above, the NOAA/CMDL wind data are often reported in estimated quadrants and this is reflected as spikes in many of the plots. For example, at HUN the recorded winds occur at discrete directions that cover the entire wind spectrum because the data were reported as S,SSE,SE,ESE,E, etc. It is important to consider the limitations of the NOAA/CMDL wind data when comparing them to the ECMWF climatology.

In comparing the two data sets, we are trying to answer the following questions: Which sites are preferentially screening out particular wind sectors? Which sites essentially match the assimilated wind for the three year period and are thus not screening for a particular "clean air sector"? Which sites may have local wind effects that are not captured in the assimilated wind data?

Re-examining the wind direction at ASC and CGO, Figure 15 shows that ASC almost exactly replicates the assimilated wind data for the 3 year period, while Figure 15 shows that the flask measurements at CGO are being carefully selected for wind in the S-W quadrant, while avoiding air from the N-E. Thus, we see the insight that is brought in by comparing the CMDL wind data to climatological winds.

Each site is grouped into one of the following categories: selection (only certain wind directions are sampled), no selection (all wind directions are sampled), and local effects (local winds differ from ECMWF). In the case of local effects, we have also tried to explain the source of the local effect and which category (selection or no selection) we believe the site falls into. These groupings are shown in Table 5. For comparison, the

53

"clean air sectors" for the 4 CMDL baseline observatories (Ferguson and Rossen, 1992)

(BRW, MLO, SMO, SPO) are also given.

Table 5: NOAA/CMDL sites grouped into 3 main categories: (1) selecting for wind direction, (2) no selection, (3) local wind effects not captured in ECMWF.

Station	Station name	Our Analysis	Fergusson and Rossen [1992]
ALT	Alert, Canada	No Selection	
ASC	Ascension Island	No Selection	
BAL	Baltic Sea, Poland	No Selection	
BME	Bermuda, West	Local / No selection	
BMW	Bermuda, East	No Selection	
BRW	Point Barrow, Alaska	No Selection	selecting: N-SE
CBA	Cold Bay, Alaska	No Selection	
CGO	Cape Grim, Tasmania	Selecting: S-W	
CMO	Cape Meares, Oregon	Local / No Selection	
CRZ	Crozet, Indian Ocean	Local / No selection	
GMI	Guam	No Selection	
HBA	Halley Bay, Antarctica	Local / No Selection	
HUN	Hegyhatsal, Hungary	No Selection	
ICE	Heimaey, Iceland	Selecting: E-N	
ITN	Grifton, North Carolina	No Selection	
IZO	Tenerife, Canary Islands	No Selection	
KEY	Key Biscayne, Florida	Local/ No selection	
KUM	Cape Kumukahi, Hawaii	Local/ No selection	
MBC	Mould Bay, Canada	No Selection	
MHT	Mace Head, Ireland	No Selection*	
MID	Midway Island	No Selection	
MLO	Mauna Loa, Hawaii	Local/ No selection	no selection
NWR	Niwot Ridge, Colorado	No Selection*	
PSA	Palmer Station, Antarctica	Local/ No Selection	
QPC	Qinghai Province, China	No Selection	
RPB	Ragged Point, Barbados	Local / No selection	
SEY	Seychelles	No Selection	
SHM	Shemya Island, Alaska	No Selection	
SMO	American Samoa	Selecting: NW-SSE	selecting NW-SE
SPO	South Pole, Antarctica	No Selection*	selecting NW-SE
STM	Station M	No Selection	
SYO	Syowa, Antarctica	Local/ No selection	
TAP	Tae-ahn Peninsula, S. Korea	Selecting: SW-NW	
UTA	Wendover, Utah	Selecting: N-NE, S*	
UUM	Ulaan Uul, Mongolia	No Selection	

* These sites have important qualifications made in the text.

55

For those sites grouped as "selecting" in Table 5, we present a brief explanation of why we believe the wind is being selected from the particular direction(s). At CGO, the samples are taken only when the wind is from the S-W in order to avoid air from nearby Australia. ICE is avoiding winds from the N-ENE quadrant, which is where local sources from the continent are located; thus, ICE is selecting from the E-N quadrants. At SMO, the winds are selected from the NW-SSE, avoiding air from the S-SW that is affected by vegetation on the island. TAP is selecting for air from the SW-NW, avoiding local sources from Korea. UTA is selecting for the N-NE and S. While these directions would be "clean air sectors," it is uncertain why air from the other quadrants is avoided. This may be a local phenomenon rather than an actual selection for wind direction. However, we have no clear reason not to group UTA as selecting based on the CMDL wind data.

We also wish to address similarities and differences for the sites where a discussion of the clean air sector was included in the CMDL report [1992]. Of the four sites, two were in agreement. At MLO, there is a diurnal wind on the mountain, but no "clean air sector" to limit sampling. In both the CMDL report and in the flask data, SMO is selecting for wind from WNW-SE. One minor difference is that our analysis showed that the flask samples collected at SMO also included wind from the SSE (Figure 15cc). At SPO (Figure 15dd), the CMDL report (1992) listed the clean air sector at SPO as NW-SE. While this is the only sector from which samples are collected at SPO, this sector accounts for the majority (>90%) of wind direction in the ECMWF assimilated winds. There is a small frequency of wind from the W, which is the area that is avoided in the field (Ferguson and Rossen, 1992). This sector is avoided because the main South Pole

56

facilities including living quarters and power plant (diesel generators) are located there. Since a model will not include any local pollution from these facilities, avoiding this sector is an unnecessary step in selecting model output. Thus, we list SPO as a non-selecting site. At BRW, we do not find that samples are being taken only when the wind is between N-SE as in Fergusson and Rossen (1992). The flask samples are taken out of this sector at approximately the same frequency that the ECMWF winds occur out of the sector (Figure 15f). Thus, we conclude that the samples are not really being taken from the clean air sector and observe that, in practice, BRW is a non-selecting site.

Two other sites warrant elaboration. At NWR, the sample collectors have been instructed to sample only during west winds. However, the wind is almost always from the general direction of the west. At MHT, the wind direction is being screened for wind from the S-W; however, samples are collected occasionally during winds from the east to sample European air. These samples are subsequently flagged as not representing "background" conditions. Thus, MHT could be considered a selecting site; however, we have left it as "no selection" category, *provided* the outliers are removed as recommended later in this chapter (Section 3.3b).

At eleven of the sites (BME, CMO, CRZ, HBA, KEY, KUM, MLO, PSA, RPB, SEY, SYO), the recorded NOAA/CMDL wind is coming from directions that occur infrequently or not at all in the ECMWF data set. For example, at KUM the wind is predominately from the north, but according to the ECMWF wind analysis, the wind is almost never from the north in that grid cell. In the case of KUM, the site is right on the

shore and is greatly affected by sea breezes. This is a local effect that the ECMWF

assimilated data would not capture due to the grid size in the assimilation process.

To ensure that the "local" effect was really due to a local phenomenon and not to a

misrepresentation of the site in the ECMWF grid, we analyzed the ECMWF wind in the

adjacent grid boxes to the north, south, east and west for all ten of these questionable

sites. We concluded that shifting grids could not explain the differences seen between the

assimilated wind fields and the NOAA/CMDL wind data. Figure 16 shows an example of

the wind from the adjacent grids for KUM.
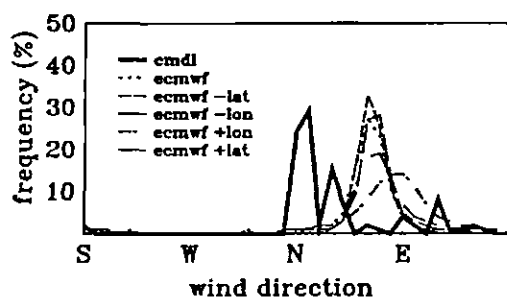
Figure 16: As in Figure 14, except wind direction from adjacent grids are shown for the
Cape Kumakahi site.

Of these eleven sites, six (BME, CRZ, KEY, KUM, RPB, SEY) are located on

islands and are experiencing sea breezes. This probably explains the local effects

observed. Three of the sites, HBA, PSA, and SYO, are coastal Antarctic locations. These

sites are most likely experiencing drainage flow and/or the ECMWF analysis is simply not capable of reproducing the winds in Antarctica. There are interpolation and spectral truncation problems in the ECMWF analysis in the Antarctic (Trenberth, 1992) in the ECMWF winds, and we believe this is why the CMDL winds at these sites differ from ECMWF. MLO is located on a mountain that has a strong diurnal wind; however, this mountain effect would not be captured in the ECMWF grid.

Finally, at CMO, we are not certain why the winds seem to be predominately from the north and south-east. However, it appears that there are no directions that are being avoided. CMO had the least number of samples (107) during the entire 3 year period, averaging only 3 samples per month. It is possible that this infrequent sampling might not be sufficient to capture the winds at the site. Since we have no reason to believe that CMO would intentionally screen for winds from the S-E, we conclude this site is local / not selecting.

A brief note about SEY. We do not believe that the wind recorded by the sample collectors actually reflects a local effect that is not captured by the model, but instead that the data recorded in the field are incorrect. At SEY the wind direction on the sample sheets was 270° for every sample between June 1993 and September 1995. This period coincides with a time when the samples were collected by a contract group. We do not feel this is accurate or representative; even as an estimation, it is improbable that the wind was from the general direction of the west every time a sample was collected for 27 straight months. While we cannot draw conclusions using this three year period, we

know of no reason that the SE quadrant (maximum frequency of wind direction) would be avoided, and thus group SEY as a non-selecting site.

There are three additional sites, Azores (AZR), Christmas Island (CHR), and Tenerife (GOZ) that were not operational throughout the entire 3 year period. These 3 sites are shown in the map in Figure 17. For AZR, the data are available for 1995; for CHR, the data are available for 1993; and for GOZ, the data are available for 1994 and 1995. The plots of the wind data at these sites, along with the ECMWF wind data for the corresponding time periods are shown in Figure 17. We group all three sites in the non-selecting category. The subsequent model comparison includes these three sites, but we emphasize that the wind distributions are based on shorter data sets.
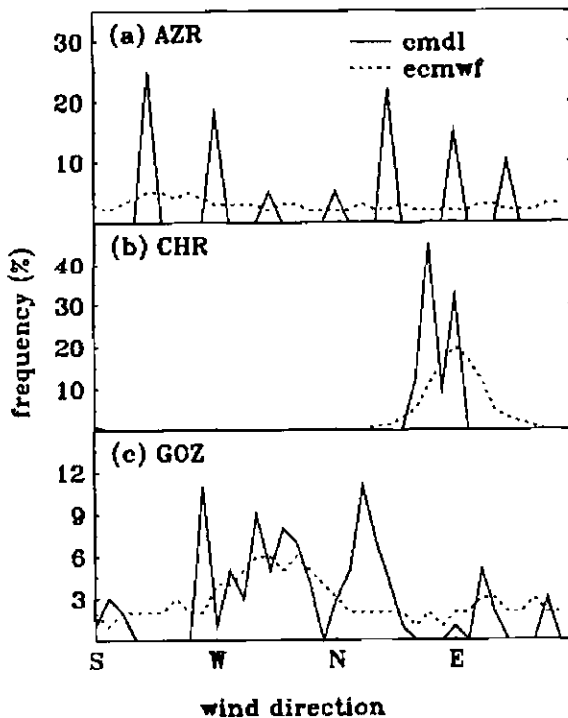
Figure 17: As in Figure 15, but each site was not operational for the full 3 year period.

3.3b Evaluating significance of filtering output in GCM.

Having defined which sites are in fact screening for wind direction, we now assess the effect of applying our proposed field consistent sampling methodology to model output. We do this in stages so that we can diagnose the relative importance of filtering for outliers, filtering for wind direction, and filtering for both outliers and wind direction. All of these comparisons will use a one year model simulation of CO_2 that was run on the offline version of the NCAR Community Climate Model version 2 (CCM2).

The method for removing outliers from the flask data has previously been described in Thoning et al. [1989], but we will provide a brief description here. A curve, $f(t)$, consisting of a quadratic trend and a series of sines and cosines to approximate the seasonal cycle is fitted to the data (or model output):

$$f(t) = a_1 + a_2 t + a_3 t^2 + \sum_{i=1}^{4} \left[a_{2i} + 2\sin(2\pi it) + a_{2i} + 3\cos(2\pi it) \right]$$

The residuals from this curve are filtered in the frequency domain with a low pass digital filter that has transmission of 0.5 at 4.5 y^{-1} and transmission of zero at ~ 7 y^{-1}. The filtered result (transformed to the time domain) is added to $f(t)$ to give the curve $F(t)$ that includes the mean and interannual variations of the trend and the seasonal cycle. Data (or model output) points lying more than ± 3 residual standard deviations (RSD) from the curve are rejected, and the curve is fit again. This process is repeated until no points exceed the 3*RSD limit.

61

Monthly mean mixing ratios and variances were calculated for each site from the model output at 0.25 day resolution with no selection (Base Case). For sites that select for wind direction, monthly means were calculated by applying the same wind selection to the model output (Filtered Set I). Another set of means was calculated for all sites after filtering for outliers (Filtered Set II). For the wind selecting sites, a set of monthly means was calculated using both outlier and wind direction filtering (Filtered Set III).

The three sets of filtered monthly means and variances were compared to unfiltered values using the student t test to evaluate the significance of the differences. The student t test equation is the following [Montgomery, 1991]:

$$ t = \frac{\overline{\chi_1} - \overline{\chi_2}}{\sqrt{S_p \left(\frac{1}{n1} + \frac{1}{n_2} \right)}} \quad \text{and} \quad S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}, \text{ where:} \qquad \text{Equation 12} $$

$\chi_1 (\chi_2)$ average concentration of data set with no filter (with a filter);

$S_1 (S_2)$ variance of data set with no filter (with a filter);

S_p common variance of both data sets; and

$n_1 (n_2)$ number of samples in data set with no filter (with a filter).

For a 95% confidence level with greater than 120 degrees of freedom, any site with t-value greater than 1.645 is statistically significant. The t-value can be positive or negative; however, for significance, the absolute value is used. We derive a further value

from this called the common standard deviation, which is the square root of the common variance. This gives the average standard deviation of both sets being compared.

We previously determined it is necessary to select for wind direction at only five sites (Table 5). When the wind direction filter was applied to the model output, four sites had monthly means that are statistically significantly different from the means derived with no filter. The only site that is not statistically different is SMO. These are shown in Table 6, along with the common standard deviation for reference.

Table 6: Sites with Statistically Significantly Different Means Using the Wind Filter Versus No Filter.

Station	months	range of csd (ppm)
CGO	5	0.31 - 1.32
ICE	3	0.67 - 2.20
TAP	5	2.86 - 6.24
UTA	3	0.73 - 1.26

Since most sites were not selecting for wind direction, we were particularly interested in the role of the outlier filter at the majority of sites. Again using the student t test, about half of the sites have at least one month that is statistically significantly different from the no selection (Base Case) mean (Table 7). The remote marine sites are minimally affected by the outlier filter, while sites that are near or on continents typically have means that are statistically different from no filtering. There are exceptions; at four

63

remote sites (CRZ, PSA, SPO, SYO), the means are found to be statistically significantly different. It should be noted that the csd values at these sites are very small (less than 0.5 ppm). This means that even small deviations will be significant since the difference in means is being divided by a very small number.

Table 7: As in Table 6, but Outlier Filter Versus No Filter.

Station	months	range of csd (ppm)
ALT	3	1.12 - 1.47
ASC	2	0.35 - 0.48
BAL	2	6.31 - 7.46
BME	1	1.63
BRW	1	4.23
CBA	2	2.24 - 2.65
CGO	5	0.40 - 1.07
CMO	1	3.53
CRZ	3	0.35 - 0.49
GOZ	2	2.06 - 2.72
ICE	2	1.39 - 1.80
KEY	3	2.06 - 3.02
MBC	3	0.78 - 3.04
MHT	3	1.38 - 2.56
NWR	2	1.81 - 2.24
PSA	1	0.23
SHM	1	5.52
SPO	1	0.11
STM	3	2.68 - 4.02
SYO	2	0.11 - 0.25
TAP	1	3.34
UTA	1	2.28

The last filtering test is the case where wind direction is screened for first, and then this filtered model output is run through the outlier filter. Of the five sites that are actually screening for wind direction, all but SMO had at least one month that was statistically significant from no filtering (Table 8).

Table 8: As in Table 6, but With Wind Filter and Outlier Filter Versus No Filter.

station	months	range of csd (ppm)
CGO	8	0.31 - 1.14
ICE	3	0.67 - 2.20
TAP	4	1.41 - 6.12
UTA	4	0.73 - 1.92

To give a sense of the relative importance of the filters, we show a time series of the model output at CGO with each filter versus no filter in Figure 18. Also shown are the resulting monthly averages with the annual mean subtracted. This shows how the monthly average is changed based on the various filtering techniques versus no filtering. From these tests, we concluded that at the majority of sites, it is important to run the model output through the wind (when selecting) and outlier filters in order to be field consistent and to avoid biasing the monthly means.
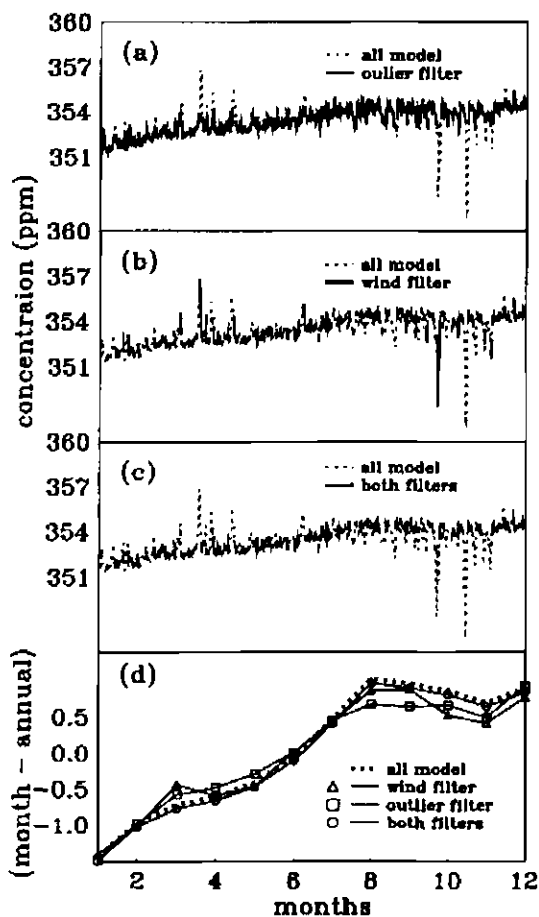
Figure 18: Time series of model output at CGO: no filtering versus (a) the outlier filter, (b) the wind filter, (c) both filters, and (d) monthly averages with the mean subtracted for no filter versus each set of filtered model data.

3.3c. Sampling Frequency

The last issue that arises is how frequently to sample the model output. Flask samples are typically collected at low frequencies, ranging from 2 times per week to 2 times per month. If we are to mimic what is done in the field, we should sample the model perhaps once per week.

All of the model output analyzed previously in this section was taken at 0.25 day intervals. The process of screening for wind direction and outliers is now repeated on model output taken once every 6 days. We have four tests cases, each selecting the model output once every six days, but with four different starting points. Set 1 starts at day 0.0, Set 2 at day 2.0, Set 3 at day 3.0, and Set 4 at day 3.5, and each set moves forward in 6 day intervals.

Our first test is to determine if these 6-day samples bias the mean. If the different cases are not statistically different from each other after the filtering, then we can conclude that the model samples should be taken at the same low frequency as the field measurements since the "true" background is reached. However, if the four cases are statistically different, we then need to address how to deal with possible errors that may arise from the relatively low frequency of flask samples.

The student t test was applied to the monthly means and variances from each of the possible combinations of the four 6-day interval sampled data sets: (A) 1-2, (B) 1-3, (C) 1-4, (D) 2-3, (E) 2-4, and (F) 3-4 (Table 9).

Table 9: Six Cases (A-F) Compare Each of the Four Subsampled Low-Frequency Data Set.

STN	Significant Cases	Significant months (#)	corresponding to max t-value csd (ppm)	difference (ppm)
ASC	CASE A	1	0.33	0.44
AZR	CASE A, B, C	4	1.65	2.66
BAL	CASE D	1	10.46	17.75
BME	CASE A, B	3	0.50	1.04
BRW	CASE B	1	0.36	0.44
CBA	CASE A, B, C	5	1.54	4.49
CGO	CASE A, D, E	6	0.16	0.24
CHR	CASE E	2	0.56	0.70
CMO	CASE C, D, E, F	6	3.03	7.51
CRZ	CASE C, D, E, F	7	0.43	0.85
GMI	CASE C, F	2	0.41	0.58
GOZ	CASE A, E	2	0.63	1.11
HUN	CASE B, C, E, F	22	3.26	8.51
ICE	CASE C	1	0.38	0.73
ITN	CASE A, B	4	0.78	1.65
KEY	CASE A, B, C, E	5	2.84	5.42
MHT	CASE A, B	2	0.78	1.45
MID	CASE E	1	0.20	0.34
NWR	CASE A	1	0.78	0.92
QPC	CASE B	1	1.02	1.73
RPB	CASE A, B, C	5	0.83	1.38
SEY	CASE B, C, D, E	7	0.30	0.61
SHM	CASE A, B, C, E	5	0.27	0.51
SMO	CASE D, E	2	0.39	0.54
SPO	CASE E	1	0.04	0.04
STM	CASE A, E	6	1.25	2.53
SYO	CASE C	1	0.03	0.05
TAP	CASE A, B, C, D, E	9	0.53	0.83
UUM	CASE A, B	2	0.92	1.30

At about 80% of the sites, the difference in the means calculated from varying the starting time of the low frequency sampling was statistically significant for at least one month. For the months that were statistically significantly different, the difference between the two means was always larger than the common standard deviation. This implies that simply accounting for the variability in the measurements (i.e. the standard deviation) is not sufficient to account for error due to sampling frequency, which could imply a large difference between the two time series. Since it is impractical to take flask samples in the field more frequently or to determine an "ideal" time to take the sample in the field, we need to address how to deal with the low frequency flask samples when compared to model output. While we cannot change the frequency of flask sampling, we can alter the frequency of model output used to calculate the mean. Thus, the next step is to compare high and low frequency model output.

We compare each of the 6-day interval sets (1-4) to high frequency (0.25 days) model output (Case I, Case II, Case III, and Case IV), and again about 80% of the sites had at least one month where the means were statistically different. Sites that had at least one month that was statistically significantly different are shown in Table 10.

Table 10: Four Subsampled Low-Frequency Data Sets Compared to the High-Frequency Model Output.

