

# **Investigating land-air carbon fluxes using a Lagrangian model and satellite retrieved carbon dioxide**

Thesis submitted for the degree of  
Doctor of Philosophy  
at the University of Leicester

by

Alan James Hewitt  
Earth Observation Science Group  
Department of Physics and Astronomy  
University of Leicester

March 2010

© Alan James Hewitt, November 27, 2010

This thesis is copyright material and no quotation from it may be published  
without proper acknowledgement.

## Declaration

I hereby declare that no part of this thesis has been previously submitted to this or any other University as part of the requirement for a higher degree. The work described herein was conducted by the undersigned except for contributions from colleagues as acknowledged in the text.

Alan James Hewitt

November 27, 2010

# Investigating land-air carbon fluxes using a Lagrangian model and satellite retrieved carbon dioxide

by Alan James Hewitt

The existing generation of satellite instruments (such as SCIAMACHY and AIRS) has allowed the retrieval of atmospheric mixing ratios of carbon dioxide. The feasibility of using these and later satellites (OCO-like or GOSAT) to investigate carbon fluxes between the terrestrial biosphere and the atmosphere, either alone or complemented by the high precision but low density network of surface measurement sites has been investigated.

A methodology to investigate regional scale carbon budgets, based on the UK Met Office Lagrangian trajectory model NAME (Numerical Atmospheric-dispersion Modelling Environment), has been developed and demonstrated.

A forward modelling methodology was developed, where top-down surface flux information from CarbonTracker was combined with the background CO<sub>2</sub> mixing ratio to obtain an atmospheric concentration. Synthetic testing of the initialisation method demonstrated that a strong correlation coefficient ( $R^2 \approx 0.9$ ) between the forward modelled and satellite observed atmospheric CO<sub>2</sub> fields can be achieved. Forward modelled CO<sub>2</sub> concentrations using CarbonTracker fluxes were demonstrated to be moderately correlated with the SCIAMACHY-retrieved CO<sub>2</sub> field ( $R^2$  varies by month, from 0.4 to 0.8).

An inverse modelling methodology was developed, where the change in carbon mass between the satellite-retrieved CO<sub>2</sub> columns and the background concentration was combined with the surface residence time from the NAME model.

Synthetic testing of the inversion method has shown that the *a posteriori* flux covariance scaled linearly to the satellite-retrieved error covariance and inversely to the NAME residence time of the ecosystem. On the regional scale, this method could improve on the carbon flux estimates from CarbonTracker and an equivalent Eulerian method using GOSAT. This thesis also presents the first carbon fluxes inverted from satellite retrieved CO<sub>2</sub> columns, which captured the seasonality of the carbon fluxes of the vegetation and negligible ocean fluxes.



# ACKNOWLEDGEMENTS

I would like to thank my Ph.D. supervisor Paul Monks who helped me shape the vision for this research project and my other collaborators: Hartmut Boesch who provided technical help and support with regards to retrievals and modelling and to Claire Witham and Alistair Manning of the UK Met Office who helped me to develop the NAME method.

A big thank you also Michael Barkley with whom I have worked regarding the analysis of SCIAMACHY/FSI retrieved columns of CO<sub>2</sub>.

I would like to thank the members of the Earth Observation Science, Atmospheric Chemistry and Geography groups at the University of Leicester, especially those who have provided a great deal of assistance to my project: Harjinder Sembhi, Rhian Evans, Gennaro Cappelluti, John Remedios, Diane Knappett, Joerg Kaduk, Zoe Fleming, Edward Comyn-Platt, Stuart Poulton and Chris Rudge. I would also like to thank the following people based at other institutions: Jadu Dash (University of Southampton), Chuixiang Yi (City University of New York), Michael Buchwitz (University of Bremen), Fikru Adamu-Lema (VNI), Paul Palmer (University of Edinburgh), Julian Meyer-Arneke and Thomas Holzer-Popp (PROMOTE).

I am grateful to all the institutions who provided data necessary for this project: The UK Met Office for providing NAME and the meteorological fields, ESA for SCIAMACHY data, BADC for the ECMWF operational data, NOAA for providing CO<sub>2</sub> raw data and the CarbonTracker product, the SCIAMACHY group at Bremen for providing SCIATRAN. Thanks also to the Natural Environment Research Council for providing funding through grant NER/S/A/2005/13330 and to GMES service element PROMOTE.

Finally, I would like to thank Manasvi for keeping me sane and my other friends and family who have helped me get through the ordeal of completing my Ph.D.

*In memory of Harjeet Kundi.*

# LIST OF ABBREVIATIONS

AAI	Absorbing Aerosol Index
AIRS	Atmospheric Infrared Sounder (NASA)
AMF	Air Mass Factor
AO	Analogue Offset
ASM	Azimuthal Scanner Module
BADC	British Atmospheric Data Centre
BU	Binary Units (detector signal is digitised)
CarbonTracker	Carbon flux product provided by NOAA
CFCs	Chlorofluorocarbons
DOAS	Differential Optical Absorption Spectroscopy
ECMWF	European Centre for Medium-range Weather Forecasts
ENVISAT	ENVironmental SATellite (ESA)
ESA	European Space Agency
ESM	Elevation Scanner Module
ESRL	Earth System Research Laboratory (NOAA)
ESTEC	European Space Research and Technology Centre
ETEX	The European Tracer EXperiment campaign (1994)
FSI (WFM-DOAS)	Full Spectral Initiation (WFM-DOAS) (retrieval algorithm developed by Barkley et al. [2006a])
FTIR	Fourier Transform Infrared spectroscopy
GCM	General Circulation Model
GISS	Goddard Institute for Space Studies
GLOBALVIEW	Product developed by ESRL (NOAA)
GOME	Global Ozone Monitoring Experiment
GOSAT	Greenhouse gases Observing SATellite “IBUKI” (JAXA)
GPP	Gross Primary Production
GWP	Global Warming Potential
HPC	High Performance Computing
IPCC	Intergovernmental Panel on Climate Change
HITRAN	HIgh resolution TRANsmision molecular absorption database
JAXA	Japan Aerospace eXploration Agency
LOWTRAN	LOW resolution TRANsmision molecular absorption database
MERIS	MEdium Resolution Imaging Spectrometer (ESA)
MODIS	MODerate resolution Imaging Spectroradiometer (NASA)

MPI-BGC	Max-Planck Institute for Biogeochemistry
MTCI	MERIS Terrestrial Chlorophyll Index [Dash and Curran, 2004]
NAME	Numerical Atmospheric-dispersion Modelling Environment
NASA	National Aeronautics and Space Administration
NCDC	National Climatic Data Centre
NERC	Natural Environment Research Council
NIR	Near Infrared
NMAT	Nighttime Marine Air Temperature
NOAA	National Oceanic and Atmospheric Administration
NEP	Net Ecosystem Productivity
NPP	Net Primary Production
NWP	Numerical Weather Prediction
OCO	Orbiting Carbon Observatory (NASA)
PBL	Planetary Boundary Layer
PDF	Probability Density Function
PMD	Polarisation Measurement Device
ppbv	parts per billion by volume
ppmv	parts per million by volume
SCA	Seasonal Cycle Amplitude
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (ESA)
SCIATRAN	Radiative transfer model developed by Rozanov et al. [2002]
SPICI	SCIAMACHY PMD Identification of Clouds and Ice/snow method [Krijger et al., 2005]
SRON	Netherlands Institute for Space Research
SST	Sea Surface Temperature
STOCHEM	A tropospheric chemistry model developed by the UK Met office
SWIR	Short Wave Infrared
SZA	Solar Zenith Angle
TM3	Tracer Transport Model [Heimann and Korner, 2003]
TOMS	Total Ozone Mapping Spectrometer
UM	Unified Model (UK Met Office)
UNFCCC	United Nations Framework Convention on Climate Change
VCD	Vertical Column Density
VMR	Volume Mixing Ratio
WFM-DOAS	Weighting Function Modified DOAS (retrieval algorithm developed by Buchwitz et al. [2000a])
WMO	World Meteorological Organisation
XCO <sub>2</sub>	Column averaged CO <sub>2</sub> concentration

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Managing Climate Change . . . . .	1
1.2	Climate change . . . . .	2
1.2.1	CFCs and the ozone layer . . . . .	3
1.2.2	Global Warming . . . . .	4
1.2.3	The Earth's radiation budget . . . . .	7
1.2.4	The Greenhouse Effect . . . . .	8
1.2.5	Radiative Forcing . . . . .	9
1.2.6	The greenhouse gases . . . . .	9
1.3	The Carbon Cycle . . . . .	14
1.3.1	Bottom-up estimates of carbon fluxes . . . . .	17
1.3.1.1	Terrestrial biosphere to atmosphere flux processes .	18
1.3.1.2	Ocean to atmosphere flux processes . . . . .	22
1.3.2	Obtaining carbon fluxes through micrometeorological methods	23
1.3.3	Top-down estimates of carbon fluxes . . . . .	25
1.3.3.1	Inverse modelling of point measurements . . . . .	26
1.3.3.2	Inverse modelling of satellite column measurements	27
1.4	Atmospheric circulation . . . . .	27
1.5	Thesis overview . . . . .	29

<b>2</b>	<b>Remote sensing of atmospheric carbon dioxide</b>	<b>31</b>
2.1	Variability of CO <sub>2</sub> in the atmosphere . . . . .	31
2.1.1	Variability in the CO <sub>2</sub> column . . . . .	33
2.2	Precision requirements for flux inversions . . . . .	33
2.2.1	CarbonTracker . . . . .	38
2.3	Remote sensing of carbon dioxide . . . . .	40
2.3.1	Thermal Infrared Techniques . . . . .	41
2.3.2	Near Infrared Techniques . . . . .	42
2.4	SCIAMACHY . . . . .	43
2.4.1	Instrument concept . . . . .	43
2.4.2	Orbit and measurement geometry . . . . .	44
2.4.2.1	Observational modes . . . . .	45
2.4.3	Calibration . . . . .	51
2.4.3.1	Icing . . . . .	51
2.4.3.2	General calibration . . . . .	51
2.5	Summary . . . . .	52
<b>3</b>	<b>Retrieval of atmospheric CO<sub>2</sub> from SCIAMACHY NIR measure-</b>	<b>54</b>
	<b>ments</b>	
3.1	Differential Optical Absorption Spectroscopy . . . . .	55
3.1.1	Introduction . . . . .	55
3.1.2	Theory . . . . .	55
3.1.2.1	Absorption of Radiation . . . . .	55
3.1.2.2	Scattering of Radiation . . . . .	56
3.1.2.3	The DOAS principle . . . . .	56
3.1.2.4	Conversion to vertical column density . . . . .	59
3.1.2.5	The DOAS spectral fitting procedure . . . . .	59
3.2	Weighting Function Modified DOAS . . . . .	60

3.2.1	WFM-DOAS Look up table . . . . .	62
3.3	Full Spectral Initiation WFM-DOAS . . . . .	62
3.3.1	Overview of the FSI algorithm . . . . .	63
3.3.2	Preprocessing of SCIAMACHY data . . . . .	65
3.3.2.1	Extraction of SCIAMACHY data . . . . .	66
3.3.3	A priori data . . . . .	67
3.3.4	Normalising the CO <sub>2</sub> vertical column density . . . . .	69
3.3.5	Factors affecting individual FSI and WFM-DOAS CO <sub>2</sub> re- trievals . . . . .	70
3.3.5.1	Cloud Coverage . . . . .	70
3.3.5.2	Cirrus Clouds . . . . .	70
3.3.5.3	Molecular line parameters . . . . .	71
3.3.5.4	Aerosols . . . . .	71
3.3.5.5	Surface reflectance . . . . .	71
3.3.5.6	Surface pressure . . . . .	73
3.3.5.7	Temperature profile and vertical distribution of CO <sub>2</sub> and water vapour . . . . .	73
3.4	Precision requirements for the investigation of carbon fluxes . . . .	73
3.5	The effect of sampled biases on the investigation of carbon fluxes .	75
3.5.1	The effect of aerosol contamination on the investigation of carbon fluxes . . . . .	76
3.6	Preliminary findings of FSI and WFM DOAS . . . . .	77
3.6.1	FSI vs AIRS . . . . .	77
3.6.2	FSI vs FTIR . . . . .	81
3.7	Summary . . . . .	83

<b>4</b>	<b>Analysis of FSI-WFM-DOAS retrievals</b>	<b>84</b>
4.1	The effect of sampling biases on the quantification of carbon fluxes	84
4.1.1	Cloudy vs cloud free pixels . . . . .	85
4.1.2	Snow/ice vs snow/ice free pixels . . . . .	86
4.1.3	Ocean vs land pixels . . . . .	86
4.1.4	Time of day sample bias . . . . .	87
4.2	Scene average CO <sub>2</sub> . . . . .	87
4.2.1	Scene average CO <sub>2</sub> biases in TM3 vs FSI comparison . . . .	88
4.2.1.1	Temporal distribution . . . . .	88
4.2.1.2	Spatial distribution . . . . .	92
4.2.1.3	Different methodologies for obtaining scene monthly average . . . . .	94
4.2.2	FSI vs ground stations . . . . .	97
4.3	Comparison with vegetation indices . . . . .	99
4.3.1	Carbon Flux Calculations . . . . .	99
4.3.2	Comparisons with MTCI . . . . .	102
4.4	Summary . . . . .	109
<b>5</b>	<b>NAME background chapter</b>	<b>111</b>
5.1	Numerical Atmospheric-dispersion Modelling Environment . . . . .	112
5.1.1	Origins . . . . .	112
5.1.2	NAME run set up . . . . .	113
5.1.2.1	Forward and backwards dispersion . . . . .	113
5.1.2.2	Tracer particles . . . . .	113
5.1.2.3	Meteorological data . . . . .	115
5.1.3	Operation of NAME . . . . .	115
5.1.4	Output of NAME runs . . . . .	117
5.1.5	Validation of NAME . . . . .	118



<b>6</b>	<b>Development of a method to investigate surface fluxes of carbon dioxide</b>	<b>120</b>
6.1	Author Developments . . . . .	120
6.1.1	Technical developments . . . . .	121
6.1.2	Scientific developments . . . . .	122
6.2	Development of NAME method . . . . .	123
6.2.1	The basics . . . . .	123
6.2.2	Thought experiments . . . . .	125
6.2.2.1	Thought experiment Case 1: Giant tube and constant fluxes . . . . .	125
6.2.2.2	Thought experiment Case 2: Giant tube and variable fluxes . . . . .	129
6.2.2.3	Thought experiment Case 3: Horizontal dispersion . . . . .	130
6.2.2.4	Thought experiment Case 4: Vertical and horizontal dispersion . . . . .	131
6.2.3	Initialisation . . . . .	133
6.2.3.1	“Edge box” initialisation method . . . . .	136
6.2.4	NAME model setup . . . . .	137
6.2.4.1	Original NAME setup . . . . .	137
6.2.4.2	Re-examining the NAME setup . . . . .	139
6.3	Offline method developments . . . . .	141
6.3.1	FSI measurement conditions . . . . .	141
6.3.2	Development of an inverse modelling technique to investigate CO <sub>2</sub> fluxes . . . . .	142
6.3.2.1	Units . . . . .	148
6.3.2.2	Majority rule for MODIS land product . . . . .	153
6.3.3	Development of a forward modelling technique to investigate CO <sub>2</sub> fluxes . . . . .	155
6.3.3.1	Mass change from coupling NAME with a pre-existing flux product . . . . .	155

6.3.3.2	Mass difference between release column and back-ground . . . . .	157
6.4	Summary . . . . .	158
<b>7</b>	<b>Synthetic testing of forward and inversion method</b>	<b>159</b>
7.1	Validation of initialisation method using CarbonTracker . . . . .	159
7.1.1	Validation of “edge box” initialisation method . . . . .	161
7.1.2	Validation of “freeze” initialisation method . . . . .	163
7.2	Validation of inversion method . . . . .	166
7.3	Summary . . . . .	172
<b>8</b>	<b>Analysis of forward and inverse modelling of FSI</b>	<b>173</b>
8.1	Analysis of FSI forward modelling results . . . . .	174
8.1.1	Comparison with individual releases . . . . .	174
8.1.2	Comparison with clustered releases . . . . .	184
8.2	Demonstration of FSI inverse modelling . . . . .	192
8.3	Summary . . . . .	195
<b>9</b>	<b>Conclusions and Future Directions</b>	<b>196</b>
9.1	Assessment of FSI WFM-DOAS retrievals . . . . .	196
9.2	Assessment of NAME method . . . . .	197
9.3	Outlook . . . . .	199
9.3.1	NAME model development . . . . .	199
9.3.2	Investigation of high northern latitude carbon fluxes in winter 199	
9.3.3	Inversion of retrieved CO <sub>2</sub> columns . . . . .	200
9.3.3.1	GOSAT . . . . .	201
9.3.4	Final Comments . . . . .	202

# List of Figures

1.1	Annual anomalies of global land-surface air temperature . . . . .	5
1.2	Variation of the Earth's surface temperature over the last 1000 years	7
1.3	Historical trace gas concentrations and radiative forcing . . . . .	12
1.4	Global mean radiative forcing . . . . .	13
1.5	The Global Carbon Cycle . . . . .	15
1.6	Fossil fuel emissions and surface uptake . . . . .	17
1.7	Regions of homogeneous ecoregions . . . . .	24
1.8	Global atmospheric circulation . . . . .	28
2.1	Seasonal cycle amplitude across latitude bands . . . . .	32
2.2	The GLOBALVIEW-CO <sub>2</sub> network . . . . .	35
2.3	Global flux inversion uncertainty vs measured column precision . .	36
2.4	The fractional change in flux uncertainty by including SCIAMACHY observations into an inversion scheme . . . . .	38
2.5	Simulated SCIAMACHY NIR spectral measurements . . . . .	41
2.6	Averaging kernels of thermal IR and NIR instruments . . . . .	42
2.7	Wave bands at which molecules measured by SCIAMACHY absorb	46
2.8	The orbital sequence of SCIAMACHY measurements . . . . .	46
2.9	SCIAMACHY viewing geometries . . . . .	47
2.10	Dimensions of SCIAMACHY nadir and limb observations . . . . .	48
2.11	Individual footprints clustered together in the swath. . . . .	49
3.1	Measurement of trace gas absorption in the open atmosphere . . . .	57

3.2	The DOAS principle . . . . .	58
3.3	Flow diagram of the FSI algorithm . . . . .	65
3.4	Flow chart of the calibration procedure . . . . .	66
3.5	The CO <sub>2</sub> vertical profiles used in the FSI algorithm . . . . .	68
3.6	Sensitivity to the vertical CO <sub>2</sub> profile . . . . .	72
3.7	SCIAMACHY XCO <sub>2</sub> vs Absorbing Aerosol Index . . . . .	77
3.8	Effect of aerosol fitting on yearly averaged retrieved CO <sub>2</sub> columns .	78
3.9	A typical example of a FSI WFM-DOAS fit . . . . .	79
3.10	Mean fit residuals . . . . .	80
3.11	SCIAMACHY and AIRS monthly averaged deviations over North America . . . . .	81
3.12	SCIAMACHY and AIRS monthly CO <sub>2</sub> anomalies over North Amer- ica during 2003 . . . . .	81
4.1	Collocated grid squares of the TM3 model and FSI retrieved CO <sub>2</sub> VMRs over Western Europe . . . . .	89
4.2	Time series of TM3 and SCIAMACHY CO <sub>2</sub> data over India . . . .	90
4.3	Time series of TM3 and SCIAMACHY CO <sub>2</sub> data over Siberia and North America . . . . .	91
4.4	Time series of TM3 and SCIAMACHY CO <sub>2</sub> data over Western Eu- rope and the Gobi desert . . . . .	92
4.5	FSI vs TM3 monthly scene averages over North America . . . . .	94
4.6	Different monthly averaging schemes . . . . .	95
4.7	Map of Siberia . . . . .	96
4.8	Time series of FSI vs ground station CO <sub>2</sub> data . . . . .	99
4.9	SCIAMACHY/FSI CO <sub>2</sub> observations versus land vegetation type over North America . . . . .	100
4.10	Simple carbon mass flux time series for Siberia 2003 . . . . .	101
4.11	Simple carbon mass flux time series for Siberian land types 2003 . .	102

4.12	Monthly averaged FSI CO <sub>2</sub> versus MTCI over grasslands for selected latitude bands. . . . .	104
4.13	Monthly averaged FSI CO <sub>2</sub> versus MTCI over woodlands for selected latitude bands. . . . .	105
4.14	Monthly averaged FSI CO <sub>2</sub> versus MTCI over deciduous broadleaf and mixed forests for selected latitude bands. . . . .	106
4.15	Maximum MTCI versus minimum FSI CO <sub>2</sub> for 1° by 1° grid squares over North American grassland 2003 and 2004. . . . .	107
4.16	Maximum MTCI versus minimum FSI CO <sub>2</sub> for 1° by 1° grid squares over North American agricultural land and woodland 2003 and 2004. . . . .	108
5.1	Output from NAME . . . . .	119
6.1	North American region divided into quantitative ecoregions . . . . .	125
6.2	Case one: horizontally constrained release particles and constant flux . . . . .	126
6.3	PBL height as a function of time of day . . . . .	128
6.4	Case two: horizontally constrained release particles and variable flux . . . . .	130
6.5	The effect of convection on NAME surface influence time . . . . .	132
6.6	Initialising the background using sun-glint . . . . .	135
6.7	Number of FSI measurements used in NAME inversion per month. . . . .	143
6.8	MODIS land classification gridded to a 1° by 1° map. . . . .	144
6.9	Overview of the inversion scheme . . . . .	149
6.10	Initialisation of the background CO <sub>2</sub> using CarbonTracker . . . . .	152
6.11	Overview of the forward modelling scheme . . . . .	156
7.1	Overview of initialisation methods . . . . .	160
7.2	Initialisation test using edge method. . . . .	162
7.3	Initialisation test using “freeze” method days 1 - 8. . . . .	164
7.4	Initialisation test using “freeze” method days 9 - 13. . . . .	165
7.5	Overview of inversion scheme validation 1 . . . . .	167
7.6	Overview of inversion scheme validation 2 . . . . .	168

7.7	Layout of the synthetic homogeneous carbon flux ecoregions . . . .	169
8.1	Mass flux plots for April 2003 . . . . .	176
8.2	Mass flux plots for May 2003 . . . . .	177
8.3	Mass flux plots for June 2003 . . . . .	178
8.4	Mass flux plots for July 2003 . . . . .	179
8.5	Mass flux plots for August 2003 . . . . .	180
8.6	Mass flux plots for September 2003 . . . . .	181
8.7	Mass flux plots for October 2003 . . . . .	182
8.8	Location of North American Release regions . . . . .	185
8.9	Area mass flux plots for April 2003 . . . . .	186
8.10	Area mass flux plots for May 2003 . . . . .	186
8.11	Area mass flux plots for June 2003 . . . . .	187
8.12	Area mass flux plots for July 2003 . . . . .	187
8.13	Area mass flux plots for August 2003 . . . . .	188
8.14	Area mass flux plots for September 2003 . . . . .	188
8.15	Area mass flux plots for October 2003 (adjusted) . . . . .	189
8.16	Area mass flux plots for October 2003 (not adjusted) . . . . .	189
8.17	Monthly mass flux for ocean ecoregion 2003. FSI was multiplied by 101.5% to adjust for the mean offset between CarbonTracker and FSI.	193
8.18	Monthly mass flux for ocean ecoregion 2003. No offset was made between CarbonTracker and SCIAMACHY/FSI. . . . .	193
8.19	Monthly mass flux for 5 ecoregions 2003. FSI was multiplied by 101.5% to adjust for the mean offset between CarbonTracker and FSI.	194
8.20	Monthly mass flux for 5 ecoregions 2003. No offset was made between CarbonTracker and SCIAMACHY/FSI. . . . .	194
9.1	The viewing geometry of GOSAT . . . . .	201

# List of Tables

1.1	Global Temperature Trends . . . . .	6
1.2	Lifetimes, radiative efficiencies and Global Warming Potential of greenhouse gases . . . . .	11
2.1	SCIAMACHY science channels . . . . .	45
3.1	Summary of the FSI retrievals and Park Falls FTIR comparison . .	82
4.1	Summary of the FSI retrievals and TM3 model comparisons . . . .	93
4.2	Summary of the comparison between SCIAMACHY and surface CO <sub>2</sub> data . . . . .	98
4.3	Summary of the validation of SCIAMACHY/FSI-retrieved XCO <sub>2</sub> .	110
6.1	Majority rule for MODIS land product . . . . .	154
7.1	Validation of linear regression inversion method . . . . .	170

# Chapter 1

## Introduction

### 1.1 Managing Climate Change

Changes to the Earth's climate are not a new phenomenon, climate changes (both local and global) have been happening throughout and before human history. Their impacts upon humanity has significantly altered the way that people have lived their lives. Often these changes have brought benefits allowing more land to be farmed, but these changes have also brought hardships to people, such as droughts, which have forced people to migrate or in other circumstances has led to collapse of civilisations such as the Mayans [Haug et al., 2003].

The development of agriculture has led to an exponential increase in the number of people living on this planet, many of them living in urban dwellings. Mass migration of people is no longer a realistic option in the adaptation to any future climate change, most inhabitable places on Earth are already settled and the potential for conflict over resources would be high in such a scenario.

For the last hundred years, humanity has had the ability to alter the climate, something it has been doing unconsciously for most of that time. A great campaign of burning fossil fuels has altered the chemical composition of the atmosphere, subtly changing the radiation balance of the Earth, which has led to climate change (discussed later).

Because the Earth's climate has, at least in global terms, been rather stable throughout much of the human population explosion and urbanisation (through the current inter-glacial period), climate change is likely to impact negatively on people's lives. There are two methods of reducing the impact of the changes on



the human population, mitigation and adaptation. Mitigation may result from a reduction in the emission of greenhouse gases that are behind the anthropogenic climate change that has been experienced or the enhancement of greenhouse gas sinks. Adaptation to climate change will take the form of ensuring that future reserves of water are protected and new land is developed for farming whilst existing land that may not support the old crops may be used for production of an alternative food group. If the mitigation efforts are weak then much stronger efforts will be needed in the adaptation to climate change, whilst significant efforts in the mitigation of anthropogenic climate change will mean that the need for adaptation is much smaller.

If adaptation to climate change is to be managed well, then understanding of the future Earth radiation budget is a requirement. To do this projections must be made about the future concentrations of greenhouse gases in the atmosphere. The purpose of this project is to improve the understanding of the land-atmosphere exchange part of the carbon cycle, which is significant as changes in atmospheric carbon dioxide have the most impact on the Earth's radiation budget of all the greenhouse gases. This understanding is also required if carbon sequestration methods are to be used to actively manage the global carbon cycle [Dilling et al., 2003].

## 1.2 Climate change

The Framework Convention on Climate Change (UNFCCC) defines climate change as “a change in climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods”. This differs from the Intergovernmental Panel on Climate Change [IPCC, 2001], who describe climate change as “the change in climate over time, whether owing to natural variability or as a result of human activities”. Climate is the long-term average of the weather. Whereas the weather can change every day, there are patterns of weather that we are familiar with, which follow a cyclical pattern through the seasons. Analysing the long term weather can show up inter-annual non-cyclical trends.

There is a vast wealth of information on previous natural patterns of climate change, in the ice ages, which occur in regular cycles lasting tens of thousands of years. Their regularity is a strong hint as to their cause: periodic variations in

the Earth's orbit around the sun, the so-called Milankovitch cycles. This must result in shock to the natural environment, although the long timescale of change probably allows adaptation to these changes to take place. Species living at certain latitudes could migrate north or south to experience similar climatic conditions. A mass extinction such as the one triggered by a meteorite impact 65 million years ago, was the result of a sudden change in the climate [Alvarez et al., 1980]. Many species were not able to adapt to the sudden change as they would have had no evolutionary experience of such a sudden catastrophe.

Mankind has been aware for quite some time of its effect on the climate, at least on a localised scale. Humanity has conducted a series of land use change programmes across the globe, creating savannas in North America, and accelerating desertification in sub Saharan Africa. In more recent years changes in atmospheric composition have been observed that are having a global effect on the climate. The two most significant are loss of stratospheric ozone and global warming.

### **1.2.1 CFCs and the ozone layer**

The Antarctic polar ozone hole was first observed in the 1980s [Farman et al., 1985], a consequence of the change in stratospheric composition. Chlorofluorocarbons (CFCs) used as refrigerants had already been identified as the likely cause of this ozone depletion [Molina and Rowland, 1974]. The consequences of stratospheric ozone loss are extremely damaging to living organisms, ozone prevents harmful ultraviolet radiation from reaching the Earth's surface. Threatened with an environmental disaster, the people of the world united in adopting the Montreal Protocol (1987). By 1994 the production of old style CFCs stopped, and further less damaging halocarbon products were also slowly phased out.

Why were CFCs so damaging to the environment? They were designed in the 1920s to replace the refrigerants of the day, gases like ammonia were unstable and posed a real danger to people using them. CFCs were designed to be extremely stable. This property allowed CFCs to travel up through the atmosphere without being broken down. Until their development, it was possible for only trace amounts of chlorine to reach the stratosphere (0.6 ppbv from natural  $\text{CH}_3\text{Cl}$ , where finally these gases could be broken down and begin a chain of ozone destruction.

Politically and economically, the problem of ozone destruction has proven far easier to solve than global warming. Its anthropogenic cause was much easier to

accept, since chlorine should not naturally occur in the stratosphere in the observed abundances. It was very clear that the effects of ozone loss would be very bad and there would be no positive benefits. It was also known that if production of CFCs ceased, the ozone layer would begin to recover. Finally, although companies producing the CFCs would be losers in this eventuality, alternative refrigerants were available. There was no contest between the two choices, the small cost of ending CFC production or the costs of skin cancer and crop damage.

This has provided modellers with a rather unusual scenario. Models are run where the anthropogenic activity suddenly stops to investigate how the environment continues to behave after such an event. For all intents and purposes, this is what has happened in the ozone layer. Ozone loss continued after 1994, as there is a lag in time between production and arrival in the stratosphere. It appears to have reached a peak in the last couple of years, and is expected to recover to pre 1980's levels by 2050 [IPCC, 2001; WMO, 2006]. It will be interesting to see how these projections work out. There are many parallels between ozone loss and global warming, and it will be interesting to see how humanity deals with the latter already having experience of the former.

### 1.2.2 Global Warming

What is the evidence that the Earth is getting warmer? The IPCC fourth report [IPCC, 2007] summarises various sources in the time frame up to 2006. They have compared several methods for obtaining air temperature above both land and sea (see Figure 1.1). There are challenges involved in obtaining a global average temperature dataset, with potential for biases arising from various sources. The locations of measurements sites are not spread homogeneously and this matters when the weather is also not homogeneous. Fewer measurements were made at the beginning of the time period.

Over the land they have compared the results of the Climate Research Unit, CRU [Brohan et al., 2006], the National Climatic Data Centre, NCDC [Smith and Reynolds, 2005], the Goddard Institute for Space Studies, GISS [Hansen et al., 2001], and the National Aeronautics and Space Administration, NASA [Lagina et al., 2005]. Over the ocean, there are results from the UK Met Office for daytime sea surface temperatures [Rayner et al., 2006] and night time marine air temperature [Rayner et al., 2003].

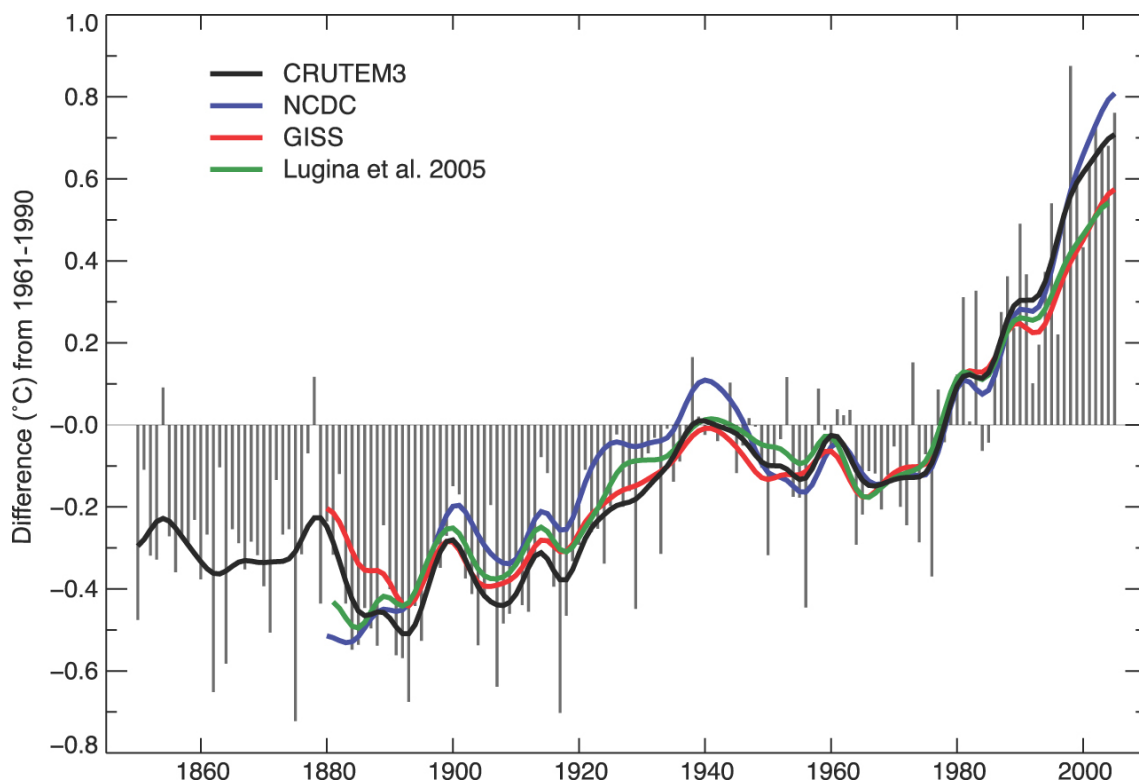


FIGURE 1.1. Annual anomalies of global land-surface air temperature ( $^{\circ}\text{C}$ ), 1850 to 2005 relative to the 1961 to 1990 mean [IPCC, 2007]. The smooth curves represent the decadal variations. The black curve from CRUTEM3 is compared with those from NCDC [Smith and Reynolds, 2005], blue, GISS [Hansen et al., 2001], red, and [Lugina et al., 2005], green. Differences arise from the diversity of spatial averaging techniques.

It is clear (see Table 1.1) that there has been a positive temperature trend over both the ocean and the land. The land has seen a greater temperature increase than the ocean. Temperature increases over the northern hemisphere have been larger than those of the southern hemisphere over both land and sea. The decadal temperature trend is increasing with the largest changes occurring in the last 25 years. This dramatic temperature increase is unusual compared with indirect temperature data from the last 1000 years [Mann et al., 1999; IPCC, 2007] as can be seen in Figure 1.2.

Models looking at the sum of non-anthropogenic forcing factors (the sum of solar and volcanic forcing) show a slight cooling trend and are increasingly separated from the observed changes. It is extremely unlikely ( $< 5\%$ ) that the pattern of warming in the past half century can be explained without anthropogenic forcing [IPCC, 2007]. It is very likely ( $> 90\%$ ), that the increase in anthropogenic greenhouse gases have caused most of the warming, and that without the cooling effect of

Table 1.1. Linear trends in hemispheric and global land-surface air temperatures, SST (shown in table as HadSST2) and Nighttime Marine Air Temperature (NMAT; shown in table as HadMAT1). For more information see IPCC [2007].

Dataset	Temperature trend ( $^{\circ}\text{C}$ per decade)		
	1850-2005	1901-2005	1979-2005
Land: Northern Hemisphere			
CRU [Brohan et al., 2006]	$0.063 \pm 0.015$	$0.089 \pm 0.025$	$0.328 \pm 0.087$
NCDC [Smith and Reynolds, 2005]		$0.072 \pm 0.026$	$0.344 \pm 0.096$
GISS [Hansen et al., 2001]		$0.083 \pm 0.025$	$0.294 \pm 0.074$
NASA [Lugina et al., 2005]		$0.079 \pm 0.029$	$0.301 \pm 0.075$
Land: Southern Hemisphere			
CRU [Brohan et al., 2006]	$0.036 \pm 0.024$	$0.077 \pm 0.029$	$0.134 \pm 0.070$
NCDC [Smith and Reynolds, 2005]		$0.057 \pm 0.017$	$0.220 \pm 0.093$
GISS [Hansen et al., 2001]		$0.056 \pm 0.012$	$0.085 \pm 0.055$
NASA [Lugina et al., 2005]		$0.058 \pm 0.011$	$0.091 \pm 0.048$
Land: Globe			
CRU [Brohan et al., 2006]	$0.054 \pm 0.016$	$0.084 \pm 0.021$	$0.268 \pm 0.069$
NCDC [Smith and Reynolds, 2005]		$0.068 \pm 0.024$	$0.315 \pm 0.088$
GISS [Hansen et al., 2001]		$0.069 \pm 0.017$	$0.188 \pm 0.069$
NASA [Lugina et al., 2005]		$0.069 \pm 0.020$	$0.203 \pm 0.058$
Ocean: Northern Hemisphere			
UKMO HadSST2 [Rayner et al., 2006]	$0.042 \pm 0.016$	$0.071 \pm 0.029$	$0.190 \pm 0.134$
UKMO HadMAT1 [Rayner et al., 2003]	$0.038 \pm 0.011$	$0.065 \pm 0.020$	$0.186 \pm 0.060$
Ocean: Southern Hemisphere			
UKMO HadSST2 [Rayner et al., 2006]	$0.036 \pm 0.013$	$0.068 \pm 0.015$	$0.089 \pm 0.041$
UKMO HadMAT1 [Rayner et al., 2003]	$0.040 \pm 0.012$	$0.069 \pm 0.011$	$0.092 \pm 0.050$
Ocean: Globe			
UKMO HadSST2 [Rayner et al., 2006]	$0.038 \pm 0.011$	$0.067 \pm 0.015$	$0.113 \pm 0.047$
UKMO HadMAT1 [Rayner et al., 2003]	$0.039 \pm 0.010$	$0.067 \pm 0.013$	$0.135 \pm 0.044$

aerosol from the same fossil fuel burning the warming would have been greater [IPCC, 2007].

In order to understand why an increase in the so-called “greenhouse gases” would cause a warming trend, the Earth’s radiation budget needs to be known.

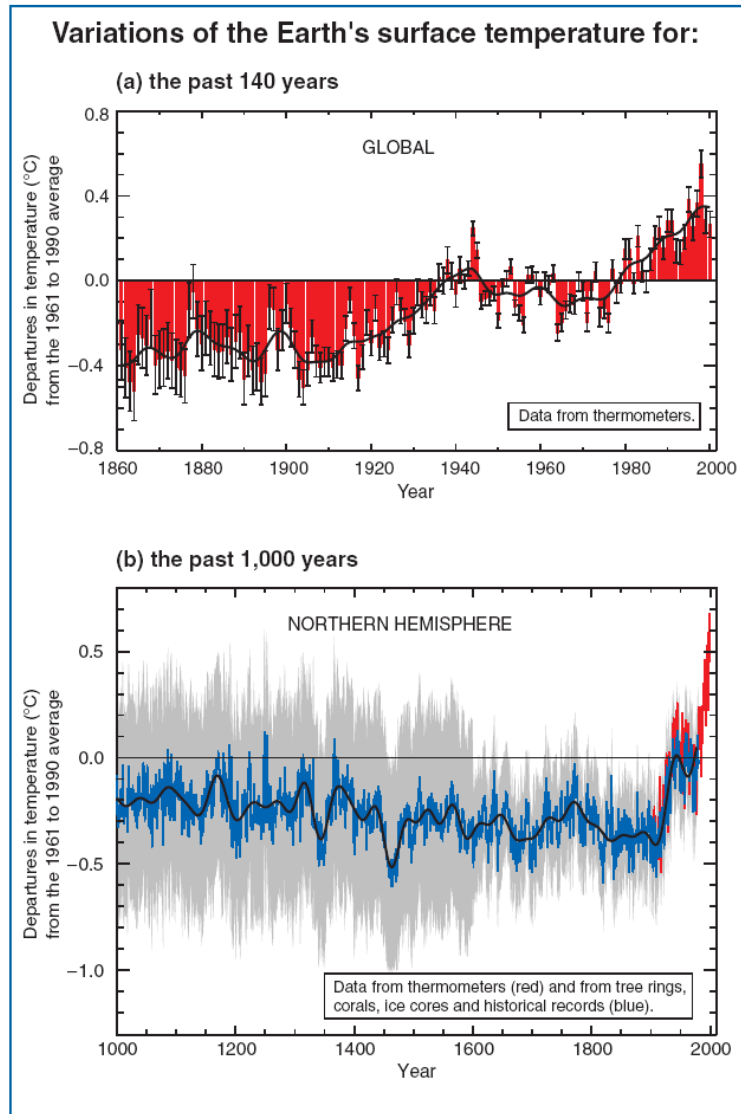


FIGURE 1.2. Variation of the Earth's surface temperature, indicated by the black lines, over the last (a) 140 years and (b) 1000 years, reconstructed from thermometer, ice core and proxy data [Mann et al., 1999; IPCC, 2007].

### 1.2.3 The Earth's radiation budget

What controls the temperature of a body in space? There is a tendency for such a body to reach thermal equilibrium. It will want to achieve a temperature at which the heat it loses to space equals the heat it acquires. The main source of heat at the Earth is solar radiation, and its main loss is thermal radiation. The Stefan-Boltzmann Law says that energy irradiated by an object is proportional to its temperature to the power of four.

$$4\pi R_E^2 \sigma T_E^4 = \pi R_E^2 (1 - A) S_\odot \quad (1.1)$$

where  $R_E$  is the radius of the Earth,  $\sigma$  is the Stefan-Boltzmann constant ( $= 5.6710 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ ),  $T_E$  is the effective temperature of the Earth (in Kelvin),  $A$  is the planetary albedo (i.e. the fraction of light reflected) and  $S_\odot$  is the Solar Constant ( $1370 \text{ W m}^{-2}$ ), although this number does depend on the output of the sun and the distance of the Earth from the sun. With a mean albedo of 0.3, the Earth's effective temperature is about 255 K. The Earth is considerably warmer with a mean surface temperature of about 288 K. This temperature difference is accounted for by the so-called “greenhouse effect”.

### 1.2.4 The Greenhouse Effect

The sun is much hotter than the Earth. The solar spectrum peaks around the visible wavelengths. The Earth's atmosphere is mostly transparent at these wavelengths so any radiation that is not reflected away will be absorbed at the Earth's surface. The Earth emits long-wave radiation, to which the atmosphere is not transparent. Homonuclear diatomic molecules ( $\text{N}_2$  and  $\text{O}_2$ ) which make up the bulk of the atmosphere and monatomic gases such as argon have no net change in their dipole moments when they vibrate and are unable to absorb the infrared radiation. However, the other gases (present only in small quantities in the atmosphere) do have a change in their dipole moment and they absorb the outgoing radiation. Some of this will be radiated back to the surface, creating a secondary source of heating to the Earth. The analogy to the actual greenhouse effect relates to glass allowing heat and light to pass through it, although the heat is prevented from escaping by convection not radiation in this case.

The surface temperature,  $T_S$ , is related to the effective temperature,  $T_E$ , by:

$$T_S = \left( \frac{T_E^4}{(1 - 0.5\beta)} \right)^{1/4} \quad (1.2)$$

where  $\beta$  is the fraction of the radiation that is absorbed by the atmosphere and the factor of a half results from half of the radiation being reradiated back to the ground. For the Earth this has the effect of increasing the surface temperature by 30 K above the effective temperature. Without this natural greenhouse effect, the Earth would be very cold and unlikely to be able to support the kind of

life we have now. Equation 1.2 gets complicated as the greenhouse gases are not evenly distributed and there are some complications about temperatures at different altitudes. If an event happened that changed the atmospheric concentration of greenhouse gases, then there would also be a change to the radiative forcing and a corresponding change in the surface temperature.

### 1.2.5 Radiative Forcing

The Earth's radiative system tends towards a steady state, where the net incoming solar radiation is balanced by the net outgoing radiation (reflected solar radiation plus emitted thermal radiation). A change in the magnitude of incoming or outgoing radiation will cause an imbalance which is referred to as radiative forcing,  $F$ , usually measured in units  $\text{Wm}^{-2}$ . Changes can happen to incoming radiation when there are changes to the Earth's orbit or to solar output. Of interest here are changes to outgoing radiation caused by changes to the concentrations of greenhouse gases. Models have, for several decades [Kiehl, 1986], shown that increases in atmospheric greenhouse gases (particularly  $\text{CO}_2$ ) have increased this radiative forcing, leading to a warming of the Earth's surface. Top of atmosphere radiation budgets can be obtained by satellites [Lin et al., 2010] to provide empirical evidence to complement the model data. However, instruments capable of measuring the incoming and outgoing radiation are only just starting to appear so there is no long-term dataset to show this.

### 1.2.6 The greenhouse gases

Different greenhouse gases will have differing abilities to affect the climate. Some will be more absorbing than others. Some have short lifetimes, so are unable to build up large quantities in the atmosphere. Non-natural greenhouse gases such as the CFCs, even in very trace amounts will have quite a large effect on the climate because they will be filling absorption lines that were previously transparent in the atmosphere, whereas an increase in  $\text{CO}_2$  will have much less of an effect as many of its lines are saturated. There is also the indirect greenhouse effect that is caused by increases in these gases, for example  $\text{H}_2\text{O}$  is very strongly absorbing, but has a very short lifetime in the atmosphere. The warming effect of increasing atmospheric  $\text{CO}_2$  allows the atmosphere to hold more moisture, thus increasing atmospheric  $\text{H}_2\text{O}$  concentrations and absorption. In this way  $\text{CO}_2$  acts as a driver



for climate change. The cooling effect of atmospheric aerosol, which has increased in concentration as a result of burning fossil fuels and increasing man-made fire events, should also be considered.

The IPCC fourth report [Manne and Richels, 2001; IPCC, 2007] adopts Global Warming Potential (GWP) as a standard for comparing the individual weighting of each gas to climate change. The index used (GWP) in the IPCC fourth report is weighted towards the final destination of climate change rather than the rate of change. This index works well in a smooth rate of increase scenario. However, it is well known that there are large pockets of methane gas in the Siberian traps and in Tropical peat land. If the steady rise in atmospheric  $\text{CO}_2$  and temperature reaches a critical tipping point, this methane could be released [Walter et al., 2006]. The GWP would be reasonable if this methane is slowly released over a hundred years, but would be meaningless if it is suddenly released over one year since there would be a great effect on the rate of climate change and adaptation would be difficult in such a short space of time.

The IPCC [2007] quote GWP for a few time horizons, twenty years, one hundred years, and five hundred years (see Table 1.2). The index is linked to the reference gas  $\text{CO}_2$  and is based on the radiative efficiency of the gas and its rate of decay in the atmosphere. In the fourth assessment report the GWP for methane includes indirect effects of enhancement of ozone and stratospheric water vapour. Three of the most important greenhouse gases,  $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{N}_2\text{O}$  are included in Table 1.2. A selected few halocarbons are also listed, whilst these products should be developed to tackle ozone destruction, consideration should also be given to their atmospheric lifetime, as many can remain for thousands of years and thus build up in the atmosphere. In the future this could make them a very significant contributor to the greenhouse effect.

The work in this thesis is concerned with the changes in atmospheric  $\text{CO}_2$ . What is so special about this one particular molecule? Changes in atmospheric concentration of the greenhouse gases since 1750 and the corresponding change in global radiative forcing will be considered. The year 1750 was chosen as the background as it marks the beginning of the industrial revolution and the beginning of anthropogenic changes to the atmosphere.

Life has thrived on the Earth for billions of years, and has had to put up with many climate changes in the past. However, human activity has reversed many of the Earth's natural feedback mechanisms, such as depositing of what we refer to

Table 1.2. Lifetimes, radiative efficiencies and direct Global Warming Potentials (GWP) relative to CO<sub>2</sub> (except for CH<sub>4</sub> which includes indirect effects from the enhancement of ozone and stratospheric water vapour) [Manne and Richels, 2001; IPCC, 2007].

Chemical Formula (or name)	Lifetime (years)	Radiative Efficiency W m <sup>2</sup> ppb <sup>-1</sup>	Global Warming Potential Given Time Horizon		
			20 years	100 years	500 years
CO <sub>2</sub>	variable	$1.4 \times 10^{-5}$	1	1	1
CH <sub>4</sub>	12	$3.7 \times 10^{-4}$	72	25	7.6
N <sub>2</sub> O	114	$3.0 \times 10^{-3}$	289	298	153
Substances controlled by the Montreal Protocol					
CFC-11	45	0.25	6730	4750	1620
CFC-12	100	0.32	11000	10900	5200
CFC-13	640	0.25	10800	14400	16400
CFC-113	85	0.30	6540	6130	2700
CFC-114	300	0.31	8040	10000	8730
CFC-115	1700	0.18	5310	7370	9990
Hydrofluorocarbons					
HFC-23	270	0.19	12000	14800	12200
HFC-32	4.9	0.11	2330	675	205
HFC-125	29	0.23	6350	3500	1100
Perfluorinated compounds					
SF <sub>6</sub>	3200	0.52	16300	22800	32600
NF <sub>3</sub>	740	0.21	12300	17200	20700
PFC-14	50000	0.10	5210	7390	11200
PFC-116	10000	0.26	8630	12200	18200

as fossil fuels in the ground. The background date of 1750 is quite reasonable in any case as ice core records (see Figure 1.3) show these concentrations to only vary gradually before this time.

Of the greenhouse gases, CO<sub>2</sub> has increased from a pre-industrial 280 ppmv to 379 ppmv in 2005. Prior to the industrial revolution, variations of less than 10 ppmv were observed on centennial timescales (see Figure 1.3). This increase has a corresponding radiative forcing of  $+1.66 \pm 0.17 \text{ Wm}^{-2}$ .

The abundance of CH<sub>4</sub> in 2005 was 1774 ppbv, a significant increase on its pre-industrial baseline of between 580 ppbv and 730 ppbv. In the last decade,

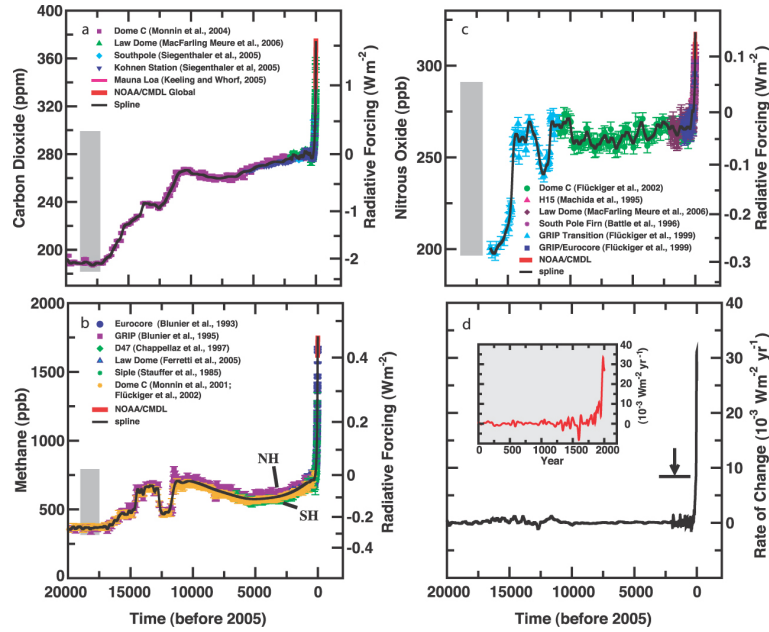


FIGURE 1.3. Historical trace gas concentrations and radiative forcing by  $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{N}_2\text{O}$  over the last 20,000 years reconstructed from ice cores (various symbols) and direct measurements (red and magenta lines) [IPCC, 2007]. The grey bars show the reconstructed ranges of the natural variability of these gases over the past 650,000 years.

atmospheric concentrations of  $\text{CH}_4$  seem to have leveled out. It is not surprising that  $\text{CH}_4$  would eventually find an equilibrium value in the atmosphere given its short lifetime (12 years). With no noticeable change in the availability of OH, the primary sink of  $\text{CH}_4$ , it was believed that the emissions of  $\text{CH}_4$  had leveled out [Dlugokencky et al., 2003]. However, unusually high arctic temperatures in 2007 and 2008 may have led to the observed increase in atmospheric  $\text{CH}_4$  [Dlugokencky et al., 2009].

$\text{N}_2\text{O}$  has also increased in the atmosphere by about 18% to 319 ppbv; the rate of growth is fairly linear. The increase of  $\text{N}_2\text{O}$  has increased the radiative forcing by  $0.16 \text{ Wm}^{-2}$ , with about 40% of emissions being anthropogenic.

The Montreal Protocol controlled gases, of which there are many, did not exist in pre-industrial times, and have a combined radiative forcing of  $+0.32 \text{ Wm}^{-2}$ . After  $\text{CO}_2$  and  $\text{CH}_4$ , CFC-12 is the third most significant long-lived gas, in terms of radiative forcing. Because these gases should be being phased out, it is expected that their concentration will decrease in the atmosphere.

Figure 1.4 shows that tropospheric ozone also has a positive effect on radiative forcing. It is short-lived but its increase in atmospheric concentration is a result of

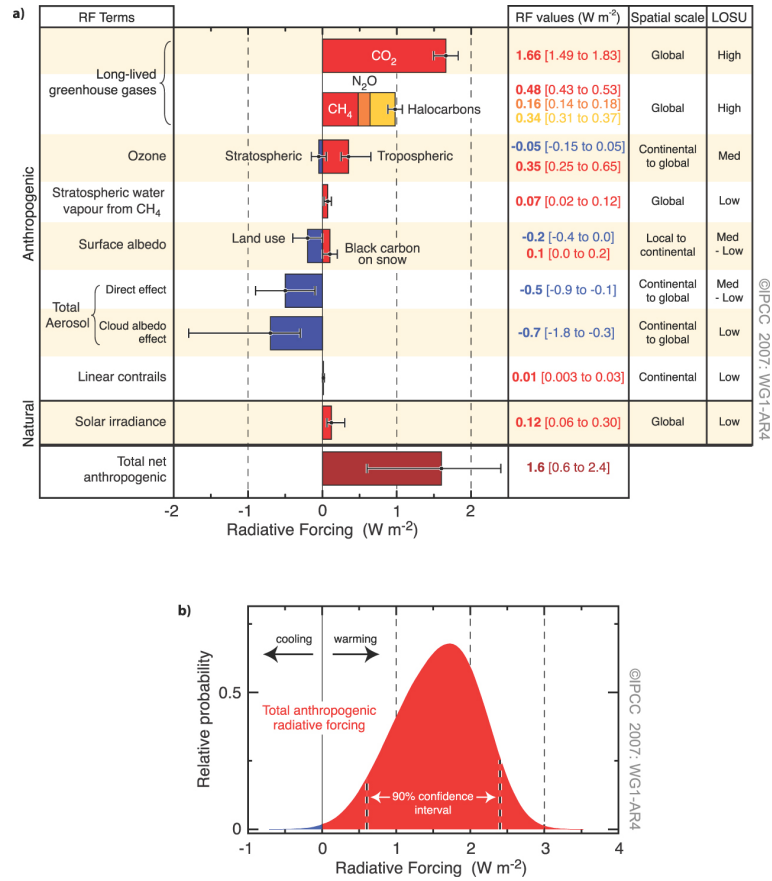


FIGURE 1.4. (a) Global mean radiative forcings of anthropogenic greenhouse gases and their 90% confidence intervals for 2005 [IPCC, 2007]. (b) The probability distribution function of the combined anthropogenic radiative forcings shown in “a” [Boucher and Haywood, 2001].

changes to the concentrations of precursor molecules.

Finally, it should be noted that the most significant greenhouse gas is water vapour, contributing to 60% of the natural radiative forcing [Kiehl and Trenberth, 1997]. It has a very short lifetime in the atmosphere before it is removed as rain, so it does not have the ability to build up. It is therefore a natural feedback in the climate system rather than a driving force. The warming effect of the other greenhouse gases could increase the rate of evaporation of water, creating a positive feedback to radiative forcing. It could also increase the presence of clouds leading to increased albedo and cooling, a negative feedback. There is no pre-industrial information about water vapour as it would be meaningless in an ice core sample. Its contribution to climate change is not well understood, but is beginning to be cited as a strong feedback to anthropogenic greenhouse gas forcing [Philipona et al., 2005].

## 1.3 The Carbon Cycle

To be able to predict the future change of temperature with any degree of accuracy it is necessary to be able to project the future radiative forcing, which in turn involves being able to project the future atmospheric concentration of carbon dioxide. If future atmospheric composition changes are to be modelled accurately, it is first necessary to understand the current global carbon cycle. This requires a deep knowledge of the stores of carbon and the transitions (or fluxes) between them. Note that the oxygen atoms that make up the carbon dioxide molecule remain in the atmosphere during both photosynthesis and respiration so there is no corresponding oxygen cycle to be concerned with.