STN	significant Cases	months (#)	corresponding to max t-value csd (ppm)	difference (ppm)
ALT	Case IV	1	0.92	0.78
ASC	Case III	1	0.56	0.44
AZR	Case I	1	1.74	1.53
BAL	Case III	2	10.59	16.96
BME	Case III	1	0.77	0.69
CBA	Case I, II, IV	6	2.08	2.96
CGO	Case I, III, IV	12	0.29	0.60
CHR	Case I, II	2	0.45	0.37
CMO	Case II, IV	3	3.20	5.67
CRZ	Case I, III, IV	3	0.29	0.71
GMI	Case I, III	3	0.88	1.35
GOZ	Case II, IV	3	1.46	1.94
HUN	Case III, IV	4	5.93	4.75
ICE	Case I	6	0.99	1.33
ITN	Case II, III	2	1.71	1.53
KEY	Case IV	3	2.18	4.15
KUM	Case II	1	0.73	0.55
MBC	Case III, IV	3	0.44	0.64
MHT	Case I	1	0.85	0.88
MID	Case III	1	0.92	1.09
MLO	Case II	1	0.73	0.58
PSA	Case IV	1	0.14	0.11
QPC	Case I	1	1.42	1.23
RPB	Case II	1	0.81	0.88
SEY	Case II, III, IV	4	0.40	0.36
SHM	Case I, III, IV	4	2.58	4.21
SMO	Case III	1	0.59	0.45
STM	Case I, II	3	1.66	3.34
SYO	Case I, IV	2	0.03	0.03
TAP	Case I, III, IV	4	4.07	8.84

To illustrate this further, Figure 17 shows monthly averaged model output minus the annual mean for CBA, GMI, HUN, and TAP, for the high frequency case and each of the 6-day sets (1-4). At each of these sites, the cases show different detrended means.
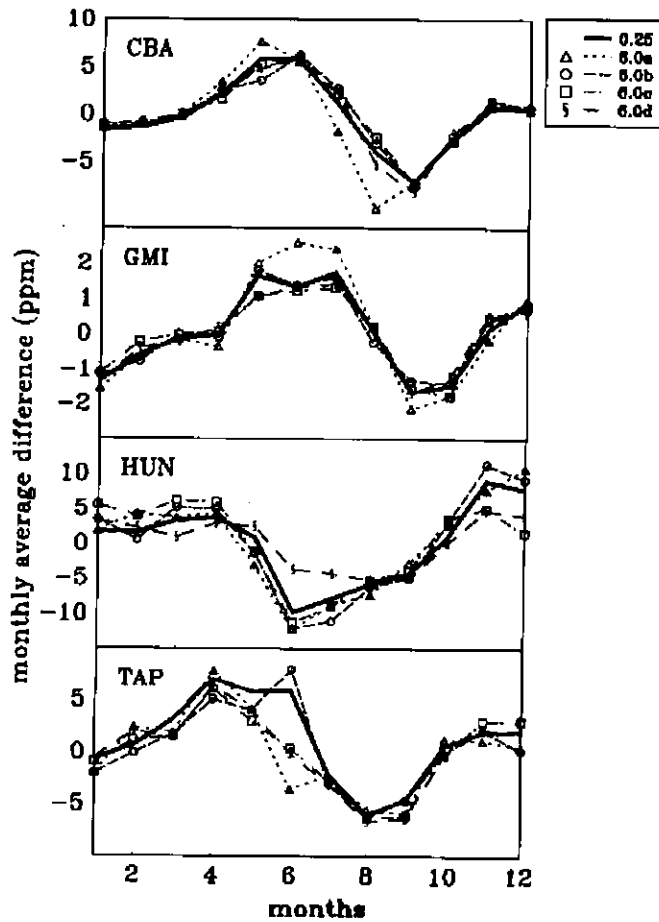
Figure 19: The monthly averages minus the annual mean are shown for the high frequency (0.25 days) model output and the four low frequency (6.0 days) cases.

72

Looking more closely at TAP, Figure 20 shows a snapshot picture for April. The monthly average for the "true" high frequency case and two low frequency cases are shown with standard deviation as error bars. If Case A (the highest) is the flask data and Case B (the lowest) is the model output, we have approximately doubled the bias from the "truth," (the high frequency output).
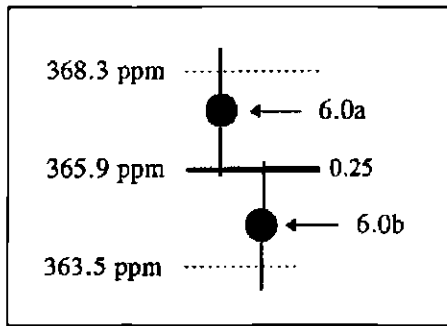
Figure 20: Monthly averages mixing ratios for April at TAP. The thick line is the mean (365.9) from the high frequency model output and the two dashed lines are one standard deviation from this mean. The two circles (6.0a and 6.0b) are means from 6-day sub-sampling of the model with different starting sample times.

From Figure 17 and Figure 20, it is evident that using high frequency model output is preferable in order to avoid exacerbating the bias in the model output mean. Using these modeling studies as an estimate of sampling frequency error in the field, we can estimate a measure of error arising from sampling frequency. To estimate this difference due to frequency (ε_f), we have created a composite average (mean absolute deviation) based on the difference in means for each month (1-12) from Cases I, II, III, and IV, as in the equation below:

$$\varepsilon_f = \frac{1}{12} \sum_{j=1}^{12} \frac{1}{4} \sum_{i=A}^{D} \left| \overline{\chi}_{0.25} - \overline{\chi}_{6.0}^i \right|_j$$, where cases A-D are 6.0 day samples at different starting

times. Thus, any biases from the low frequency sampling in the field can be included in

the error estimation when compared to model output. For inverse modelers, we

recommend augmenting the noise matrix by this value. A summary of these differences

(ε_f) is shown in Table 9, along with the average standard deviation from flask samples for

comparison. We see that in most cases, the flask sample standard deviations are larger

than ε_f . Recalling that if we were to sample the model at low frequencies, this bias

would double, the ε_f would then approach or outweigh the standard deviations. Thus, by

using the high frequency data, we are minimizing the additional error as a result of low

frequency flask samples.

Table 11: The average difference in monthly mean due to sampling frequency based on model output and the average standard deviation from flask samples in 1993.

STN	ε_f (ppm)	σ (ppm)	STN	ε_f (ppm)	σ (ppm)
ALT	0.22	1.22	KUM	0.12	0.82
ASC	0.08	0.46	MBC	0.29	1.10
AZR	0.28	1.28	MHT	0.26	1.28
BAL	1.79	3.01	MID	0.17	0.90
BME /BMW	0.26	1.15	MLO	0.12	0.56
BRW	0.29	1.09	NWR	0.34	0.98
CBA	0.67	1.22	PSA	0.07	0.16
CGO	0.17	0.20	QPC	0.42	0.93
CHR	0.17	0.37	RPB	0.21	0.59
CMO	0.73	1.75	SEY	0.19	0.47
CRZ	0.10	0.34	SHM	0.54	1.32
GMI	0.25	0.69	SMO	0.19	0.35
GOZ	0.47	1.40	SPO	0.03	0.14
HBA	0.03	0.11	STM	0.55	1.40
HUN	2.99	4.95	SYO	0.02	1.02
ICE	0.68	1.12	TAP	1.28	2.53
ITN	0.51	2.99	UTA	0.41	1.84
IZO	0.21	0.59	UUM	0.44	1.24
KEY	0.53	0.77			

We use two tests to ensure that the model derived ε_f is representative of an observation bias. First, we compare the average residual standard deviations for one year of *in-situ* measurements to model output at the 4 baseline sites (BRW, MLO, SMO, SPO) (Table 10). Overall, the model has slightly more variability than the measurements, although close enough to support using model output as a surrogate for the real measurements. Secondly, we sub-sampled the *in-situ* data using the same methodology at

the four continuous sites. The ε_f calculated from the model agrees well with the value from *in-situ* data (Table 12). This further supports our use of the model to calculate ε_f.

Table 12: Average of RSD from 1 Year Model Output Compared to 1 Year of In Situ Data.

	Average RSD (ppm)		ε_f (ppm)	
STN	**Model**	**NOAA/CMDL**	**Model**	**NOAA/CMDL**
BRW	0.66	0.62	0.29	0.42
MLO	0.32	0.27	0.12	0.19
SMO	0.37	0.14	0.19	0.10
SPO	0.04	0.06	0.03	0.03

3.4 Summary: Accounting for Errors in the Kalman filter

We find that there are three main issues to minimize and account for error when using model output and comparing to measurements. First, the model error (δ) should be accounted for as described using the modified Junge's equation. Secondly, when using flask measurements, the model should be treated in a similar way to the collection of samples. In the case of carbon dioxide, this includes identifying which sites are screening for wind direction (Table 5) and doing the same in the model, and removing the outliers from the model output (Table 7). This should make the comparison between model output and flask samples more consistent.

Finally, an additional error should be accounted for that arises from the low frequency with which flask samples are collected, since we found that sub-sampling the data biases the mean. First, the model output should be taken at high frequency to avoid

further biasing the comparison (Figure 17 and Figure 20). While this may seem counter-intuitive, it avoids the potential of doubling the bias. The error that arises is characterized as ε_f in Table 11 and this should be included in the noise matrix. Thus, we revise equation 8 to include both of these errors as follows:

$$N_t = [N_{ij}] = [\sigma_{it}^{obs}\sigma_{jt}^{obs}] + [\delta_{it}^{obs}\delta_{jt}^{obs}] + [\varepsilon_{it}^{obs}\varepsilon_{jt}^{obs}] \qquad \text{Equation 13}$$

CHAPTER 4

RESULTS

The first part of this chapter (section 4.1) will show pseudo data results using the full three-dimensional GCM. This section will consider both the optimum year and the definition of regions for which to invert based on the available measurements sites.

After showing satisfactory results in the pseudo data section, section 4.2 will present the inverted fluxes when the optimization technique is applied to real observational data. Recall from the initial cost function (Equation 1) that there are two components that are minimized in the solution. The first component is the change in emissions (ΔE) and the second is the difference in observations and model output (χ^{obs} - χ^{model}). Both of these components will be discussed in section 4.2

To determine if this inverse solution is robust, a series of sensitivity studies are conducted. These studies provide a framework for understanding the limitations of the inverted solution. These tests and results are described in section 4.3.

4.1 Pseudo Data Tests with CCM2

There are several components required to invert for surface fluxes, and each of these components is described here. First, an atmospheric model is required to generate model output. We use the same offline version of the Community Climate Model 2

(CCM2) described in section 3.3a. The inverse requires initial surface flux estimates for carbon dioxide (section 4.1a) and observations of CO_2 concentrations (section 4.1b). It is also important to divide the world into meaningful regions for the inversion (section 4.1b).

4.1a Initial Flux Estimates

One of the required inputs for the inverse method is an initial estimate of surface fluxes. Each estimate also requires an associated confidence level. Although the general surface fluxes and IPCC "best" estimates were previously described in Chapter 2, the actual inventory that will be used for initialization is described here.

To estimate fossil fuel sources, country by country estimates were taken from Energy Information Association (EIA). These estimates were then distributed within each country using gridded population as a proxy (Figure 21). The Consortium for International Earth Science Information Network (CIESIN) provided the necessary population data. The net fossil fuel source used in the inversion for 1992 is 6.0 Gt C/yr, as compared to the IPCC estimate for the 1980's of 5.5 ± 0.5 Gt C/yr. The EIA does not supply an error estimate; however, it is reasonable to apply the same percentage of error (10%) given by IPCC for the 1980's.. The estimate is an annual value and it is evenly distributed over each month; likewise, the error estimate is also evenly distributed.

79

Contours from 0 to .11 by .005 (kg/m2/mth)
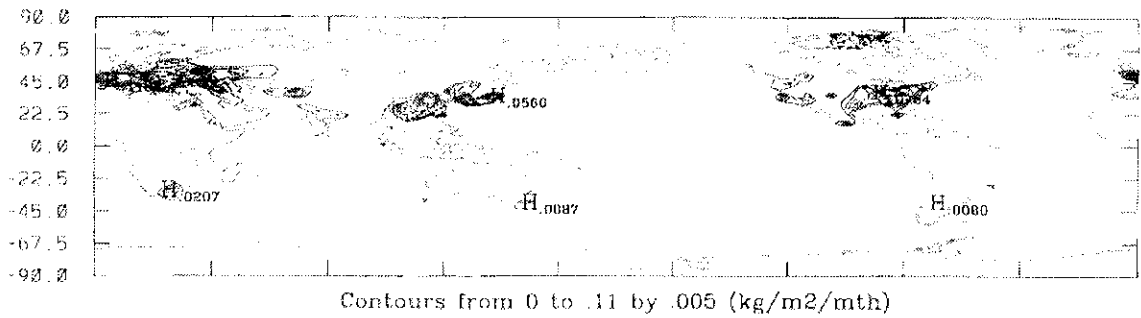
Figure 21: Initial Fossil Fuel Estimate (after EIA, 1997)

Gridded land use change data was estimated by Houghton et al., 1987 (Figure 22) and is also publicly available from Goddard Institute Space Science (GISS) and the Carbon Dioxide Information Analysis Center (CDIAC). The net source is 0.3 Gt C/yr, which is certainly lower than the estimated 1.6 ± 1.0 Gt C/yr given by IPCC; however, the GISS data is the only publicly available data set of its kind. Like the fossil fuel estimate, the land use estimate is an annual source and is evenly divided among the months, as is the associated error (order of magnitude).

Contours from 0 to .005 by .001 (kg/m2/mth)
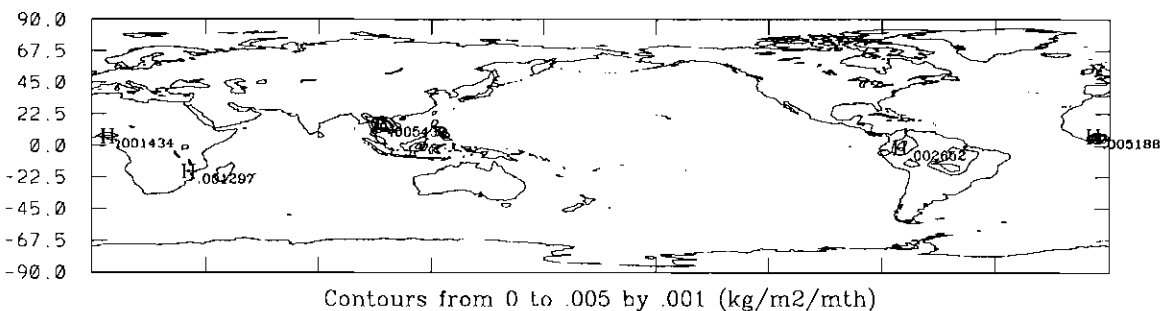
Figure 22: Initial Land Use Change Estimate (after Houghton et al., 1987).

Gridded monthly vegetative fluxes were estimated by Fung et al., (1987) and are publicly available from GISS and CDIAC. The annual net vegetative flux is -0.03 Gt C/yr. To show some of this initial vegetative flux estimate, Figure 23 shows (a) January and (b) July respectively.

Contours from −.03688 to .07312 by .01 (kg/m2/mth)

Contours from −.1981 to .0519 by .05 (kg/m2/mth)
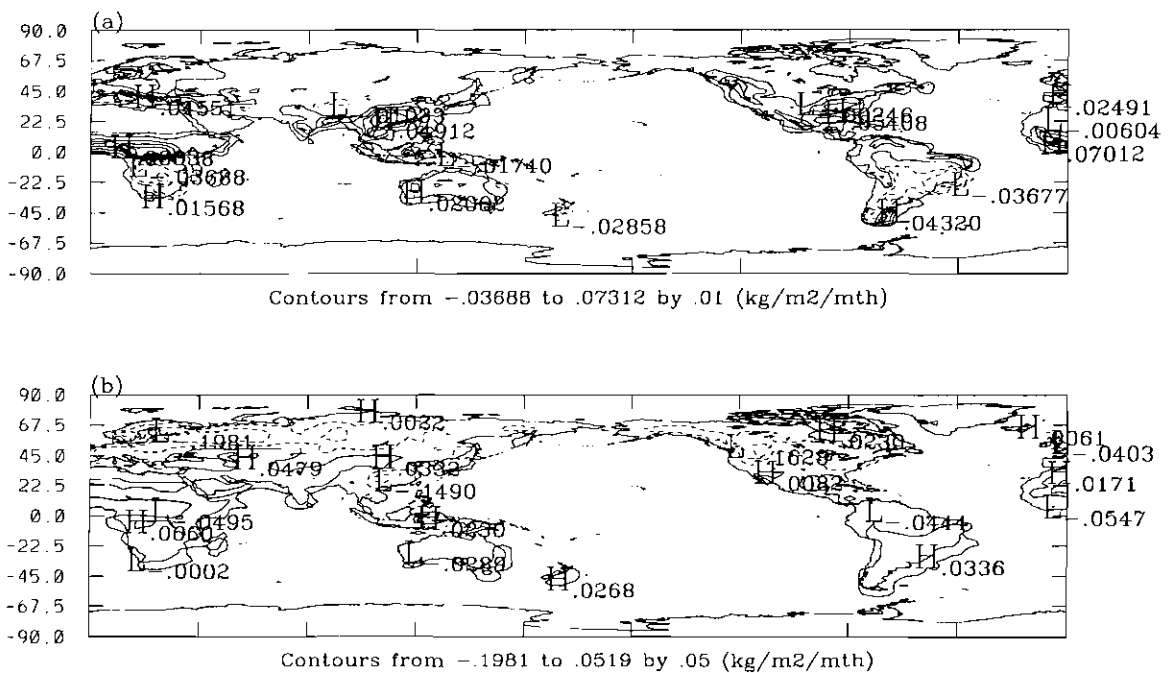
Figure 23: January and July Vegetative Flux (after Fung et al., 1987)

While the net sum is uncertain within several orders of magnitude, assigning monthly error estimates is not as straight forward as with fossil fuel emissions. Vegetative (and ocean) fluxes vary monthly by large negative and positive values, so that the net flux is the sum of large negatives and positives. Since no monthly error estimate is provided, there are two ways to distribute the error estimate. One way is to divide the

total error evenly over the months, as is done for the constant fossil fuel source. The problem with assigning error this way is that it will artificially limit the amount that each month is allowed to change. For example, the net vegetative flux in the northern hemisphere is given as -0.00058 Gt C/yr. From Chapter 1.2, we know that according to the literature there is a range of estimates from approximately zero to -2 Gt (Table 3), so the uncertainty is about 2 Gt. Divided evenly, this is an uncertainty of 0.16 Gt C / month. However, the initial estimate for July alone is -2.3 Gt C /month. Thus, assigning an uncertainty of only 0.167 Gt (or 7% of the initial estimate), the inverse would be so constrained that it would essentially be unable to change. This is not ideal since we know that there is a large amount of uncertainty with the vegetative flux in the Northern Hemisphere.

The other way to assign uncertainty is to assume that each month is uncertain by some percentage of the initial estimate for the month. This is the method that is employed here, with an uncertainty of one order of magnitude given to each month (initial monthly estimate*10). This allows the inverse the freedom to adjust each month within a reasonable amount. However, the initial net uncertainty will be considerably larger than the actual net uncertainty. In this case, the initial uncertainty for the Northern Hemisphere vegetation is 12.4 Gt compared to an uncertainty in the literature of 2.0 Gt. However, the inverse calculates a statistical value for the uncertainty based on the first update of the covariance matrix (Equation 7). Recall that this is a conservative estimate of the error (Young, 1984), so the calculated error will be an upper limit of uncertainty.

The evenly distributed error would severely limit the inverse optimization, but provide a realistic initial error estimate. However, in this case, the flux (and subsequently the error estimate) would hardly change at all from the initial condition and nothing would be gained from this exercise. Instead, the inverse is allowed the freedom to change and the inverted error is statistically calculated, but the initial error is overestimated. This approach is also taken for the ocean flux estimated error.

The only available month by month gridded ocean flux estimates is from Erickson et al. (1996) at NCAR. The Erickson data set has a net flux of -0.25 Gt C/year, which is lower than the IPCC estimates (Table 13). January and July are shown in Figure 24.

Contours from −.02973 to .01527 by .005 (kg/m2/mth)

Contours from −.02486 to .02514 by .005 (kg/m2/mth)

Figure 24: (a) January and (b) July initial ocean estimate (from Erickson et al., 1996)

It is also important to note that the monthly flux from this data set does not agree qualitatively with satellite measurements of chlorophyll, but instead tends to be solely based on temperature controls. As previously described (Chapter 2), there are many competing variables that control the flux of carbon in the oceans and their combined effects are highly non-linear. For example, chlorophyll is at its maximum intensity in the northern hemisphere oceans during summer, which suggests that this should be the maximum uptake season. From a thermal standpoint, colder water can hold more carbon than warmer water. Apparently emphasizing the latter rather than the former mechanism, the Erickson estimate is a net source during the summer. Keeping these potential errors in the initial distribution in mind, sensitivity studies will be presented in section 4.3 that estimate the influence of a regional distribution error on the inverted solution. As a summary, Table 13 compares the initialized estimate with the IPCC (1995) estimate.

Table 13: Summary of balance of fluxes (Gt C/yr) with the different flux inventories.

Sources	Initial 1992 estimate	IPCC 1980's estimate
fossil fuels	6.0 ± 0.5	5.5 ± 0.5
land use change	0.3 ± 1.0	1.6 ± 1.0
Sinks		
oceans	0.25 ± 1.2	2.0 ± 0.8
northern hemisphere vegetation	0.0006 ± 12.4	0.5 ± 0.5
Atmospheric increase	1.4 ± 0.1	3.3 ± 0.2
Implied "Missing" Sink		
??	3.15 ± 6.4	1.3 ± 1.5

4.1b Defining regions and selecting the optimum year for inversion

As in the study using the ESR model (section 3.2), we need to test the method with a known solution. Several issues need to be resolved using pseudo data studies. First, is the method capable of deducing regional fluxes with the available observations? What is the best year to simulate given the number of observations available each year? What is the maximum number of meaningful regions that can be resolved with the available data? All of the questions are interrelated. A general discussion describing how the world was divided into regions and how the optimum year was chosen is given below.

The first attempt at dividing the world into meaningful regions was based on trying to keep "like" regions together (Figure 26). The term "like" regions means regions of similar vegetation or regions in the ocean that generally have the same sign of flux throughout. For example, South America was divided into two regions to keep the rain forest region (7) separate from the rest of the continent (region 4). As another example, Europe was divided into east (21) and west (20) regions. Since the two differ from each other politically, it follows that their fossil fuel usage (or at least reporting of it) could be different. As an initial year to test the inversion, 1990 was selected for several reasons, including an overlap with other studies (Ciais et al., 1995), and a general lack of peculiarities during the year. For example, it is not an El Niño or La Niña year; there were no major volcanoes; and the trend in increase was similar to the rest of the 1980s. The available January 1990 observation sites are shown in Figure 25.
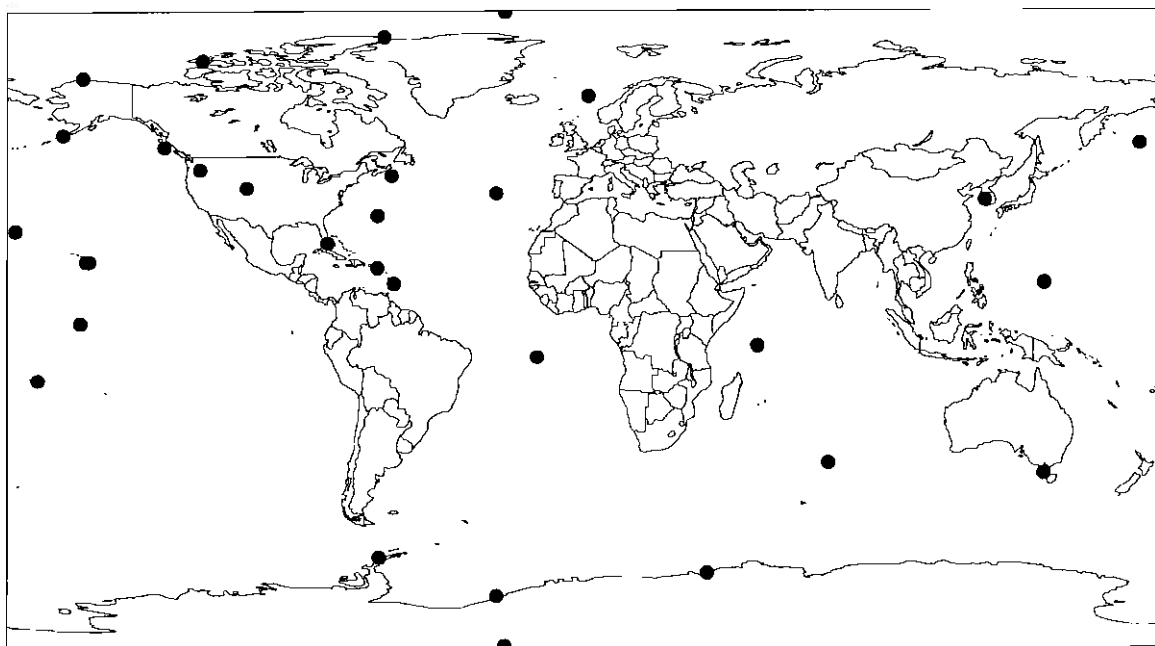
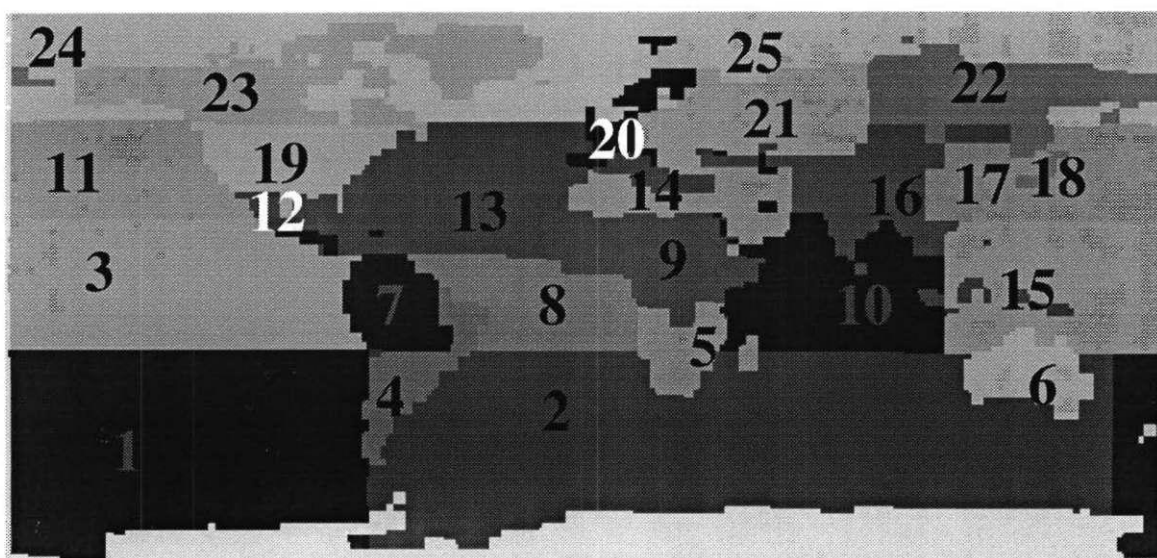Figure 25: January 1990 observation sites.

Figure 26: Initial division of regions for CO_2 inverse

Using the grid points that represent the NOAA/CMDL network in January 1990, a pseudo data inverse was performed with the twenty five regions shown in Figure 26 and the results are shown in Figure 27. To perform the pseudo data inverse, the initial flux estimates were input into the model, and the model output at the same grid location as actual observation sites were retained for each month. Then, the model was run with a perturbed flux scenario (1.5*initial). Because each region was perturbed by the same factor, a uniform error estimate was used in the covariance matrix. The method was able to invert for the net flux to within 3% of the "correct" total.

While this is extremely good from a global standpoint, the performance within each region must be examined more closely. After examination, it is apparent that the inverse has difficulty distinguishing between East and West Europe (regions 20 and 21), with one over-predicting and one under-predicting. Both Africa (regions 5, 9, and 14) and Central/South America (regions 4, 7, and 12) have difficulty converging on the "correct" solution in at least one of their respective divisions. In Africa, only the middle region (9) performed well, with the most southern (region 5) and the most northern (region 12) regions having difficulty. In Latin and South America, only the northern part (region 7) converged well, with the southern region of South America (region 4) and Mexico (region 14) both failing to converge sufficiently on the "correct" solution. The other weak area in this pseudo data inverse was Japan (region 18).
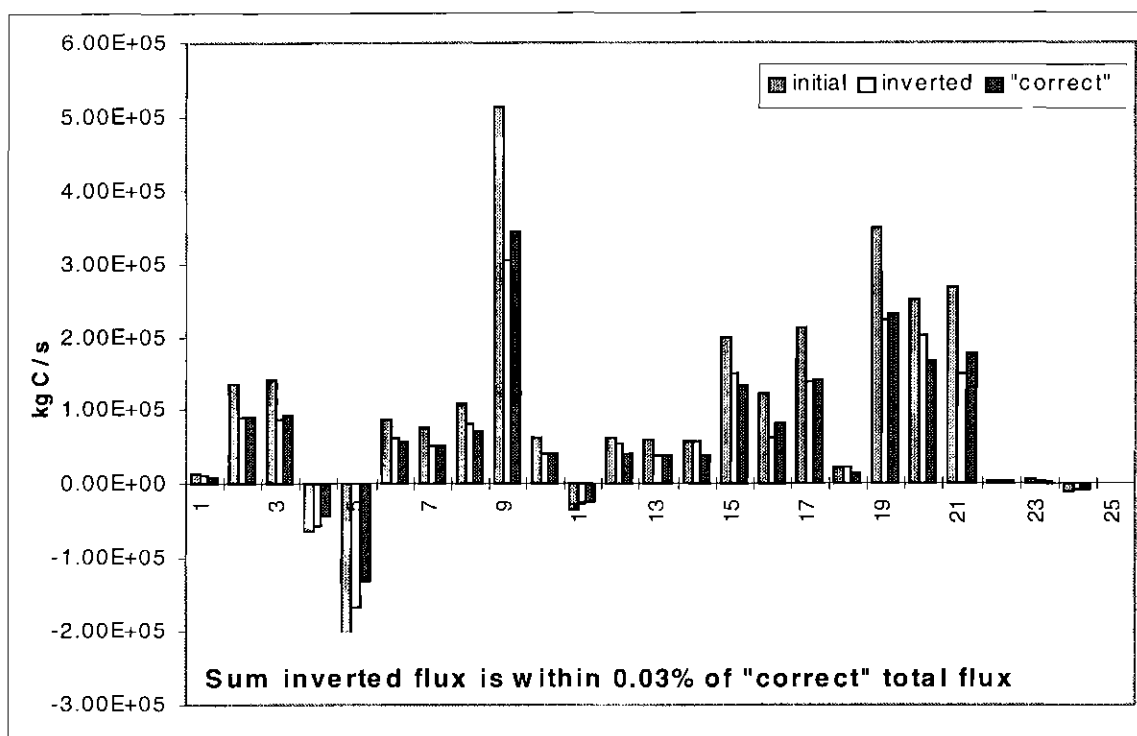
87

Figure 27: January 1990 pseudo data results using initial regional divisions.

The main factor affecting the convergence of a given region is the existence of an observation site that is sensitive to the region. For the weaker areas, we wanted to determine if there were any other available data sites that could improve the performance of the inverse. Thus, we looked at inverting for other years, as well as including data from other measurement groups. After searching for all available data, we wanted to determine the year that would allow the inverse to deduce the largest number of regions. To determine this, a list containing the most sensitive site for each region was created. Then, we assessed how many of these key sites were operational from 1990 to 1994 and for how many months. If the first choice of a site was not available, we determined

whether a second, third, or fourth most sensitive site was available. With this process of elimination, 1992 was selected as the best year to simulate. A map of the sites is shown in Figure 28.

One of the main reasons for selecting 1992 was the important role that the South China Sea shipboard data had on deducing much of the Asian continental and oceanic regions. Although operational since 1991, the only full year of data is 1992 due to various problems in the field. Another key site discontinued after 1992 is the South Georgia Island site in South America. Other key sites (e.g. Mace Head, Ireland; N. Carolina, USA; Ryori, Japan) were added in late 1991 and early 1992, making the previous years less than ideal for inverting.
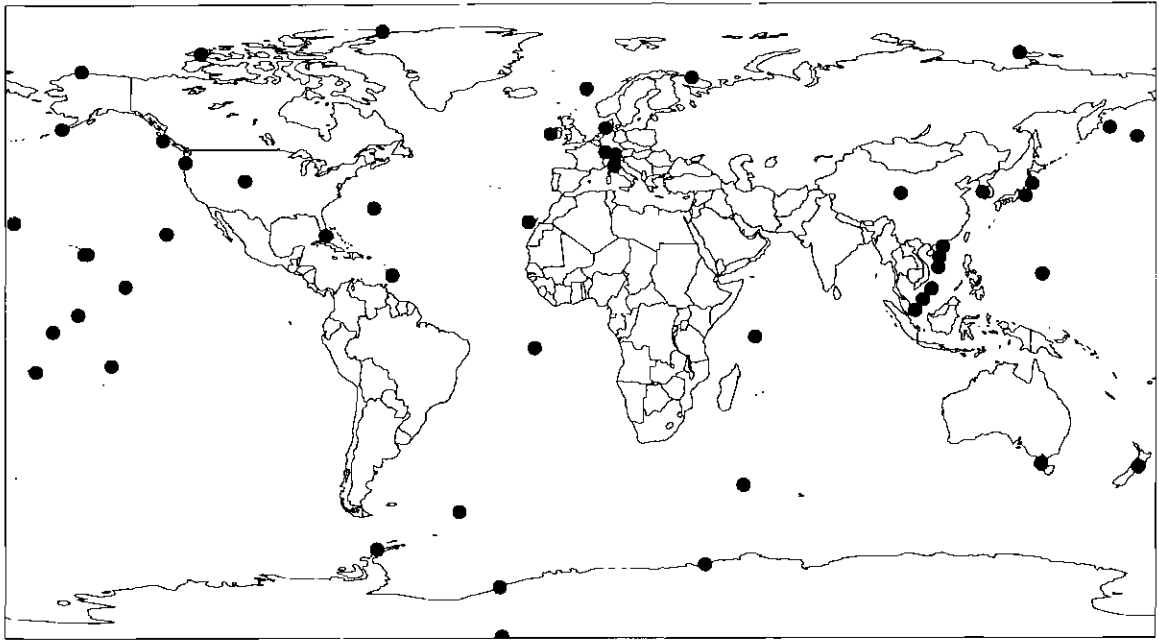
Figure 28: Observation sites in 1992.

Having selected the optimum year, the pseudo data inverse was performed again. Of the original regions that were poorly constrained, the 1992 pseudo data inverse is capable of distinguishing between West and East Europe, with the addition of several European sites, and of constraining Japan, with the addition of Ryori. However, there were no additional sites to help constrain Africa or South America. Thus, these regions must be regrouped. Since India and Western China (region 16) converged well and are similar in climate to Northern Africa, these regions were grouped together. South and mid-Africa were grouped into one region; similarly, all of South and Central America were combined into one region. The regroupings are shown in Figure 29.
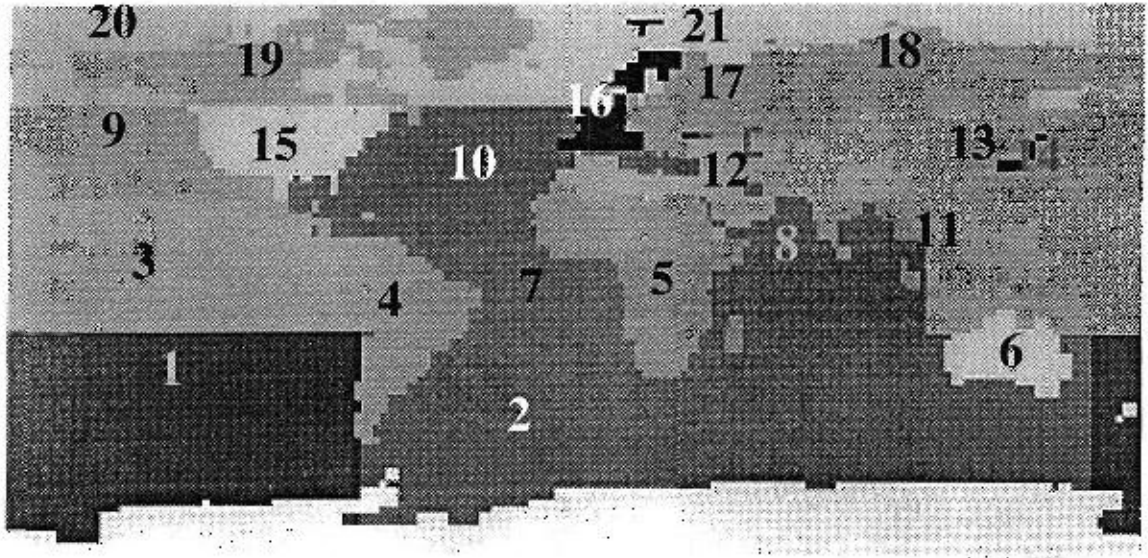
Figure 29: Regions based on 1992 data to constrain them.

After consideration, we also determined that the ocean region divisions needed to be altered. Instead of keeping them in one grouping throughout the year long inverse, we decided that their divisions should be more fluid to account for seasonally changing signs in the ocean flux. By subdividing them into positive and negative fluxes, as the seasons change, there will not be large positives and negatives within one region.

At this point, a final reevaluation of the regions was conducted. We wanted to determine if the northern hemisphere could be refined further, keeping in mind that the ultimate goal of this research is to distinguish between vegetative and oceanic regions (the two possible fluxes that could account for the "missing" sink). To do this, we

separate the fossil fuel component from the vegetative flux. By separating the fluxes in this manner, the strengths of the Kalman filter technique are better utilized. The Kalman filter allows weighting of various error components, including an initial estimate of the constraints on the initial state vector (fluxes). Thus, a much tighter constraint can be placed on the Northern Hemispheric fossil fuel source component, and looser constraints are given to the vegetative and oceanic fluxes. As with the ocean fluxes, when possible, we also further separate the vegetative flux into positive and negative regions that change seasonally.

Thus, after the initial attempt at separating regions, the regions were refined until satisfactory convergence of the pseudo data was achieved for all regions. To ensure that our finalized regions are in fact an improvement over the first attempt at dividing regions, some comparisons are included here. Recall that we are comparing the initial pseudo data inverse (January 1990) results (Figure 27) to the revised pseudo data inverse (January 1992) results (Figure 30a). From a net standpoint, the overall flux in both cases comes to within 3% of the "correct" total. However, we were not initially concerned with the total net inversion, but instead with the performance of the individual regions. To look more closely at the performance of the individual regions, the following statistical measure called the percent aggregate error (PAE) is applied, where N is the number of regions (25 in both cases) and E is the emissions (or flux):

$$PAE = \frac{\sqrt{\sum_{i=1,N}^{N}\left(E_i^{correct} - E_i^{inverted}\right)^2}}{E_{net}^{correct}} * 100\%$$

The percent aggregate error in the initial pseudo data inverse is 4.9% and this reduces to 2.7% in the revised pseudo data inverse. Furthermore, in the initial pseudo data inverse, only 10 regions are within ± 10% of the "correct" value, compared with 16 in the revised inverse. Additionally, in the initial pseudo data inverse, 10 regions have greater than ± 20% difference from the "correct" solution, compared with only 5 in the revised regional inverse. Thus, while the revised inverse still has some troublesome areas, the overall improvement within the regions is evident. Note that both the initial and revised divisions have a total of 25 regions, so comparing the numbers of regions is a fair comparison.

Since the borders of the positive and negative flux regions change seasonally and the fossil fuel regions are superimposed on top of the northern vegetative flux regions, there was no feasible way to recreate a revised Figure 29 with the new regions. Instead, a summary and description of the finalized regions is shown in Table 14. We carried out the final pseudo data test for the entire year of inversion. The January and July pseudo data inverse are shown graphically in Figure 30. Two statistical measures summarize each regions performance in Table 15. The first is the PAE, now calculated for each region with N being the number of months in a year. The second is the net average difference, calculated as the sum inverted flux divided by the sum "correct" flux. The

initial estimate is 1.5 (or 50% greater than the "correct" value); therefore, a perfect inverse would have a normalized average of 1.00. This does not mean that every month is exactly 1.00, but that over the year, the inverted solution averages out to be perfect.

Table 14: Twenty-five Regions with Description.

| Region	Location	Notes / Description
1	Southern Ocean - positive flux	Defined by continuous ocean
2	Southern Ocean - negative flux	Defined by continuous ocean
3	Mid-South Pacific - positive flux	varies by season - same sign flux
4	Mid-South Pacific & Indian Ocean-negative flux	varies by season - same sign flux
5	Australia and New Zealand	vegetation and fossil fuel
6	South and Central America	vegetation, land use change and fossil fuel
7	Indian Ocean - positive flux	varies by season - same sign flux
8	Mid Atlantic - positive flux	varies by season - same sign flux
9	Mid Atlantic - negative flux	varies by season - same sign flux
10	Mid-South Africa	vegetation, land use change and fossil fuel
11	North Africa to Western China	vegetation, land use change and fossil fuel
12	Pacific Islands and Thailand	vegetation, land use change and fossil fuel
13	E. China & Japan - vegetative flux	vegetation only
14	U.S. and Canada - negative flux	vegetation only
15	North- Mid Pacific - positive flux	varies by season - same sign flux
16	U. S. and Canada - positive flux	vegetation only
17	Eurasia - negative flux	vegetation only
18	Eurasia - positive flux	vegetation only
19	North Pacific	varies by season - same sign flux
20	North Atlantic	varies by season - same sign flux
21	Western Europe	fossil fuel only
22	Eastern Europe / Eurasia	fossil fuel only
23	United States and Canada	fossil fuel only
24	E. China	fossil fuel only
25	Japan	fossil fuel only
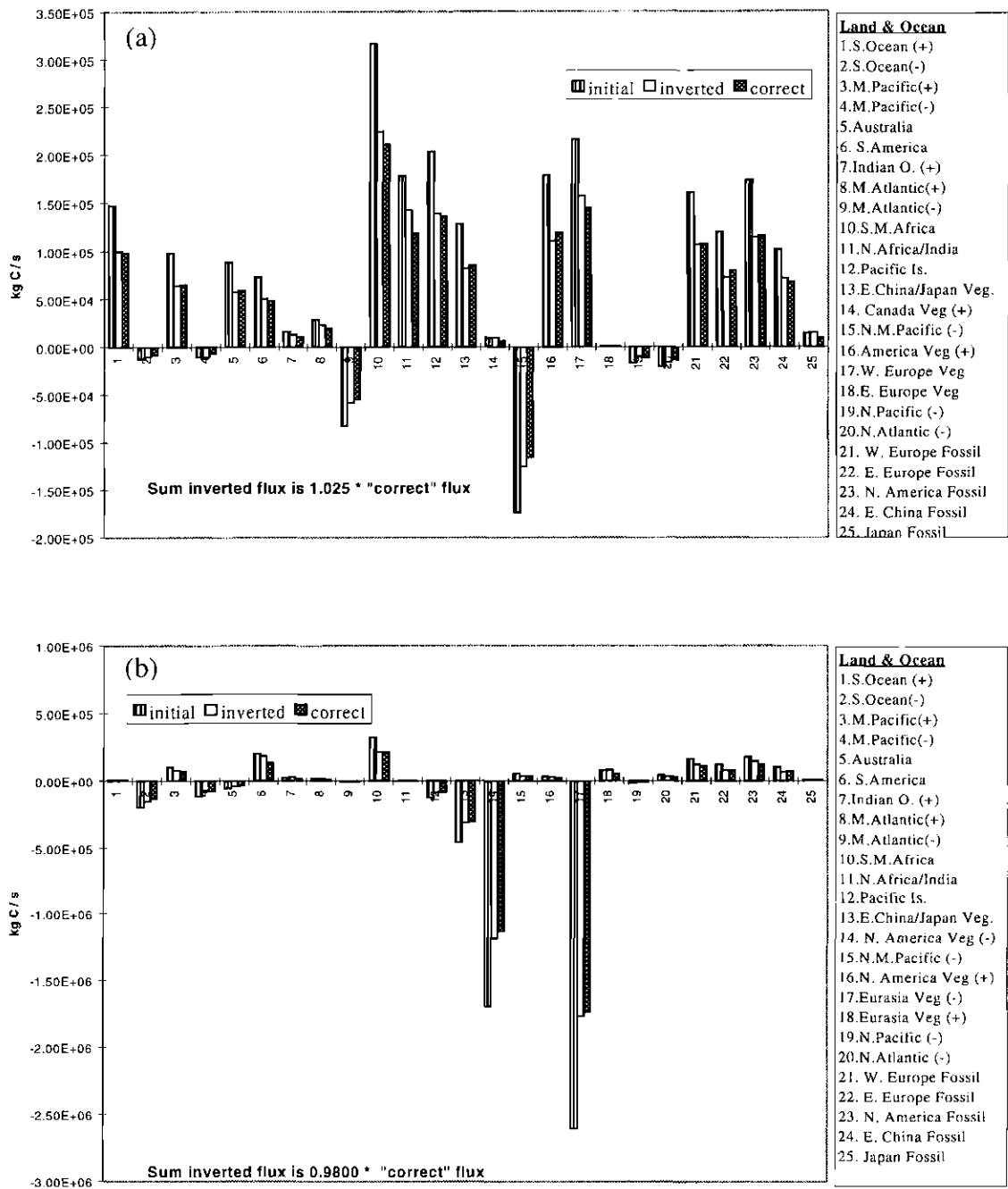
Figure 30: Pseudo data results for each region in (a) January and (b) July.

Table 15: Summary of Pseudo Data Results, where PAE is in percent (0% is perfect) and the net average with possible values of 1.5 (initial) to 1.00 ("correct").

Region	Location	PAE (%)	Net Average
1	Southern Ocean - positive flux	5.1	1.13
2	Southern Ocean - negative flux	5.4	1.15
3	Mid-South Pacific - positive flux	1.8	1.03
4	Mid-South Pacific & Indian Ocean-negative flux	4.9	1.08
5	Australia and New Zealand	4.1	1.05
6	South and Central America	13.6	1.32
7	Indian Ocean - positive flux	15.2	1.38
8	Mid Atlantic - positive flux	10.3	1.32
9	Mid Atlantic - negative flux	4.9	1.13
10	Mid-South Africa	5.3	1.12
11	North Africa to Western China	1.8	1.03
12	Pacific Islands and Thailand	5.4	1.12
13	E. China & Japan - vegetative flux	3.1	0.99
14	U.S. and Canada - negative flux	2.2	1.05
15	North- Mid Pacific - positive flux	13.6	1.07
16	U. S. and Canada - positive flux	2.3	0.96
17	Eurasia - negative flux	1.4	1.02
18	Eurasia - positive flux	1.9	1.00
19	North Pacific	4.3	1.08
20	North Atlantic	6.4	1.15
21	Western Europe	4.9	1.09
22	Eastern Europe / Eurasia	1.9	1.01
23	United States and Canada	3.9	1.06
24	E. China	4.4	1.08
25	Japan	10.2	1.31

These pseudo data results show that in each region, the technique converges on or towards the "correct" solution. As an overall summary, the pseudo data inverse results are within 3.07 % of the total "correct" flux for the entire year long inverse.

4.1c Distinguishing Land Vs. Ocean

Although the pseudo data tests presented in the previous section suggest that the inverse is capable of inverting for the large number of northern hemispheric regions, we want to further demonstrate how the observation sites can distinguish between land and ocean. This section is presented not to highlight the inverse capabilities, which have already been demonstrated, but actually to ensure that there is sufficient data to distinguish the signals between and land and ocean.