It is not possible to have a complete understanding of the carbon cycle, to do so would require knowledge of the location of every carbon atom, of which there are an immeasurably large number of places where carbon is stored, and an even larger number of possible exchanges between them. To simplify the system and to allow the behaviour of the carbon cycle to be modelled, many small systems are banded together into larger systems. For example, all the organisms in a forest are treated as a single ecosystem; instead of having lots of individual stores of carbon there is a single ecosystem biomass and their individual carbon fluxes are pooled together as the gross primary production (GPP), net primary production (NPP) and net ecosystem production (NEP).

Because carbon dioxide has a long atmospheric lifetime, after it is released into the atmosphere it becomes mixed and eventually becomes part of the background. Given the magnitude of the flux signatures and the precision of future space-borne CO<sub>2</sub> instruments (OCO failed launch 2009 and GOSAT launched 2009 with measurement precision of better than 1%), biospheric signals should be indistinguishable from the background after 3 or 4 months, and fossil fuel signatures within one month [Palmer, 2008]. It is thus reasonable to consider the whole Earth in terms of biomass, atmospheric, fossil fuel, ocean and sediment storage and exchange (see Figure 1.5).

Some of these exchanges are fairly easy to understand. Going from fossil fuel to atmospheric CO<sub>2</sub> requires understanding the combustion process. The only limit to this exchange is the availability of oxygen in the atmosphere. There is no limit to the ability of the atmosphere to store carbon (we will see later that there is a limit to vegetation storage of carbon).

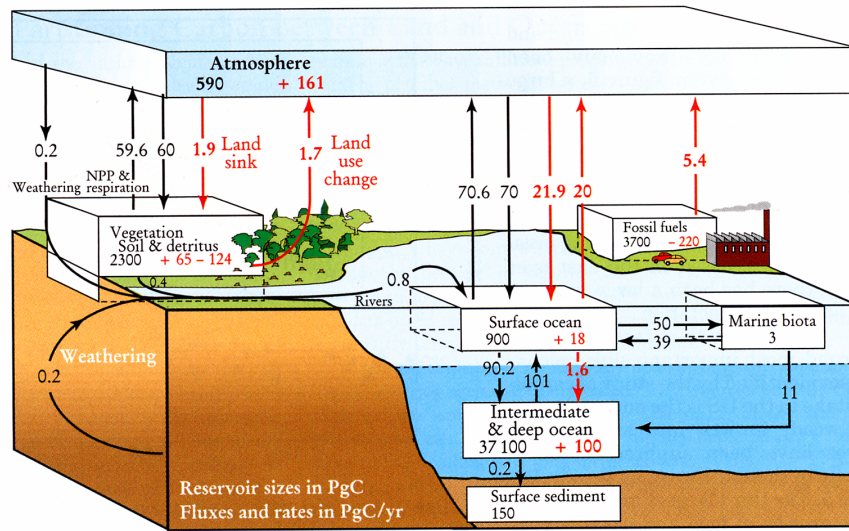


FIGURE 1.5. The Global Carbon Cycle: Fluxes are denoted by arrows in petagrams of carbon per year whereas boxes represent total storage in each reservoir (also petagrams). Pre-industrial stores and fluxes are shown in black whilst anthropogenic changes are shown in red. The first red number in the land reservoir box indicates the terrestrial uptake and the second reflects the decrease from deforestation. The net flux between natural processes is balanced whilst it is unequal for anthropogenic processes. Values are averaged over the 1980s by Sarmiento and Gruber [2002].

The rate of exchange between fossil fuels and the atmosphere before the industrial revolution is insignificant, and the number shown in black (Figure 1.5) as fossil fuel storage will have stayed the same for a very long time. The rate of formation of fossil fuels is incredibly slow; the fossil fuel burning should be considered a one way process. All of the fossil fuel carbon burnt is now stored in the atmosphere (where it was originally released), or the biosphere or the oceans (where some of it has now made its way). The amount and rate that fossil fuel is burnt is the best understood part of the cycle, since there are records of financial transactions between the people extracting it and the people paying for it (and presumably burning it).

To be able to estimate future atmospheric concentrations of carbon dioxide, the following should be well understood:

- Projected future fossil fuel emissions. This is not something that can be understood using physical science but with social sciences such as economics. It will be affected by economic growth, development of low energy technologies, take up of lower or zero carbon fuels and deliberate coordinated attempts to reduce fossil fuel emissions (such as the Kyoto protocol or more generally

public information about saving energy).

- The rate of exchange between the oceans and the atmosphere. The exchange happens in both directions and tends to equilibrium. The exchange happens at the ocean skin, where  $\text{CO}_2$  dissolves in the water. The rate depends on the partial pressures of  $\text{CO}_2$  between the air and sea, and can vary depending on temperature, ocean salinity and air pressure.
- The rate of exchange between the terrestrial biosphere and the atmosphere. This is rather more complicated than the ocean exchange case. It is often said that older ecosystems are net carbon neutral. This is assumed because soils have a limited volume, and thus have a limited ability to store carbon, there is also a limit to the amount of living biomass that can be supported above ground. Currently this is not the case, with observed  $\text{CO}_2$  uptake by the land surface. Of great interest is whether this trend will continue or perhaps reverse in the future.

The question of projecting future fossil fuel emissions will not be tackled in this thesis; for simplicity the emissions scenarios A1, A2 and B2 from the IPCC special report of Nakicenovic et al. [1998] will be referred to. The past and current fossil fuel emissions of  $\text{CO}_2$  are well known. It is clear that a significant proportion of emitted  $\text{CO}_2$  is no longer present in the atmosphere [Hansen and Sato, 2004] (see Figure 1.6), and so must have been taken up by the terrestrial biomass or the oceans.

There are two methods used to assign surface (both ocean and land) to atmosphere carbon fluxes: the bottom-up approach where localised information is used to model biospheric behaviour (including carbon flux magnitudes) and these are scaled up to larger regions and the top-down approach (inverse modelling) where measured variations in atmospheric concentrations of  $\text{CO}_2$  are used to identify sources and sinks of atmospheric carbon. At the beginning of this research project there had not been an example of satellite measured atmospheric  $\text{CO}_2$  being inverted to identify carbon fluxes. The development of a method for the inversion of SCIAMACHY/FSI retrieved  $\text{CO}_2$  is therefore of significant interest.

Another area of interest is to develop a framework to combine the top-down and bottom-up approaches to investigate carbon fluxes, as these two methods were rarely used together in the investigation of carbon fluxes. A brief description of

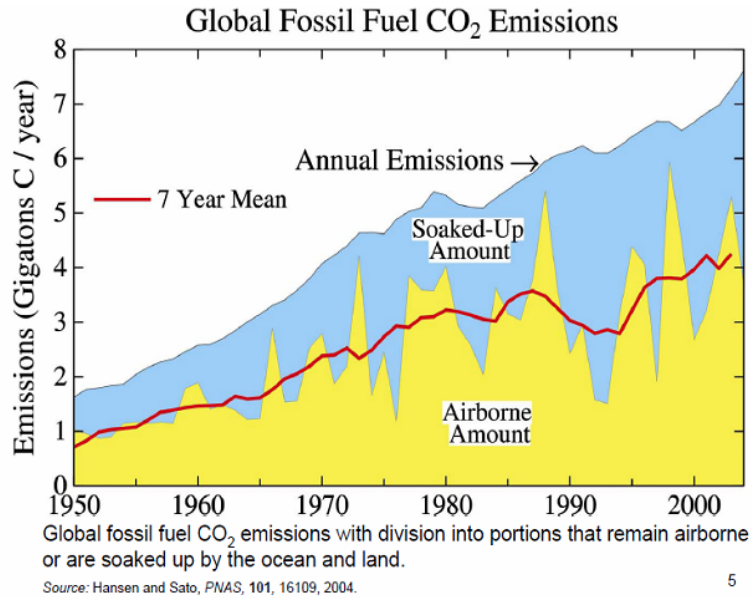


FIGURE 1.6. Plot of the global fossil fuel emissions and the fraction absorbed by the atmosphere and surface sinks (ocean plus terrestrial) since the 1950s [Hansen and Sato, 2004]. Whilst there is little inter-annual variation in the global emissions of fossil fuels (noting the steady increase), there is significant inter-annual variation in the absorption of carbon by the terrestrial biosphere.

both approaches to investigating carbon fluxes follows, as well as some examples being developed by other research groups.

### 1.3.1 Bottom-up estimates of carbon fluxes

On the whole bottom-up carbon flux estimates are produced purely as model output. Carbon cycle models have been developed with a set of parameterisations to govern how carbon can be stored and how it can be transferred between stores and a set of parameters are used to determine the rate of transition and where applicable (such as soils) the capacity of storage. *A priori* information is used to determine the rules of the model, for example a limited number of soil and vegetation types may be described and assuming that the model is grid based, each grid square (and in three dimensions layer) may be assigned a type. Some models designed to look at future carbon cycles may even allow these factors to change to allow for processes such as desertification.

A successful model when inputted with current weather information should be able to replicate the real world carbon flux magnitudes. A measure of this success can be made by including the surface fluxes into a transport model, and comparing



the resulting CO<sub>2</sub> field with Earth measurements [Chan et al., 2008]. However, there is always a risk that tuning the model to fit the real world will mean that a model will not properly deal with (the anticipated) future climate change.

Global carbon cycle models are fairly complex and usually built up with different modules to perform the different transition calculations (clearly the air-sea fluxes are fundamentally different from the biospheric processes which are different from the fossil fuel burning). When a model is upgraded, a comparison between the older and newer versions can yield different results [Cox et al., 1999].

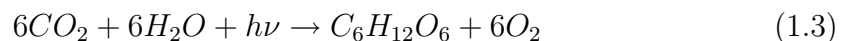
As these global chemical models become more advanced they are being applied to investigate future carbon cycle behaviours. In order to do this the models must couple radiative forcings and the corresponding climate information back into the carbon cycle model (a coupled climate model). This is very difficult to get right, especially including tipping points, and a recent inter-comparison of 11 such coupled carbon cycle climate models using the A2 emission scenario produced a huge spread of possible outcomes [Heimann and Reichstein, 2008].

The next two sections describe the physical processes upon which these global carbon flux models have been developed.

#### 1.3.1.1 Terrestrial biosphere to atmosphere flux processes

The terrestrial biosphere does contain a significant amount of carbon, around 2300 Gigatonnes (Gt), three times that stored in the atmosphere (see Figure 1.5). This is shared between the vegetation and the soils. The figure shows the storage and transitions between the atmosphere, the vegetation, litter and soils.

The dry weight of organic matter is composed of about 45-50% carbon. The term vegetation is used to refer to the whole biosphere as animals account for less than 0.1% of carbon in living organisms. Carbon is fixed into organic matter by vegetation using solar energy in the process of photosynthesis.



The amount of carbon taken up by photosynthesis is referred to as the Gross Primary Production (GPP), which globally is estimated to be about 120 GtC/yr [Ciais et al., 1997]. This is a significant proportion of atmospheric carbon and

means that all the atmospheric carbon gets turned over by the vegetation every few years.

Photosynthesis occurs in chlorophyll, a substance that gives plants their green colour. For most plants this is contained in the leaves. Photosynthesis requires sunlight, carbon dioxide and water to take place. There are three known methods used by plants to “fix” atmospheric carbon dioxide (binding the gaseous molecule to dissolved compounds inside the plant). All three methods of photosynthesis include the Calvin cycle stage. The first step of the Calvin cycle involves the  $\text{CO}_2$  molecule combining with ribulose-1,5 biphosphate (a five carbon sugar) in the presence of the enzyme RuBisCO to form two molecules of 3-phosphoglycerate (a pair of triple carbon molecules).

To begin the process of photosynthesis,  $\text{CO}_2$  has to enter the plant through the stomata (a pore in the plant leaves that can be opened and closed). The majority of plants are of type  $\text{C}_3$ , where the stomata typically stays open throughout the day and night. The gaseous  $\text{CO}_2$  molecules pass through the stomata, dissolve in the liquid water in the plant and react directly with the ribulose-1,5 biphosphate and enzyme [Calvin and Benson, 1948] (the term  $\text{C}_3$  refers to the fact that the initial product of the reaction produces a pair of triple carbon molecules). The stomata let water evaporate away from the plant leaves to the detriment of the plant, therefore the  $\text{C}_3$  process is the least water efficient method of photosynthesis.  $\text{C}_3$  plants thrive in areas of moderate sunlight intensity, moderate temperatures and plentiful ground water.

The RuBisCO enzyme also fixes oxygen and is relatively inefficient at fixing  $\text{CO}_2$ . A second enzyme (PEP carboxylase), which catalyzes the reaction of phosphoenolpyruvate (PEP) with  $\text{CO}_2$  to form the quadruple carbon molecule oxaloacetate, is present in  $\text{C}_4$  plants [Hatch and Slack, 1966]. The product is converted to malate which is transported to the bundle-sheath cells, where it is de-carboxylated to release  $\text{CO}_2$  and begin the Calvin cycle. The  $\text{C}_4$  process is more energy intensive but more water efficient, and can only out-compete  $\text{C}_3$  in regions of more intense sunlight and where ground water is limited. CAM (crassulacean acid metabolism) photosynthesis follows the same chemical process as  $\text{C}_4$ . However, CAM plants are adapted for the most arid conditions and only open the stomata during the night to conserve water. The  $\text{CO}_2$  molecules are trapped at night and take part in the photosynthetic reactions during the day. CAM significantly reduces the amount of  $\text{CO}_2$  available for photosynthesis, thus CAM type plants grow very slowly.

There are two patterns of cyclical behaviour to the rate of photosynthesis. One is the diurnal cycle and the other is the annual cycle. The diurnal cycle is very pronounced; photosynthesis does not take place at night for  $C_3$  and  $C_4$  plants because there is no sunlight. Then at sunrise, the rate of photosynthesis increases roughly proportionally to the amount of sunlight at first, then reaches a plateau where any further increase in solar irradiance has no effect [Blackman and Matthaei, 1901]. This plateau can be raised with an increase in temperature.

The stomata of  $C_3$  and  $C_4$  plants can close in times of drought, leading to a decrease in GPP and transpiration. Later the net carbon fluxes will be discussed, describing how the summer should see a net uptake of carbon from the atmosphere. However, if climate change results in less rainfall we may see net sources of carbon during the summer [Ciais et al., 2005].

The annual pattern of photosynthesis varies because of changes to the environment and to changes in the plants themselves. In temperate regions, the amount of sunlight hours varies greatly between summer and winter. There is also a significant temperature difference between the seasons which affects another process in the photosynthesis cycle. This means that the rate of photosynthesis is very low in the winter. Some plants lose their leaves in the autumn months, so as to conserve water and energy. In the tropics, the difference between seasons is less pronounced and the seasonal variation of fluxes is much less.

The carbon is returned to the atmosphere by the chemically identical processes of respiration or combustion. It is the reverse of the reaction that takes place with photosynthesis, with chemical energy released in the process. For most ecosystems, respiration tends to dominate over combustion as wild or man made fires are uncommon. There is no physical reason for the rate of respiration to be linked to that of photosynthesis in the biosphere; they are affected by different conditions. However, once an ecosystem is established there are good reasons why they will reach an equilibrium.

About half of all ecosystem respiration takes place in the vegetation. This process is called autotrophic respiration  $R_A$ . Unlike photosynthesis, respiration takes place throughout the whole of the organism. The difference between GPP and  $R_A$  is the Net Primary Production (NPP) which can be seen physically as the amount of plant matter produced. NPP is a useful measure when considering crop growth where the vegetation is not consumed in the same place as it is grown.

Heterotrophic respiration,  $R_H$ , is the amount of carbon respired by organisms

other than the primary producers. The plant matter may be eaten by animals, or it may fall as litter and be decomposed by microbial organisms above ground. A small amount may make its way into the soils where it will be broken down over a much longer timescale.

The diurnal cycles for both types of respiration are similar, which makes it hard to distinguish the two using a top-down approach. The rate of respiration is mainly related to temperature. During the night when it is cooler, the rate of respiration slowly reaches a minimum, whereas during the day it gently rises to a maximum when it is warmest. The day/night difference is small. The summer/winter cycle is more pronounced, since biotic activity virtually stops during the winter in cold regions.

The difference between GPP and the two forms of respiration is the Net Ecosystem Productivity (NEP). Over the course of a year, this could be positive (a net sink of carbon from the atmosphere), negative (a net source) or neutral. By adding this up for every ecosystem, we can find the global exchange between the atmosphere and the terrestrial biosphere.

Established ecosystems tend to be carbon neutral (although this is open to debate, there are examples of mature ecosystems acting as carbon sources or sinks [Chiti et al., 2010]). There are a number of feedbacks that would encourage this. The rate of photosynthesis find an equilibrium because there have been no noticeable trends in solar intensity or temperature (before the industrial revolution). The other factor that would influence this is leaf area index, and this does not change past a certain age of forest.

Whilst leaf area does not change, the tree continues to grow, which means that it will have to respire more. Trees cannot grow indefinitely; there is a maximum height which they can attain, which is set because the osmosis pressure that is used to take water from the roots to the top of the tree cannot take it above a certain height. Above ground, vegetation finds a natural limit to the biomass that can be supported. This sets a limit to the rate at which litter falls and consequently sets the balance for exchanges with the soils as well.

In the case of land use change, there will be a change in NEP. The land use element is often separated from NEP (see Figure 1.5) although it is the same process. If a forest turns to grassland because of climate change or to cropland because of deliberate human interference, there will be a change in carbon mass stored as vegetation, owing to the fact that forests store more carbon.

Alternatively, trees may be planted and a new forest created. The first few years of any forest will be a carbon source, even if the trees are growing as planting will disturb the soil.

Because respiration and photosynthesis occur in the same place and use or produce  $\text{CO}_2$ , it is difficult to distinguish the magnitude of the two from atmospheric concentrations, though not impossible [Yi et al., 2004]. The overall flux is gently negative during the night with a sudden rise to positive sometime after sunrise, which then plateaus during the mid-afternoon, before falling back to negative before sunset.

There is strong evidence of a current terrestrial carbon sink in the high northern latitudes. From the current understanding there has been a noticeable declining trend in the terrestrial take up of atmospheric carbon. A range of studies, with different emission scenarios, suggest a minimum uptake of less than 10% to a maximum uptake of 34% of emissions sequestered by the terrestrial biosphere with peak uptake before mid century with a decline or even a reversal [IPCC, 2007].

Some studies project increases in soil decomposition will balance increases in plant productivity in the next 80 years [Smith et al., 2006] but there are a lot of uncertainties, such as the temperature sensitivity of decomposition [Davidson and Janssens, 2006] and precipitation patterns [Mitchell et al., 2004].

#### **1.3.1.2 Ocean to atmosphere flux processes**

In the case of fluxes between the atmosphere and the ocean it is clear that increasing atmospheric  $\text{CO}_2$  will result in a net flux to the ocean as the partial pressures attempt to reach equilibrium. The solubility of  $\text{CO}_2$  in sea water is affected by temperature (being more soluble in cold water), salinity and pressure. Since this is a skin effect, wind speed and waves will also be a factor. Although oceans have a very large capacity for storing carbon, the very slow turnover of ocean currents means that most of the anthropogenic dissolved carbon is still near the surface resulting in a much smaller partial pressure gradient than would be expected if ocean dissolved carbon was well mixed.

Some processes such as the carbonate pump (where the shells of deceased phytoplankton drop into the deep ocean) may accelerate the oceanic take up of atmospheric  $\text{CO}_2$ . The oceans are the likely end destination for anthropogenic  $\text{CO}_2$ ,

assuming all known fossil fuel reserves are burnt then anthropogenic CO<sub>2</sub> will have an average atmospheric lifetime of 1800 years [Montenegro et al., 2007].

The increasing trend of the airborne fraction of anthropogenic CO<sub>2</sub> implies that carbon emissions have been growing at a faster rate than the carbon sinks on the land and oceans. As the land and ocean are mosaics of regions that are gaining and losing carbon, this trend could be the result of any of the following scenarios: weakening sink regions (either absolutely or relatively to growing emissions), intensification of source regions or sink regions that have transitioned to source regions [Canadell et al., 2007].

A recent inter-comparison of 10 ocean circulation models has been used to identify a small (0.3 Pg of carbon per year) CO<sub>2</sub> sink of atmospheric carbon in the southern ocean [Gruber et al., 2009]. It is thought that the southern ocean sink has become weaker in recent years, owing to the poleward displacement and intensification of westerly winds which have enhanced the ventilation of carbon-rich waters (isolated from the atmosphere at least since 1980). The weakening of the southern ocean sink is responsible for half of the fractional change in the oceans uptake of anthropogenic CO<sub>2</sub> [Canadell et al., 2007].

### **1.3.2 Obtaining carbon fluxes through micrometeorological methods**

The distinction between the top-down and bottom-up approaches is clear but there are some methods that incorporate elements of both. In particular, ecologists have developed a micrometeorological method (commonly referred to as eddy covariance) for determining very local scale biosphere-atmosphere fluxes of carbon [Baldocchi et al., 1988]. On this localised scale, eddy covariance is a top-down method. However, hundreds of these sites have been set up, and they are categorised by vegetation type, soil type, etc. These measurements sites take a sample of flux information from a large cross-section of biomes. This flux information can then be used in a vegetation and soil model (possibly coupled with the weather), where the carbon flux from the most appropriate eddy covariance site is incorporated. This soil/vegetation model should be considered to be a bottom-up approach.

Regional scale fluxes can be obtained by scaling up the localised fluxes over a larger region. The bottom up approach is particularly powerful when used in transport models, run with current or projected future weather. By considering

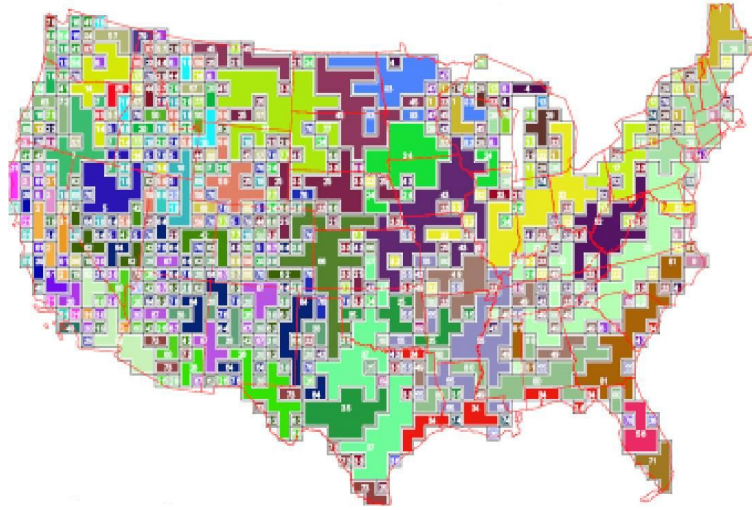


FIGURE 1.7. Regions of homogeneous ecoregions [Hargrove and Hoffman, 2005]. Multivariate clustering was used to identify an arbitrary number (in this case 50) of the most diverse ecoregions, after which the gridded regions are assigned to the ecoregion (represented by the different colours) that it matches best.

different parameters that would affect fluxes, multivariate clustering can be performed to identify homogeneous flux ecoregions [Hargrove and Hoffman, 2005] (see Figure 1.7).

Eddy covariance has also been adapted in the investigation of ocean to atmosphere  $\text{CO}_2$  fluxes [Miller et al., 2010].

Micrometeorological methods have been used to explain unusual patterns in the observed atmospheric  $\text{CO}_2$  field. For example, the extremely large seasonal cycle amplitude of atmospheric  $\text{CO}_2$  measured at high northern latitude stations was found to be caused by Siberian  $\text{CO}_2$  efflux during winter [Zimov et al., 1996], a process that was not thought to happen on account of the extreme cold preventing soil respiration during this period.

Another significant finding was that during the European heat-wave and drought in the summer of 2003, primary production fell to such an extent that Europe became a net source of atmospheric  $\text{CO}_2$  [Ciais et al., 2005], which is of great concern if in future years such events are expected to become more frequent. On the other hand, the increasing levels of atmospheric  $\text{CO}_2$  could be encouraging net take up of carbon in African tropical forests [Lewis et al., 2009].

### 1.3.3 Top-down estimates of carbon fluxes

The top-down approach, the basis for this research project, involves measuring concentrations of carbon dioxide in the atmosphere. If there were no fluxes to the atmosphere, the concentration everywhere would be the same, there would be no variation from the background. The measured variations in the CO<sub>2</sub> field are the result of previous fluxes to and from the atmosphere. It is possible to trace these variations back to the source. In a very simple case you may see smoke from a fire over the horizon; you know where the fire is even if you cannot see it.

There are usually many solutions to an inverse problem and the task is to find the best one. A good inversion solution can be obtained if there is a great contrast between the background and measured quantity (which is not the case for the atmospheric CO<sub>2</sub> field), if it incorporates a reliable atmospheric transport model, if there is a great degree of spatial (true of satellites) and temporal (not so true of sun-synchronous satellites) sampling, if the measurements are made close to the time/space of the source/sink and finally if *a priori* knowledge of the system is used in the inversion.

Previously, a variety of measurement types (from ground stations, tall towers, aircraft) and a variety of inversion techniques have been used to invert atmospheric CO<sub>2</sub> measurements to obtain surface flux information. The most successful of the inversion techniques used to produce global carbon flux maps are based on data assimilation techniques such as 4D-VAR, one such example is CarbonTracker [Peters et al., 2007].

Most models contain unknown conditions, perhaps the initial conditions (CO<sub>2</sub> field) and other parameters (such as the CO<sub>2</sub> fluxes). Four-dimensional variational data assimilation “4D-VAR” allows these parameters to be tuned to produce a realistic result. This is done by optimising the fit between the model solution and the set of observations (in the case of CarbonTracker the CO<sub>2</sub> ground station measurements). In 4D-VAR this is done iteratively, the fourth dimension in 4D-VAR reflecting the fact that the observations span across both space and time. 4D-VAR is one of the main tools used in numerical weather prediction (NWP).

Data assimilation techniques seem to be the favoured approach for the inversion of satellite retrieved CO<sub>2</sub>, with several methods being developed (in parallel to this research project) using both 4D-VAR [Engelen et al., 2004, 2009] and similar methods such as ensemble Kalman filters [Feng et al., 2009].



### 1.3.3.1 Inverse modelling of point measurements

Whether collected in a flask for later analysis or sampled instantaneously using a spectroscopic method, such measurement will be referred to as point (or *in-situ*) measurements. The quality of point measurements (taken at a ground station, on board a plane, at a tall tower, on a ship or buoy) is expected to be much higher than that of remotely sensed (such as satellite) measurements of atmospheric CO<sub>2</sub>, since many factors affecting the retrieved spectra can be controlled (pressure, presence of clouds and aerosol). In the point measurement case, the uncertainties in the measured CO<sub>2</sub> are insignificant compared to the time variation of the measured atmospheric CO<sub>2</sub>; whereas the uncertainties in remotely sensed atmospheric CO<sub>2</sub> concentrations are on the same order of magnitude as the expected changes. The optimisation in a 4D-VAR inversion of point measurements is heavily weighted towards the measurement over the model. A network of high precision CO<sub>2</sub> measurement locations [ESRL, 2008b], are used in the basis of most inversions (shown in Figure 2.2).

Results from the TransCom 3 project, an inter-comparison of 16 inversion schemes where GLOBALVIEW CO<sub>2</sub> was used [Gurney et al., 2002], has shown that for the majority of the worlds oceans (with the notable exception of the Southern ocean) top-down inversion estimates of carbon fluxes agree with the bottom-up carbon inversion estimates based on partial pressure models. For land regions with a high density of ground stations, such as Europe and North America, Gurney et al. [2002] also shows that the inversion uncertainties are small, so that total regional flux estimates are very reliable in these regions. The carbon flux uncertainties over Asia, Africa and South America are large however, owing to the sparsity of measurement sites and (in tropical regions) poorer performance of the transport models.

Since a key objective of top-down modelling of carbon fluxes is to understand how the flux behavior changes with localised conditions (to be able to model these changes in the future), a much more localised understanding of the carbon cycle is needed. This could be achieved, for example, by nesting a high resolution regional model (and high density of CO<sub>2</sub> measurements) into a courser global model [Rodenbeck et al., 2009].

### 1.3.3.2 Inverse modelling of satellite column measurements

A thought experiment devised by Rayner and O’Brien [2001] was used to demonstrate that at a certain measurement precision (about 2.5 ppmv or 1%) the inversion of a vast number of such satellite column retrievals would result in global inverted carbon fluxes of similar quality to that of the current high precision, low density ground network. However, the distribution of ground stations is not homogeneous and the inversion of lower measurement precision CO<sub>2</sub> columns in the Asian, African or South American continents would still result in an improvement in the precision of regional flux estimates. A more detailed description of the precision requirements of satellite retrieved CO<sub>2</sub> columns is given in Section 2.2.