As already discussed, there must be an observation site sensitive to a change in flux for each region in order for the inverse to sucessfully converge on the solution. This information is all contained in the partial derivative matrix (Equation 5). The point that we wish to illustrate here is that given two sites at approximately the same latitude, one over land and one over ocean, the signals are different enough to distinguish between vegetative flux and oceanic flux. To demonstrate that there is different information, we can look at components of the partial derivative matrix. In Figure 31, the base case and the two north American vegetative flux perturbations are shown for a site in North Carolina (ITN) and a site in Bermuda (BME). Notice that the ITN site is far more sensitive to vegetative perturbations than Bermuda. In Figure 32, the Atlantic Ocean

perturbation is shown at the same two sites. In this case, Bermuda is far more sensitive than ITN to a change in ocean flux.
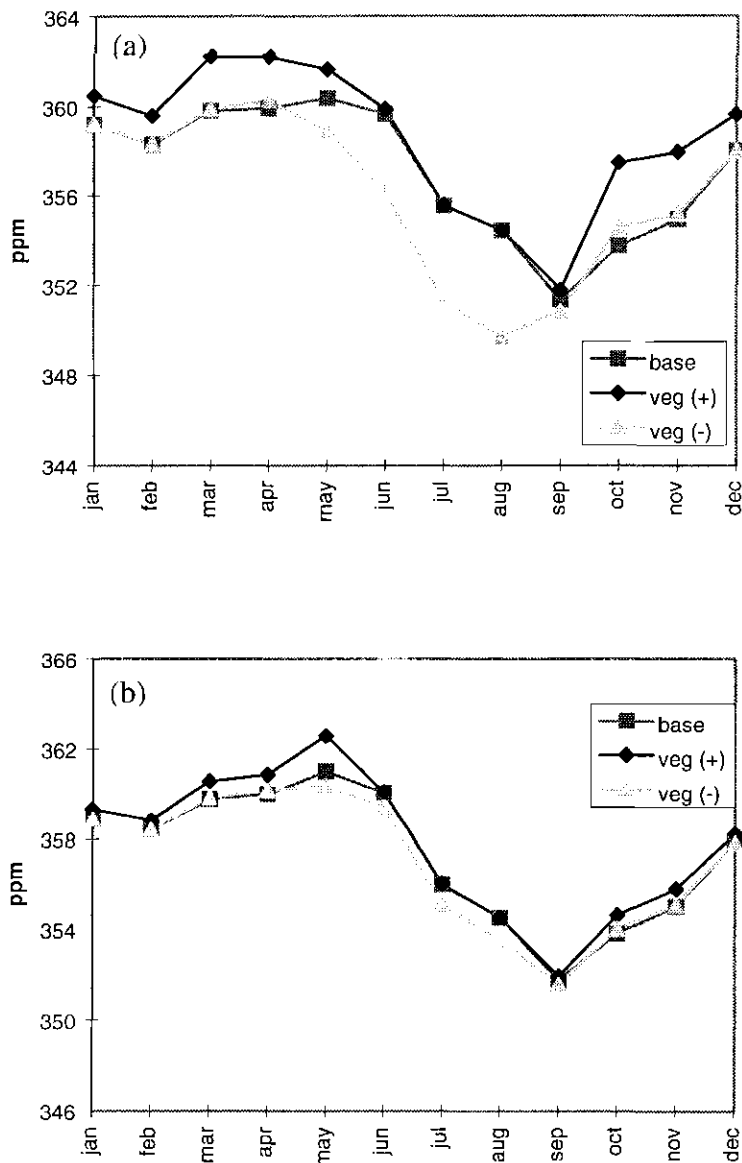

Figure 31: Base and North America Vegetative Components of the Partial Derivative Matrix at (a) North Carolina and (b) Bermuda Sites.
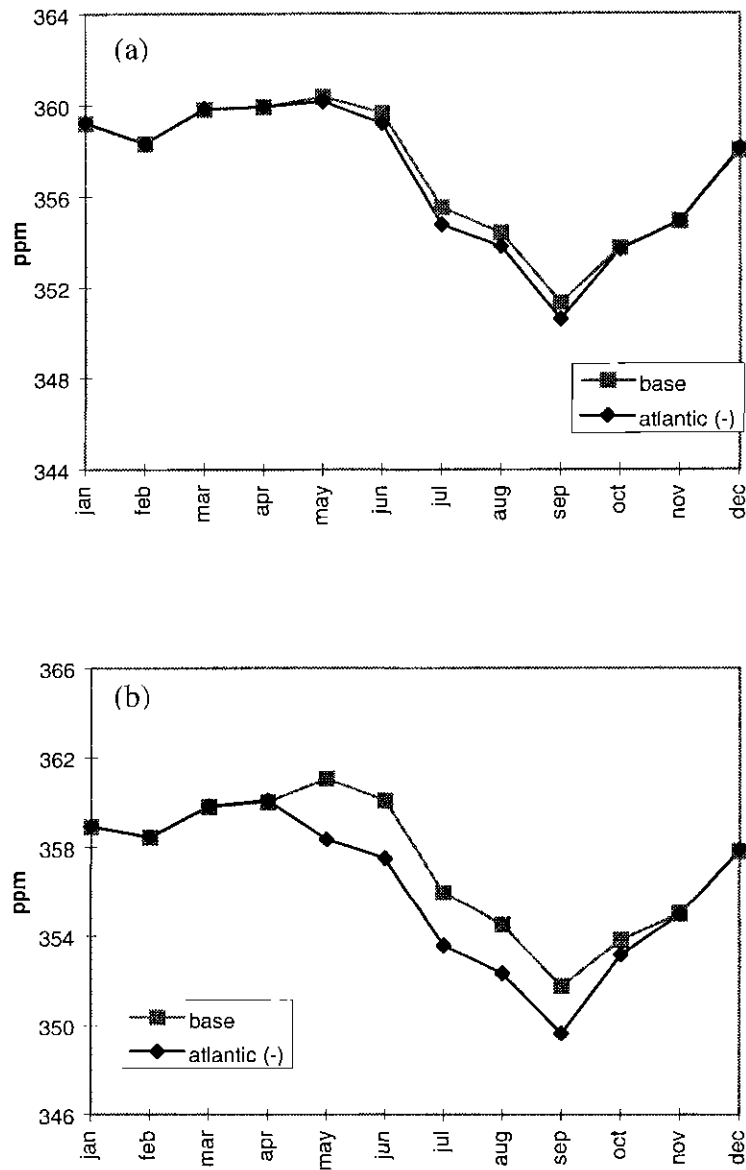
Figure 32: Base and Atlantic Ocean Components of the Partial Derivative Matrix at (a) North Carolina and (b) Bermuda Sites.

Given the different sensitivity of each of these sites to the vegetative and oceanic flux perturbations, the inverse should be able to successfully distinguish between land and oceanic signals. This has already been demonstrated in the previous pseudo data study; however, a final rigorous test is conducted here. In this case, the "poor" guess at emissions is set to be equal in every region, except for the fossil fuel components (only a 10% change). This is an extreme case that is much worse than any realistic scenario. This test is applied for one month, and the results are shown in Figure 33
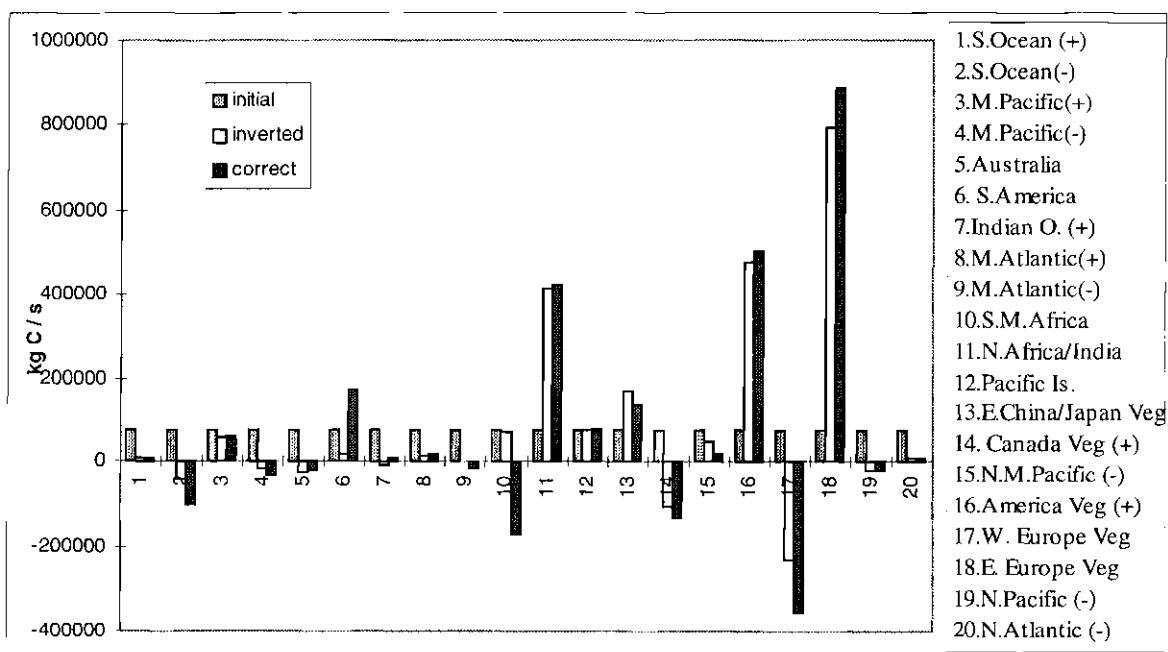

Figure 33: Pseudo Data Results Where all Regions Have the Same Initial Estimate. Values are shown in terms of gross error, where zero indicates a perfect solution.

To summarize these pseudo data results, the net solution is able to get within 11% of the total correct solution, and the average gross error for the regions is reduced from 225% to 44%. The northern hemispheric regions perform better than the southern hemispheric regions, but this is to be expected from the original pseudo data testing. The handling of the southern hemisphere will be addressed in the following section. Importantly though, this final test illustrates that even in an extreme case, the inverse solution is capable of distinguishing between land and ocean in the northern hemisphere.

4.1d Treatment of poorly constrained regions in southern hemisphere

While differentiation of vegetative and fossil fuel fluxes would be theoretically applicable to the southern hemisphere, the regions there are not as well constrained. In essence, the first division of regions was intended to do this by keeping vegetatively similar regions together; however, this inverse yielded unsatisfactory results (Figure 27) and the regions were regrouped to include different vegetative types. Therefore, it is not possible to divide the fluxes in the Southern Hemispheric continents in this manner.

There is a potential problem with regions that contain both positive and negative fluxes. If the inverse solution finds that the net flux for a region should change signs, then a non-physical situation arises. To explain this, consider the following hypothetical example, where region "A" is depicted in Figure 34. Assume that the right side is sub-tropical rainforest (sink for CO_2) and the left is plains and farmland (source of CO_2). Furthermore, there is only one measurement site that is nearby and it is marked with the

symbol: ⊗. The initial estimate (Figure 34a) shows a net positive for the whole region. The "correct" flux is somehow known (Figure 34b), and although the negative component had been initially estimated correctly, the positive component had been overestimated. Thus, the net flux should really be a sink. If the model were rerun with this new corrected flux scenario, the measurement site should show a slight decreased concentration, since it would be mostly affected by the east side (which stayed constant), but somewhat by the west (which decreased).

Now consider how the inverse would correct the initial estimate. The way that the inverse is posed, the solution is *only* for the net flux, which we will assume arrived at the "correct" net flux shown in Figure 34b. Once the solution is found, the initial distribution of fluxes is maintained within each region and simply multiplied by the new net flux. Therefore, when the net is redistributed in Figure 34c, the positive region becomes negative and vice versa. Now, the observation site will likely see an increased concentration since it will be disproportionately affected by the east side of the region. Furthermore, with the non-physical change of signs, as the inverse moves in time, one would expect increased errors arising as the model output at the observation site is disproportionately changed.
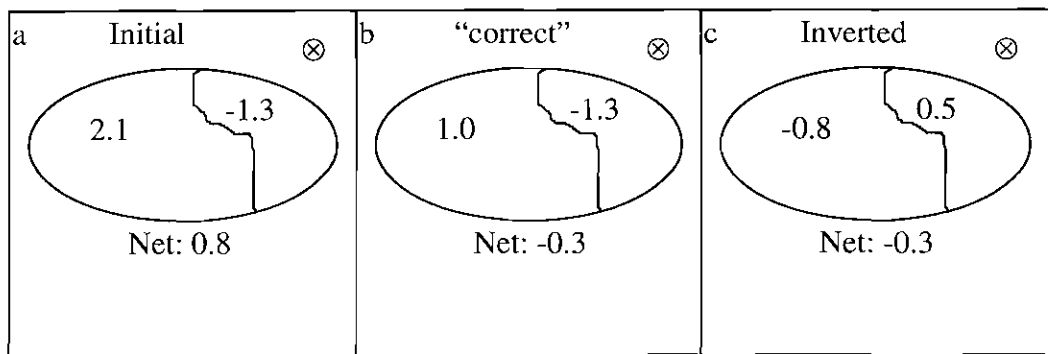
Figure 34: Hypothetical Region "A" with (a) initial distribution; (b) "correct" distribution; and (c) typical inverted distribution.

Clearly, it would be best to separate positive and negative flux regions (as done for the oceans and the northern hemisphere); however, there is not enough measurement information to separate these regions. The best that can be done is to ensure that the fluxes are not allowed to change so much that the net sign will be allowed to change. The only mechanism available to prevent this type of non-physical solution is to shrink the covariance matrix in these regions to limit the allowable step size of ΔE. Thus, the initial estimates are not as well known as the covariance matrix will suggest, but the optimized solution will not include these potential non-physical results. Instead, in these regions, the values will be constrained to the shape of the initial estimates. By limiting the step size, the inverted flux regions now retain the same sign on each side. As an example of how this would work, Figure 35c is an example of how the inverse with a tightened covariance matrix would look. In this case, the solution is an improvement over the inverted flux in Figure 34c; however, it is still not ideal.
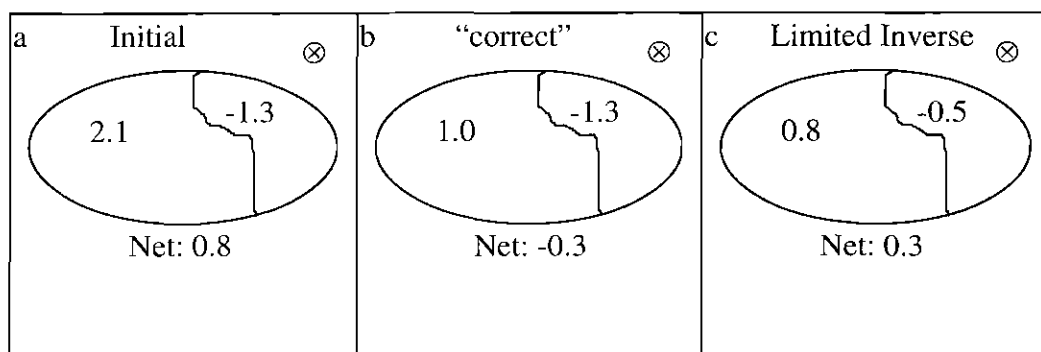
Figure 35: Hypothetical Region "A" with (a) initial distribution; (b) "correct" distribution; and (c) limited inverse distribution (ΔE constrained)

Despite this limitation, there are ways to improve future flux estimations. To do this, we tested the placement of various potential sites in order to increase the spatial resolution of the inverse in these regions. We find that for the northern part of Africa, a site in Israel that NOAA/CMDL began in late 1995 increases our capability for deducing this region. For the southern part of Africa, a new site around Port Elizabeth, South Africa would improve the inverse. To increase our capabilities in South America, it is crucial to re-open the SGI site that NOAA/CMDL discontinued after 1992. The addition of a site on the western coast of Chile will also increase capabilities for deducing South America, and a site in San Jereniano, Mexico (near Acapulco) will increase the resolution for separating central America from South America. In summary, the key to deducing a region is to have a site downwind that is sensitive to changes in flux from the region. If no site is sensitive to a region, it is impossible to invert for its flux. Thus, it is no surprise that the sparse CO_2 network in the southern hemisphere makes it impossible to separate regions into the same kind of resolution we can deduce in the northern hemisphere .

105

4.2 1992 inverted fluxes

Using the initial estimates described in section 3.3, we performed an actual inverse using measured mixing ratio data from 1992. In summary, the net flux of the inverted sources and sinks is much more balanced than the initial estimate. Overall, the inverted net ocean sink is 2.4 Gt C for the year, and net land (including all sources and sinks) is a source of 3.1 Gt C. This changed from the initial distribution of -0.25 Gt C for the ocean and 4.8 Gt C on land. A summary of the inverted fluxes compared to the initial estimate and the IPCC (1995) estimate is in Table 16. Aside from changes in the net flux, the regional distributions also changed considerably. Whereas the northern hemisphere was originally estimated as a very weak sink for vegetative uptake, the strength of the sink increased by 4 orders of magnitude. This change suggests that the "missing" carbon sink lies in the northern hemisphere biosphere. Below, each of the regions will be discussed in detail.

Table 16: Summary of balance of fluxes (Gt) with the different flux inventories.

Flux inventory	net ocean	net land	net N.H.veg.
initial	-0.25 ± 1.2	4.78 ± 6.7	-0.0006 ± 12.4
IPCC, 1995	-2.0 ± 0.8	6.6 ± 1.1	-0.5 ± 0.5
inverted	-2.43 ± 0.5	3.09 ± 1.1	-2.15 ± 1.2

4.2a Northern hemisphere vegetative flux

Summing together the inverted vegetative flux in Canada, America, Europe and Russia, the net flux is -1.6 Gt C/yr. This is a change from the initial estimate of -0.00059 Gt C/yr. Adding in the vegetative flux in Eastern China and Japan, the net northern hemisphere vegetative flux is -2.15 Gt C/yr. This represents a significant net difference from current estimates, attributable to the "missing sink" for carbon dioxide in the northern hemisphere. This net northern hemispheric vegetative flux is shown for each month in Figure 36.
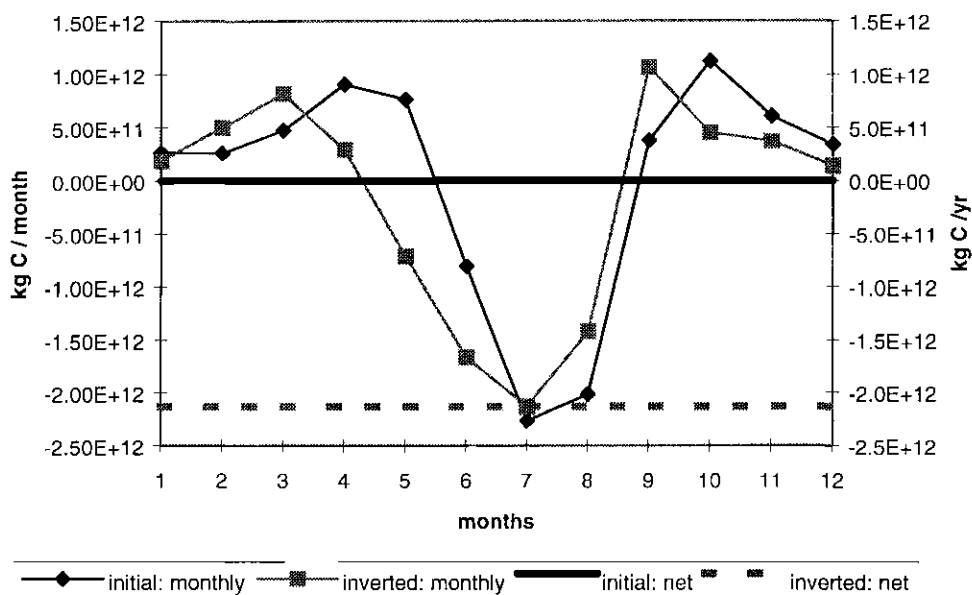
Figure 36: Net Northern Hemispheric Vegetative Flux

Notice that the vegetative uptake begins in May rather than June, as initially estimated. This indicates an earlier growing season than originally assumed. The month of May alone accounts for a net difference of 1.5 Gt difference. The initial estimate was a release of 0.76 Gt, and the inverted solution shows an uptake of approximately 0.72 Gt.

These results are supported by recent interpretations of satellite measurements of leaf index, where Myneni et al. (1997) found that the growing season has been extended by 10 days from 1981 to 1994, and that the magnitude of the leaf index has increased every year for the last 13 years, from 0.334 in 1981 to 0.361 in 1994. While this method cannot yet be converted into a flux of CO_2, taken in combination with our inverse results,

108

there is strong evidence that the "missing" sink is in fact the northern hemisphere biosphere. Other recent studies also find increased vegetative activity in the northern hemisphere (Chapin et al., 1996; Keeling et al., 1996).

Table 17 shows the net inverted flux for each of the main regions in the northern hemisphere. Notice that the majority of the net increased sink lies in the continental U.S. (60%), with the second largest component of the net sink in Eastern China (24%).

Table 17: Net Initial and Inverted Northern Hemisphere Vegetative Flux.

Region	initial	inverted
Canada + Alaska	0.00005	-0.23
United States	-0.00003	-1.34
West Europe	-0.00007	0.12
East Europe	0.00007	-0.18
Russia	-0.00009	0.02
Japan	-0.00001	-0.02
E. China	0.00008	-0.52
Sum	0.00001	-2.15

Taking a closer look at each of these regions (Figure 37), there are two main differences between the initial and the inverted solution. First, in almost every region (except E. China), the earlier growing season is evident. Second, in the two largest sink regions (the U.S. and E. China), there is a decreased winter

source. This is not true of any of the other regions. These are the only two regions that contain subtropical latitudes. This suggests that the effect of year round photosynthesis in subtropical areas may have been underestimated previously.
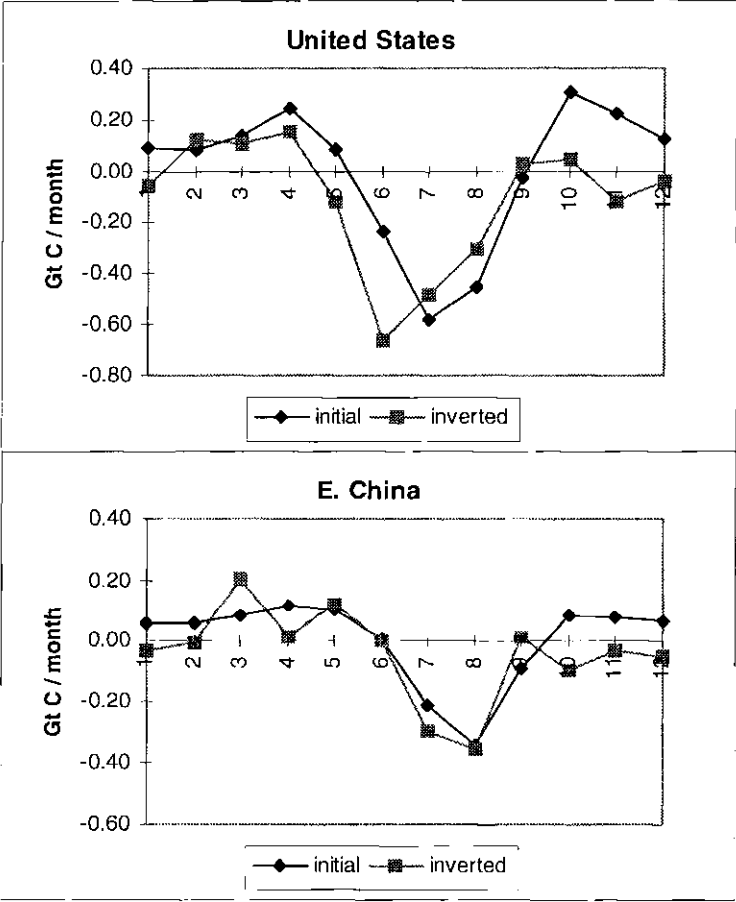

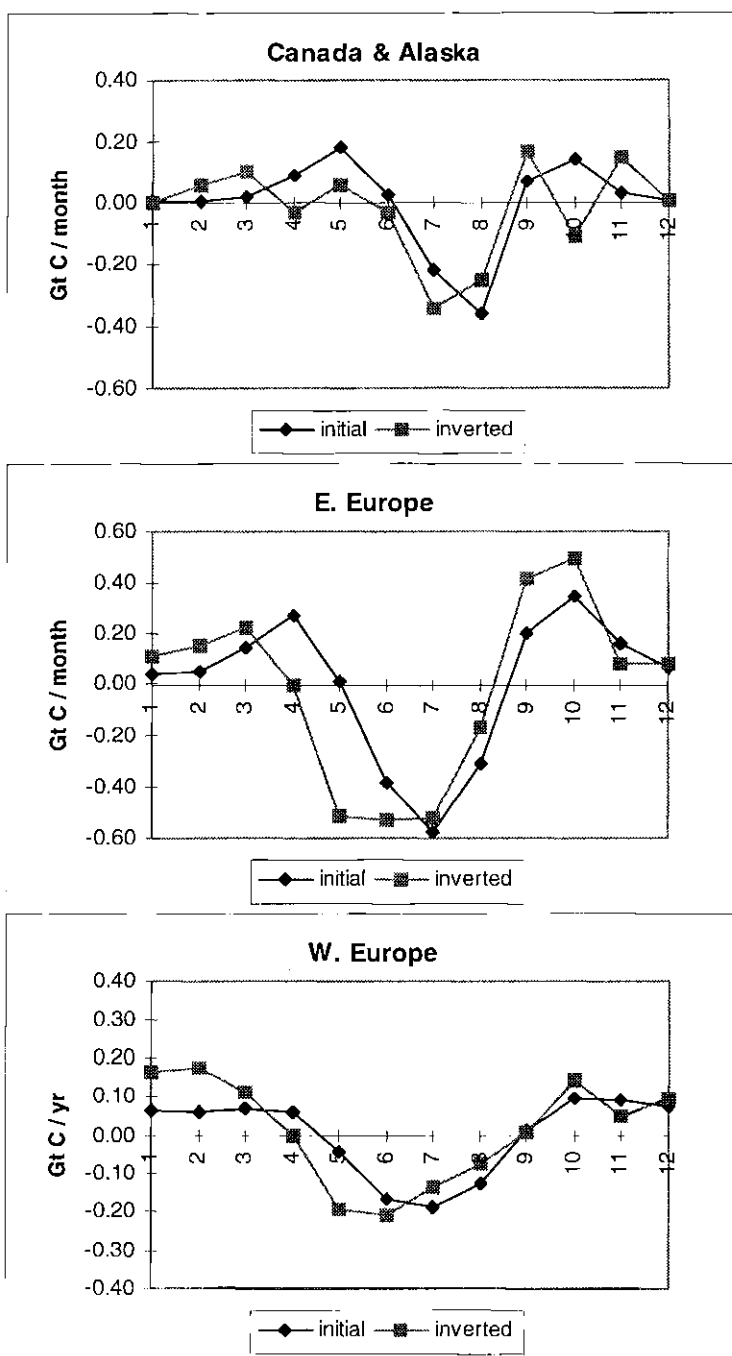Figure 37: Northern Hemisphere Vegetative Flux Shown by Region.
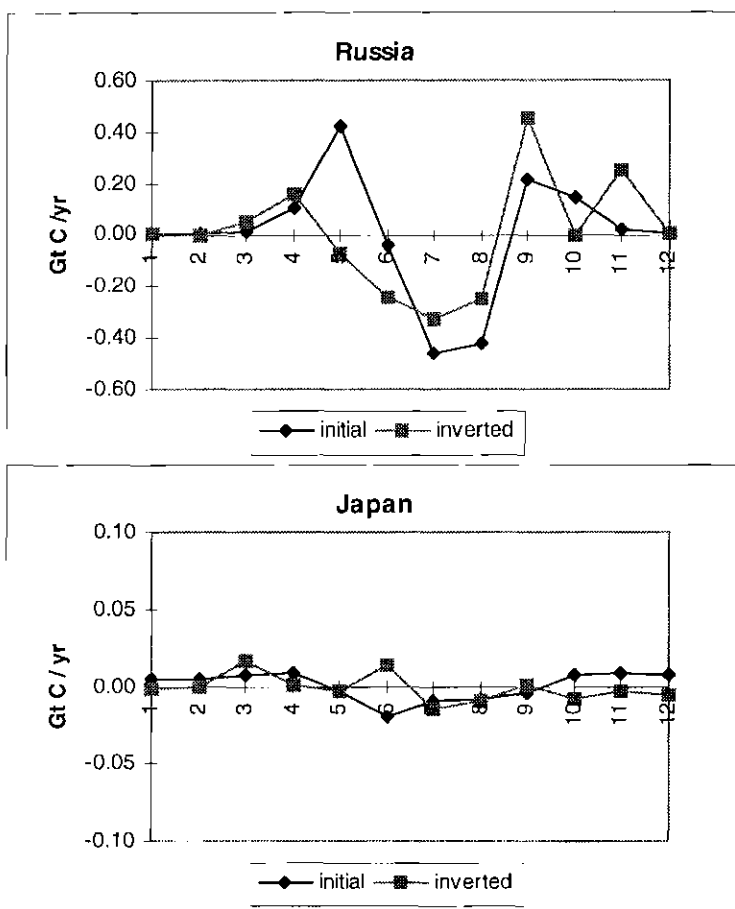
Figure 37 (continued)

111

Figure 37 (continued).

To ensure that the increased vegetative uptake in the month of May is not an effect of a large covariance matrix, the inverse was repeated giving an identical weighting to all of the regions (except northern hemisphere fossil fuel). The resulting solution changed the vegetative sink by ~6%. To put this in perspective, the change from the initial to the inverted estimate was an order of magnitude (194%). Thus, the increased sink in the vegetation is not an artificat of a large uncertainty in the covariance matrix.

4.2b Northern hemisphere fossil fuel source

The net fossil fuel estimate in the northern hemisphere increased by a negligible amount (less than 1%). However, we assumed that the emissions were initially well known (within 10%) and did not give these regions the flexibility to change significantly. Thus, we can draw no conclusions about the fossil fuel estimate based on the inverse.

4.2c Ocean flux

The oceans started at a net flux of -0.25 Gt C for the year, which we noted earlier was about an order of magnitude smaller than most commonly accepted estimates (IPCC, 1995). The net inverted flux is -2.43 Gt C / year. This is much more inline with the "best" estimates. However, as discussed previously, the initial data set used here was the only available data set that varied by month, and it started off about an order of magnitude too small. Thus, the ocean now agrees with the best estimates, but is not contributing significantly to the "missing sink."

The only other available gridded ocean flux estimate is available from GISS and it is an annual average by Broeker et al. (1986), hereafter refered to as GISS-OCN. While this is not a good data set to use as an initial condition, it is useful to compare the net annual inverted flux to this annual data set. In absolute magnitude, the two are very close, differing by only ~0.1 Gt C/yr; the net inverted flux is much closer to this estimate than the initial Erickson estimate. Regionally, however, there are some interesting differences. Table 18 compares the annual net flux of the initial estimate (Erickson), the

113

GISS-OCN annual estimate, and the inverted solution. Figure 38(a-c) shows the net inverted ocean flux, the GISS-OCN estimate, and the difference of the two data sets respectively. Also shown in Figure 38(c) are shaded regions of statistical significance from a t-test (from Equation 12).

Table 18: Summary of Ocean Regions for Initial estimate (Erickson), GISS-OCN estimate and Inverted solution.

Ocean Region	Erickson (Gt C/yr)	GISS-OCN (Gt C/yr)	Inverted (Gt C/yr)
South Ocean (> 60°)	-0.22	-2.9	-0.83
Temperate S. Pacific	-0.13	-0.23	-0.51
Temperate S. Atlantic	-0.02	0.03	-0.04
Tropical Pacific	0.49	1.0	0.07
Tropical Atlantic	0.09	0.18	0.04
Temperate N. Pacific	-0.14	-0.01	-0.22
Temperate N. Atlantic	0.01	-0.21	-0.16
N. Pacific	-0.01	0.20	-0.19
N. Atlantic	-0.01	-0.46	-0.25
Indian	-0.04	-0.14	-0.32

90.0
67.5
45.0
22.5
0.0
-22.5
-45.0
-67.5
-90.0

(a)

Contours from −.04212 to .00788 by .01 (kg/m2/yr)

90.0
67.5
45.0
22.5
0.0
-22.5
-45.0
-67.5
-90.0

(b)

Contours from −.1042 to .0458 by .01 (kg/m2/yr)

90.0
67.5
45.0
22.5
0.0
-22.5
-45.0
-67.5
-90.0

(c)

Contours from −.0522 to .0877 by .02 (kg/m2/yr)
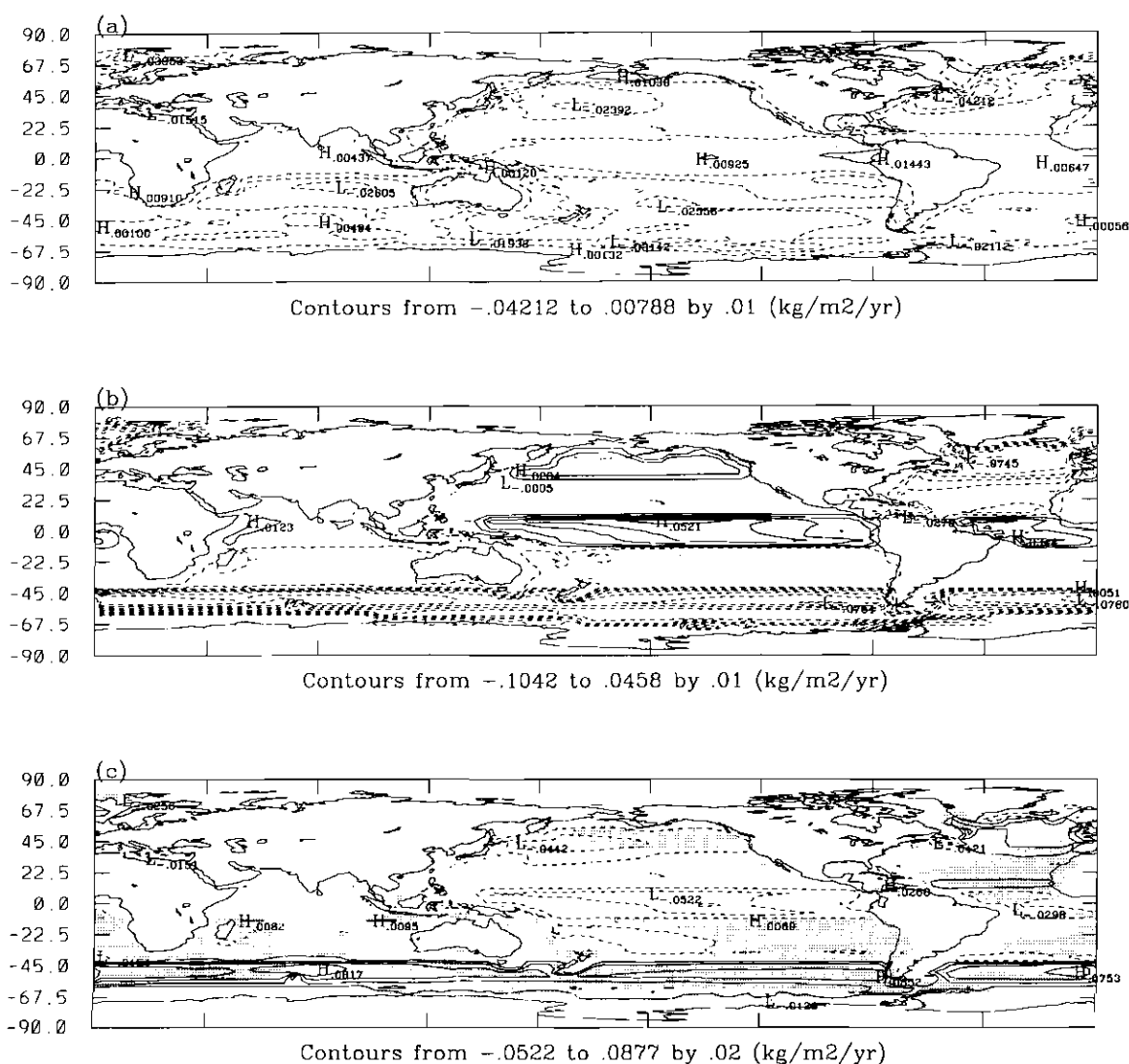
Figure 38: Annual ocean fluxes, with (a) the inverted solution, (b) the GISS-OCN estimate, and (c) the difference (a-b). (Note: solid lines are positive and dashed are negative, shaded values are statistically significant differences from a t-test.)

Looking at the differences in the various ocean regions, the inverted solution of the north Pacific (monthly inverted values are in Figure 39) shows a net sink of -0.40 Gt C/yr, while the GISS-OCN estimate shows the region as a net source of 0.19 Gt C/yr.

The north Atlantic (monthly inverted fluxes are in Figure 40) in both estimates shows an net sink; however, the inverted solution is a weaker sink (-0.41 Gt C /yr) than the GISS-OCN estimate (-0.67 Gt C /yr). Note that in both Figure 39 and Figure 40, the GISS-OCN estimate is shown for reference.
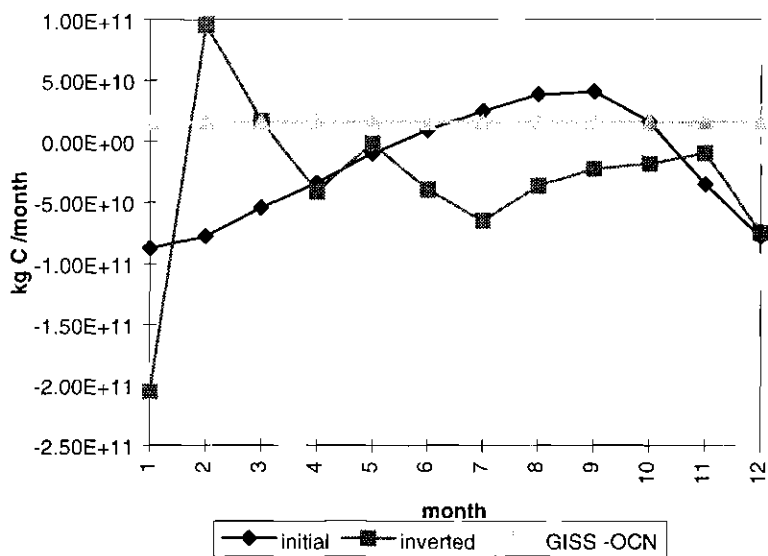
Figure 39: North Pacific inverted fluxes.
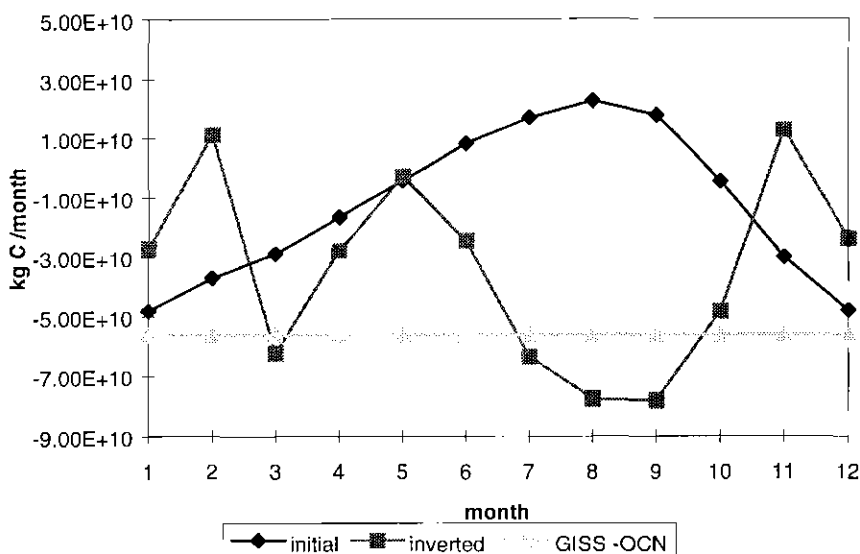
Figure 40: North Atlantic inverted fluxes

Both of the inverted solutions in these regions seem to oscillate in the winter months, with the north Atlantic oscillating into spring. There are several possible causes of this oscillation. First, the regions are not ideally constrained, and so in subsequent months the filter may be over-correcting for the estimates from previous months. However, from Table 15, the north pacific and north Atlantic (regions 8, 9, 15, 19, and 20) regions except for the mid north Atlantic positive component show substantial improvements from the initial guess. Second, it is possible that the fluxes are actually oscillating due to the non-linear uptake effects in the ocean. Unfortunately, there are no other monthly estimates to verify this possibility. Finally, the third possibility is that the initial distribution may be in error and biasing the outcome. This possibility will be explored in detail in section 5.3 using sensitivity studies.

117

In the tropics, both the inverted solution and the GISS-OCN estimate show sources, but the inverted region is a weaker source. The inverted Indian ocean (monthly values shown in Figure 41) shows a stronger sink (-0.32 Gt C/yr) than estimated by GISS-OCN (-0.14 Gt C/yr). The southern ocean is a net sink in both estimates, but the GISS-OCN data set estimates a much stronger sink (-2.9 Gt C /yr) than the inverted estimate of -0.82 Gt C/yr.
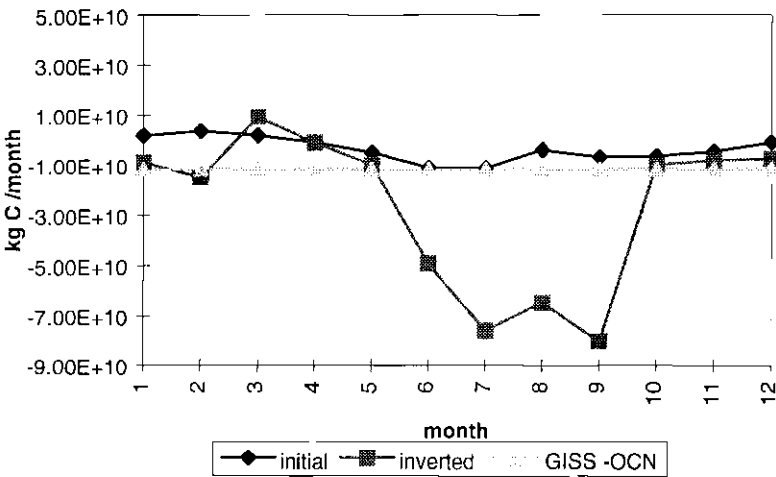

Figure 41: Indian Ocean inverted fluxes.

Table 18 is a summary of the net Erickson (initial) ocean fluxes, the GISS-OCN ocean flux, and the inverted solution. Notice that the GISS-OCN and inverted solution agree more closely to each other (both in sign and magnitude) than either to the initial Erickson estimate.

4.2d Southern hemisphere land flux

The other initial flux believed to be underestimated when compared to IPCC (1995) estimate was the land use change / biomass burning source. The initial estimate used was believed to be about an order of magnitude too small. Unfortunately, we were not able to separate that component since it is predominately occurs in Africa, Asia, and South America. Furthermore, as discussed in section 4.1, we put a limit on how much we would allow these regions to change so that the sign of the flux would reverse. There are however some interesting qualitative results. The sum flux increased by 37% in South America, as did Northern Africa and India by 33%; however, the flux in Mid-South Africa decreased by 98%. Thus, the combined net increase in these three regions was only 14% (or 0.2 Gt C). While this hardly makes up for our assumed underestimation of 1 Gt from deforestation, the regions had to be constrained to small changes due to the lack of sensitive observations nearby.

4.2e Net balance of fluxes

Adding up the net land source and the net ocean sink, the sum is 2.12 Gt C for the inverted solution versus 4.54 Gt C for the initial estimate. Looking at the inverse from a net balance perspective, the sum of sources, sinks, and atmospheric increase should balance to zero. Although the IPCC (1995) estimate for atmospheric increase is 3.3 ± 0.2 Gt C/yr, this estimate is for the 1980s. There was an observed slowing in the trend of atmospheric growth rate of carbon dioxide in the early 1990s (Sarmiento, 1994; Conway

et al., 1994). There have been several suggestions to explain this slowed rate of increase, including the eruption of Mt. Pinatubo (Sarmiento, 1994), changes in fossil fuel use in Eastern Europe after the fall of communism (Grub, 1994), and increased vegetative uptake due to global warming (Myneni et al., 1997). While the cause is uncertain, it is clear that the rate of atmospheric increase during the simulated time period is much smaller than the IPCC estimate. Using NOAA/CMDL data, Conway et al. (1994) estimate the atmospheric increase for 1992 to be 1.39 ± 0.11 Gt C.

When using the Conway et al. (1994) estimate, the initial flux values are over-predicting by 3.2 Gt, or a "missing sink" of 3.2 Gt C. The inverted estimate is now underpredicting by -0.7 Gt C. Since the sum should in theory be close to zero, the net balance for the 1992 inverse optimization is significantly improved over the initial estimates. As shown above (section 5.2a), this change is largely due to increased vegetative uptake in the northern hemisphere. The flux from the oceans also increased by an order of magnitude compared to the initial estimate; however, the initial value was approximately an order of magnitude smaller than the accepted value for oceanic uptake. Thus, we infer that the "missing" sink actually lies in the northern hemisphere biosphere.

According to the inverse, there is now a "missing" source, albeit relatively small by comparison. One probable explanation is that there is still an underestimated land use change source. Since this source lies mostly in the poorly constrained southern hemisphere land, and the inverted solution was limited in these regions, it follows that there is still a large uncertainty here.

120

4.2f Model output compared to observations

One of the main reasons to perform the inverse is because the initial flux estimates do not agree well with observations. Using these initial estimates, the model output overpredicts when compared to observations; hence, the "missing" sink. Thus, another way to measure the success of the inverse is to compare the model output from the inverse solution to real observations. For each site that had observations for a majority of the inverted year (1992), a month by month comparison can be found in Figure 42.

Statistically, there are two measures that one can use for comparison. The first is the gross error, defined by the following equation:

$$\% \, gross \, error = \sum_{i=1}^{N} \left| \frac{\left(\chi_i^{obs} - \chi_i^{model} \right) / \chi_i^{obs}}{N} \right| * 100$$

However, in inverse studies, a better measure of performance is a normalized bias (Chang et al., 1996), since this shows whether the estimates are over or underpredicting when compared to observations. This can be calculated with the following equation:

$$\% \, normalized \, bias = \sum_{i=1}^{N} \frac{\left(\chi_i^{obs} - \chi_i^{model} \right) / \chi_i^{obs}}{N} * 100$$

Using these statistical measures, Table 19 shows the performance of the inverse and of the initial inventory when compared to observations. In terms of a gross error, the initial estimate and inverted solution have errors of 0.64% and 0.29%, respectively. In terms of

121

a normalized bias, the initial estimate under-predicts by -0.47% and the inverted results over-predicts by 0.0034%. This shows a substantial improvement in the inverted solution over the initial in terms of agreeing with observations.