Although used with point (aircraft) measurements [Lin et al., 2007], Lagrangian (“air mass flowing”) methods have largely been ignored for the purposes of inverting satellite retrieved CO<sub>2</sub> measurements. This is rather surprising, considering it provides a better description of regional scale atmospheric transport. Its major drawback is that it is very difficult to initialise the CO<sub>2</sub> background in a Lagrangian model. Very recently, Rodenbeck et al. [2009] has suggested nesting a Lagrangian model inside a global chemical transport model, to obtain the benefits of improved transport from Lagrangian modelling whilst retaining initialisation information from the GCM.

## 1.4 Atmospheric circulation

Global atmospheric circulation is the large scale movement of air and is a consequence of the uneven distribution of heat over the Earth’s surface. There exists a global-scale general circulation pattern that, although subject to some inter-annual variation, remains fairly constant. It consists of six cells of atmospheric circulation that exist along six latitude bands (three in the Northern Hemisphere and three in the Southern Hemisphere) that circumnavigate the Earth. The three southern cells mirror the three northern cells.

The northern Hadley Cell extends roughly between the Equator and the 30<sup>th</sup> parallel. The northern and southern Hadley Cells are separated by the Intertropical Convergence Zone, along the line of the Solar Zenith (which is not fixed owing to the tilt in the Earth’s rotational axis, relative to the Sun). The intense heating at this region drives convective movement, lifting air upwards and creating a region of low

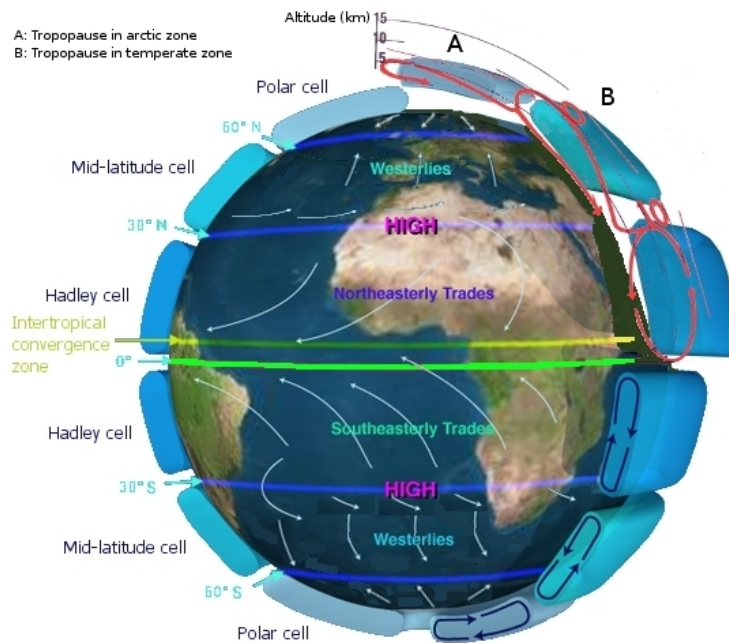


FIGURE 1.8. An idealised diagram of global atmospheric circulation. In the Northern Hemisphere, warm air rises near the equator, traveling north and falls at the high pressure region around  $30^\circ$  north creating the Hadley Cell. To complete the cell the near surface air travels as the Northeastly Trade Winds. The behaviour of the Polar Cell is very similar, where the extreme cold at the pole drags air down, creating a cell extending from about  $60^\circ$  north up to the pole. In between these two cells is the Ferrel Cell, a secondary circulation feature driven by the other cells. The near surface air in this region typically travels from West to East, hence the name Westerlies. Since the region of interest, North America, mainly falls within the region between the  $30^{th}$  and  $60^{th}$  parallels, there is likely to be a general westerly trend in the atmospheric transport. The Subtropical Jet Stream is located between the Hadley and Ferrel Cells and the Mid-latitude Jet Stream is located between the Ferrel and Polar Cells. The general atmospheric circulation pattern is mirrored in the Southern Hemisphere (image courtesy of NASA).

pressure. This air moves north at high altitude until it reaches around  $30^\circ$  north, where it falls to the surface and the returns as the Northeastly Trade Winds (in the Southern Hemisphere the Trade Winds are Southeasterly. The Hadley Cells dominate atmospheric transport in this latitude region excluding almost all other weather systems. Wind becomes very weak in the Intertropical convergence Zone creating the Doldrums (see Figure 1.8).

The Polar Cell extends roughly from the  $60^{th}$  parallel to the pole. The general pattern of warm air rising, moving northwards, sinking and traveling southwards follows the same pattern as the Hadley Cell. The driving force of the Polar cell is the

intense cold at the pole, creating a High pressure area pushing the air southwards. The Polar Cell dominates all other weather systems in this latitude band to such an extent as to exclude them. Preventing the Hadley and Polar Cells joining is the effect of the Coriolis Force which deflects air to the left of its direction of motion in the Northern Hemisphere and to the right in the Southern Hemisphere. At the northern boundary of the Hadley Cell and the southern boundary of the Polar Cell are the Subtropical and Mid-latitude Jet Streams respectively, both westerly wind features that circumnavigate the Earth (see Figure 1.8).

The Ferrel Cell is a secondary circulation feature driven by the Polar and Hadley Cells extending roughly from the 30<sup>th</sup> parallel to the 60<sup>th</sup> parallel. It is a significantly weaker feature than the two other cells and the prevailing winds (the Westerlies) are in competition with other weather systems (anticyclones in the Northern hemisphere). At the lower limit of the Ferrel Cell are the high pressure horse latitudes and at the upper limit exists the low pressure Polar Front (see Figure 1.8).

There are two significant effects of the general circulation pattern on the carbon cycle research carried out in this thesis. Firstly, the region of interest (North America) falls mainly within the Ferrel Cell and there is a general westerly transport of near-surface air (where the evidence of land-atmosphere carbon exchange will be greatest). Secondly, air travels relatively freely along lines of latitude but with great difficulty across the line of latitude and especially across cells. This, combined with variable land-atmosphere carbon exchange, leads to a latitudinal gradient in atmospheric CO<sub>2</sub> (see Figure 2.1).

## 1.5 Thesis overview

In Chapter 2, the reasons for choosing the SCIAMACHY instrument over the AIRS instrument for retrieving atmospheric CO<sub>2</sub> columns are explained. A brief description of its functions and orbit are given, followed by an explanation about how the instrument orbit constrains the way that the carbon cycle can be investigated.

Chapter 3 describes how remote sensing can be used to determine atmospheric concentrations of CO<sub>2</sub>. It also describes the development of the FSI retrieval scheme, and highlights the sources of uncertainty of individually retrieved CO<sub>2</sub> columns.

The authors contribution to the validation of FSI retrieved CO<sub>2</sub> columns is presented in Chapter 4. Of particular interest is the uncertainty in the atmospheric CO<sub>2</sub> field produced from an ensemble of measurements. This information was used to assess the suitability of using SCIAMACHY/FSI data to validate bottom-up model fluxes in a forward scheme. It was also used to assess the suitability of using SCIAMACHY/FSI data in an inversion scheme, using the criteria set out by Rayner and O'Brien [2001].

Chapter 5 gives a brief introduction to the functioning of the NAME model before any changes were made by the author. Chapter 6 describes the development of the Lagrangian method developed in this research project. This model was designed with many of the science questions set out in this introduction chapter in mind. The model included a framework for relating bottom-up flux maps with top-down estimates, which could allow the investigation of homogeneous ecoregion carbon emissions.

Chapter 7 uses synthetic retrievals to investigate how successfully the model could perform forward and inverse modelling. Real SCIAMACHY data was used to investigate carbon cycle behaviour in Chapter 8. A summary of the research project and outlook for further research are provided in Chapter 9.

## Chapter 2

# Remote sensing of atmospheric carbon dioxide

### 2.1 Variability of CO<sub>2</sub> in the atmosphere

A brief description of carbon cycle processes was given in Section 1.3. In order to use variations in the atmospheric CO<sub>2</sub> field to investigate these carbon cycle processes, the nature of these interactions should be understood. In the case of the air-sea transfer of carbon, the entire process occurs at the boundary between the atmosphere and the ocean (at the very lowest level of the atmosphere). In the case of the terrestrial biosphere to atmosphere exchange of carbon, all transfers occur below the maximum canopy height (between 0 and 100 metres), which is certainly below the planetary boundary layer (the turbulent lowest level of the atmosphere, air is quickly vertically mixed within this turbulent layer). Making up the smallest proportion of the carbon cycle is the burning of fossil fuels, which in most cases happen close to the surface. In some cases this air will be released from a tall chimney and being hotter than surrounding air may rise quickly to higher altitudes.

At a time soon after the emission to or absorption from the atmosphere of carbon, there will be a noticeable increase or decrease in the atmospheric mixing ratio of CO<sub>2</sub> in the lowest level of the atmosphere. As time continues to pass, the evidence of this emission or absorption event to the atmosphere becomes weaker as atmospheric mixing dilutes the “plume” of CO<sub>2</sub>. The global atmospheric circulation patterns of air (see Section 1.4) help set up latitude bands, so that atmospheric concentrations of CO<sub>2</sub> vary quite significantly when measured across latitudes but

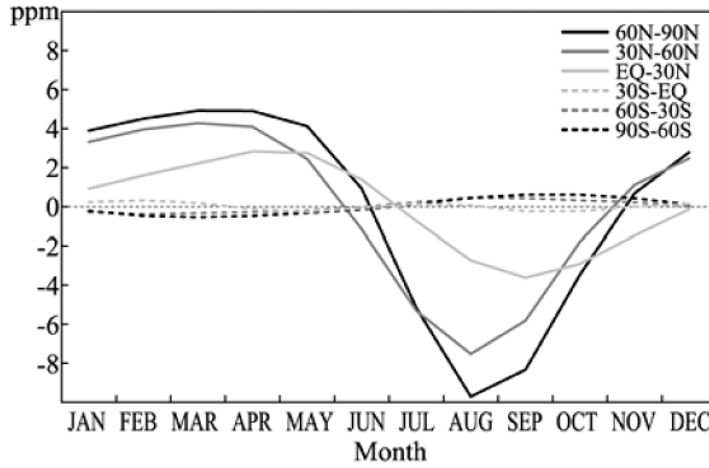


FIGURE 2.1. The seasonal cycle  $\text{CO}_2$  amplitude for each  $30^\circ$  latitudinal zone minus the long term growth rate. Figure reproduced from WMO [2005].

much less when measured across longitudes on the same latitude band. Vertical mixing of a plume may take several days or weeks within the troposphere, and the time taken for mixing to occur across the tropopause is even longer. Mixing of air across the equator is extremely slow and since most fossil fuel burning occurs in the northern hemisphere, the increase in the atmospheric mixing ratio of  $\text{CO}_2$  in the southern hemisphere lags several years behind that of the north. Another difference across the equator is the reversal of seasons, so that the maximum and minimum occur at different times. The seasonal cycle amplitude of atmospheric  $\text{CO}_2$  is greater in the northern hemisphere (see Figure 2.1), owing to there being more land mass and thus terrestrial biosphere in the north.

It is desirable to measure the variations in the  $\text{CO}_2$  field to understand the spatial and temporal variability of the exchange of carbon between the terrestrial biosphere and the atmosphere. Carbon fluxes can be investigated directly, for example using the eddy covariance method [Baldocchi et al., 1988] which uses micrometeorological techniques, though this can only be applied to a very localised area. The alternative approach is the inversion of measured atmospheric concentrations of  $\text{CO}_2$ . There already exists a network of local  $\text{CO}_2$  measurement sites [ESRL, 2008b], characterised by sparse spatial and high temporal coverage with very high measurement precision and accuracy, which has been inverted to produce a global carbon flux product, CarbonTracker [Peters et al., 2007]. It is desirable to complement this network with the high spatial density of measurements that can be provided by satellite measurements.

### 2.1.1 Variability in the CO<sub>2</sub> column

One of the difficulties with using satellite remote sensing to investigate the carbon cycle is that most of the variation in CO<sub>2</sub> concentration occurs at the lowest part of the atmosphere whereas the instrument is looking for variation across the entirety of its line of sight. Simulations by Olsen and Randerson [2004] have shown that the diurnal cycle of total column CO<sub>2</sub> rarely exceeds 1 ppmv, that the seasonal cycle amplitude of retrieved column CO<sub>2</sub> is half of that measured at the surface, and that the mean column north-south gradient is also half that measured at the surface. The diurnal cycle measured at the surface can be as large as 40 ppmv in some regions, whilst the instrument precision is 0.1 ppmv or better. The combination of location of measurement and instrument precision results in a disparity in the signal-to-noise ratio of ground measurements compared to satellite measurements.

## 2.2 Precision requirements for flux inversions

The term “precision requirements” when discussing the possibility of flux inversions is somewhat loaded. Even with a very low precision ground-based instrument it should not be a challenge to recognise that there is a diurnal cycle, which is in essence just a product of local fluxes. By taking high precision measurements at a location not affected by local sources over a long enough time period, it should be possible to reproduce the annual cycle and the current increasing trend shown by Keeling [1960] at Mauna Loa, in the famous Keeling Curve.

A single measurement station can produce enough information to demonstrate the total CO<sub>2</sub> exchange between the atmosphere and the other main stores of carbon, but it will tell us little about the spatial variation of these fluxes or the magnitude of the exchange between the individual stores. To investigate in more detail, it is useful to have measurements from many locations and to have a high frequency of measurements at those locations.

Using a suitable inversion scheme, such as 4D-VAR data assimilation, it should be possible to combine these different measurements together. The required inputted information would be the measurement value, its precision and the time and location of the measurement. It should not matter whether the measurement is from a flask CO<sub>2</sub> concentration or from a CO<sub>2</sub> column measured from a satellite instrument. Each individual measurement should help improve the understanding



of the exchange of CO<sub>2</sub> with the atmosphere.

Rightly or wrongly then, most published literature on precision requirements of space-based measurements of CO<sub>2</sub> refer to the limits at which satellite instruments can provide flux precisions comparable with those of existing ground networks, rather than the precision required to understand surface processes, to meet Kyoto protocol regulations, or to measure anthropogenic emissions. There are also discrepancies between these results.

The surface network [ESRL, 2008*b*], is coordinated by the Carbon Cycle Greenhouse Gases Group of the National Oceanic and Atmospheric Administration, Earth System Research Laboratory (NOAA ESRL), although it is a cooperative effort of the many national organisations that provide the data. The network has grown to more than a hundred sites, with measurements made from towers, balloons, aircraft, buoys, and ships. A few sites are capable of measuring continuously with infrared analyzers, but most are simple with flask collections taken at regular or irregular intervals to be sent away.

The cooperative nature of the project means that there is data from a lot of sites that no longer record information. Much of the world is not sampled by these networks, especially in the southern hemisphere, (see Figure 2.2). Geographical and temporal gaps in measurements need to be smoothed out to make a useful product for studying the carbon cycle. This can be done using data extension and data integration techniques [Masarie and Tans, 1995]. This approach is used to create the GLOBALVIEW-CO<sub>2</sub> product, that is then used to model the carbon cycle.

Using the same network of measurements an entirely different approach to obtaining carbon fluxes is obtained through data assimilation, [ESRL, 2008*a*]. Individual measurements are used to reanalyze the recent flux history of CO<sub>2</sub>, by coupling a transport model to an ensemble Kalman filter [Peters et al., 2007]. Efforts are being made to expand this method to other observation types, such as satellite retrievals and eddy flux measurements.

A brief theoretical paper by Rayner and O'Brien [2001] set the standard for all future satellite precision versus ground network precision. In this paper Rayner and O'Brien [2001] set out to calculate the global flux uncertainty based on the uncertainties of 26 regions. The data uncertainties of the regions was obtained from the variability of the sample locations to the monthly mean fit (assuming the network of 56 GLOBALVIEW stations, at the time of publication). The data

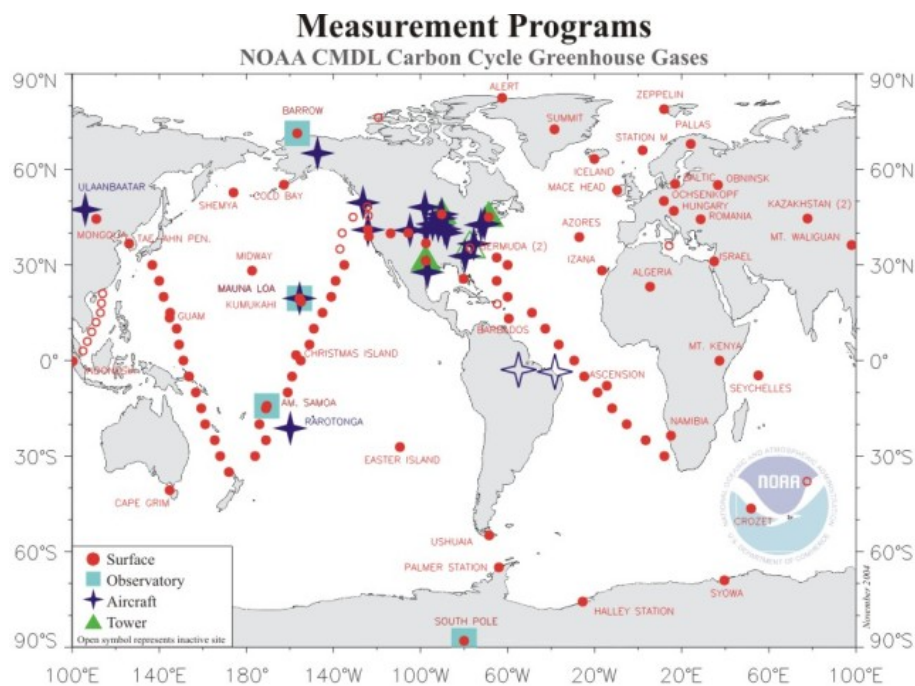


FIGURE 2.2. The GLOBALVIEW-CO<sub>2</sub> network (image courtesy of NOAA). The NOAA CMDL group operates four measurement programs: Measurements from tall towers and aircraft began in 1992, in-situ measurements are made at the CMDL baseline observatories and from fixed sites and commercial ships as part of the cooperative sampling network.

uncertainty was also calculated as a function of measurement precision, for a set of pseudo-satellite measurements.

The results from the paper are rather abstract, in that the precision of monthly averaged column data needs to be better than 2.5 ppmv on an 8° by 10° footprint for comparable performance with the existing surface network, or 1.5 ppmv for ocean only sampling (this is assuming samples using sun glint) (see Figure 2.3). The global coverage, the quantity of data available, and the dilution of surface fluxes (there are major problems in comparing GLOBALVIEW stations at the tropics with the rest of the network) are the advantages that suggest the integration of satellite retrievals can provide a comparable constraint to the current surface network. The ground network strengths are high precision measurements, the lack of dilution of surface fluxes (except at the tropics), and the temporal resolution. It is clear from the study the column measurements are much better suited than point measurements for inclusion in an atmospheric transport model.

Other studies examine the regional improvements to carbon cycle understanding that can be made with varying degrees of satellite instrument precision. In

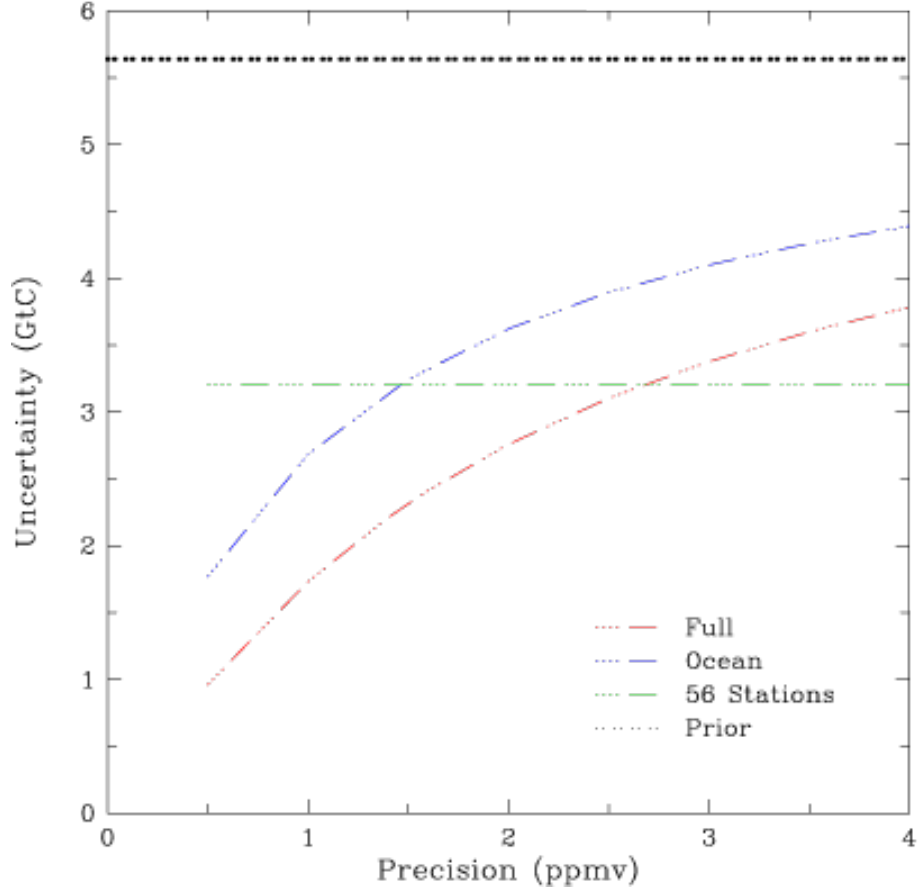


FIGURE 2.3. The global flux inversion uncertainty (GtC/yr) as a function of column precision for both the ocean-only case (blue) and for global coverage (red), compared with the uncertainty of the ground network (green) [Rayner and O’Brien, 2001].

their discussion, Houweling et al. [2004] relates the sensitivities to surface fluxes of SCIAMACHY, OCO, NOAA/CMDL, AIRS (in its current form, referred to as AIRS(L)), and a thermal infrared instrument with a more complete wavelength coverage, which they refer to as AIRS(U). SCIAMACHY is not able to measure over the oceans in this study, and its footprint size means 10% of measurements should pass a cloud filter (they also examine where they will encounter more and fewer cloud free regions). An individual measurement precision of 3.6 ppmv is assumed. There are typically 15 measurements to an ensemble, so there is an ensemble uncertainty of 1.65 ppmv. The  $1.6\ \mu\text{m}$  absorption band is treated as having a constant vertical weighting function [Dufour and Breon, 2003].

OCO is given a more conservative individual measurement precision of 3.6 ppmv, although with its small footprint size means there are many more available measurements, with an ensemble precision almost to the minimum of 1% (an approximation

of systematic errors). Both cases for thermal infrared produced near 1% ensemble precision.

First a pulsed emission in the model, from the location of a ground site was analyzed [Houweling et al., 2004]. The expected response of the ground station was a factor of 10 greater than that for OCO, which was in turn a factor of 10 greater than the response of AIRS in its current form. However, the ground station response to fluxes tails off very quickly as the distance to the pulse increases, whereas the satellite instruments are still sensitive to the pulse over some distance and time.

When the world is divided into much smaller flux regions than those used by Rayner and O’Brien [2001], the improvements to flux uncertainties from *a priori* are shown for the different instruments and also the ground network. SCIAMACHY performs very well over all the continents in this respect (see Figure 2.4), which is understandable as this is both where SCIAMACHY sampling is best, and where the *a priori* understanding of fluxes performs the worst. Potentially, it could also outperform the ground network.

Houweling et al. [2004] conclude that satellite performance to reduce flux uncertainties improves with increasing sensitivity to the near surface. The thermal infrared instrument AIRS, has a slight advantage in global coverage over SCIAMACHY, although the use of sun glint functions by an OCO-like instrument should swing the balance back towards the near infrared instruments.

Flask networks have a high precision relative to the background, not just because you can measure the concentration to about 0.1 ppmv but also because the diurnal variation at the surface is much larger than the column. Precision of individual measurements from SCIAMACHY FSI retrievals are approaching 1% or around 4 ppmv [Barkley et al., 2006a], and for OCO-like retrievals this would be greatly improved to around 0.7 ppmv for high sun conditions and around 1.5 to 2.5 ppmv for low sun (winter) conditions [Connor et al., 2008]. This precision can be improved using an ensemble of measurements. The single measurement precision can be improved by a factor of  $\sqrt{1/n}$ . This method for improving the precision is limited by the spatial variation of atmospheric CO<sub>2</sub> concentrations and it does not improve the accuracy of the soundings.

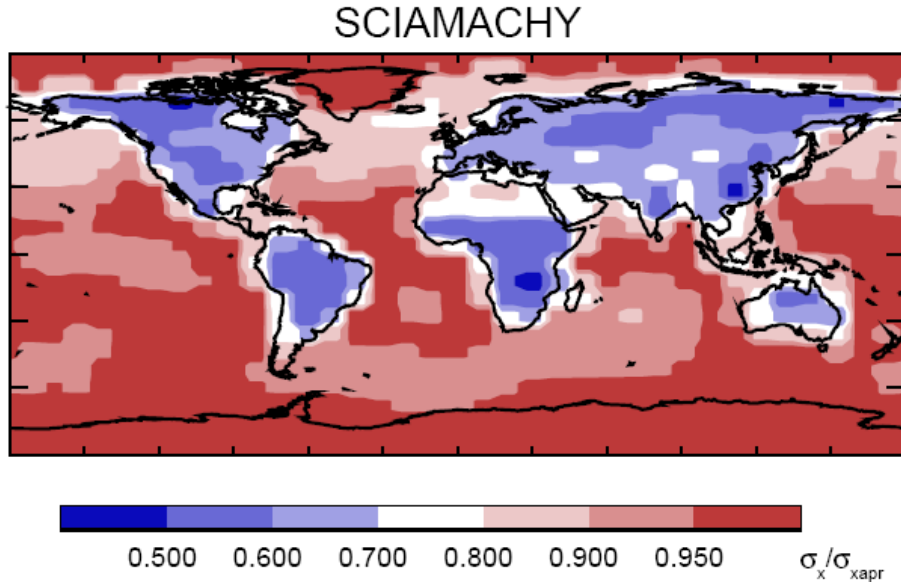


FIGURE 2.4. Fractional change in flux uncertainty ( $\sigma_x / \sigma_{x_{\text{apriori}}}$ ) gained through the inversion of retrieved SCIAMACHY CO<sub>2</sub> columns (assuming an individual measurement precision of 3.6 ppmv) per year and per model grid box [Houweling et al., 2004]. Blue areas indicate regions where SCIAMACHY can significantly reduce flux uncertainties.

### 2.2.1 CarbonTracker

In the latter chapters of this thesis, it is necessary to compare remotely sensed atmospheric columns of CO<sub>2</sub> with a different product representing the atmospheric concentration of CO<sub>2</sub>. The CarbonTracker CO<sub>2</sub> flux (and weather) product [ESRL, 2008a] provides a very useful comparative tool, being both easily available and a very good approximation of the true atmospheric CO<sub>2</sub> concentration.

The atmospheric CO<sub>2</sub> measurements of the ESRL [2008b] network (see Figure 2.2) are assimilated in a 4D-VAR scheme to produce the flux and CO<sub>2</sub> weather products [ESRL, 2008a]. Individual measurements are used to reanalyze the recent flux history of CO<sub>2</sub>, by coupling a transport model to an ensemble Kalman filter [Peters et al., 2007].

Peters et al. [2007] evaluated the success of CarbonTracker by comparing the product with a set of 13,000 independent flask samples of CO<sub>2</sub> taken in the free troposphere. The annual mean and standard deviation of the distribution is  $0.07 \pm 1.91$  ppmv (the bias is somewhat larger when it was assessed by season, with a model overestimate of  $0.27 \pm 2.67$  ppmv in the summer and underestimate of  $0.15 \pm 1.47$  in the winter). Column average mole fractions of CO<sub>2</sub> were measured by a Fourier transform spectrometer (FTIR) at Park Falls, Wisconsin. Differences

between CarbonTracker and the FTIR instrument were small ( $\Delta = 0.5 \pm 0.9 \text{ ppmv}$ ) with a high linear correlation  $R^2 = 0.93$  with strong agreement in the diurnal and seasonal cycles [Peters et al., 2007]. Similar efforts to obtain column precision of the CarbonTracker  $\text{CO}_2$  weather product involved comparisons with vertical profiles obtained through aircraft measurements with strong agreement ( $\Delta = 0.05 \pm 2.7 \text{ ppmv}$  and  $R^2 = 0.83$ ) [Peters et al., 2007].