Table 19: Statistical measure of model output compared to observations.

flux inventory	gross error (%)	normalized bias (%)
initial	0.64	-0.47
inverted	0.29	0.003

123
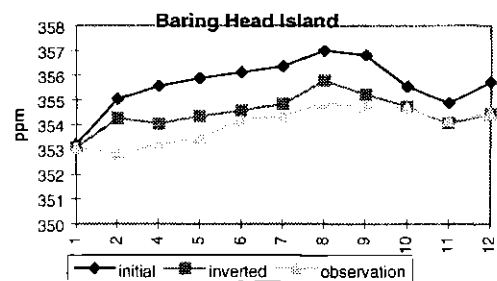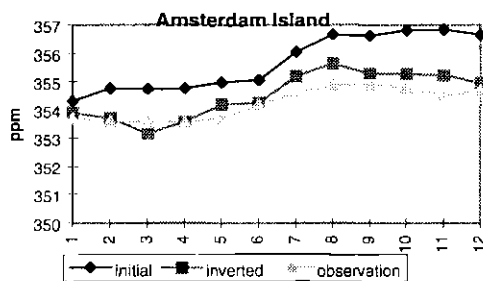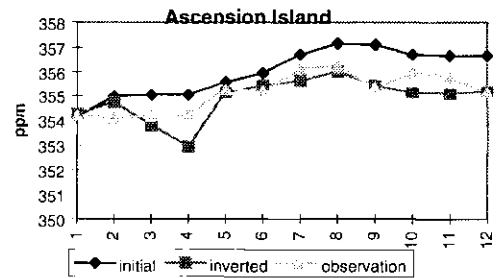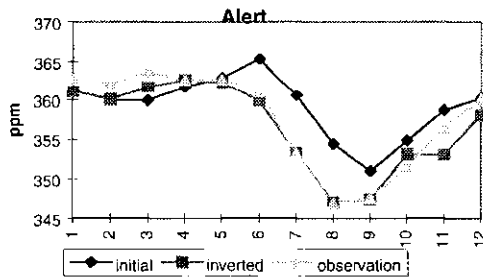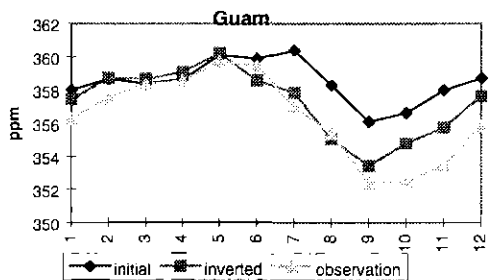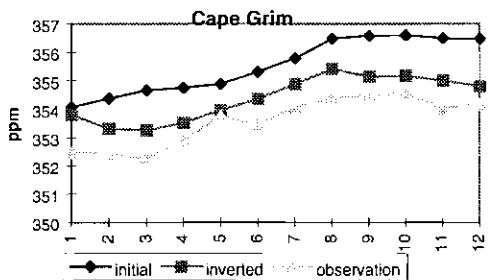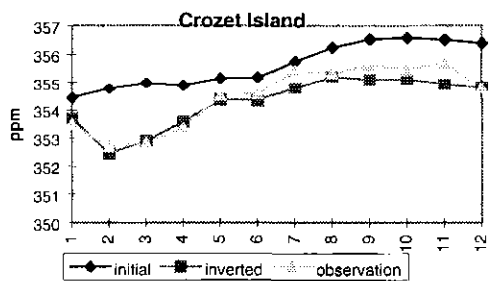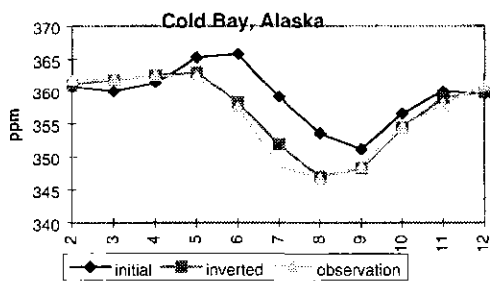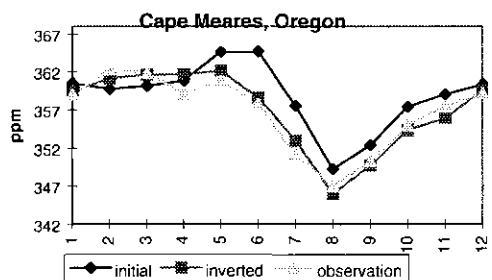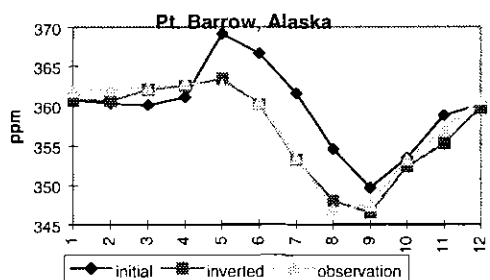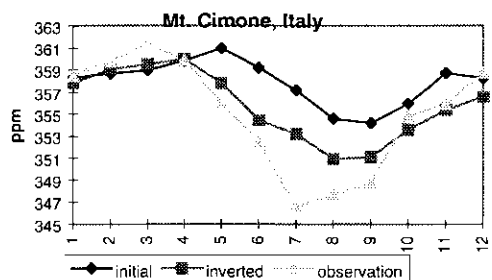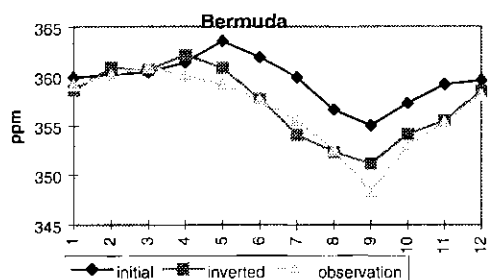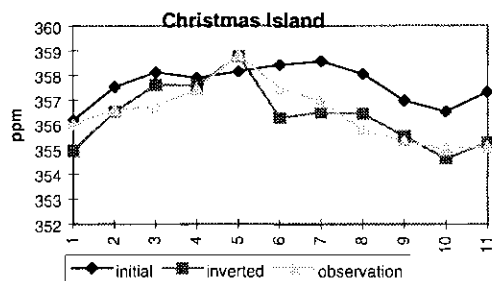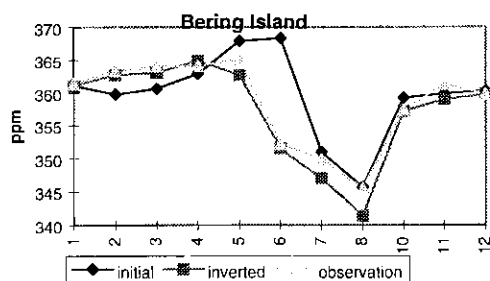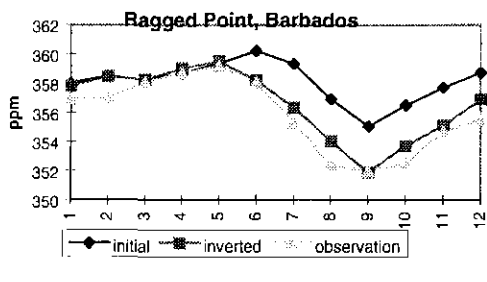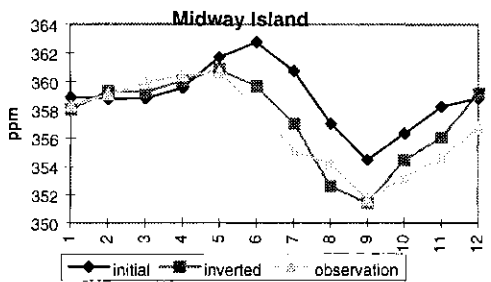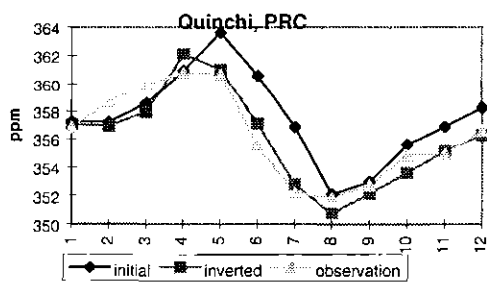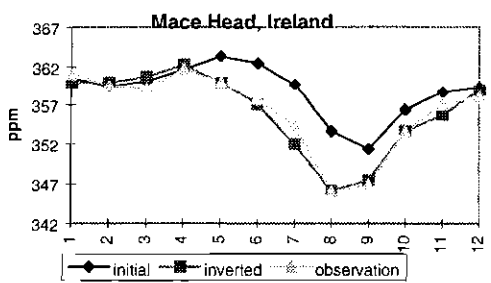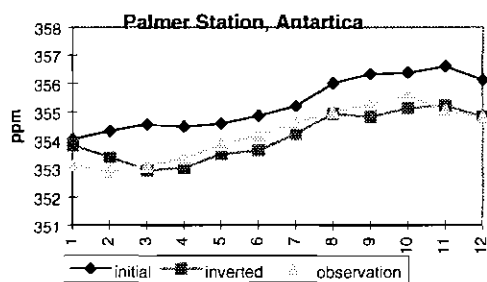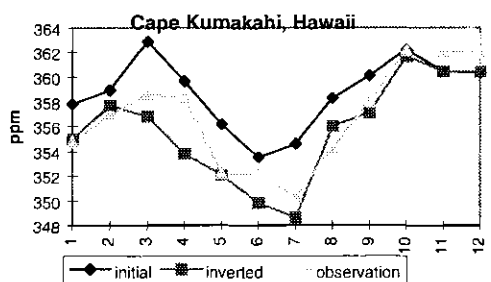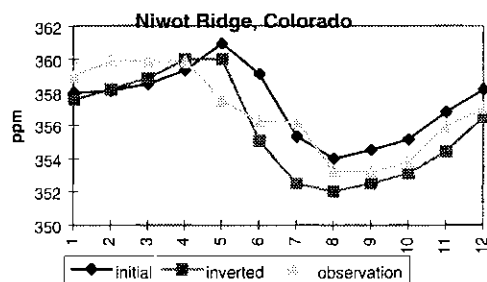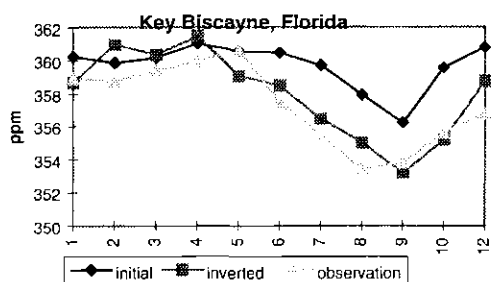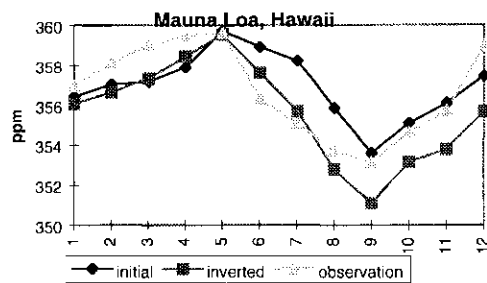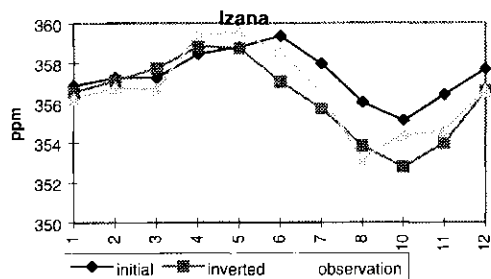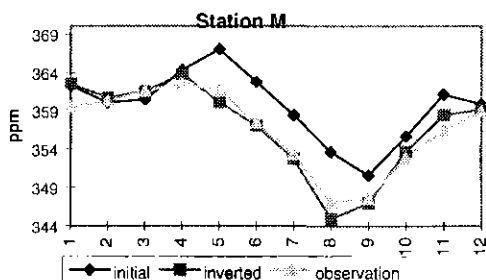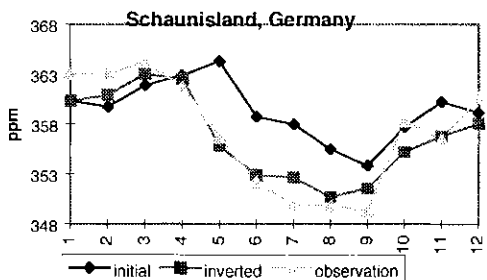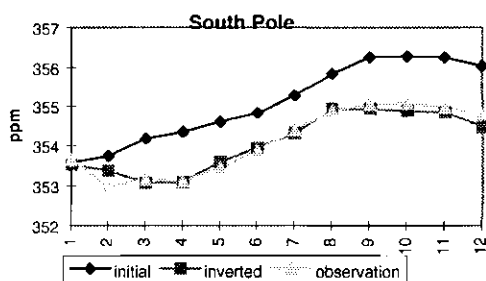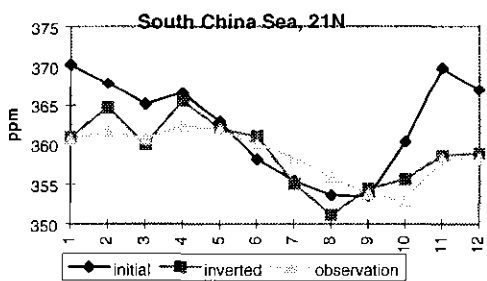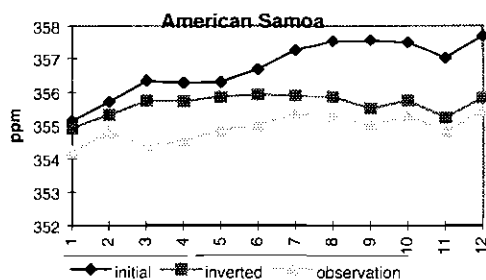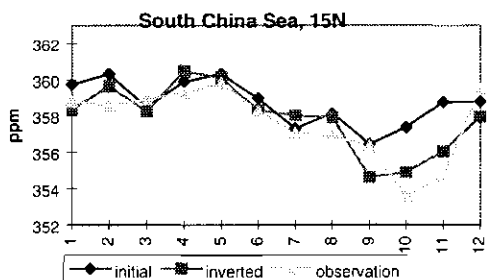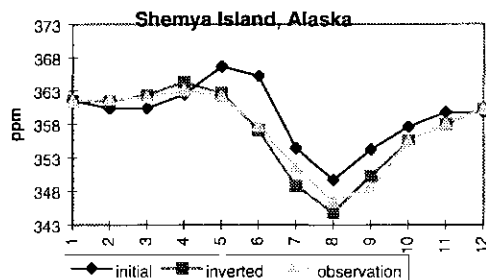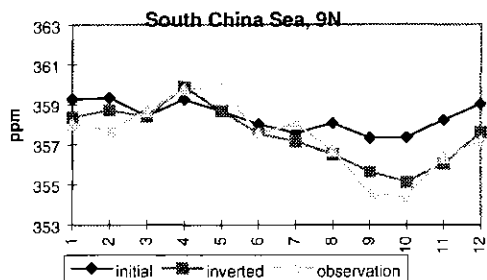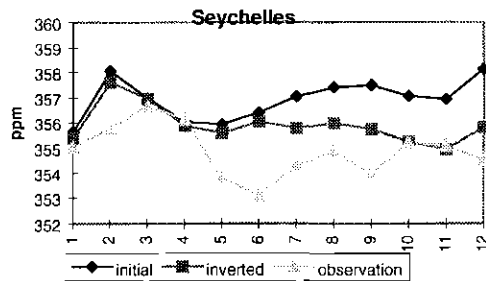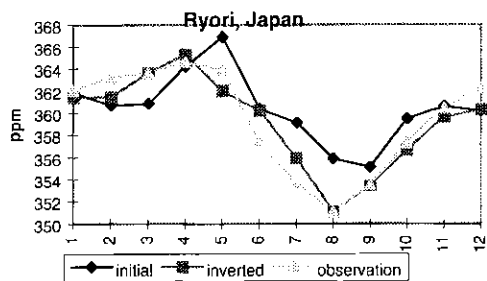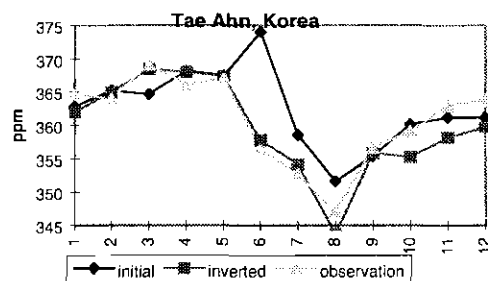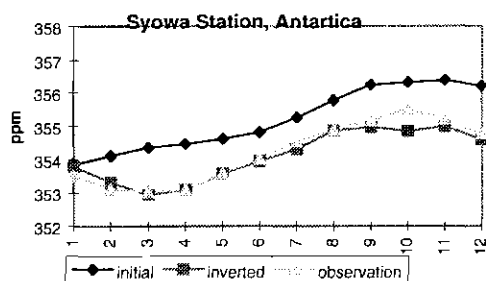
Figure 42: Model output compared to observations using initial and inverted fluxes.

Looking at Figure 42, note that at almost all observation sites, the January initial estimate is much closer to observations than any other time of the year. This is because the values were optimized to agree in the spin-up and error from a poor estimates has not had time to accumulate. Recalling that one of the possible explanations for the winter ocean oscillations was a poor initialization, one can see from these plots that this is not the case.

4.2g Linearity of the system

Although the inverse greatly increases agreement with observations, one final assurance that it is working properly is to test whether it behaves linearly. We assume linearity in the system, which is to say that a change in the state vector should have an associated linear change in the measurement. This is an easy relationship to test. Since we perform the inverse off-line, we use a partial derivative matrix to predict what the model output will be as a result of the inverted solution. However, as described in section 2.2, we then re-simulate each month with the new inverted flux estimates. This

126

insures that we do not propagate a "bad" guess in time (discussion in section 2.2). By comparing the predicted model output to the actual model output in the rerun, we can test if the system is behaving linearly. Throughout the one year simulation, the average difference in the predicted model output versus actual was 0.014ppm. Thus, the system is behaves linearly.

4.3 1992 Inverted Fluxes: Sensitivity Studies

This section addresses the sensitivity of the inverse solution to its various input parameters. Although the filter accounts for errors when calculating the solution, it is important to test how robust the solution given realistic errors in the inputs. First, as described in detail in section 3.2, the observations contain errors due to various factors. These factors include measurement error, natural variability, and sampling frequency errors. To address how sensitive the inverse is to error in the observations, the first test (section 4.3a) adds Gaussian white noise to the observations and examines how the inverse solution changes as a result.

The other main input is the initial estimate of fluxes, which consists of two parts. The first is the absolute magnitude of the net sum of fluxes and the second is the distribution of the flux within each region. From pseudo data studies, we know that the inverse is much more sensitive to distribution than to the net flux. This makes intuitive sense. If the initial distribution shifted the boreal forest in Russia by 10 degrees latitude, one would expect a difference in the sensitivity of the Bering Island site to Russia in the

127

partial derivative matrix. However, assuming a linear system (already shown in section 4.2g), a change in the net magnitude should not change the partial derivative matrix. We ran two tests that alter the initial distribution within a region, maintaining the same net value, to determine how sensitive the inverse is to errors in the distribution.

4.3a Random noise added to observations

To determine how the inverted solution changes due to measurement errors, random Gaussian white noise was added to the measurements. The random values were generated by a computer. Since a computer is incapable of generating truly random numbers, they are actually pseudo-random. The amount of noise that we added was Gaussian, meaning that 60% falls within one standard deviation, 95% within two standard deviations, and 99.6% within three standard deviations. As the goal was to vary the observations within the range of the noise matrix, we set the standard deviation to be half of the actual standard deviation of the measurement. Thus, 95% of the values fall within the actual standard deviation of the measurement.

The net results using these measurements are summarized in Table 20, with the initial and regular inverted fluxes included for comparison. For the sensitivity study, the northern hemisphere flux is of particular interest, since this is where the inverse suggests the "missing" sink lies. Thus, we have included a column in Table 20 with the net vegetative uptake in the northern hemisphere.

128

Table 20: Summary of balance of fluxes (Gt) with the different flux inventories.

Flux inventory	net ocean	net land	net N.H.veg.
inverted	-2.43 ± 0.54	3.09 ± 1.1	-2.15 ± 1.2
inverted with noise	-2.30 ± 0.54	3.27 ± 1.1	-2.20 ± 1.2

From a net standpoint, there is not a significant change in the allocation of fluxes between ocean and land. However, when individual regions are examined, some interesting results arise. First, in regions that are well constrained and influenced by multiple sites, the fluxes change insignificantly. For example, Figure 43 is equivalent to Figure 36, with the new inverted flux (with noise) added. The difference in the two inverted fluxes is almost imperceptible. This attests to the robustness of the solution and supports the conclusion that the missing sink found in the northern hemisphere vegetation is a real effect.
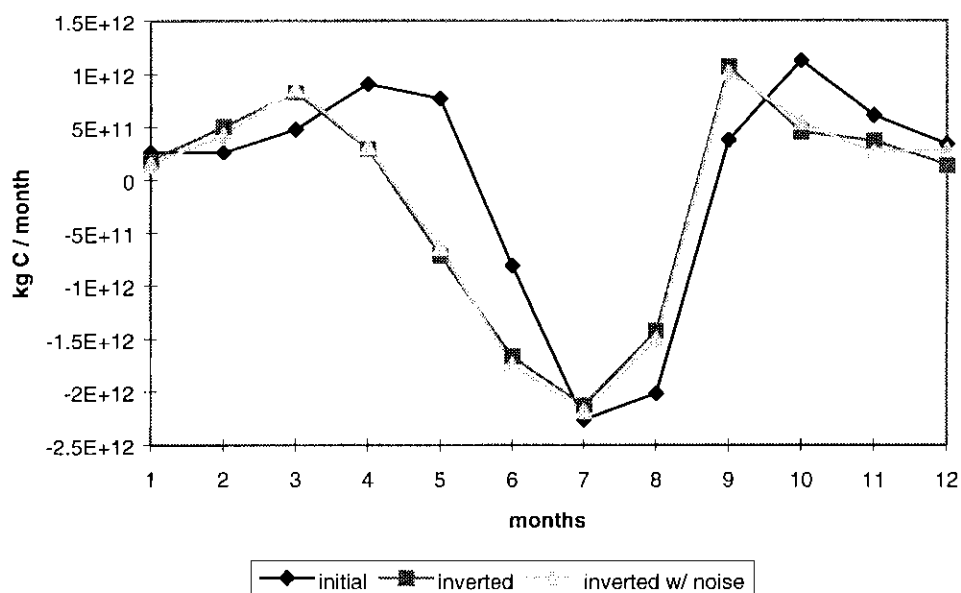
Figure 43: Northern Hemisphere Vegetative Flux including initial, inverted, and inverted with noise.