It should be noted that the independent flask comparison was carried out over the USA, where the density of measurement locations is highest. It is far easier for the model to interpolate between many nearby measurement stations over the USA than it is to produce accurate model values at locations very distant from measurement stations. An error map, if it existed, of the CarbonTracker product would most likely show North America and Europe to have the highest model precision and accuracy, with vastly poorer performance over the other continents which are barely sampled (see Figure 2.2). The findings of Peters et al. [2007] show that North America is an ideal location to validate column retrieved  $\text{XCO}_2$  using the CarbonTracker product.

In their study [Schneising et al., 2008], the Bremen WFM-DOAS retrieval scheme mixing ratios were compared with modelled mixing ratios from CarbonTracker. It should be remembered that Bremen retrievals use  $\text{O}_2$  columns instead of pressure to define path length and use different aerosol scenarios to FSI. Firstly, the model  $\text{CO}_2$  mixing ratio profile (CarbonTracker has  $\text{CO}_2$  mixing ratios gridded in the vertical as well as horizontal directions) was multiplied by the SCIAMACHY/FSI averaging kernel, which negates the biases between retrieved and true  $\text{XCO}_2$ . This is very convenient for comparisons. In theory it is desirable to instead modify the SCIAMACHY retrieval to bring this closer to the true  $\text{XCO}_2$  column concentration, since this would lead to a true column mass difference. This cannot be done in practice, although retrieval schemes for OCO and GOSAT are close to a constant averaging kernel. The difference between applying or not applying the SCIAMACHY averaging kernel, results in a difference of typically less than 1 ppmv in the CarbonTracker column.

Individual measurement precision error is estimated to be around 5-6 ppmv for the WFM-DOAS retrievals, whilst Schneising et al. [2008] suggests a standard deviation of within 2.7 ppmv in the summer and 1.5 ppmv in the winter. To account for systematic biases between the two data sets, the WFM-DOAS columns are scaled up by a factor of 1.015, which could be used for FSI also. The northern hemisphere seasonal cycle amplitudes of SCIAMACHY (6 ppmv) are greater than

those of CarbonTracker (4 ppmv). On a regional scale, the variability of WFM-DOAS is typically 4 ppmv larger than CarbonTracker, although this can be as much as 8 ppmv.

Further validation of CarbonTracker with FSI-WFM-DOAS is described in Section 6.2.3.

## 2.3 Remote sensing of carbon dioxide

Atmospheric carbon dioxide interacts with electromagnetic radiation, which in turn is detected by the satellite instrument. In a perfect measurement, the magnitude of this interference is directly related to the number of CO<sub>2</sub> molecules in the column through which the satellite looks. However, many sources of uncertainty creep into the measurement which affects both the accuracy and the precision of the retrieved CO<sub>2</sub> mixing ratio.

SCIAMACHY is not the ideal instrument to investigate CO<sub>2</sub> fluxes as it has a number of sub-optimal elements:

- The sun-synchronous orbit, means that all measurements are made 10 am local time. It is not possible to obtain information about the diurnal cycle from SCIAMACHY alone, and thus is impossible to separate out the source and sink terms in an inversion scheme [Yi et al., 2004].
- The weak CO<sub>2</sub> feature at 1.61  $\mu\text{m}$  is not going to produce the highest precision space-based measurements.

Since every space instrument capable of measuring atmospheric CO<sub>2</sub> is also in a polar orbit, none will be able to give information about the diurnal cycle. It would be technically quite difficult to design a geostationary satellite instrument capable of measuring atmospheric CO<sub>2</sub> with a high precision. This is because the intensity of reflected sunlight falls by the square of the distance, creating a much lower signal-to-noise ratio.

The Atmospheric Infrared Sounder (AIRS) instrument, measuring in the thermal infrared (3.7  $\mu\text{m}$  - 15.4  $\mu\text{m}$ ), has also been used to measure atmospheric concentrations of CO<sub>2</sub> [Engelen et al., 2004]. The term sounder refers to the fact that AIRS can also measure water vapour and temperature as a function of height.

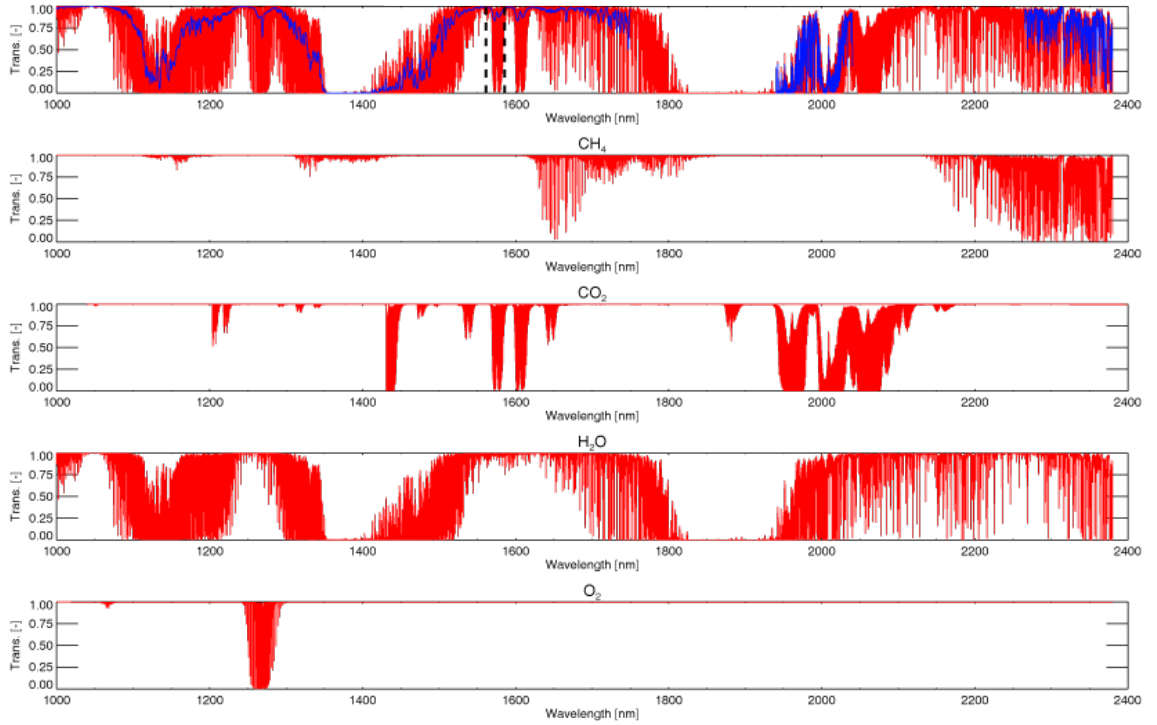


FIGURE 2.5. Simulated SCIAMACHY NIR spectral measurements [Barkley, 2007].

An additional problem for SCIAMACHY remote sensing is that the weak  $\text{CO}_2$  feature at  $1.61 \mu\text{m}$  overlaps with water vapour features at this wavelength, (see Figure 2.5). Figure 2.5 shows simulated NIR transmission in the  $1.0$  to  $2.4 \mu\text{m}$  window [Barkley, 2007]. The top panel shows the total transmission, followed by  $\text{CH}_4$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{O}$  and  $\text{O}_2$  respectively. The vertical black lines show the spectral fitting window used by the WFM-DOAS retrieval scheme [Buchwitz et al., 2000a] and the FSI-WFM-DOAS retrieval scheme [Barkley et al., 2006a] described in the following chapter. A novel solution, such as the correlated k distribution scheme, could be used to separate out overlapping gases [Buchwitz et al., 2000b].

There are some advantages of near infrared (NIR) retrievals over the thermal infrared when investigating the carbon cycle, which is why the one was chosen in this study over the other. A brief description of the two is given below.

### 2.3.1 Thermal Infrared Techniques

Carbon dioxide absorbs a significant amount of thermal infrared radiation in the  $15 \mu\text{m}$  band from  $600$  to  $800 \text{ cm}^{-1}$ . This spectral region corresponds to the maximum intensity of a black body Planck radiance curve at  $300 \text{ K}$  (similar to the Earth's



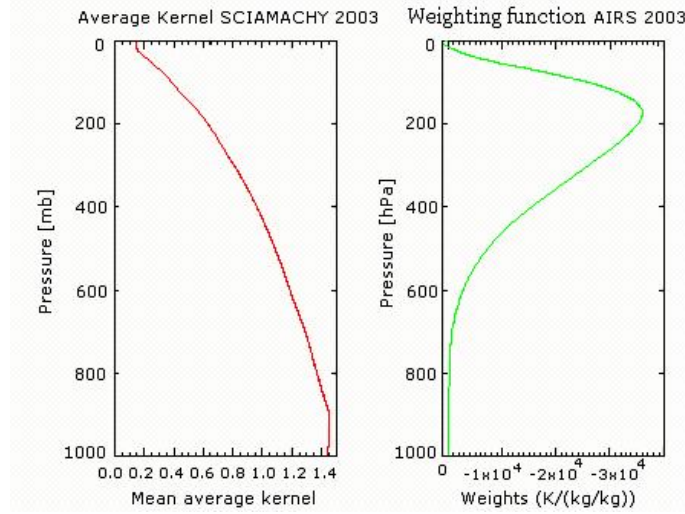


FIGURE 2.6. Averaging kernels of thermal IR (such as AIRS) and NIR (such as SCIAMACHY) instruments [Barkley, 2007].

surface). The emission by  $\text{CO}_2$  of radiation in this wavelength region is also very strong. Since these absorption features are optically thick, a thermal infrared instrument such as AIRS which retrieves atmospheric  $\text{CO}_2$  is not sensitive to the emission and absorption that happens at the lowest part of the atmosphere (see Figure 2.6).

The AIRS instrument is capable of measuring both over ocean (since thermal emission processes in the atmosphere are unaffected by surface processes) and also at night (since the energy detected by the instrument originates from the earth and not the Sun). The application of retrieved thermal infrared  $\text{CO}_2$  columns such as those of Engelen et al. [2004] are greatly limited by the fact that the instrument is not sensitive to the lowest parts of the atmosphere where the carbon absorption and emission processes are taking place. A very recent study by Chevallier et al. [2009] demonstrates that although inversion of AIRS-retrieved  $\text{CO}_2$  columns can improve the *a priori* fluxes, it is not capable of improving on the flux inversion of the ground network.

### 2.3.2 Near Infrared Techniques

There are several  $\text{CO}_2$  absorption bands in the near infrared which could be exploited by a passive instrument to determine the atmospheric concentration of  $\text{CO}_2$ . The strongest band from the major transition  $00^0_1$  at  $4.3 \mu\text{m}$  is in the tail of the solar Planck function and cannot be used. The next strongest major transition

$10^01$  at  $2.7\ \mu\text{m}$  overlaps with a water vapour absorption band and is unsuitable. The third strongest major transition  $20^01$  at  $2.0\ \mu\text{m}$  and fourth strongest major transition  $30^01$  at  $1.6\ \mu\text{m}$  are both within the spectral range of the SCIAMACHY instrument (see Section 2.4).

Since water is highly absorbing in this wavelength region, it is not possible to retrieve over oceans (and for that matter ice and snow) in the near infrared. Since the origin of the radiation detected by the instrument is the Sun, there is no possibility of retrieving  $\text{CO}_2$  columns on the night side nor is it possible to retrieve at very high latitudes during the winter months, which makes investigation of the Siberian  $\text{CO}_2$  efflux during winter (described by Zimov et al. [1996]) impossible. However, the optically thin nature of the absorption bands means that measurements by instruments (such as SCIAMACHY) in the near-infrared are sensitive to variations throughout the atmosphere (see Figure 2.6), particularly the lowest levels where the carbon fluxes are taking place.

## 2.4 SCIAMACHY

### 2.4.1 Instrument concept

The SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) instrument is a passive hyper-spectral UV-VIS-NIR grating spectrometer [Bovensmann et al., 1999]. It is designed to measure sunlight transmitted, reflected and scattered from the Earth's atmosphere or surface.

In response to an ESA call for experiments to fly on-board the polar platform, the SCIAMACHY proposal [Burrows et al., 1988] was submitted as a potential part of the atmospheric chemistry payload. SCIAMACHY was selected as one of the ten instruments that would fly on the renamed ENVISAT (environmental satellite), which launched on March 1<sup>st</sup>, 2002. The heritage of SCIAMACHY lies in the ground-based measurements using Differential Optical Absorption Spectroscopy (DOAS) [Platt et al., 1979], which is discussed in Chapter 3.

The development of the SCIAMACHY concept was intertwined with that of ENVISAT. Remote sensing of trace gases using satellites was still in its infancy at the time of the call, and there were advantages in having multiple instruments measuring the same air or surface. The bigger risks, lack of orbital flexibility, and

more specialised missions have meant that big multi-instrument missions have fallen out of favour.

The name SCIAMACHY is a Greek expression, literally meaning chasing shadows, equivalent to doing the impossible task. The mission objectives were bold, it has been tasked with improving our knowledge of the global atmospheric composition, and the change in composition in response to natural and anthropogenic activity [Gottwald, 1987]. It was also tasked with investigating the related global issues of importance to the chemistry and physics of our atmosphere such as:

- the impact of anthropogenic activity and natural processes on tropospheric ozone, air quality and global warming,
- exchange processes between the stratosphere and troposphere,
- the interaction of stratospheric chemistry and dynamics,
- natural modulations of atmospheric composition resulting from volcanic eruptions, lightning, solar output variations (e.g. solar cycle), or solar proton events.

The following gases were targeted for measurement:  $O_2$ ,  $O_3$ ,  $O_4$ ,  $NO$ ,  $NO_2$ ,  $NO_3$ ,  $CO$ ,  $CO_2$ ,  $HCHO$ ,  $CH_4$ ,  $H_2O$ ,  $N_2O$ ,  $SO_2$ ,  $BrO$ ,  $OCIO$ . Information on clouds and aerosol can also be deduced from the SCIAMACHY measurements. To determine all the listed species, SCIAMACHY observes the wavelength ranges 214-1773nm, 1934-2044nm and 2259-2386nm (see Figure 2.7) across its eight channels (see Table 2.1). The retrieval of these gases depends on being able to measure their absorptions precisely.

Retrieving total columns concentrations of minor trace gases with an accuracy of 1-5% (or 5-10% for their profiles) requires observing fractional intensity changes of  $10^{-3}$  to  $10^{-4}$  with respect to the optical depth. This can only be achieved with an instrument providing measurements with a high signal-to-noise ratio and a good radiometric calibration [Bovensmann and Burrows, 1998].

## 2.4.2 Orbit and measurement geometry

SCIAMACHY (on board ENVISAT) flies in a polar, sun-synchronous orbit with a daytime descending mode crossing time of 10 am. This is equivalent to fixing the

Table 2.1. SCIAMACHY science channels (1 & 2 = UV, 3 & 4 = VIS, 5 = NIR, 6-8 = SWIR)

Channel	Spectral Range (nm)	Resolution (nm)
1	214 - 334	0.24
2	300 - 412	0.26
3	383 - 628	0.44
4	595 - 812	0.48
5	773 - 1063	0.54
6	971 - 1773	1.48
7	1934 - 2044	0.22
8	2259 - 2386	0.26

position of the sun to the left of the orbital plane of the satellite. It flies at a mean altitude of about 800 km above sea level and takes 100.6 minutes to complete an orbit of the Earth. The satellite follows a 501 orbit cycle, so that the 502<sup>nd</sup> orbit is the same as the 1<sup>st</sup>. It takes 35 days to complete the repeat cycle, so for every day on Earth, SCIAMACHY makes just over 14 overpasses [Gottwald, 1987].

SCIAMACHY operates in several measurement modes, which take place in a set order for an individual orbit. On the descending (day) side, SCIAMACHY alternates between Nadir and Limb operational modes (see Figure 2.8), taking advantage of the strong reflection of sunlight from the Earth. When the geometry allows, SCIAMACHY will perform solar or lunar occultation measurements (the instrument looks directly at the sun or moon), which allows for very high signal to noise measurements that can be used to identify very low abundance trace gases other than those targeted for measurement (see Figure 2.10). The sun and moon are also used for calibration. On the ascending (night) side, the lack of sunlight makes science measurements impossible. This part of the orbit is used for calibration and monitoring (using the internal calibration lamps or dark signal).

All parts of the SCIAMACHY instrument are fixed, except for the two scanner modules, the Elevation Scanner Module (ESM) and the Azimuthal Scanner module (ASM). These two mirrors (and the orbit of the satellite) are used to control the viewing geometries.

#### 2.4.2.1 Observational modes

##### Nadir Mode

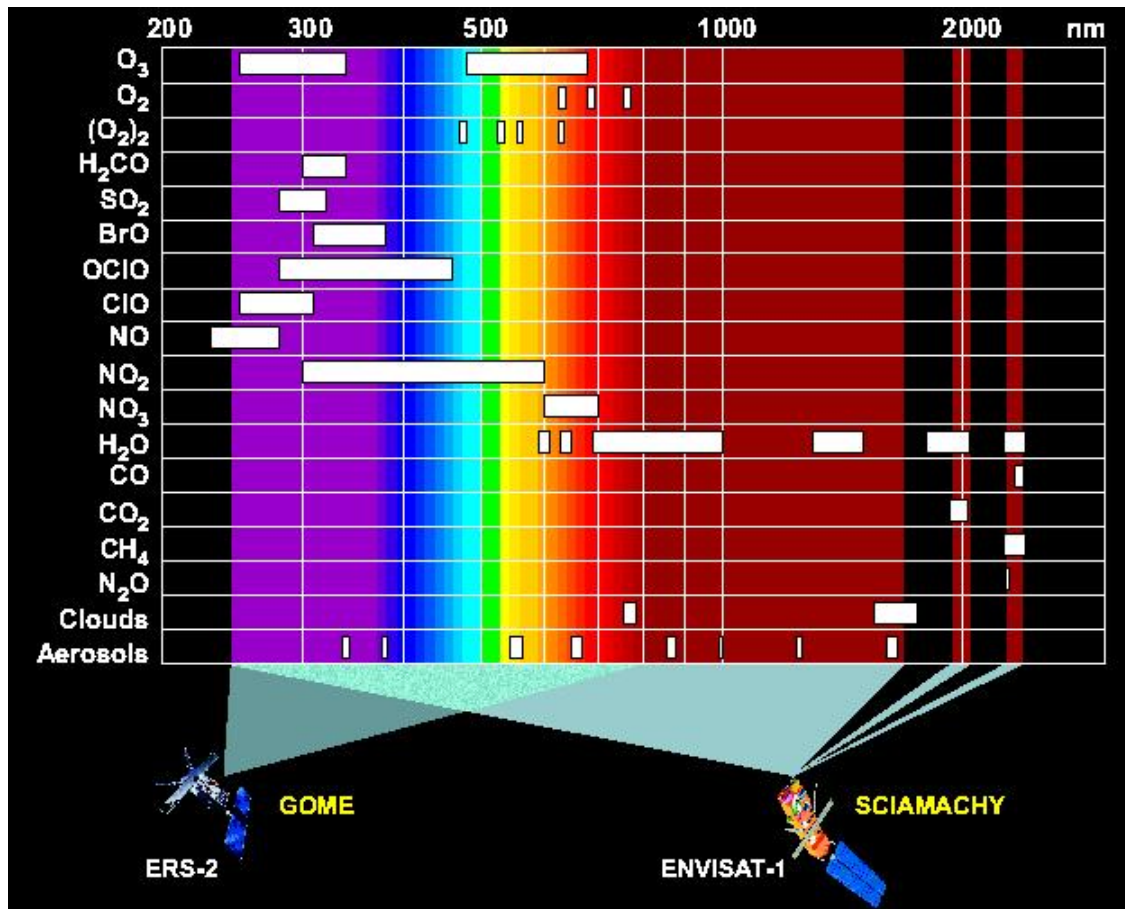


FIGURE 2.7. Wave bands at which molecules measured by SCIAMACHY absorb (image courtesy of DLR).

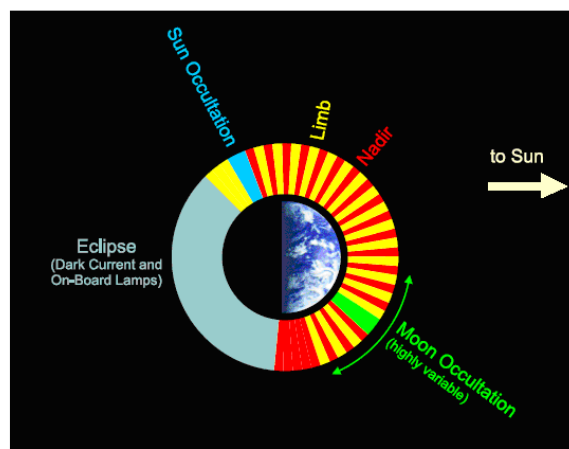


FIGURE 2.8. The orbital sequence of SCIAMACHY measurements (image courtesy of DLR).

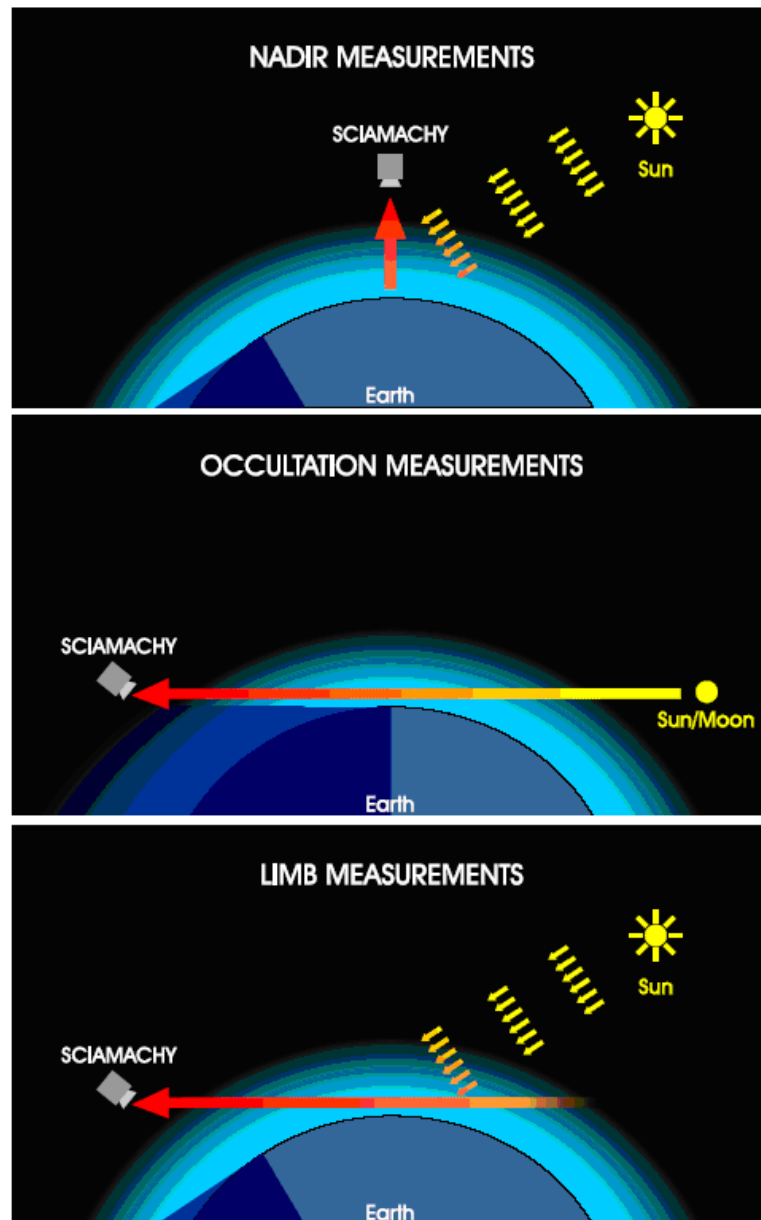


FIGURE 2.9. The viewing geometries of SCIAMACHY (image courtesy of DLR).

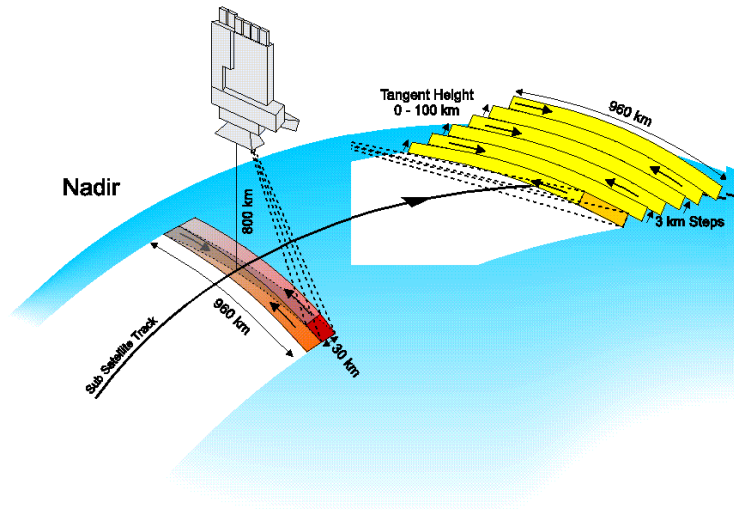


FIGURE 2.10. The swath and footprint dimensions for SCIAMACHY nadir and limb observations (image courtesy of DLR).

In nadir mode, the instrument looks through the atmosphere to the ground below it. Since the instrument measures the reflected sunlight in this geometry, the absorption spectra will be influenced by the atmosphere between the Sun and the Earth, as well as between the Earth and the instrument. The vertical resolution of the nadir mode is poor, since it is difficult to separate out the absorptions at different altitudes in the column. The instrument scans about  $32^\circ$  to the left and right as it moves southwards, giving the nadir mode a swath width of 960 km across track.

A typical scan by the ESM mirror is 4 seconds across track followed by a 1 second backscan. The sequence is repeated for either 60 or 85 seconds depending on the orbital region. For channel six (the channel used to retrieve  $\text{CO}_2$ ) each “footprint” has an integration time of 0.25 seconds. Thus, if shown on the ground these footprints are gridded together 16 across track and either 12 or 17 along track, see (see Figure 2.11). The nominal footprint size is 60 km across track by 30 km along track. The actual exposure time is variable, as a function of the solar zenith angle (SZA), in order to obtain the best signal to noise ratios.

In the alternating mode, complete global coverage is obtained every 6 days (every 3 days if only operating in nadir mode), although coverage is more frequent at higher latitudes as the swaths get closer together. The typical gap between the nadir grids along track (whilst the instrument is in limb mode) depends on the amount of time spent in limb mode and satellite velocity (about  $7 \text{ km s}^{-1}$ ) is a

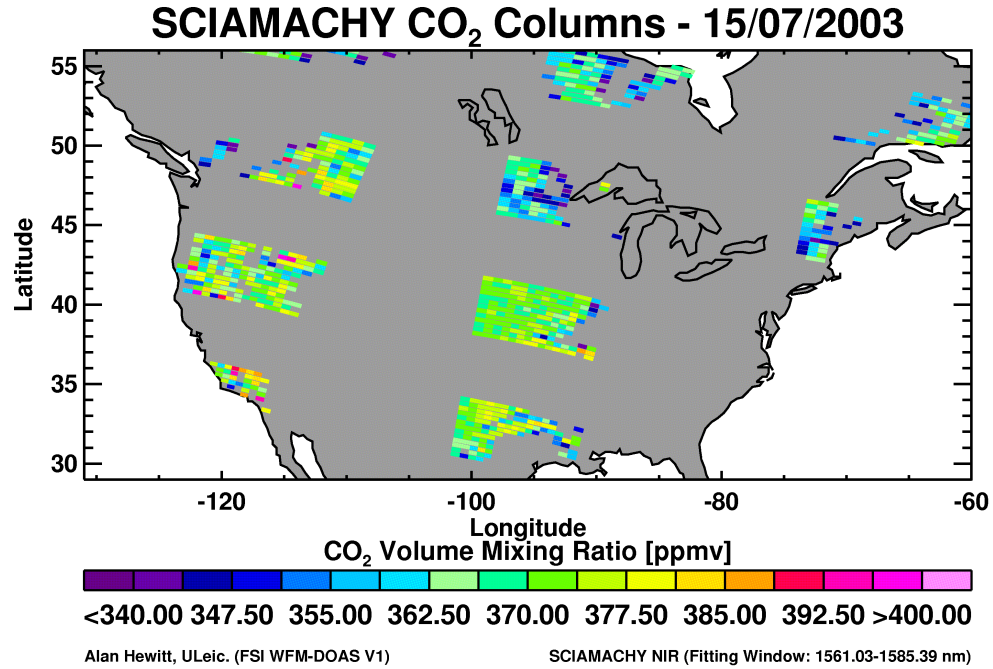


FIGURE 2.11. Individual footprints clustered together in the swath.

little less than the length of the across track grids.