However, in regions that were not as well constrained (the southern hemisphere) the difference between the two inverted inventories is not as trivial. The differences for the entire world are shown in Figure 44. Although most of the differences can be seen in Africa, Asia, and South America, none of these differences were statistically significant when a t-test was performed as in Equation 12. This shows that the inverse solution is robust.
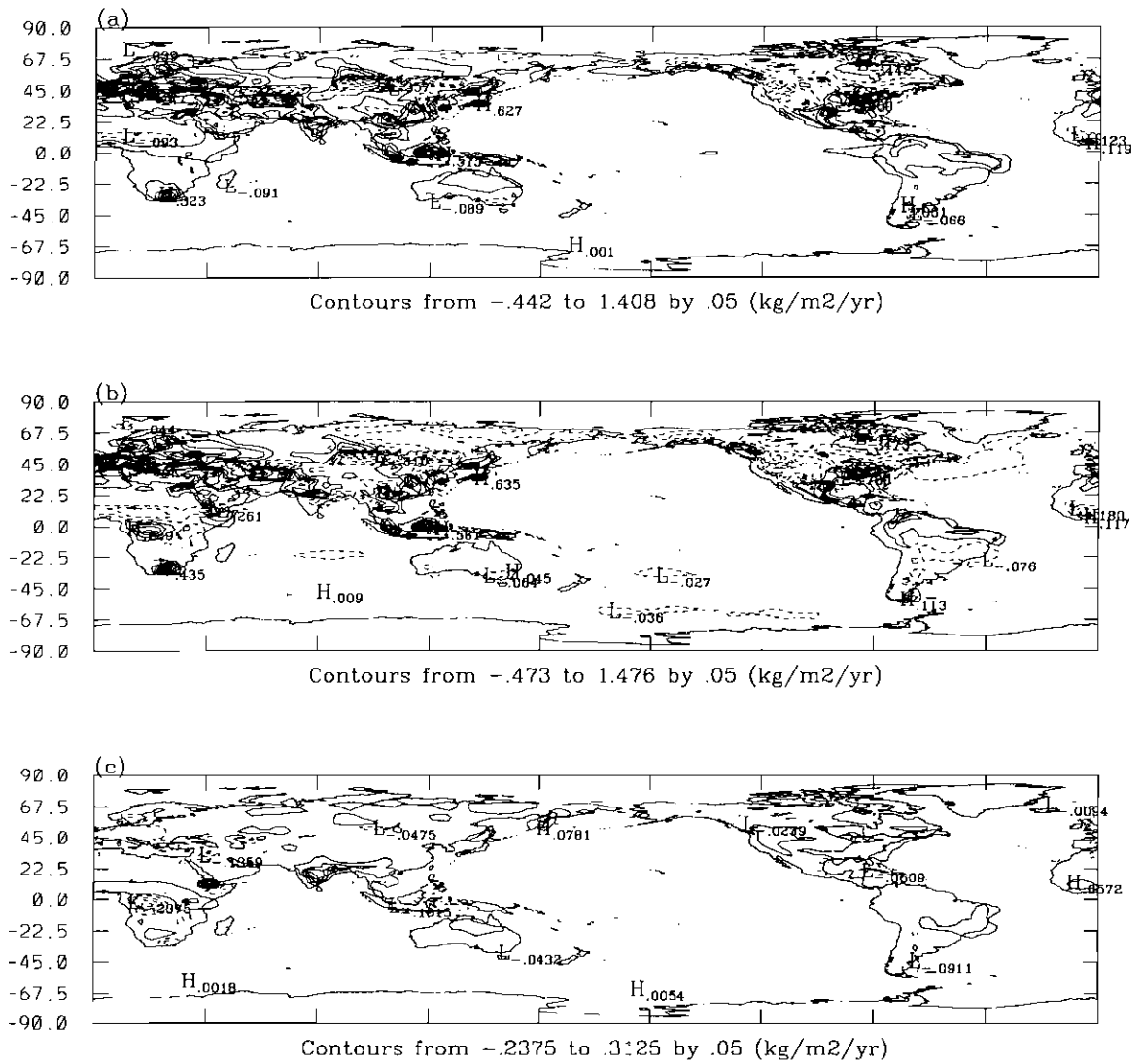
Figure 44: Sensitivity to white noise, where (a) is inverted flux, (b) is inverted flux with white noise, and (c) is the difference (a-b).

We already knew that the inverted results in these regions should be considered with caution (Section 4.1), but this sensitivity test shows the importance of improving the data network in the region if we ever wish be perform a meaningful inverse for these regions. It is interesting to note that the net balance actually improves with the added

noise. Accounting for most of the difference, the net source in Africa, India, and South America changes from 0.2 to 0.33, due only to the added Gaussian white noise to the observations. This illustrates how sensitive these regions are to changes in observations due to their poor constraint. This also adds support to the hypothesis that the underestimation of fluxes lies in an underestimated source in these regions.

4.3b Shifted initial distributions within a region

The other potential source of error is the flux distribution within a region. Only the net flux value is inverted for, and then the inverted solution is redistributed according to the initial distribution (refer back to Figure 34). To test how sensitive the inverse is to errors in the initial distribution, two tests were developed. The first assumes a slight error exists in the initial distribution, which is relevant to vegetative distributions. From remotes sensing techniques, the location of forests is fairly well constrained; and thus, the distribution of vegetative flux is fairly accurate. This is not to say that the actual initial magnitude of fluxes are correct, but that they are spatially representative. For example, if the initial distribution put a large vegetative flux in Arizona, clearly something would be amiss. It is more probable that the initial vegetative distribution contains small errors.

To simulate this type of error, the vegetative flux in America was shifted by one grid cell from the initial distribution. After rerunning the partial derivative matrix with this new distribution, the inverted fluxes were recalculated. This process was performed

for January and July. The difference in the regular inverse and the inverse from the

sensitivity test can be seen in Figure 45 for January and Figure 46 for July.

Contours from −.3918 to .7082 by .1 (kg/m2/mth)

Contours from −.3758 to .6242 by .1 (kg/m2/mth)

Contours from −.1348 to −.0347 by .1 (kg/m2/mth)

Figure 45: January sensitivity test, where (a) is inverted solution, (b) is inverted with shift, and (c) is difference.

Contours from -1.908 to .592 by .5 (kg/m2/mth)

Contours from -1.792 to .708 by .5 (kg/m2/mth)

Contours from -1.075 to .425 by .5 (kg/m2/mth)

Figure 46: July sensitivity test, where (a) is inverted solution, (b) is inverted with shift, and (c) is difference.

As evident in Figure 45, there is almost no effect in January, with only small differences showing up in North America. However, this is not the active season for vegetation. Not surprisingly, July, the month of maximum uptake, shows greater sensitivity (Figure 46). Most obvious, is the slight northward shift of the fluxes, which is

134

the spatial change imposed by the test. In terms of actual fluxes, the net uptake changes from -0.83 Gt in the regular inverse to -0.97 Gt C/month with the shifted region, or a 14% change. Interestingly, the impact is felt elsewhere in a slightly reduced Eurasian vegetative flux from -0.99 Gt to -0.83. Thus, the net effect on the northern hemisphere vegetative uptake is negligible, changing from -2.13 to -2.09 Gt. While this amount of error is well within the reasonable limits, it reminds us that the inverted solution is not exact. However, this test gives support for the "missing" sink remaining in the biosphere.

4.3c Equal initial distributions within a region

While the initial distributions of vegetative flux are strongly correlated to where the vegetation exists, this is not necessarily true of the oceans. As discussed in Section 2.1, there are a multitude of non-linear factors that regulate oceanic uptake beyond the location of phytoplankton. Furthermore, there are vast variations in the partial pressure of carbon dioxide in relatively small areas (Watson et al., 1991). Thus, there is a large amount of uncertainty in the initial distribution of the oceanic fluxes. To test how much difference this could make in the inverted solution, we ran a second sensitivity test.

To simulate this type of uncertainty, instead of using a small variation, the net flux within the north Pacific ocean region was redistributed such that each grid had an equal mass, while maintaining the same net flux. Running this type of vast variation should give us a sense of the type of errors that are associated with the inverted oceanic flux.

Again, January and July were run with this sensitivity test and the results are shown in

Figure 47 and Figure 48 respectively.
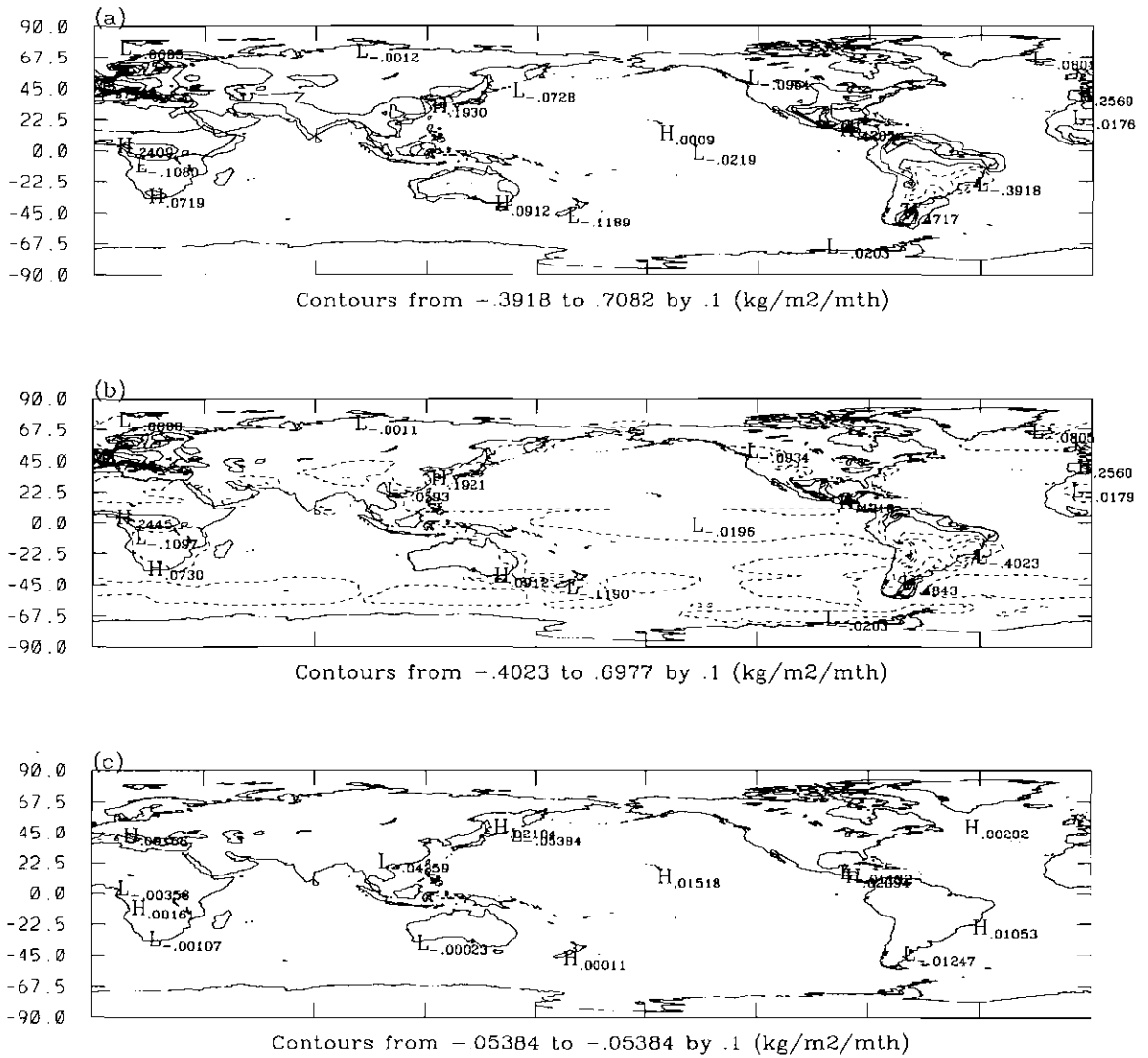
Figure 47: January sensitivity test 2, where (a) is inverted solution, (b) is inverted with shift, and (c) is difference.

136

Contours from −1.908 to .592 by .5 (kg/m2/mth)

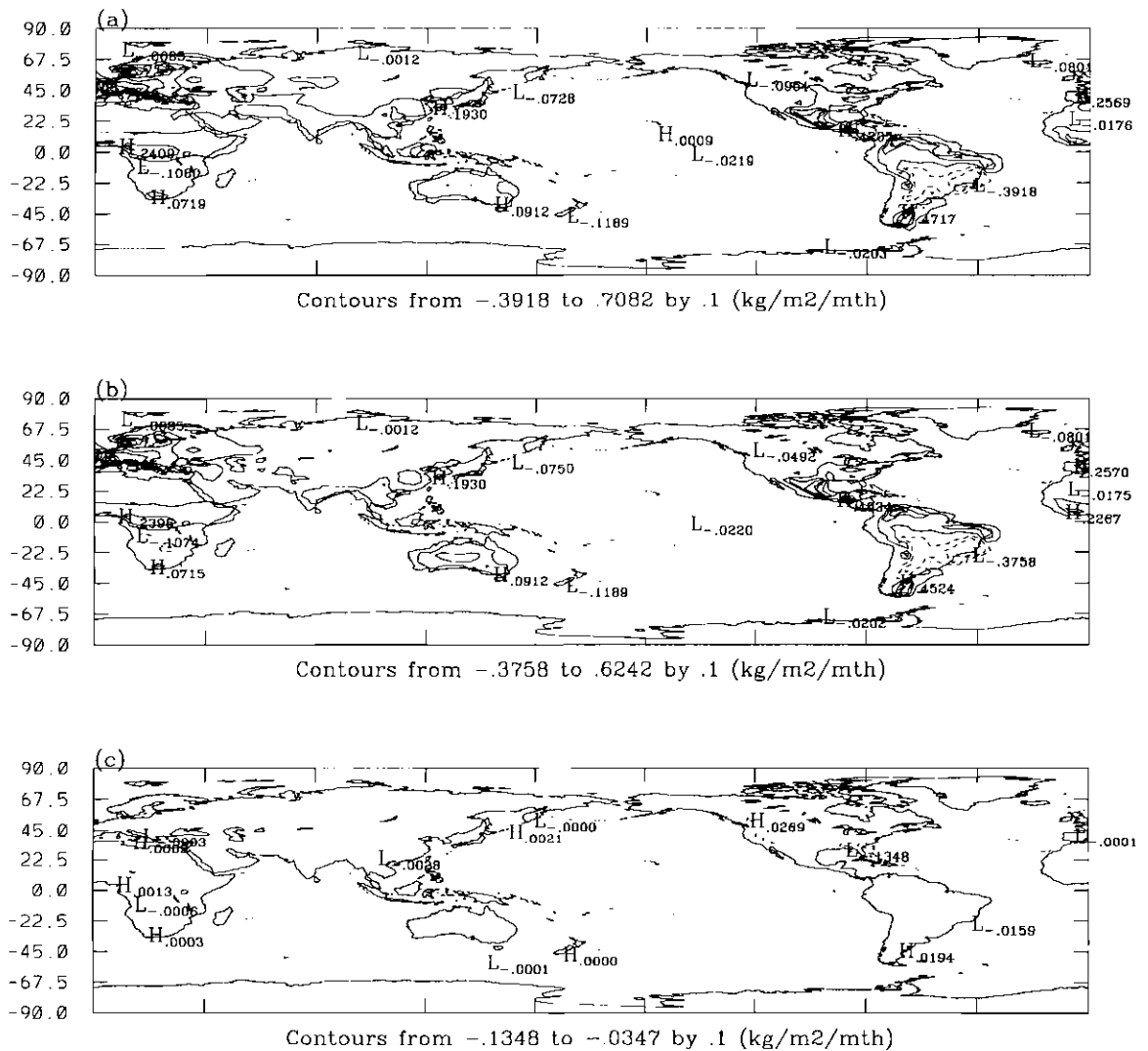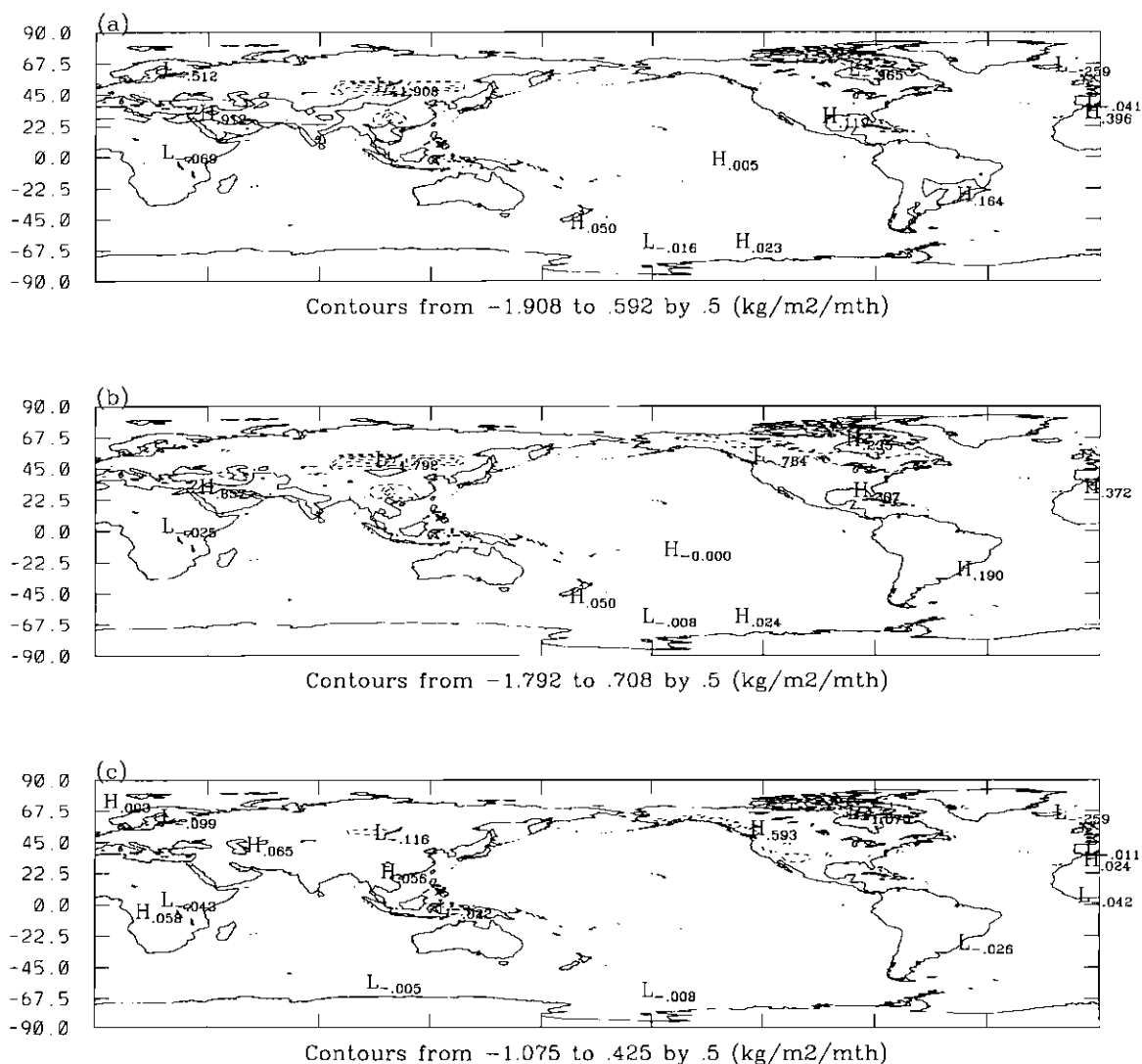Contours from −1.933 to .567 by .5 (kg/m2/mth)

Contours from −1.063 to −.062 by .5 (kg/m2/mth)
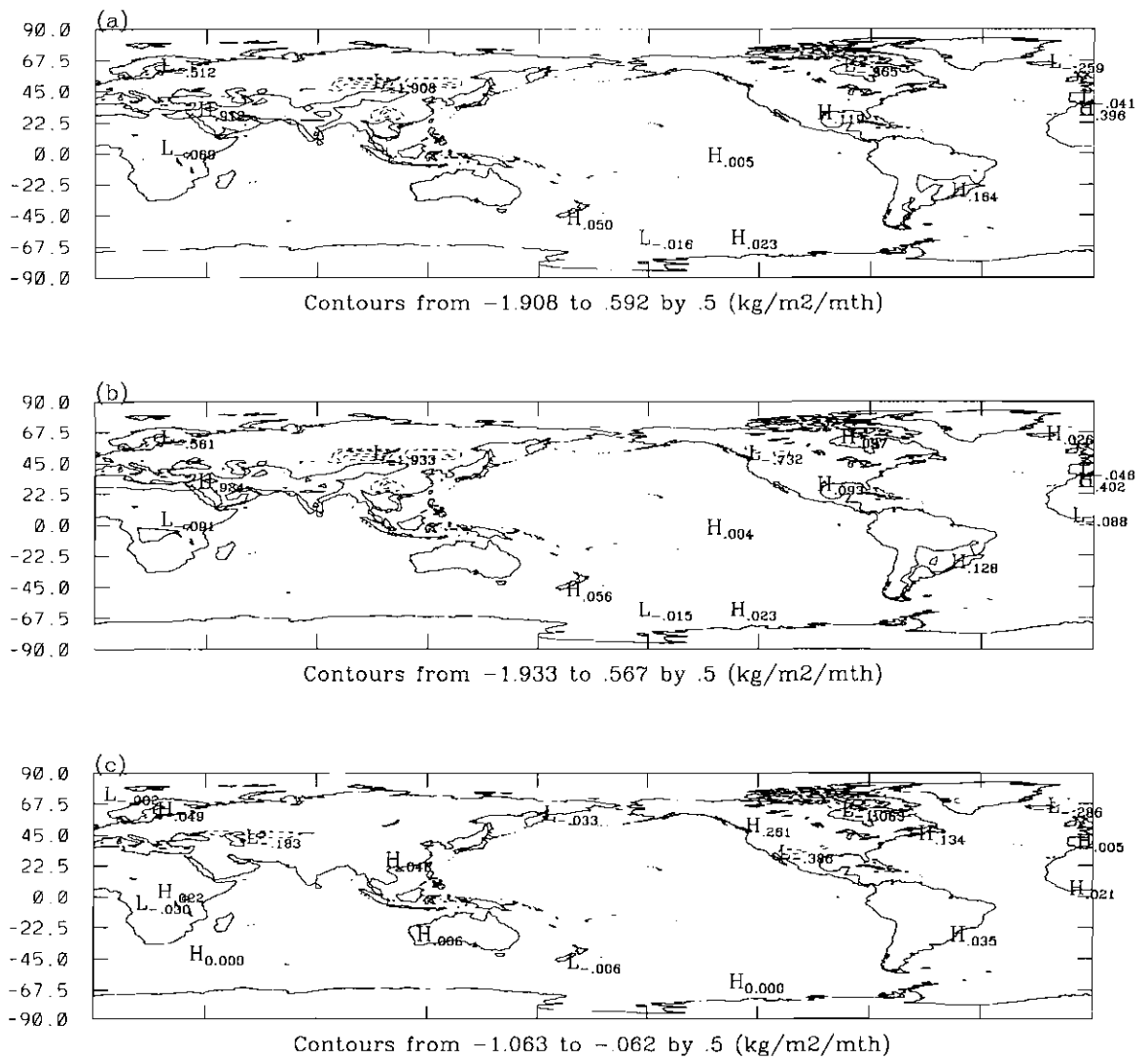
Figure 48: July sensitivity test 2, where (a) is inverted solution, (b) is inverted with shift, and (c) is difference.

The main effect of the January test is the shape redistribution in the Pacific ocean with the flux in the region changing from -0.21 to -0.23 Gt C. July is slightly more sensitive to changes in the distribution. The inverted flux changes from -0.07 to -0.03 Gt C in the region. While not an insignificant change, the inverted solution still moves in

the same direction from an initial estimate of +0.02 Gt C. This weakened sink is mainly countered by an increased sink in North America of -0.09Gt. Having significantly altered the distribution in a large region of the ocean, it is reassuring that the general direction and magnitude of the inverted solution is not significantly altered. This suggests that the inverse solution is fairly robust.

CHAPTER 5

CONCLUSIONS / RECOMMENDATIONS

The goal of this research was to develop and test an inverse method capable of improving the flux estimates for carbon dioxide. By inverting for the solution using the iterative-adaptive Kalman filter, the net balance of carbon dioxide fluxes improved over the initial estimate and the IPCC (1995) estimate. Furthermore, the model output agreed significantly better with observations than the run with the initial estimate. Much of this improvement can be attributed to the larger sink found in the inverted northern hemisphere vegetative flux.