Only nadir mode measurements are used in the CO<sub>2</sub> retrieval algorithm and thus the inversion scheme demonstration. The column CO<sub>2</sub> air above each of the measurement footprints is likely to have travelled a great distance together and encountered very similar fluxes along the way. If a high precision sample of the mixing ratio of CO<sub>2</sub> were desired, it would be possible to improve the precision of the individual retrievals by taking an ensemble mean (the precision improves by a factor of the square root of the number of individual measurements).

The size of the nadir footprints produces a fine scale spatial resolution for most trace gases, and could even be used to study emissions from megacities. However, it is difficult to filter out cirrus clouds from measurements. For a footprint size of 60 km by 30 km, the amount of cloud-free measurements may be about 15% which improves to 36% for 3 km<sup>2</sup> footprints from OCO. Most measurements do not have a very strict cloud filter, so the majority of retrievals will have some cloud contamination [Boesch et al., 2006].

Note that the terminology used in most publications refers to “pixels” instead of “footprints”. Unfortunately this term is commonly used in another separate part



of this research, and the distinction here will save confusion later in this thesis.

### **Limb Mode**

In limb mode, the spectrometer slit is projected parallel to the horizon in-flight direction using the ESM and ASM mirrors. The instrument observes the atmosphere tangentially to the Earth's surface taking into account effects such as the curvature of the Earth [Bovensmann et al., 1999]. A typical limb scan consists of 34 horizontal scans at different tangent heights. The first scan is made 3 km below the horizon (taking 1.5 seconds), after which there is a step of 3 km in elevation and the next scan begins. The satellite looks through a long column of air, and so the horizontal resolution of each measurement is around 400 km, which is poor compared with 3 km vertical resolution. The fast movement of the satellite causes the individual scans to be displaced along the track.

The first scan, which looks at the ground, covers the same footprint as the following nadir scanning section (see Figure 2.10). This rather novel set up, allows stereo viewing of the tropospheric columns, which (with great difficulty) would allow for retrieval of tropospheric concentrations of some trace gases.

### **Occultation Mode**

With each orbit, the instrument will encounter sunrise, which occurs before the instrument passes over the north pole, (see Figure 2.8). In the northern hemisphere summer, this sunrise can appear to happen when the instrument is as low as 27° North on the night side rising to 75° North in the winter. This is due to the tilt of the Earth's axis. For a brief time in this part of the cycle, the two mirrors are used to point at the sun, which is observed at first through the atmosphere and then directly (which is essentially another method of calibration).

The solar occultation offers the best signal-to-noise ratio of any science measurements that SCIAMACHY could make, and can be used to investigate trace gases not listed on the original mission. This allows for investigating the unusual polar chemistry, which leads to ozone destruction, but only over the North Pole. During the full moon, lunar occultation measurements can be made which are the equivalent of the solar occultation, except they are made over the South Pole.

### **Other Measurement Modes**

SCIAMACHY is stored with various sequences which control the motion of its mirrors, allowing the switching between modes. The sequences can be modified, although for any pointing this would require forward planning. It may be beneficial

to have a limb-only or nadir-only geometry, if a specific event occurred (such as a volcanic eruption).

Certain wavelengths of sunlight are absorbed by surface water and ice (in particular the wavelengths that are used to investigate CO<sub>2</sub>) which means that nadir measurements over ocean have very low signal-to-noise ratios in these wavelength regions. However, at a certain geometry, where the angle of incidence and reflection are the same, the surface acts like a mirror and reflects a significant amount of sunlight. This effect, known as sun glint, is occasionally picked up by SCIAMACHY in the equatorial regions (as the angle of incidence here is small, and could match the reflected angle i.e. the nadir looking angle by coincidence). A dedicated CO<sub>2</sub> mission, the Orbiting Carbon Observatory (OCO) which suffered a launch failure in 2009, was meant to take advantage of sun glint to allow measurements over the oceans.

### **2.4.3 Calibration**

#### **2.4.3.1 Icing**

Channel 7 of SCIAMACHY was expected to be used to measure atmospheric CO<sub>2</sub> on account of the large absorption feature at 2.06  $\mu\text{m}$  (see Figure 2.7). However, it became clear during in-flight calibration that the detectors in channels 7 & 8 were contaminated by a build up of ice, which caused a loss of radiative sensitivity of up to 80% [Wuttke et al., 2004]. This problem could not be overcome by heating the instrument, a solution originally foreseen only to decontaminate the radiative coolers.

Despite the throughput correction developed for the channels, it was decided that retrievals at these wavelengths would be abandoned and the weak CO<sub>2</sub> band at 1.61  $\mu\text{m}$  would be used instead.

#### **2.4.3.2 General calibration**

The goal of the calibration is to convert the electronic signal of detectors into physical units, in this case the light intensity at different wavelengths. This is done by running a sequence of individual calibration steps in the right order on the measurement data, see [Lichtenberg et al., 2006].

Lessons learnt from GOME (the Global Ozone Monitoring Experiment), meant that SCIAMACHY had relatively few calibration issues. SCIAMACHY has been fortunate in having a slow rate of optical performance degradation, this has been very useful when building a long-term data set. Apart from the icing problems and stray light problems in channel 7, SCIAMACHY has performed well, up to the standards it was expected to do so before launch.

## 2.5 Summary

Before launch, the SCIAMACHY team carried out sensitivity studies of the precision estimates of SCIAMACHY trace gas measurements [Bovensmann et al., 1999]. The radiation entering SCIAMACHY was simulated using a radiative transfer model coupled with a line-by-line model of the atmospheric absorptions where necessary. The theoretical precision of atmospheric CO<sub>2</sub> concentrations are about 1% for nadir and solar occultation modes and 10% in limb mode. The theoretical precision of atmospheric O<sub>2</sub> concentrations are about 1% for nadir and solar occultation modes and 10% in limb mode. Simultaneous retrieval of CO<sub>2</sub> and O<sub>2</sub> can be used in the retrieval scheme to verify that the whole column is being examined (not just measurement above cloud tops).

These values compare favourably with the requirements for performing a “useful” inversion (i.e. one that could add knowledge to the existing ground network). However, the values quoted by Bovensmann et al. [1999] should really be regarded as theoretical limits to the precision. In the following chapters about the retrieval algorithm and analysis of its results, it is clear that we are getting much more than 15% of measurements pass the cloud control filter [Boesch et al., 2006]. Without a very strict cloud filter, the theoretical precision limits will be not be as good as quoted.

In a clean air situation (no aerosol or clouds), this precision could be improved by taking the mean of a number of individual measurements, although this method cannot improve the accuracy of the measurements.

The separation of the weak CO<sub>2</sub> line at 1.61  $\mu\text{m}$  from the oxygen A-band (O<sub>2</sub>) line at 0.76  $\mu\text{m}$  could lead to some discrepancies, as the refractive and reflective effects of aerosol and clouds (changing the path lengths of light at these frequencies) will differ somewhat. However, this method is still more reliable than comparing with *a priori* surface pressure alone.

The chosen wavelength to retrieve carbon dioxide columns is also in a water vapour absorption band. The H<sub>2</sub>O column must be understood before separation of the two spectral signatures can take place. This could involve direct retrieval and/or *a priori* model information of water vapour.

## Chapter 3

# Retrieval of atmospheric CO<sub>2</sub> from SCIAMACHY NIR measurements

In order to quantify the exchange of carbon between the atmosphere and the surface, an indirect method is required. The principle is simple, make many measurements of the atmospheric CO<sub>2</sub> concentration in time and space, and associate high and low atmospheric CO<sub>2</sub> concentrations with emission and absorption of carbon by the land or ocean surface. This is complicated by the transport of atmospheric gases, which is discussed in Chapter 6. A high background concentration of atmospheric CO<sub>2</sub> combined with small variations owing to air surface exchange means that a very high measurement precision is required. Remote sensing satellite instruments provide one way of obtaining these measurements. This chapter will discuss the Full Spectral Initiation - Weighting Function Modified - Differential Optical Absorption Spectroscopy (FSI-WFM-DOAS) retrieval method [Barkley et al., 2006a], used to demonstrate real SCIAMACHY measured CO<sub>2</sub> concentrations in the inversion scheme (see Chapter 6). The work in this thesis is not dedicated to improving measurement precision, and the author made only modest coding corrections to this retrieval scheme. For a complete description of the retrieval scheme see [Barkley, 2007]. The purpose of this chapter is to demonstrate the basis of the data used and the sources of uncertainties.

## 3.1 Differential Optical Absorption Spectroscopy

### 3.1.1 Introduction

The understanding of the energy levels of atoms and molecules comes from their interaction with electromagnetic radiation. The discrete changes of energy levels in molecules are due to three different excitation schemes: electronic transitions with very large energy transitions (on the order of 1 eV), vibrational transitions (on the order of 0.1 eV) and rotational transitions (on the order of 0.001 eV). A single transition could involve a change in both the vibrational and rotational energy levels. Each molecule has a unique set of discrete (or quantised) energy jumps, thus the signature from these interactions can be used to identify which gases are present in a sample. Unsaturated absorption spectral line depths can be used to determine the density of a particular molecule between the light source and the detector. This calculation is complicated in the real atmosphere with the presence of scattering particles such as aerosols. In order to determine the density of a particular molecule in the real atmosphere, the scattering and absorption features need to be separated. One method that works particularly well for passive satellite measurements is the Differential Optical Absorption Spectroscopy (DOAS) method [Platt et al., 1979], a method which led to the development of the SCIAMACHY instrument. The OCO (launch failure) and GOSAT (launched 2009) missions designed specifically to measure CO<sub>2</sub>, whose future measurements will be combined with the inversion scheme (chapter 5), also use a DOAS scheme.

### 3.1.2 Theory

#### 3.1.2.1 Absorption of Radiation

The radiance  $I(\lambda)$  will be reduced by the amount  $dI_a(\lambda, \mathbf{r})$  after traversing an absorbing layer of (infinitesimal) thickness  $dr$  at position vector  $\mathbf{r}$  is given by:

$$dI_a = -I(\lambda, \mathbf{r}) \cdot \varepsilon_a(\lambda, \mathbf{r}) dr = -I(\lambda, \mathbf{r}) \sum_i \rho_i(\mathbf{r}) \sigma_i(\lambda, \mathbf{r}) dr \quad (3.1)$$

where  $\varepsilon_a(\lambda)$  is the absorption coefficient,  $\sigma_i(\lambda, \mathbf{r})$  is the (wavelength, temperature and pressure-dependent) absorption cross-section of the absorber (molecule)  $i$ , and  $\rho_i(\mathbf{r})$  is the density of each absorbing molecule ( $i$ ) in the volume. Equation 3.1

may be integrated to yield the Beer-Lambert law, with  $I_0(\lambda)$  denoting the initial radiance with optical density  $D(\lambda)$ :

$$I(\lambda) = I_0(\lambda) \exp \left\{ - \int_0^L \sum_i \rho_i(\mathbf{r}) \sigma_i(\lambda, \mathbf{r}) dr \right\} = I_0(\lambda) \exp \{ -D(\lambda) \} \quad (3.2)$$

The slant column density,  $S_i$ , of an absorber is defined as the integrated concentration of trace gas  $i$  along the light path:

$$S_i = \int_0^L \rho_i(\mathbf{r}) dr = \frac{D(\lambda)}{\sigma_i(\lambda)} \quad (3.3)$$

The column density,  $S$ , can be obtained from the ratio of the optical density of absorber  $i$  to its absorption cross section.

### 3.1.2.2 Scattering of Radiation

The radiance will be reduced by the amount  $dI_s(\lambda, \mathbf{r})$  after traversing a scattering layer of (infinitesimal) thickness  $dr$  at position vector  $\mathbf{r}$  (in analogy to absorption) can be given by:

$$dI_s = -I(\lambda, \mathbf{r}) \cdot \varepsilon_s(\lambda, \mathbf{r}) dr = -I(\lambda, \mathbf{r}) \cdot \frac{d\sigma_s(\lambda)}{d\Omega} \cdot N d\Omega dr \quad (3.4)$$

where  $\varepsilon_s(\lambda)$  is the scattering coefficient,  $\sigma_s(\lambda)$  is the scattering cross section of a scattering centre, and  $N$  is the number of scattering centres per unit volume. Thus the total extinction by the radiation after traversing a layer of thickness  $dr$  is given by the sum of absorption and scattering:

$$dI = -dI_a - dI_s = -I(\lambda) \cdot (\varepsilon_a(\lambda) + \varepsilon_s(\lambda)) dr \quad (3.5)$$

### 3.1.2.3 The DOAS principle

In laboratory conditions, it is relatively straightforward to obtain the column density,  $S_i$ , from equation 3.3. In the real atmosphere, this is made complicated by the influence of scattering, Rayleigh extinction,  $\varepsilon_R(\lambda)$ , Mie extinction  $\varepsilon_M(\lambda)$ , and instrumental effects and turbulence,  $A(\lambda)$ . The concept of differential absorption

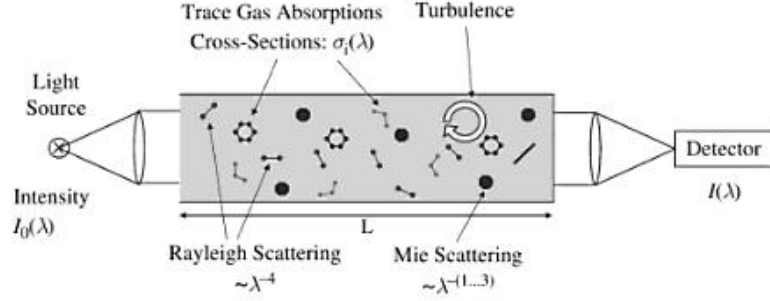


FIGURE 3.1. Measurement of trace gas absorption in the open atmosphere [Platt and Stutz, 2008].

is used to overcome this problem, the absorption cross section is split into high frequency  $\sigma'_j(\lambda)$  and broadband  $\sigma_{j0}(\lambda)$  components:

$$\sigma_j(\lambda) = \sigma'_j(\lambda) + \sigma_{j0}(\lambda) \quad (3.6)$$

The complete description of the Beer-Lambert law for this setup (Figure 3.1) is given by:

$$I(\lambda) = I_0(\lambda) \cdot \exp \left[ -L \cdot \left\{ \sum (\sigma_j(\lambda) \cdot c_j) + \varepsilon_R(\lambda) + \varepsilon_M(\lambda) \right\} \right] \cdot A(\lambda) \quad (3.7)$$

where  $c$  is the average trace gas concentration given by:

$$c = \frac{\ln \frac{I_0(\lambda)}{I(\lambda)}}{\sigma(\lambda) \cdot L} = \frac{D}{\sigma(\lambda) \cdot L} \quad (3.8)$$

Inserting equation 3.6 into equation 3.9 as described in the principles of DOAS [Platt and Stutz, 2008], gives the following:

$$I(\lambda) = I_0(\lambda) \cdot \exp \left[ -L \cdot \left\{ \sum_j (\sigma'_j(\lambda) \cdot c_j) \right\} \right] \cdot \exp \left[ -L \cdot \left\{ \sum_j (\sigma_{j0}(\lambda) \cdot c_j) \varepsilon_R(\lambda) + \varepsilon_M(\lambda) \right\} \right] \cdot A(\lambda) \quad (3.9)$$

The first exponential function in equation 3.9 describes the effect of the structured differential absorption of a particular trace species and the second exponential term describes the slow varying absorptions and the influence of scattering terms. In many atmospheric DOAS setups the intensity  $I_0(\lambda)$  at the light source may not



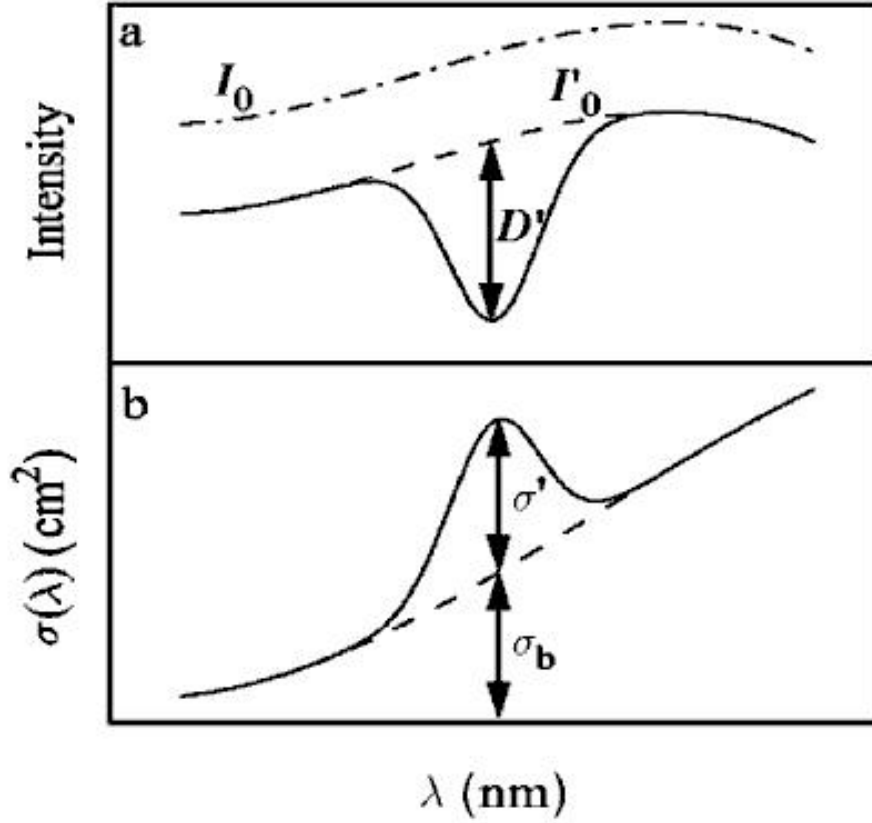


FIGURE 3.2. The DOAS principle:  $I_0(\lambda)$  and  $\sigma$  are separated into narrow ( $D'$  and  $\sigma'$ ) and broad band ( $I'_0$  and  $\sigma_b$ ) [Platt and Stutz, 2008].

be known, to get around this problem an intensity  $I'_0(\lambda)$  is defined as the intensity in the absence of differential absorption (see [Platt and Stutz, 2008]) as shown:

$$I'_0(\lambda) = I_0(\lambda) \cdot \exp \left[ -L \cdot \left\{ \sum_j (\sigma_{j0}(\lambda) \cdot c_j) \varepsilon_R(\lambda) + \varepsilon_M(\lambda) \right\} \right] \cdot A(\lambda) \quad (3.10)$$

The differential optical density,  $D'$ , is thus defined:

$$D' = \ln \frac{I'_0(\lambda)}{I(\lambda)} = L \cdot \sum_j \sigma'_j(\lambda) \cdot c_j \quad (3.11)$$

Figure 3.2 demonstrates how the narrow absorption features can be separated from the gradually changing broadband scattering features. In the case of SCIAMACHY passive DOAS measurements,  $I_0$  gets measured during the solar calibration. By scaling the Earth reflected sunlight to the direct solar measurement, the solar absorption lines can be identified and removed before the DOAS process.

### 3.1.2.4 Conversion to vertical column density

The setup of passive DOAS in nadir geometry of solar radiation reflected back from the Earth to the instrument produces a complicated light path. The strength of the absorption features are used to obtain the density of the absorbing gases. If the incoming solar radiation is at an angle  $\theta$  relative to the nadir, then the total light path can be given as:

$$LightPathLength = NadirLightPath + [NadirLightPath \times (\cos\theta)^{-1}] \quad (3.12)$$

A more useful result is the Vertical Column Density (VCD),  $V$  (in units of molecules  $\text{cm}^{-2}$ ), defined as the integrated concentration of the trace gas as a function of altitude  $z$ :

$$V = \int_0^z \rho(z) dz \quad (3.13)$$

The VCD can be obtained by dividing the slant column density,  $S$  (see Equation 3.3) by the air mass factor (AMF) which is calculated by a radiative transfer model, in this case SCIATRAN [Rozanov et al., 2002] and is a function of wavelength,  $\lambda$ , and the observation geometry,  $\Phi$ . A more detailed description of air mass factors can be found in the DOAS textbook [Platt and Stutz, 2008].

$$V = \frac{S(\Phi, \lambda)}{AMF(\Phi, \lambda)} \quad (3.14)$$

The vertical column density is later converted to the volume mixing ratio (VMR) which is referred to  $\text{XCO}_2$  (WFM-DOAS does this using the co-retrieved  $\text{O}_2$  column, whereas FSI uses *a priori* pressure information). This is the average VMR in the column. The nadir measurement geometry makes it difficult to extract any vertical distribution information which would be useful in the investigation of surface fluxes.

### 3.1.2.5 The DOAS spectral fitting procedure

Both the WFM (Section 3.2) and the FSI (Section 3.3) retrieval spectral fitting procedures are based on a least squares approach to minimise the difference between the observed optical density,  $D(\lambda_k)$ , and the absorption cross sections,  $\sigma(\lambda_k)$ , multiplied by their respective slant column densities,  $S_i$ , after the subtraction of a low

order polynomial,  $c_m \lambda_k^m$ , to compensate for the broadband continuum [Barkley, 2007]:

$$\chi^2 = \sum_k \left[ D(\lambda_k) - \sum_i \sigma_i(\lambda_k) S_i - \sum_m c_m \lambda_k^m \right]^2 \rightarrow \min \quad (3.15)$$

Equation 3.15 is solved from the partial derivatives of  $\chi^2$  with respect to the fit parameters; the slant column densities,  $S - I$ , and the polynomial coefficients,  $c_m$ :

$$\frac{\partial \chi^2}{\partial S_i} \rightarrow 0 \quad \text{and} \quad \frac{\partial \chi^2}{\partial c_m} \rightarrow 0 \quad (3.16)$$

The standard DOAS retrieval technique described here is suitable for near UV and visible spectra, and has been demonstrated for the GOME instrument [Eisinger and Burrows, 1998]. It makes several assumptions relating to the pressure, temperature and wavelength dependence of trace gas absorption cross sections. Another important assumption is that the trace gas absorption is weak, since the logarithm of the radiance depends linearly on the trace gas vertical column amount (from the Beer-Lambert law). The trace gas absorption cross section in the near infrared has a much stronger dependence on temperature, wavelength and pressure, and many of these absorptions are strong. Selecting an absorption cross section for a specific altitude introduces large total column errors. The development of Weighting Function Modified (WFM) DOAS [Buchwitz et al., 2000a] attempted to compensate for these problems using a look up table approach. These ideas were further developed as Full Spectral Initiation (FSI) WFM-DOAS [Barkley et al., 2006a], which was developed for the purpose of investigating the exchange of carbon between the terrestrial biosphere and the atmosphere (see Chapter 6).

## 3.2 Weighting Function Modified DOAS

The WFM-DOAS algorithm is based on approximating the logarithm of the measured sun normalised radiance ( $I_i^{meas}$ ) by a corresponding linearised model quantity ( $I_i^{ref}$ ) (i.e. using the mean radiance and its derivatives calculated from a model atmosphere), plus a quadratic polynomial. The standard DOAS trace gas absorption cross section reference spectra are replaced by trace gas total column weighting functions (applying an altitude independent scaling factor to the entire profile).

Buchwitz et al. [2000a] modified the spectral fitting procedure (equation 3.15) to take the form:

$$\left\| \ln I_i^{meas}(\mathbf{V}^t) - \left[ \ln I_i^{ref}(\bar{\mathbf{V}}) + \frac{\partial \ln I_i^{ref}}{\partial \bar{V}_{CO_2}} \cdot (\hat{V}_{CO_2} - \bar{V}_{CO_2}) + \frac{\partial \ln I_i^{ref}}{\partial \bar{V}_{H_2O}} \cdot (\hat{V}_{H_2O} - \bar{V}_{H_2O}) + \frac{\partial \ln I_i^{ref}}{\partial \bar{V}_{Temp}} \cdot (\hat{V}_{Temp} - \bar{V}_{Temp}) + P_i(a_m) \right] \right\|^2 \equiv \|RES_i\|^2 \rightarrow \min \text{ w.r.t } \hat{\mathbf{V}} \text{ \& } a_m \quad (3.17)$$

where the subscript  $i$  refers to each detector pixel of wavelength  $\lambda_i$ . The polynomial  $P_i(a_m)$  (which has coefficients  $a_0$ ,  $a_1$  and  $a_2$ ) is included to account for the spectral continuum and broadband scattering. The true, modelled and retrieved vertical columns are represented respectively:

- $\mathbf{V}^t = (V_{CO_2}^t, V_{H_2O}^t, V_{Temp}^t)$
- $\bar{\mathbf{V}} = (\bar{V}_{CO_2}, \bar{V}_{H_2O}, \bar{V}_{Temp})$
- $\hat{\mathbf{V}} = (\hat{V}_{CO_2}, \hat{V}_{H_2O}, \hat{V}_{Temp})$

The column weighting functions in Equation 3.17 replace the absorption cross sections found in the standard DOAS methods. Water vapour and temperature column weighting functions are also required for  $CO_2$  retrievals, as water vapour features and temperature effects overlap the  $CO_2$  features. The fit parameters are trace gas columns  $\hat{V}_{CO_2}$ ,  $\hat{V}_{H_2O}$ , the temperature scaling factor  $\hat{V}_{TEMP}$  and the polynomial  $a_m$ . The retrieved column is obtained by adjusting the model intensity to the measured intensity at several spectroscopic points  $\lambda_i$  simultaneously.

For both WFM and FSI, the reference spectra and weighting functions were computed using SCIATRAN [Rozanov et al., 2002], a forward model covering the spectral range of the SCIAMACHY instrument. The WFM-DOAS algorithm (equation 3.17) is quickly solved using a least squares method:

$$\|\mathbf{y} - \mathbf{Ax}\|^2 \equiv \|RES_i\|^2 \rightarrow \text{minimise with respect to } \mathbf{x} \quad (3.18)$$

equation 3.18 is solved using by matrix multiplication described by Press [1992]:

$$\hat{\mathbf{x}} = \mathbf{C_x} \mathbf{A}^T \mathbf{y} \quad (3.19)$$

where  $\mathbf{C}_x = (\mathbf{A}^T \mathbf{A})^{-1}$  is the solution covariance matrix. The components of vector  $\mathbf{y}$  are  $y_i = (\ln I_i^{meas} - \ln I_i^{mod})/\sigma_i$ , where  $\sigma_i$  denotes the standard deviation of the intensity divided by the mean intensity. Vector  $\mathbf{x}$  contains the fit parameters (the dimensionless total column fit parameter  $(\hat{V} - \bar{V})/\bar{V}$  and the polynomial coefficients). Each column of matrix  $\mathbf{A}$  refers to the fit parameters and each row to one of the spectral fitting points [Buchwitz et al., 2000a].

### 3.2.1 WFM-DOAS Look up table

Processing vast quantities of satellite data is a very computationally expensive process. The measurement of atmospheric  $\text{CO}_2$  is meaningful only in the context of investigating the surface exchange of carbon with the atmosphere. With this target in mind Buchwitz et al. [2000a] considered that maximising the number of available measurements (i.e. global coverage) would provide the best constraint to the magnitude and error on the surface fluxes.

Rather than run SCIATRAN for every SCIAMACHY sounding to provide the model spectrum, the WFM-DOAS algorithm has a list of about 500 possible scenarios that combine combinations of solar zenith angle, viewing angle and surface pressures, with the closest match being compared to the measured line spectra. An iterative approach can be introduced if the solution to Equation 3.17 is not sufficiently linear, in which case the results of the initial linear regression are used as the first guess model parameters. The look up approach takes seconds to process each retrieval, enabling global processing (and reprocessing when improvements are made to the algorithm). Various factors affect the precision of the WFM-DOAS algorithm, and are described in the next section.

## 3.3 Full Spectral Initiation WFM-DOAS

The desire to reduce as much as possible the errors introduced to the total column volume mixing ratio (described in Section 3.3.5) was the driving force behind the Full Spectral Initiation (FSI) WFM-DOAS technique [Barkley et al., 2006a]. By including as much *a priori* information that is available into the retrieval scheme, a more realistic reference spectra can be used to compare with the measured spectra. Rather than using a look up table approach, this scheme produces a new reference spectra for every measured spectra. This technique should produce more accurate

reference spectra, and thus reduce the retrieval error of the CO<sub>2</sub> columns. The main drawback of this method is that it is computationally expensive. Some methods used by WFM-DOAS were not used in the FSI scheme, such as not performing iterative steps on the retrieval and not co-retrieving O<sub>2</sub> columns, in order to cut down retrieval time.