5.1 Northern Hemisphere Vegetative Fluxes

Of particular interest for the carbon dioxide inverse was to locate the "missing" sink. The inverse solution for the northern hemisphere indicated that the "missing sink" lies in the vegetative flux. There were two components that account for this increased vegetative sink. First, the growing season begins one month earlier than the initial estimation. Note that since we are only inverting on time scales of one month, this does not necessarily mean that the growing season is actually extended by a full month, but that the net magnitude over the two months suggest uptake rather than release.

At least qualitatively, these findings are supported by satellite interpretation of leaf index. Myneni et al. (1997) found that the growing season has been extended by 10 days (6 in the beginning and 6 at the end) from 1981 to 1994. Furthermore, they find that the magnitude of the leaf index has increased every year for the last 13 years, from 0.334 in 1981 to 0.361 in 1994. While this method cannot yet be converted into a flux of CO_2, taken in combination with our inverse results, there is strong evidence that the "missing" sink is in fact the northern hemisphere biosphere. Other recent studies also find increased vegetative activity in the northern hemisphere (Chapin et al., 1996; Keeling et al., 1996). Although we can only speculate, possible explanations for this increased vegetative activity are warmer temperatures (Myneni et al., 1997) and nitrogen deposition due to anthropogenic sources of nitrogen (Hudson et al., 1994).

The second component of the increased vegetative sink is a decreased source in the continental United States and Eastern China during the fall/winter season. Since these are the only two regions containing sub-tropical vegetation, this suggests that the effect of year round photosynthesis in subtropical areas has previously been underestimated. This is a virtually ignored in previous work, and yet this helps explain the full amount of uptake that accounts for the "missing" sink.

Further support for the credibility of the inverse solution come from three sensitivity studies conducted to test the robustness of the solution. Errors were introduced to each of the input variables and in every case the net sink remained in the northern hemisphere biosphere.

It seems that satellite's provide a potentially powerful tool in giving global coverage of leaf index. However, the algorithms to interpret the leaf index from a qualitative to quantitative measurements still need development. As a recommendation for improving future vegetative flux estimates, one idea is to include the results found here as constraints when developing an algorithm to interpret satellite leaf index into actual flux values.

5.2 Southern Hemisphere Fluxes

While the northern hemisphere remained relatively unaffected by the introduction of errors into the input variables, the southern hemisphere proved more sensitive. From pseudo data studies, Africa and South America were found to be poorly constrained by observations. Future work should include another inverse study after the addition of the latest NOAA/CMDL sites have been operational for at least a year. We have already told our recommendations to NOAA/CMDL of the future sites that were determined to be useful in improving the constraint of these regions. If these suggestions are taken, we will hopefully be able to repeat the inverse in the future and gain more information about South America and Africa. Both of these regions are becoming increasingly important, particularly with the current trends in deforestation and biomass burning rising at the present alarming rates. Without good measurement sites in place, it will be difficult to identify and help control future changes that may arise.

5.3 Ocean Fluxes

The inverse solution shows a net ocean sink of -2.4 Gt C, which is in relatively good agreement with current IPCC estimates of oceanic net sink values (-2.0 $\pm$ 0.8 Gt C/yr). The sign of the ocean flux is dominated by blooms rather than temperature (as suggested by the initial estimate). Uptake begins in late spring and continues through the summer season. This is in agreement with field study results by Watson et al. (1991). As a final comparison, the total inverted flux was compared to the annual flux estimate from GISS (Broecker et al., 1986). The significant differences were a weaker sink in the southern ocean (mid-high latitudes), a net sink rather than source in the eastern north Pacific, and a weaker sink in the northern mid Atlantic (refer to Figure 38). There is a current effort to measure the ΔpCO_2 in the Joint Ocean Global Flux Study (JGOFS) and the World Ocean Circulation Experiment (WOCE), as well as to develop algorithms to interpret satellite chlorophyll data into CO_2 flux estimates. When these estimates of oceans fluxes are complete, it would be interesting to repeat the inverse. However, sensitivity studies altering the distribution of the ocean flux found that the solution was fairly robust, with only minor changes resulting.

5.4 Applicability of Method to Other Trace Gas Fluxes

The inverse technique presented has proven capable of deducing fluxes for carbon dioxide with reasonable success and robustness. With these promising results, this

142

technique should be employed to help improve flux estimates of other important trace gases. Of particular concern, carbon monoxide, methane, and nitrous oxide have all seen increased concentrations since the industrial revolution (Table 21). Improving the estimations of these fluxes is a very important next step. However, further refining of the method may be necessary to account for the chemistry involved with the other carbon cycles gases. While CO_2 is essentially inert in the atmosphere, both carbon monoxide and methane are oxidized by the hydroxyl radical (OH). Thus, to perform an inverse for these gases, one has to have a reasonable budget of OH. Since OH is the main oxidizer in the atmosphere, there are many competing mechanisms for reacting with OH. Thus, it has a very complicated associated chemistry. This is not to say that inverting for carbon monoxide and methane is unrealistic, it is only to caution that the chemistry adds an additional uncertainty and additional testing would be required to account for this type of uncertainty. In fact, this is a challenging and exciting use for inverse methods in the future. This thesis lays the groundwork for a potentially powerful tool to be used in identifying more complicated sources and sinks in the atmosphere. In addition to utilizing the method described in this thesis, we recommend a detailed study of the errors associated with the OH chemistry.

Table 21: Greenhouse Gas Concentrations pre and post industrial (IPCC, 1995).

Gas	Pre-Industrial	1994
CO_2	~280 ppmv	358 ppmv
CH_4	~700 ppbv	1720 ppbv
N_2O	~275 ppbv	312 ppbv

5.5 Summary of Recommendations:

1. To improve CO_2 inverse:

   • Add measurement stations to improve estimation in Southern Hemisphere, particularly South America and Africa.

   • Repeat inverse for CO_2 with the 1996 measurements, since the newest NOAA/CMDL sites will have been in operation for a year. This should improve some of the difficulty in the Southern Hemisphere, particularly Northern Africa with the addition of the WIS site in Israel. Additionally, it would be interesting to compare the results with interannual variations.

   • Repeat inverse with updated ocean flux estimates once the JGOFS and WOCE experiments are complete.

   • Use inverse results in conjunction with satellite estimates of leaf index to improve algorithms for estimating vegetative uptake of CO_2.

2. Apply methodology to other trace gases:

- As nitrous oxide emissions continue to increase with continuing land use change, understanding its budget becomes increasingly important. The inverse method developed here should be directly applicable. However, the regions will need to be redefined specific to N_2O and its measurement network.

- Apply the methodology to methane and carbon monoxide. While the methodology developed here is applicable, additional work will need to be done to assess errors arising from uncertainties in the OH budget. As with N_2O, the regions will also have to be refined to fit with the measurement network for CH_4 and CO.

APPENDIX A

Assumptions:

1. $\chi^{obs} = PE$ P is some relationship and E is emissions

2. $E^{est} + \Delta E = \hat{E}$ where $\hat{E}$ is the best estimate of emissions

                  and $\Delta E = E_{t+1}^{est} - E_t^{est}$

3. $E_t^{est} = \hat{E}_t + \sigma_{E,t}$ where $\sigma_{E,t}$ is the error in emissions estimation at time t

Want to minimize two components in a cost function.

1. $\chi^{obs} - P\hat{E}$ the difference in observed concentrations and model calculated emissions

2. $\Delta E = E_{t+1}^{est} - E_t^{est}$ the change in emissions step size

The cost function, J, is the following:

$$J = (\chi_t^{obs} - P_t \hat{E}_t)^T N_t^{-1} (\chi_t^{obs} - P_t \hat{E}_t) + (\Delta E)^T C_t^{-1} (\Delta E)$$

Minimizing with respect to emissions,

$$\frac{\partial J^{est}}{\partial E_{t+1}^{est}} = 0 = 2P^T N^{-1} (\chi^{obs} - PE^{est} - P\Delta E) + 2C^{-1}\Delta E$$

substitute in the relationship that $PE^{est} = \chi^{model}$:

$$0 = P^T N^{-1} (\chi^{obs} - \chi^{model}) - P^T N^{-1} P\Delta E - C^{-1}\Delta E$$

rearrange, isolating ΔE:

146

$$\Delta E = \left(P^T N^{-1} P + C^{-1}\right)^{-1} P^T N^{-1} (\chi^{obs} - \chi^{model})$$

for simplicity, call

$$\Delta \chi = \chi^{obs} - \chi^{model}$$

want to rearrange so that $\Delta \chi$ is on left hand side:

$$\Delta E \Delta \chi^{-1} = \left(P^T N^{-1} P + C^{-1}\right)^{-1} P^T N^{-1}$$

using the theorem that $\left(AB\right)^{-1} = B^{-1} A^{-1}$, we invert the equation

$$\Delta \chi \Delta E^{-1} = N(P^T)^{-1} \left(P^T N^{-1} P + C^{-1}\right)$$

multiply the right hand terms together:

$$\Delta \chi \Delta E^{-1} = N(P^T)^{-1} P^T N^{-1} P + N(P^T)^{-1} C^{-1}$$

By the definition of inverse matrices, where $A(A)^{-1} = I$ (the identity matrix), the right hand side simplifies to $P + N(P^T)^{-1} C^{-1}$, where the whole eqn is now:

$$\Delta \chi \Delta E^{-1} = P + N(P^T)^{-1} C^{-1}$$

The following is a series of steps is taken to rearrange the terms so that they are in terms of a gain matrix.

First, multiply the equation by C:

$$\Delta \chi \Delta E^{-1} C = PC + N(P^T)^{-1}$$

Second multiply the equation by P^T:

147

$$\Delta\chi\Delta E^{-1}CP^T = PCP^T + N$$

Multiply the equation by $(PCP^T + N)^{-1}$

$$\Delta\chi\Delta E^{-1}CP^T (PCP^T + N)^{-1} = I$$

Multiply the equation by $\Delta\chi^{-1}$

$$\Delta E^{-1}CP^T (PCP^T + N)^{-1} = \Delta\chi^{-1}$$

Multiply by $\Delta\chi$

$$\Delta E^{-1}CP^T (PCP^T + N)^{-1}\Delta\chi = \Delta\chi^{-1}\Delta\chi$$

Multiply by ΔE

$$CP^T (PCP^T + N)^{-1}\Delta\chi = \Delta E$$

Substitute G, the gain matrix for $CP^T (PCP^T + N)^{-1}$

$$G\Delta\chi = \Delta E$$

Rearrange to get the equation into the form used in this paper:

$$\Delta E = G\Delta\chi$$

Substitute in for the delta terms:

$$E_{t+1}^{est} - E_t^{est} = G_t(\chi_t^{obs} - \chi_t^{model})$$

Rearrange into final form :

$$E_{t+1}^{est} = E_t^{est} + G_t(\chi_t^{obs} - \chi_t^{model})$$

$$E_{t+1}^{est} - E_t^{est} = G_t(\chi_t^{obs} - \chi_t^{model})$$

REFERENCES

Bellgardt, K.-H., W. Kuhlmann, H.-D. Meyer, K. Schugerl, and M. Thoma, Application of an extended Kalman filter for state estimation of a yeast fermentation, *IEE Proceedings*, **133 pt. D**, 226-234, 1986.

Bennett, A.F., W.P. Budgell, Ocean Data Assimilation and the Kalman Filter: Spatial Regularity, *J.Phys. Oceanogr.*, **17**, 1583-1601, 1987.

Broecker, W.S., J.R. Ledwell, T. Takahashi, R. Weiss, L. Merlivat, L. Memery, T-H Peng,B. Jahne, and K.O. Munnich, Isotopic verseus micrometerological ocean CO_2 fluxes: a serious conflict, *J. Geophys. Res.*, **91**, 10517-10527, 1986.

Brown, M. Deduction of Emissions of Source Gases Using an Objective Inversion Algorithm and a Chemical Transport Model, *Journal of Geophysical Research*, **98**, 12639-12660, 1993.

Ciais, P., P.P. Tans, J.W. C. White, M. Trolier, R. J. Francey, J.A. Berry, D.R. Randall, P.J. Sellers, J.G. Collatz, and D.S. Schimel, Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}C$ measurements from the NOAA climate monitoring and diagnostics laboratory global air sampling network, *J. Geophys. Res.*, 100, 5051-5070, 1995.

Chang, M., C. Cardelino, D. Hartley, and W.L. Chang, Inverse Modeling of Biogenic Isoprene Emissions, *Geophy. Res. Let.*, **23**, 3007-3010, 1996.

Chameides, W.C. and E.M. Perdue, Biogeochemical Cycles, Chapter 6: Carbon Cycle, Oxford Univ. Press, 1997.

Chapin, F.S., III, S.A. Zimov, G.R. Shaver, and S.E. Hobble, CO_2 fluctuation at high latitudes, *Nature*, **383**, 585-586, 1996.

Conway, T.J., P.P. Tans, L.S. Waterman, K.W. Thoning, D.R. Kitzis, K.A. Masarie, and Ni Zhang, Evidence for interannual variability the carbon cycle from the National Oceanic and Atmospheric Administration / Climate Monitoring and Diagnostics Laboratory global air sampling network, *J. Geophys. Res.*, 99, 22831-22855, 1994.

Conway, T.J., P.P. Tans, L.S. Waterman, K.W. Thoning, K.A. Masarie, and R.H. Gammon, Atmospheric carbon dioxide measurements in the remote global troposphere, 1981-1984, *Tellus*, 40B, 81-115, 1988.

Cunnold, D.M., R.G. Prinn, R.A. Rasmussen, P.G. Simmonds, F.N. Alyea, C.A. Cardelino, A.J. Crawford, P.J. Fraser, and R.D. Rosen, Atmospheric lifetime and annual release estimates for CFCl₃ and CF₂Cl₂ from 3 years of ALE data, *J. Geophys. Res.*, **88**, 8379-8400, 1983.

Cunnold, D.M., P.J. Fraser, R.F. Weiss, R.G. Prinn, P.G. Simmonds, B.R. Miller, F.N. Alyea, and A.J. Crawford, Global trends and annual releases of CFCl₃ and CCl₂F₂ estimated from ALE/GAGE and other measurements from July 1978 to June 1991, *J. Geophys. Res.*, **99**, 1107-1126, 1994.

Dai, A. and I. Fung, Can climate variability contribute to the "missing" CO_2 sink?, *Global Biogeo. Cycles,* **7**, 599-609, 1993.

Daley, R., R. Menard, Spectral Characteristics of Kalman Filter Systems for Atmospheric Data Assimilation, *Monthly Weather Rev.,* **121**, 1554-1565, 1993

Dlugokencky, E.J., J.M. Harris, Y.S. Chung, P.P. Tans, and I. Fung, The Relationship between the methane seasonal cycle and regional sources and sinks at Tae-Ahn Peninsula, Korea, *Atmos. Envir. Part A*, **27A**, 2115-2120, 1993.

Enting, I.G., and J.V. Mansbridge, Seasonal sources and sinks of atmospheric CO_2: Direct inversion of filtered data, *Tellus,* **41b**, 111-126, 1989.

Enting, I.G., and J.V. Mansbridge, Latitudinal distribution of sources and sinks of CO_2: Results of an inversion study, *Tellus,* **43b**, 156-170, 1991.

Enting, I.G. and G.N. Newsam, Inverse problems in atmospheric constituent studies: II. Sources in the free atmosphere, *Inverse Problems*, **6**, 349-362, 1990.

Erickson, D.J., P.J. Rasch, P.P. Tans, P. Friedlingstein, P. Ciais, E. Maier-Reimer, K. Six, C.A. Fischer, and S. Walters, The seasonal cycle of atmospheric CO_2: A study based on the NCAR Community Climate Model (CCM2), *J. Geophys. Res.*, 101, 15079-15097, 1996.

Ferguson, E.E. and R.M. Rosson, eds., Climate Monitoring and Diagnostics Laboratory, Rept. No. 20, Summary Report 1991, U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, and Environmental Research Laboratories, 1992.

Fung, I.Y., C.J. Tucker, and K.C. Prentice, Application of advanced very high resolution radiometer vegetation index to study atmosphere-biosphere exchange of CO_2, *J. Geophys. Res.*, **92**, 2999-3015, 1987.

151

Gauthier, P., P. Courtier, and P. Moll, Assimilation of simulated wind lidar data with a Kalman Filter, *Monthly Weather Rev.,* **121**, 1803-1820, 1993

Gelb, A. Applied Optimal Estimation, Chapter 5: Optimal Linear Smoothing; MIT Press: Cambridge, 1994.

Ghil, M. Meteorological Data Assimilation for Oceanographers. I. Description and Theoretical Framework, *Dynam. of Atmos. and Oceans,* **13**, 171-218. 1989.

Grubb, M., Not so puzzling, *Nature,* **368**, 489, 1994.

Gutman, P. and M. Velger, Tracking targets with unknown noise variance using adaptive Kalman filtering, Proceedings of IEEE conference on Decision and Control, 1988.

Hack, J.J., B.A. Boville, B.P. Briegleb, J.T. Kiehl, P.J. Rasch, and D.L. Williamson, Description of the NCAR Community Climate Model (CCM2*), NCAR Technical Note TN-382+STR,* NCAR: Boulder, Colorado, 1993.

Hartley, D., Deducing Trace Gas Emissions Using an Inverse Method in Three-Dimensional Chemical Transport Models. MIT Center for Global Change Science, Report No. 17, 1992.

Hartley, D. and R.G. Prinn, Feasibility of Determining Surface Emissions of Trace Gases Using an Inverse Method in a three-dimensional Chemical Transport Model, *J. Geophys. Res.,* **98**, 5183-5197, 1993.

Hesshaimer, V., M. Heimann, and I. Levin, Radiocarbon evidence for a smaller oceanic carbon dioxide sink than previously believed, *Nature,* **370**, 201-203, 1994.

Houghton, R.A., R.D. Boone, J.R. Fruci, J.E. Hobbie, J.M. Melillio, C.A. Palm, B.J. Peterson, G.R. Shaver, G.M. Woodwell, B. Moore, Skole, and N. Myers, The flux of carbon from terrestrial ecosystems to the atmosphere in 1980 due to changes in land use: Geographic distribution of the global flux, *Tellus, Ser. B,* **39**, 122-139, 1987.

Houghton, R.A., Is Carbon Accumulating in the Northern Temperate Zone?, *Global Biogeo. Cycles,* **7**, 611-617, 1993.

Hudson, R.J.M, Gherini, S.A., and Goldstein, R.A., Modeling the global carbon cycle: Nitrogen fertilization of the terrestrial biosphere and the "missing" CO_2 sink, *Global Biogeo. Cycles,* **8**, 307-334, 1994.

Intergovernmental Panel on Climate Change (IPCC), *Climate Change 1995: The Science of Climate Change,* 572 pp., Cambridge University Press, New York, 1996.

Junge, C.E. Residence time and variability of tropospheric trace gases, *Tellus*, **26**, 477-487, 1974.

Kalman, R.E. A new approach to linear filtering and prediction problems, J. of Basic Eng., 35-45, 1960.

Kauppi, P.E., K. Mielikainen, and K. Kuusela, Biomass and carbon budget of European forest, 1971-1990, *Science*, **256**, 70-74, 1992.

Keeling, C.D., J.F.S. Chin, and T.P. Whorf, Increased activity of northern vegetation inferred from atmospheric CO_2 measurements, *Nature*, **382**, 146-149, 1996.

Khalil, M.A. and R.A. Rasmussen, Atmospheric Methane: Recent Global Trends, *Environ. Sci. Techol.*, **24**, 549-553, 1990.

Komhyr, W.D., R.H. Gammon, T.B. Harris, L.S. Waterman, T.J. Conway, W.R. Taylor, and K.W. Thoning, Global Atmospheric CO_2 Distribution and Variations from 1968-1982 NOAA/GMCC CO_2 Flask Sample Data, *J. Geophys. Res.*, 90, 5567-5596, 1985.

Mahowald, N. Development of a 3-dimensional Chemical Transport Model Based on Observed Winds and Use in Inverse Modeling of the Sources of CCl_3F, Center for Global Change Science, Report No. 42, April, 1996.

Melillo, P.E., J.R. Fruci, R.A. Houghton, B. Moore, and D.L. Skole, Land-use change in the Soviet Union between 1850 and 1980: Causes of a net release of CO_2 to the atmosphere, *Tellus, Ser. B*, **40**, 116-128, 1988.

Montgomery, D.C. Design and Analysis of Experiments, John Willey and Sons, Inc.: New York, Chapter 2. p. 14-49, 1991.

Myneni, R.B., C.D. Keeling, C.J. Tucker, G. Asrar, R.R. Nemani, Increased plant growth in the northern high latitudes from 1981-1991, *Nature*, **386**, 698-702, 1997.

Prather, M., Chemistry of the global troposphere: fluorocarbons as tracers of air motion., *J. Geophys. Res.*, 92, 6579-6613, 1987.

Prinn, R.G., Toward an Improved Global Network for Determination of Tropospheric Ozone Climatology and Trends, *J. of Atmos. Chem.* **6**, 281-298, 1988.

Quay, P., B. Tilbrook, and C.S. Wong, Carbon sink: the role of the oceans, *Science*, **256**, 74-79, 1992.

Ramonet, M. and P. Monfray, CO_2 baseline concept in 3-D atmospheric transport models, *Tellus, Ser. B*, **48**, 502-502,1996.

Rasch, P.J., X. Tie, B.A. Boville, and D.L. Williamson, A three-dimensional transport model for the middle atmosphere, *J. Geophys. Res.*, 99, 999-1119, 1994.

Sastri, T., An Adaptive Autoregressive Model, *Comput. and Indus. Engng*, **9**, 9-27, 1985.

Schlessinger, W.H., Biocheochemistry: An analysis of global change, Academic Press, San Diego, pp. 443, 1991.

Sedjo, R.A., Temperate forest ecosystems in the global carbon cycle, *Ambio*, **21**, 274-277, 1992.

Siegenthaler, U. and J.L. Sarmiento, Atmospheric carbon dioxide and the ocean, *Nature*, **365**, 119-125, 1993.

Surendran, S. and R.J. Mulholland, Estimation of atmospheric CO_2 concentration using Kalman filtering, *Int. J. Systems Sci.,* **17**, 897-909, 1986.

Takahashi, T., Tans, P.P., Fung, I., Balancing the Budget: Carbon Dioxide Sources and Sinks, and the Effects of Industry, *Oceanus,* **35**, 18-28, 1992.

Tans, P.P., T.J. Conway, and T. Nakazawa, Latitudinal distribution of the sources and sinks of atmospheric carbon dioxide derived from surface observations and an atmospheric transport model, *J. Geophys. Res.,* **94D**, 5151-5172, 1989.

Taylor, J., The mutable carbon sink, *Nature,* **366**, 515-516, 1993.

Thompson, M.L., I.G. Enting, G.I. Pearman, and P. Hyson, Interannual Variation of Atmospheric CO_2 Concentration, *J. of Atmos. Chem.,* **4**, 125-155, 1986.

Thoning, K.W., P.P. Tans, and W.D. Komhyr, Atmospheric carbon dioxide at Mauna Loa Observatory 2. Analysis of the NOAA GMCC data, 1974-1985, *J. Geophys. Res.,* 94, 8549-8569, 1989.

Trenberth, K., Global Analyses from ECMWF and Atlas of 1000mb to 10mb Circulation Statistics, NCAR Technical Note, NCAR/TN-373+STR, Climate and Global Dynamics Division, Boulder, CO, 1992.

Warneck, P., Chemistry of the natural atmosphere, Academic Press, San Diego, 1988.

Watson, A.J., C. Robinson, J.E. Robinson, P.J. le B. Williams, and M.J.R. Fasham, Spatial variability in the sink for atmospheric carbon dioxide in the North Atlantic, *Nature,* **350**, 50-55, 1991.

Young, P., <u>Recursive Estimation and Time-Series Analysis</u>, Chapters 7 and 10; Springer-Verlag: Berlin, Heidelberg, 1984.