### 3.3.1 Overview of the FSI algorithm

The FSI algorithm is applied to calibrated radiances in the fitting window 1561.03 to 1585.93 nm. The window was chosen to minimise interference from water vapour and not to extend into channel 6+, where the InGaAs detectors were doped with higher amounts of Indium leading to different behavioural characteristics [Lichtenberg et al., 2005]. The number of fitting points within this micro-window is usually thirty-two, with each detector pixel spanning a wavelength interval of 0.7 nm. Ideally, CO<sub>2</sub> lines from channel 7 of SCIAMACHY would have been used but problems with ice build up on the detectors has prevented this. The channel 6 detector has been stable during the period of interest (2003 to 2005) with very few problems owing to thermal and radiation degradation [Kleipool, 2004a]. Dead or bad pixels are checked for each orbit, and this information is used to update the pixel mask [Frankenberg et al., 2005]. Detector pixels are also discarded if erroneous spikes occur in the measured radiance. All measurements have been corrected for non-linear effects [Kleipool, 2003b] and the dark current [Kleipool, 2003a]. The FSI algorithm also uses a solar reference spectrum with improved calibration, provided to the SCIAMACHY community by ESA, courtesy of Johannes Frerick (ESA, ESTEC), in preference to that in the official L1C product. An optimised systematic shift, based on the inspection of the fit residuals, of 0.15 nm is applied to the observed spectra to align it to the synthetic radiances calculated by SCIATRAN. To improve the quality of the FSI spectral fits, the latest version of the HITRAN molecular spectroscopic database has been implemented in the radiative transfer model [Rothman et al., 2005].

A reference spectrum is created for each SCIAMACHY observation. The CO<sub>2</sub> profile is selected from the climatology devised by Remedios et al. [2006] according to the time of observation and the latitude band in which the ground pixel falls. To obtain the closest atmospheric state to the observation temperature, pressure and water vapour profiles, derived from operational 6-hourly ECMWF data (1.125° × 1.125° grid), are interpolated onto the local overpass time and centre of the of the

SCIAMACHY pixel.

A look-up table of mean radiances was generated as a function of both the surface reflectance and solar zenith angle, in order to determine the linearisation point of the surface albedo. Three aerosol scenarios were included in the FSI retrieval scheme, maritime was used over ocean, and rural over land. In both these cases boundary layer visibility is 23 km. Where a major city (population greater than a quarter of a million) falls within the SCIAMACHY footprint, the third aerosol scenario, urban, is selected with a reduced visibility of 5 km. These scenarios are very simplistic, and have not been adopted by the updated WFM algorithm. For all cases the LOWTRAN aerosol model is employed with relative humidity set to 80% in summer and 70% in winter.

Running SCIATRAN with an off-nadir line of sight introduces many more fourier terms and dramatically increases the running time. Because there is very little multiple scattering in the near infrared, it is possible to instead modify the real spectra by adjusting it to a nadir path length. The correction method used in FSI was developed by Buchwitz et al. [2000a].

Prior to retrieval all SCIAMACHY observations are filtered for quality. Cloud contaminated pixels are checked with the SPICI cloud detection method devised by Krijger et al. [2005]. A second levels of filtering was included in the FSI algorithm that measures polarisation and ignores pixels that exceed a certain threshold. A certain degree of cloud contamination will in some circumstances pass the filters, which is problematic for an algorithm such as FSI that does not co-retrieve  $O_2$  columns, and thus has reduced path lengths. Retrievals with large solar zenith angles (greater than  $75^\circ$ ) and SCIAMACHY back scans are ignored. Columns are normalised with ECMWF surface pressures to produce column volume mixing ratios (VMR).

Post processing filtering takes place to ensure that only the highest quality retrievals are used. Some measurements will be affected by undetected cloud or aerosol scattering, and these columns can have a VMR significantly outside of that expected for the column. Roughly one in twenty of the retrievals fall outside of the range 340 - 400 ppmv, often quite significantly, which suggests that FSI would benefit from co-retrieval of  $O_2$  being used alongside surface pressure. The retrieval fitting error can also be used to filter out bad pixels, some analysis of the output data compares the same measurements but with different minimum filtering of this value. Figure 3.3 shows a schematic diagram of the FSI-WFM-DOAS algorithm.

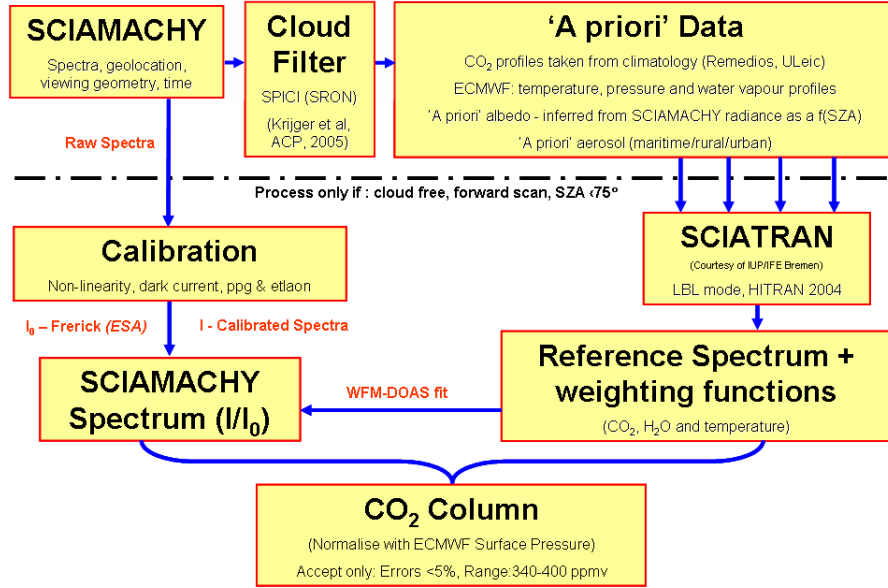


FIGURE 3.3. Schematic flow diagram of the FSI algorithm [Barkley, 2007].

### 3.3.2 Preprocessing of SCIAMACHY data

The SCIAMACHY level 1b raw spectra needs to be calibrated before the useful measurement data sets can be extracted. After omitting level 1b files smaller than 200 megabytes (orbit files smaller than this tend to be corrupted), the remaining orbit files are calibrated using the SciaL1C tool [Frerick, 2006] following the two routes shown in Figure 3.4. One route has all the calibration options switched on to produce a complete 1C file (denoted  $L1C_{All}$ ) and the second makes use of just the wavelength calibration (denoted  $L1C_{\lambda}$ ). The surface albedo is taken from the fully calibrated spectrum, whilst FSI fits performed only with the partially calibrated spectra. This is implemented because the data and calibration corrections for stray-light, polarisation and radiance sensitivity are not yet optimal in the SciaL1C tool. Each of the stages of Figure 3.4 are described briefly here and in more detail in Barkley [2007].

The non-linearity correction is applied because SCIAMACHY's NIR detectors exhibit nonlinear behaviour (described in [Kleipool, 2003b]) in the measured signal rather than being proportional to the observed intensity and to the integration time Friess [2001]. A correction scheme developed by SRON is used to compensate for this. Non-linearity of the detectors response to observed intensity can cause uncertainties of up to 200 BU (Binary Units).

The dark signal is made up of two components, an exposure time-independent



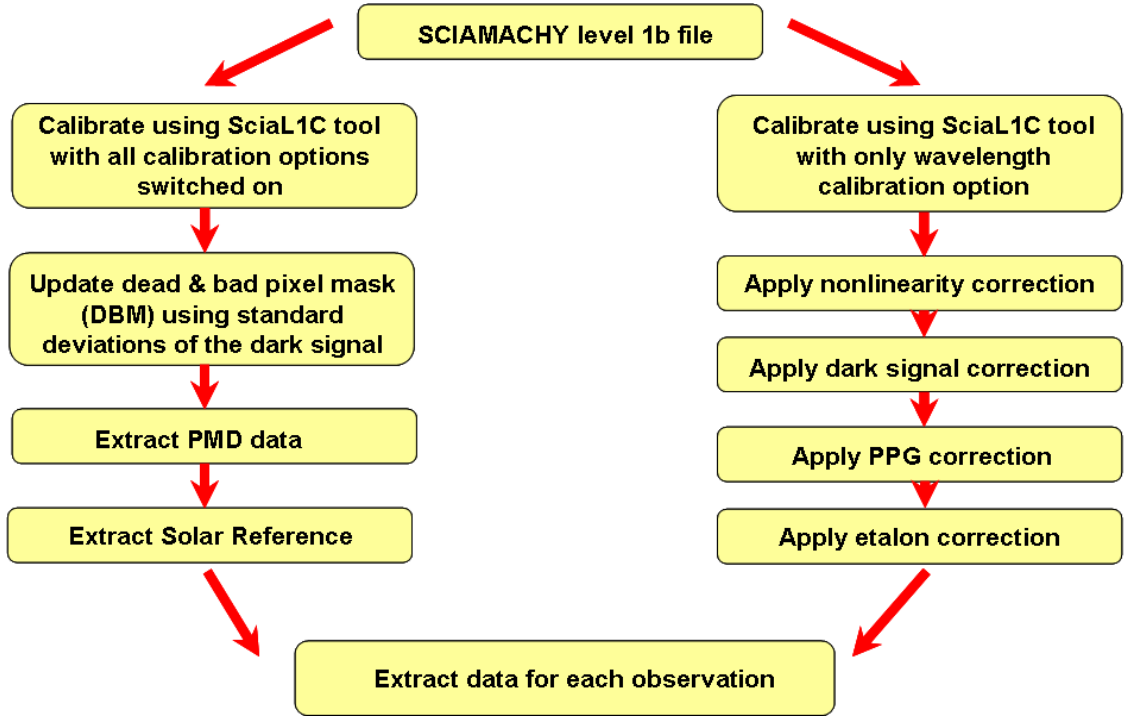


FIGURE 3.4. Flow chart of the calibration procedure [Barkley, 2007].

contribution called the analogue offset (AO) and an exposure time-dependent contribution called the dark current (DC) which scales linearly with time [Kleipool, 2004b]. SCIAMACHY has five dedicated dark current orbital states which are monitored on the eclipse side of the orbit. These measurements combined with modelled dark current behaviour are used to correct for the dark signal. The dark signal can provide uncertainties of up to 20 BU.

The effects of pixel-to-pixel gain and the etalon effect can both be corrected by dividing the signal of each pixel element (after nonlinear and dark signal corrections have been applied) by the respective PPG and etalon calibration spectra provided in the level 1b product.

### 3.3.2.1 Extraction of SCIAMACHY data

The next step is to cloud screen the SCIAMACHY data using the SCIAMACHY PMD Identification of Clouds and Ice/snow (SPICI) cloud detection method devised by Krijger et al. [2005]. This method is fast and can distinguish between cloud and ice/snow. A saturation threshold is used to define the “whiteness” of a scene by calculating the maximum and minimum of the simultaneous measurement of the weighting of PMDs 2, 3 and 4.

The signals ( $S_{PMD_n}$ ) from the three PMD channels are first individually weighted:

$$\begin{aligned} W_4 &= S_{PMD_4}/A_r \\ W_3 &= S_{PMD_3}/A_g \\ W_2 &= S_{PMD_2}/A_b \end{aligned} \tag{3.20}$$

where the weights  $A_r = 0.795$ ,  $A_g = 1.000$  and  $A_b = 0.750$  have been derived by normalising the readouts of PMDs 2 and 4 to PMD 3 over cloudy scenes. For each scene the saturation (or “whiteness”) can be calculated using:

$$\text{Saturation} = \frac{\max(W_4, W_3, W_2) - \min(W_4, W_3, W_2)}{\max(W_4, W_3, W_2)} \tag{3.21}$$

The saturation is low when the three PMDs are equally bright and high when they differ. A PMD ground pixel is cloud free if the saturation exceeds 0.35 [Krijger et al., 2005]. This strict threshold limit was chosen to allow high precision CO<sub>2</sub> measurements from FSI. The main drawback of only using a saturation threshold to distinguish cloud pixels is that it cannot discriminate snow covered pixels in the visible wavelengths. In the NIR, snow covered pixels appear much darker, and so do not provide sufficient quality spectra to measure CO<sub>2</sub> anyway.

### 3.3.3 A priori data

Time independent biases are introduced into the retrieval scheme as discussed in section 3.3.5.7. Using *a priori* information about the CO<sub>2</sub> vertical profile will reduce these biases. FSI-WFM-DOAS [Barkley et al., 2006a] introduced CO<sub>2</sub> profiles from a climatology [Remedios et al., 2006] developed at the University of Leicester, comprising of a profile for each of the 30° latitude bands, for each month processed (from 2003-2005). The profiles for 2003 are shown in Figure 3.5.

The temperature and pressure profiles are obtained directly and the water vapour indirectly from the specific humidity using the ECMWF operational data set courtesy of the British Atmospheric Data Centre (BADC). The trace gas profiles are included in the SCIATRAN radiative transfer model.

The surface albedo is estimated using the mean radiance of the observation within the FSI spectral fitting window. During retrieval processing the mean sun-normalised radiance of each SCIAMACHY measurement, calibrated using SciaL1C

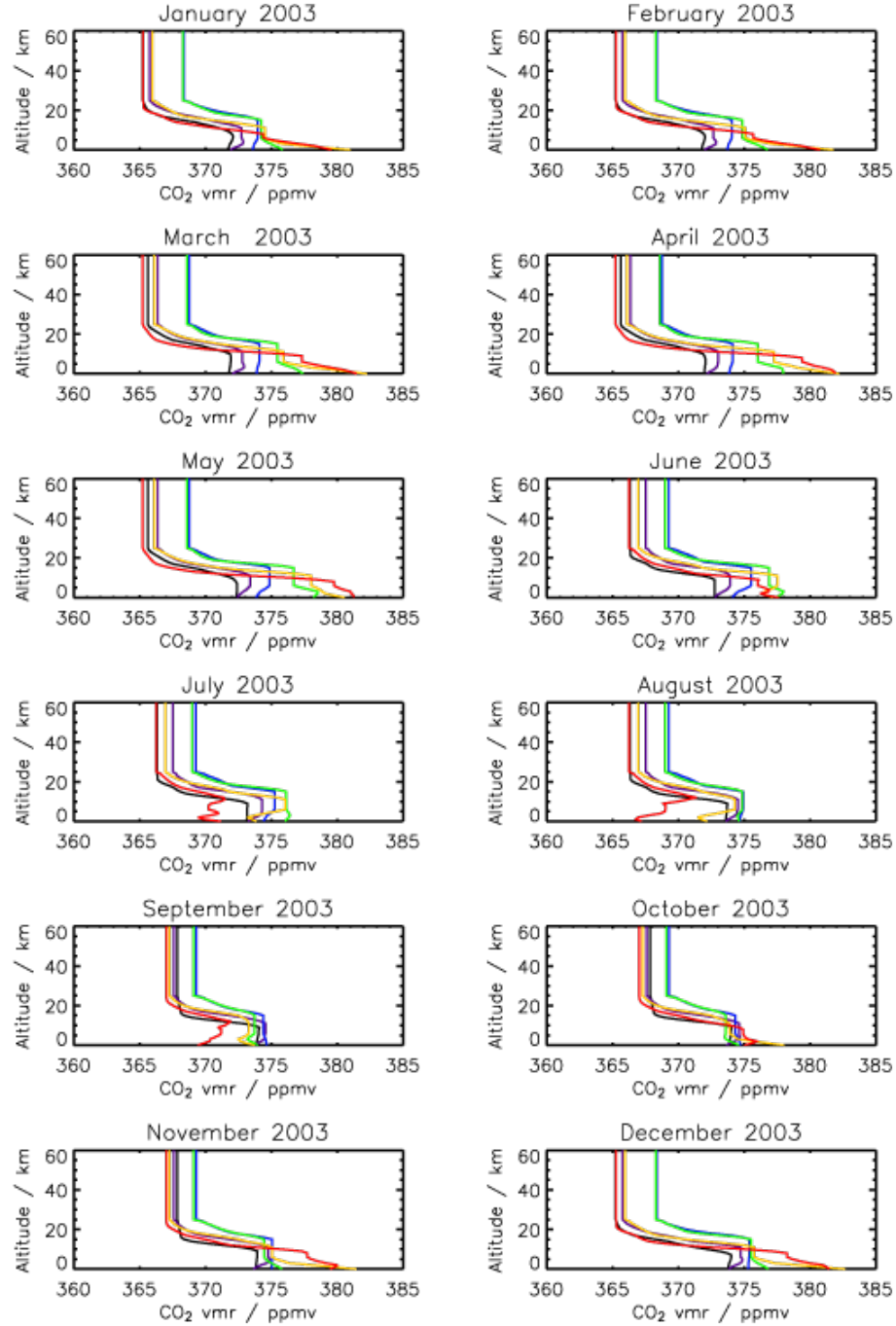


FIGURE 3.5. The vertical  $\text{CO}_2$  profiles, for the year 2003, used in the FSI algorithm [Barkley, 2007]. This data set, for 2003, contains a total of 72  $\text{CO}_2$  profiles, constructed from the flask measurements made by the GLOBALVIEW network. Each month has six profiles centered on the latitudes  $75^\circ\text{N}$  (red),  $45^\circ\text{N}$  (yellow),  $10^\circ\text{N}$  (green),  $10^\circ\text{S}$  (blue),  $45^\circ\text{S}$  (purple) and  $75^\circ\text{S}$  (black), each representative of a  $30^\circ$  latitude band. Changes in the  $\text{CO}_2$  VMR above 20 km have little effect on the retrieval as the instrument is not sensitive there. Since high altitudes are also not affected by surface fluxes the model  $\text{CO}_2$  VMR is nominally constant above this height.

(from the file  $L1C_{All}$ ) and excluding spikes, is determined and an estimate for the albedo obtained from the set of pre-calculated radiances, using the solar zenith angle of the observation as a secondary constraint. This method, although only an approximation, tries to minimise the retrieval errors owing to imperfect knowledge of the surface albedo.

Three aerosol scenarios are used to represent urban, rural and ocean aerosol scenarios. This simplistic labelling of the aerosol scenario is not used by the Bremen group. These scenarios are implemented in the LOWTRAN model [Kneizys et al., 1996]. A city database was constructed identifying urban areas with a population in 2003 of greater than 250,000 people using information from the world gazetteer [Gazetteer, 2003].

Running SCIATRAN off nadir greatly increases the run time for each measurement. A correction is applied to each reference spectra for the solar zenith angle and the scan angle [Barkley, 2007]. For angles less than  $45^\circ$ , the error on this correction can be kept below 1%.

### 3.3.4 Normalising the $\text{CO}_2$ vertical column density

To obtain the vertical mixing ratio (VMR) from the retrieved vertical column density (VCD) the column must be normalised. WFM-DOAS used  $\text{O}_2$  co-retrieved columns as a proxy for the total air column since oxygen is well mixed in the atmosphere. FSI uses surface pressure as a proxy for the total air column, since it was desirable to reduce extra processing time. The best proxy would be one that is well mixed and ideally have spectral features close to the features of the molecule of interest. The best case would be to use both surface pressure and a proxy molecule, where any discrepancy is likely to be owing to partial cloud cover below the threshold used to disqualify a pixel. Using surface pressure as a proxy brings unusual results over “hilly” areas, perhaps because the centre pixel altitude (used to produce surface pressure) gives a different total column of air to that actually present. An alternative suggestion is that aerosol particles are trapped as they attempt to climb over mountainous regions leading to an enhancement of their effects.

### **3.3.5 Factors affecting individual FSI and WFM-DOAS CO<sub>2</sub> retrievals**

The following factors can affect the precision of individual satellite CO<sub>2</sub> measurements. Some of these errors will be introduced by the local or regional geography, the knock-on effect of this is regional scale biases which could appear in the inversion of CO<sub>2</sub> satellite data. The consequences of these regional scale biases are discussed in Sections 4.1 and 3.5.

#### **3.3.5.1 Cloud Coverage**

The path length of radiation from the sun, then through the atmosphere, finally arriving at the SCIAMACHY instrument can be altered by the presence of cloud in the instrument field of view. One possibility is single scattering with reflection off the cloud top, which can happen in overcast conditions for example. In this scenario, the light path through the atmosphere is greatly reduced as it does not sample the air close to the surface. It is also possible in some situations for multiple scattering to occur, which may increase the light path. The main effect of altering the light path of the spectra is for the light to experience fewer (or greater) numbers of absorbers, if not considered this could lead to the incorrect conclusion of a lower (or higher) volume mixing ratio for the trace gas of interest. Both the FSI and WFM-DOAS algorithms use a cloud filter, which ignores measurements where partial or total cloud cover is detected (the threshold for detecting clouds is low, it is much better to ignore good measurements than to process bad ones).

#### **3.3.5.2 Cirrus Clouds**

These high altitude clouds are much harder to detect than other tropospheric clouds, and occur most frequently in the tropics. Made up of ice crystals, their scattering and absorption properties are also difficult to model owing to their variable sizes and shapes. In their study, [Buchwitz and Burrows, 2004] showed that the error created through cirrus contamination is likely to be of the order 1%. Recent upgrades to the WFM-DOAS algorithm were specifically designed to account for cirrus cloud contamination [Reuter et al., 2010], these have yet to be adapted by the FSI algorithm.

### 3.3.5.3 Molecular line parameters

The synthetic spectra are generated using the SCIATRAN radiative transfer model. Parameters such as line position, absorption strength and broadening coefficients need to be quantified accurately. SCIATRAN implements the latest version of the HITRAN spectroscopic database [Rothman et al., 2005] to ensure that these parameters are as close as possible to the measured spectra.

### 3.3.5.4 Aerosols

These are micron or sub-micron sized particles suspended in the atmosphere, affecting the optical path length by scattering and absorbing incoming solar radiation. Aerosols are generally localised to where there are produced, occurring naturally from desert dust, volcanoes, sea spray and biogenic emissions as well as from anthropogenic sources such as biomass burning and pollution. However, aerosol can be transported over great distances and can affect large areas with the right weather conditions such as Saharan dust storms. The original WFM algorithm could not reliably detect over the Sahara, but recent improvements [Schneising et al., 2008] include a dust storm filter (not available for the FSI retrievals). Aerosols are present to a greater or lesser extent in every measurement, so cannot be filtered out completely. The effects of aerosol, like cirrus clouds are very difficult to model.

Without using the TOMS data aerosol filter [Herman et al., 1997], the unfiltered Bremen retrieval code showed variations in the XCO<sub>2</sub> columns around the Sahara region of 10% (column mean ratios of 370 - 415 ppmv) compared with variations at the ground station at Assekrem of just 1% [Houweling et al., 2005]. Even ignoring regions of high aerosol loading, when trying to compare with SCIAMACHY measurements the fluxes from the surface monitoring network [Gurney et al., 2002] exceeds the uncertainty of the inverse flux estimates. The global average effect of aerosol contamination is to add an uncertainty of 3% to the retrieved XCO<sub>2</sub> columns [Houweling et al., 2005].

### 3.3.5.5 Surface reflectance

A high signal-to-noise ratio in the spectral measurements will result in a better spectral fit; the signal strength being related to the albedo. Water is particularly bad at reflecting in the near infrared, except in the case of sun glint. The operation

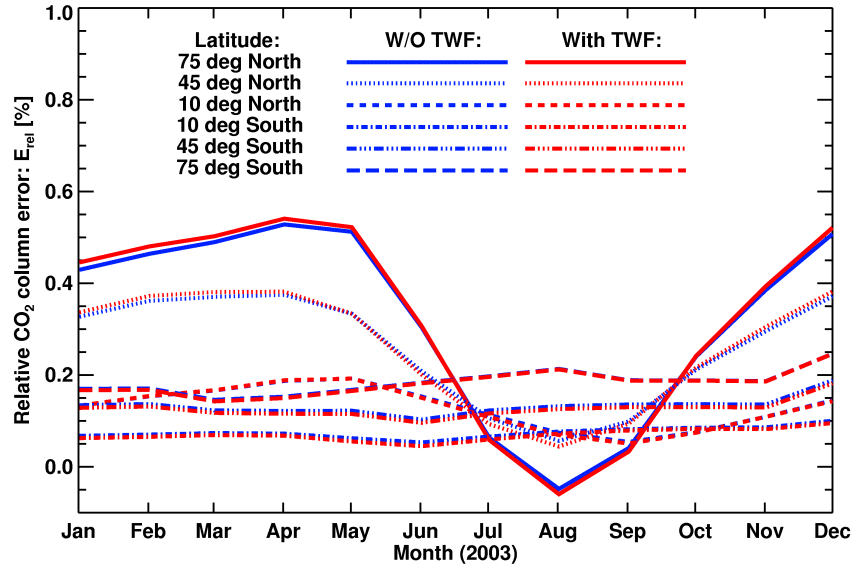


FIGURE 3.6. The relative errors in the retrieved  $\text{CO}_2$  column for  $75^\circ\text{N}$  (solid),  $45^\circ\text{N}$  (dotted),  $10^\circ\text{N}$  (dashed),  $10^\circ\text{S}$  (dash dot),  $45^\circ\text{S}$  (dash dot dot) and  $75^\circ\text{S}$  (long dashes) produced by using  $\text{CO}_2$  profiles other than the US Standard atmosphere (associated with the reference spectrum), with the simulated retrievals performed both with (red) and without (blue) the temperature weighting function [Barkley et al., 2006a].

of SCIAMACHY in nadir mode results in very few sun glint measurements being made over the oceans, leading to very poor signal to noise ratios. The albedo of the sampled surface within any 60 km by 30 km pixel is very likely to be inhomogeneous and it is not unreasonable for two neighbouring pixels to have quite different spectral fitting errors and possibly as a result quite different measured  $\text{CO}_2$  columns not being the result of an actual difference in the concentration. In contrast to its high albedo in the visible, snow cover typically reduces the albedo over most vegetation types in the NIR.

Measurements over water and snow (regions of weak  $\text{CO}_2$  surface exchange) could be very useful in determining the background  $\text{CO}_2$  concentration from which the  $\text{CO}_2$  concentration measured over land has changed, making it possible to estimate the surface exchange of  $\text{CO}_2$ . The OCO instrument (failed launch 2009) had the capability to direct the sensor to pick up sun glint, greatly improving the signal-to-noise ratio of measurements over the oceans.

#### 3.3.5.6 Surface pressure

The amount of air that is traversed by radiation on the way to the detector will influence the strength of the absorption pattern. This total column amount of air can be affected by meteorology (high and low pressure systems) and by the topography. WFM-DOAS uses a dual retrieval scheme with CO<sub>2</sub> and O<sub>2</sub> being simultaneously measured. The atmospheric O<sub>2</sub> concentration is assumed to be invariable, thus the retrieval of O<sub>2</sub> is really a measure of the surface pressure or the total mass of air in the retrieved column. The dual retrieval in the WFM-DOAS system is possible because of the speed at which the retrievals are performed. The spectral lines used to identify CO<sub>2</sub> and O<sub>2</sub> are separated and some broadband features may differ between them. Though dual retrieval with O<sub>2</sub> does not perfectly describe the total column of air through which to obtain the CO<sub>2</sub> concentration it does prove particularly useful in cases of uneven topography.

#### 3.3.5.7 Temperature profile and vertical distribution of CO<sub>2</sub> and water vapour

The temperature profile and vertical distribution of CO<sub>2</sub> and water vapour affect the radiative transfer through the atmosphere. These distributions will affect the shape of the measured spectral features. Using *a priori* information about these profiles could result in better matching SCIATRAN simulated spectra. The original WFM-DOAS algorithm did not include *a priori* information such as vertical profiles of CO<sub>2</sub>, these were later included as a response to the improvements of the FSI-WFM-DOAS algorithm [Barkley et al., 2006a] which will be discussed in the next section. A CO<sub>2</sub> climatology [Remedios et al., 2006] was used to simulate sensitivity to the vertical profile of CO<sub>2</sub>. Taking a uniform climatology tends to lead to an overestimation of the true CO<sub>2</sub> concentration [Barkley et al., 2006a]. The magnitude of this overestimation has a seasonal and latitudinal pattern (see Figure 3.6)

### 3.4 Precision requirements for the investigation of carbon fluxes

As previously stated, the goal of this research is to investigate what remote sensing can tell us about the carbon cycle. In order for active management of the global



carbon cycle [Dilling et al., 2003] it is imperative that there is a better understanding of the underlying dynamics controlling atmospheric exchange, and that better carbon cycle monitoring capabilities are in place.

It has been shown using the existing network of in-situ measurement stations (the global view network) that the terrestrial biosphere and oceans have absorbed almost half of the anthropogenic CO<sub>2</sub> emitted during the last 40 years. However, the sparsity of this network means that the nature, geographic distribution and temporal variability of these sinks are not adequately understood which would preclude accurate predictions of their responses to future climate change [Cox et al., 2000; Fung et al., 2005; Friedlingstein et al., 2006].

There is a compelling case of a northern hemisphere terrestrial sink from the inversion of global view CO<sub>2</sub>, but the network is too sparse to quantify the distribution of the sink over the north American and Eurasian biospheres or to estimate fluxes over the southern ocean [Gurney et al., 2002, 2003, 2004; Law et al., 2003; Baker et al., 2006].

The ground network also has difficulty explaining the variation of the accumulation of atmospheric CO<sub>2</sub> (1-7 Gigatonnes of carbon per year) in response to steadily increasing emission rates (see Figure 1.6) [Hansen and Sato, 2004].

Space-based remote sensing of atmospheric CO<sub>2</sub> has the potential to deliver the data needed to resolve many of the uncertainties in the spatial and temporal variability of carbon sources and sinks. Several sensitivity studies have evaluated the improvement in carbon flux inversions that would be provided by precise, global space based column XCO<sub>2</sub> data. The consensus of these studies [Miller et al., 2007] is that satellite measurements yielding the column averaged CO<sub>2</sub> dry air mole fraction, XCO<sub>2</sub> with bias free precisions in the range of 1 - 10 ppmv (0.3 - 3 %) will reduce uncertainties in CO<sub>2</sub> sources and sinks owing to uniform and dense global sampling (the theoretical basis for the range of measurement precision was discussed further in Section 2.2).

The precision requirements for space based XCO<sub>2</sub> data vary depending on the spatial and temporal resolution of the data and the spatiotemporal scale of the surface flux inversion [Houweling et al., 2004].

The highest precision XCO<sub>2</sub> data (less than 1ppmv retrieval error) would address the largest number of carbon cycle questions. On the other hand, lower precision XCO<sub>2</sub> data (10ppmv retrieval error) will add no new information about the carbon

cycle compared with the current ground measurement network (except perhaps in the very poorly covered areas).

Since SCIAMACHY was not a dedicated CO<sub>2</sub> mission, little thought was put in place to achieve sub 1% XCO<sub>2</sub> data. The satellite operation of alternating limb and nadir is unhelpful, as it reduces the temporal resolution of the mission. Also the local time of retrieval at 10 am is not ideal. Much better information about the surface fluxes would be obtained from inverting data retrieved near the diurnal minimum at around 2 pm. The most significant drawback of SCIAMACHY XCO<sub>2</sub> retrievals is the inability to retrieve over ocean, owing to the poor surface reflectivity of water in the NIR.

Validation of column measurements has been difficult as there are very few similar measurements (looking at the whole vertical column) to compare against. Some of the later SCIAMACHY retrievals have been compared to the FTIR sites set up to validate the OCO / GOSAT missions, although these are few and far between.

The OCO science team analysed a broad range of modelled and measurement data to define the science requirement for space-based XCO<sub>2</sub> data precision [Miller et al., 2007]. They came up with two fundamental questions:

- What precision does the XCO<sub>2</sub> data product need to be to significantly improve our understanding of CO<sub>2</sub> surface fluxes (sources and sinks)?
- Does the measurement/retrieval/validation approach adopted provide the needed XCO<sub>2</sub> precision?

The first question is somewhat theoretical, using sophisticated modelling techniques, Miller et al. [2007] described the sort of carbon flux behaviours that they want to be able to remotely study, and then examine how the instrument sees this behaviour. The second question can be answered using the fitting error of SCIATRAN to each SCIAMACHY spectra.

### **3.5 The effect of sampled biases on the investigation of carbon fluxes**

Some of the various factors that lead to either un-retrieved or biased individual satellite measurements are discussed in Section 3.3.5. There will be a knock-on

effect to the forward or inverse modelled carbon fluxes as a result of these individual measurement biases, particularly since the cause of these biases may affect an entire region rather than just an individual measurement. Section 3.5.1 examines the consequences of measured biases to the investigation of the carbon fluxes, summarising work not conducted by the author. This section is twinned with Section 4.1, examining the biases introduced with regards to sampled vs unsampled retrievals, which should be regarded as an original contribution of the author. Both types of bias were considered when developing the forward and inverse modelling tools described in Chapter 6.

### **3.5.1 The effect of aerosol contamination on the investigation of carbon fluxes**

A sampling bias is introduced by including pixels that should in fact have been discarded, as in the case of aerosol contaminated pixels. The Bremen group re-examined their retrieval scheme after the development of the FSI algorithm. The latest retrieval scheme [Schneising et al., 2008] uses an aerosol detection method developed for TOMS [Herman et al., 1997]. In Figure 3.7 the monthly averaged unfiltered CO<sub>2</sub> is plotted on the left hand side showing June 2003 on the top and November 2003 on the bottom, with the right hand side showing the corresponding monthly averaged Absorbing Aerosol Index (AAI) from TOMS. It is clear that for a biologically inactive region such as the Sahara desert, this region should not be a source which could be interpreted from the June plot. It is clear that the high concentrations in this scene are a result of aerosol contamination instead. The uncontaminated November scene corresponds to an inactive carbon cycle in the Sahara region.

Schneising et al. [2008] also shows the annual average for three years of WFM-DOAS CO<sub>2</sub> showing unfiltered (left) and aerosol filtered (right) for the whole globe (see Figure 3.8). The Americas appear to be less affected by aerosol contamination which would make them a better place to examine carbon fluxes than parts of Africa or Asia.

Inverse modelling calculations confirm that the aerosol-induced errors in the measured CO<sub>2</sub> total column are too large to allow meaningful source and sink estimates [Houweling et al., 2005]. For North Africa several petagrams (Pg) of carbon are needed to bring the model in agreement with the measurements. Even

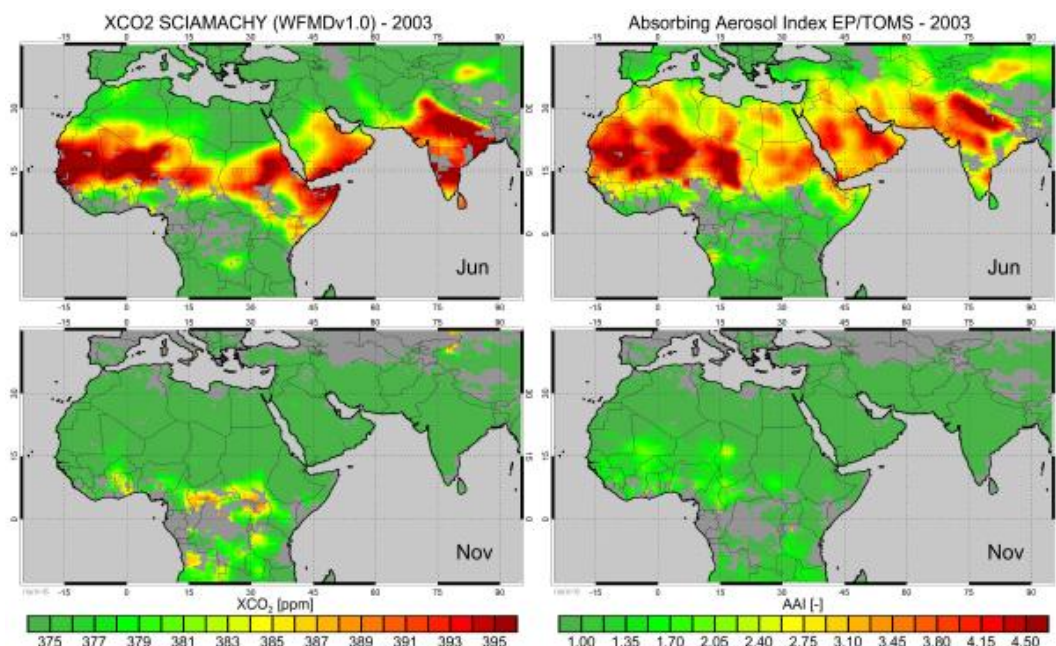


FIGURE 3.7. The left panels shows SCIAMACHY XCO<sub>2</sub> not filtered for aerosol contamination (top: June 2003, bottom: November 2003) [Schneising et al., 2008]. The two panels on the right show the Absorbing Aerosol Index (AAI) data product from TOMS/Earth Probe. The data have been smoothed using a boxcar function.

for Temperate North America (the contiguous United States) an error of 0.4 Pg/yr is found, exceeding the uncertainty of inverse modelling estimates for this region on the basis of the surface monitoring network reported by Gurney et al. [2002].

## 3.6 Preliminary findings of FSI and WFM DOAS

### 3.6.1 FSI vs AIRS

A comparison was made with CO<sub>2</sub> retrieved from AIRS by Engelen et al. [2004] again examining seasonal patterns over large regions [Barkley et al., 2006b]. As previously mentioned in Chapter 2, the AIRS instrument is most sensitive to variations in CO<sub>2</sub> in the upper troposphere and lower stratosphere. Since AIRS cannot see down to the surface, any spatial features detected in the CO<sub>2</sub> field by AIRS cannot be written off as differences in surface reflectance, which has been suggested as a possible cause of the spatial features in SCIAMACHY CO<sub>2</sub>. In fact, Barkley et al. [2006b] found similar spatial patterns in the CO<sub>2</sub> field of both instruments, albeit with AIRS having roughly half the deviation from the monthly average CO<sub>2</sub>

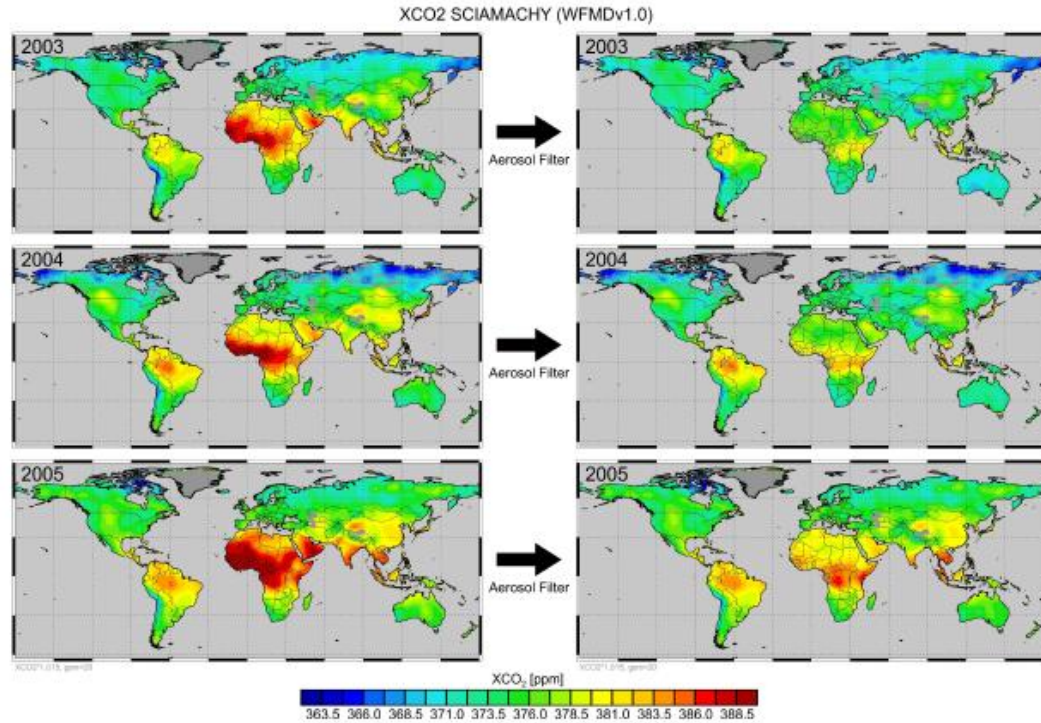


FIGURE 3.8. Yearly averages of SCIAMACHY  $\text{CO}_2$  showing the data without (left) and with (right) aerosol (AAI) filtering [Schneising et al., 2008]. When the two are matched with the corresponding CarbonTracker annual averages the filtered provides a much closer pairing.

(see Figure 3.11). The seasonal cycle of AIRS  $\text{CO}_2$  has a smaller magnitude and is slightly delayed to that of SCIAMACHY/FSI (see Figure 3.12), both to be expected as surface flux signatures take time to reach the high parts of the atmosphere and will have been diluted in that time. The monthly averaged  $\text{CO}_2$  fields are again obtained using the active grid square method.

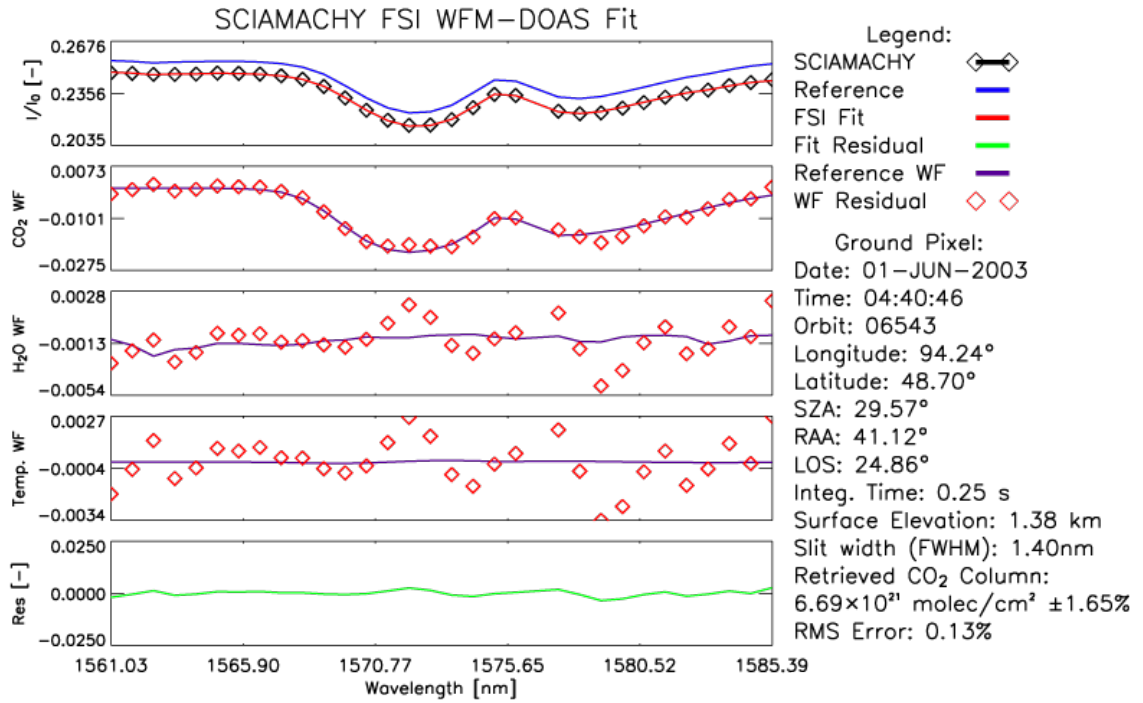


FIGURE 3.9. A typical example of a FSI WFM-DOAS fit [Barkley, 2007]. Top panel: The reference spectrum (blue) generated by the retrieval (from the *a priori* data), the sun-normalised radiance measured by SCIAMACHY (black diamonds) and the FSI WFM-DOAS fit (red) to this measurement. Following three panels: (i) The CO<sub>2</sub> fit. This shows CO<sub>2</sub> total column weighting function (WF) and the CO<sub>2</sub> fit residuum (red diamonds), which is the CO<sub>2</sub> WF plus the difference between the measurement and fit (i.e. the fit residual) (ii)-(iii) Similar but for H<sub>2</sub>O and temperature. Bottom Panel: The fit residual with a root-mean-square (RMS) difference of 0.13%.

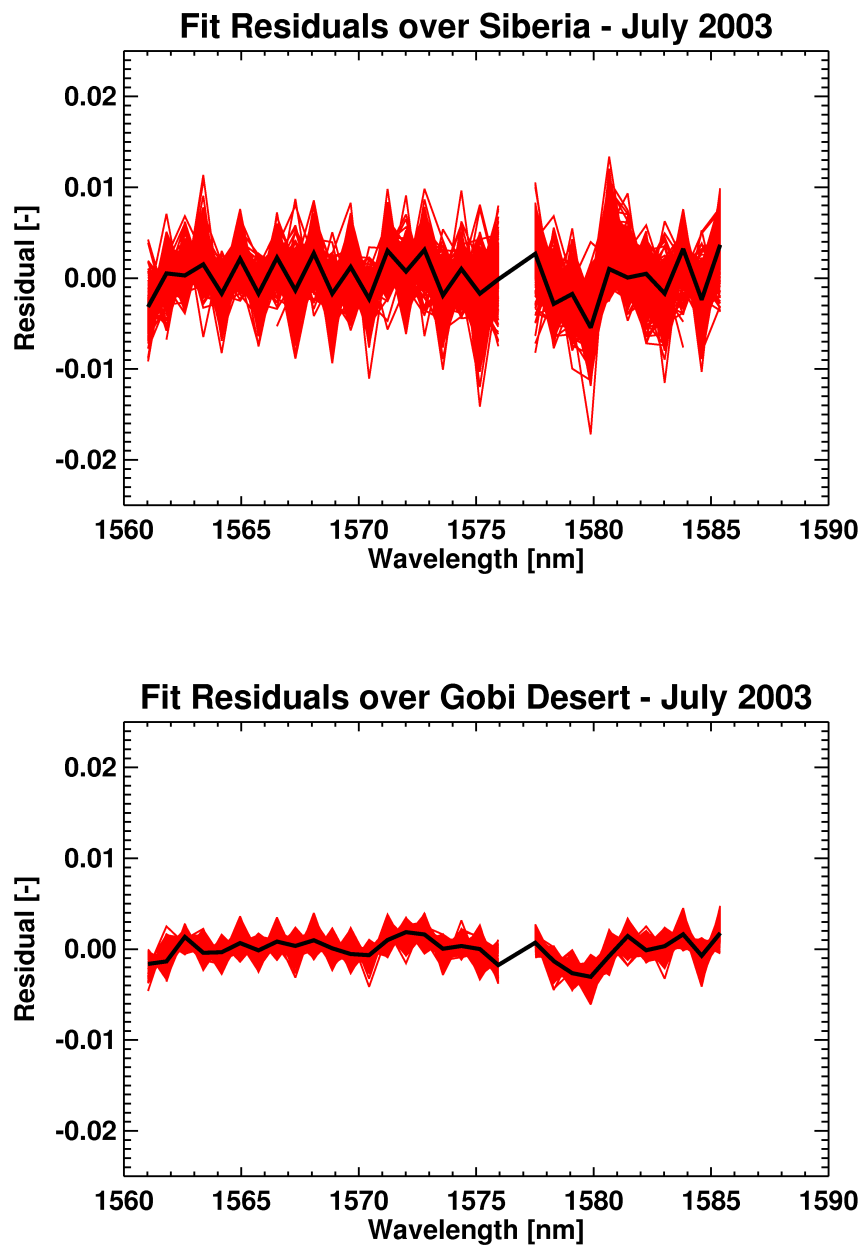


FIGURE 3.10. FSI WFM-DOAS fit residuals [Barkley, 2007] over Siberia (top) and the Gobi Desert (bottom) during July 2003, over-plotted with the mean fit residual (black). The mean RMS errors are 0.237% and 0.140% respectively. The detector pixel, of wavelength 1576.72 nm, is omitted in the retrieval as this always worsens the quality of the fit and increases the error on the retrieved column.

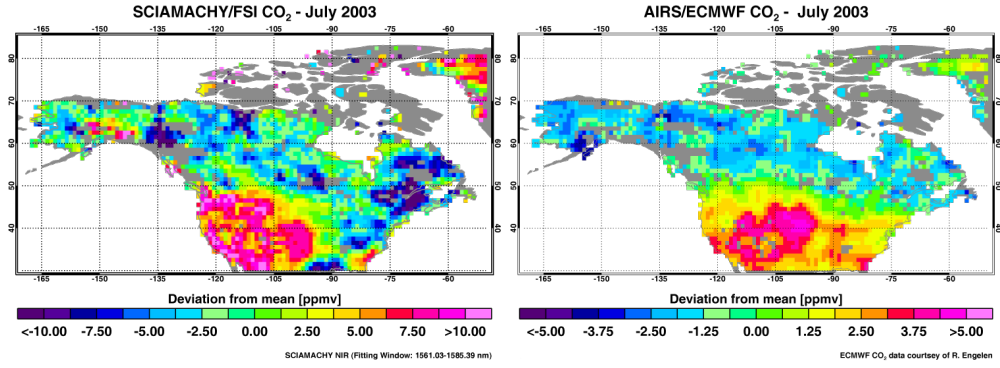


FIGURE 3.11. The SCIAMACHY monthly averaged deviations over North America in July 2003 (left) and the corresponding AIRS observations (right) [Barkley et al., 2006b].

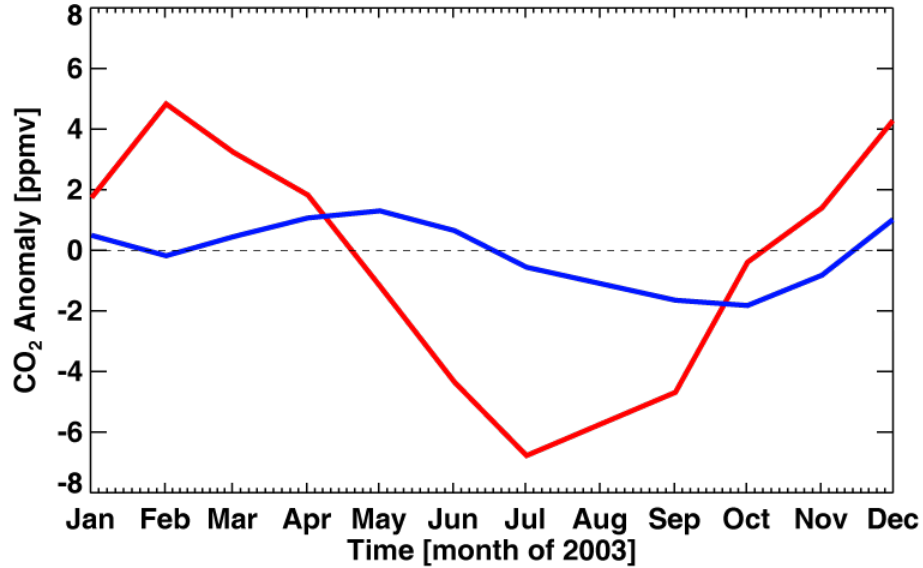


FIGURE 3.12. The  $\text{CO}_2$  anomaly over North America detected by SCIAMACHY/FSI (red) and AIRS (Blue) [Barkley et al., 2006b].

### 3.6.2 FSI vs FTIR

Barkley et al. [2006c] compared FSI retrievals with measurements at the FTIR site in Egbert, Canada. The site is 70 km away from Toronto and some potential comparisons between the data sources may be subjected to aerosol contamination. In this initial test, FSI retrievals were said to be co-retrieved within the same day's FTIR measurement when samples were collected with  $10^\circ$  longitude and  $5^\circ$  latitude of the Egbert station. Using the methodology described by Dils et al. [2006], it was shown that FSI-retrieved  $\text{CO}_2$  is negatively biased to the FTIR station by 4.1%



Table 3.1. Comparison of daily averaged FSI measured CO<sub>2</sub> falling within various collocation limits of the FTIR site at Park Falls [Barkley et al., 2007]. For each collocation limit the following is given: the number of days with collocations,  $N_c$ , the mean bias of FSI to Park Falls,  $B$ , the annual mean of the collocated days for Park Falls,  $\text{Mean}_{PF}$ , and for SCIAMACHY,  $\text{Mean}_{FSI}$ , and their corresponding  $1\sigma$  standard deviations as well as the correlation of the mean of the collocated mean time series,  $r$ .

Collocation limits (lon $\times$ lat)	$N_c$ [-]	$B$ [%]	$\text{Mean}_{PF}$ [ppmv]	$\sigma_{PF}$ [ppmv]	$\text{Mean}_{FSI}$ [ppmv]	$\sigma_{FSI}$ [ppmv]	$r$ [-]
$1.0^\circ \times 1.0^\circ$	20	-2.1	374.5	2.1	366.7	9.3	0.36
$3.0^\circ \times 3.0^\circ$	29	-1.5	374.5	2.7	369.1	7.8	0.73
$5.0^\circ \times 5.0^\circ$	34	-1.1	374.4	2.6	370.2	6.8	0.71
$10.0^\circ \times 10.0^\circ$	40	-0.9	374.4	2.8	371.0	6.2	0.68

with a standard deviation of 2.8% (the TM3 model CO<sub>2</sub> was negatively offset by 1.8% with a standard deviation of 1.9%).

There were concerns that aerosol contamination of both measurement sets was affecting the results of this comparison. Before comparing to a second location at Park Falls some consideration of the nature of the comparison was made. Taking large collocation limits is a good tool for observing the overall offset bias between SCIAMACHY/FSI and the FTIR location as it is assumed that for large limits the random noise variability of individual FSI measurements are canceled out and that the mean of a days measurements would be representative of the regional CO<sub>2</sub> field. It also has to be assumed that the FTIR site measurements are representative of the regional CO<sub>2</sub> field, since they are not dominated by local sources and sinks. It would be interesting to examine if this is the case, which could be done by calibrating two FTIR stations next to each other (so that they monitor the same air) then moving one a few kilometers away so they sample the same regional air but different local air.

Taking very narrow collocation limits allows a much more direct comparison of FSI against FTIR, which could be used to examine whether the apparent random noisiness of individual FSI measurements is related to the noisiness of the CO<sub>2</sub> field as a result of local fluxes. Unfortunately, for both large and narrow collocations, the temporal sparsity of diurnal measurements at Park Falls meant that a diurnal average was used as the daily average for Park Falls rather than a small cluster of measurements close to the SCIAMACHY overpass time [Barkley et al., 2007]. This is likely to affect the offset bias (shown in Table 3.1) between the two measure-

ments and make the narrow collocation limit experiment impossible. The strong correlation coefficient of the time series of 20, 29, 34 and 40 days for FTIR and FSI is likely to be mainly influenced by the seasonal cycle of the CO<sub>2</sub> field.

### **3.7 Summary**

To aid the flow of this thesis, Chapter 3 was originally intended to be background material describing the development of FSI and analysis of individual retrievals and Chapter 4 would continue to describe the analysis of clusters or scenes of FSI measurements. The thesis has been redrafted so that Chapter 3 contains work not carried out by the author (including some work on the analysis of FSI scenes) and Chapter 4 only contains work with a significant contribution from the author. To retain flow, the summary of Chapter 4 will include some results from this Chapter.

## Chapter 4

# Analysis of FSI-WFM-DOAS retrievals

This research project overlapped with the work of Barkley [2007]. Much of the work in this Chapter was carried out jointly by the author and Michael Barkley following similar lines of inquiry. The development of differing methodologies for the analysis of FSI-WFM-DOAS reflected our differing interests in the CO<sub>2</sub> product. Of particular interest to this author was the change of mass of CO<sub>2</sub> with time, the development of which eventually led to the adaptation of NAME to analyse the carbon cycle (see Chapter 6). All work presented in Chapter 4 is either an original idea of this author or analysis results where this author has made a significant contribution.

### 4.1 The effect of sampling biases on the quantification of carbon fluxes

SCIAMACHY samples many hundreds of footprints each orbit, which can be fed into the FSI-WFM-DOAS algorithm. For a variety of reasons only a small fraction of these retrievals are used to obtain column XCO<sub>2</sub> concentrations. The reasons for excluding pixels are justified in allowing only the most precise measurements to be carried forward into any future inversion scheme.

Some of the various factors that lead to either un-retrieved or biased individual satellite measurements were discussed in Section 3.3.5. There will be a knock-on effect to the forward or inverse modelled carbon fluxes as a result of these

individual measurement biases, particularly as the causes of the biases may affect an entire region rather than just an individual measurement. In this section, some of the possible consequences to the investigation of carbon fluxes introduced through selective sampling of satellite CO<sub>2</sub> retrievals are discussed. The literature makes no mention of the possible consequences of sampling biases in satellite retrieved XCO<sub>2</sub> to the forward and inverse modelled carbon fluxes.

For obvious reasons, a comparison between sampled and unsampled measurements was not possible, as a result these biases have been described qualitatively.

Consideration of both the sampling biases described here and the sampled biases described in Section 3.5 were considered when developing the forward and inverse modelling tools described in Chapter 6.

#### **4.1.1 Cloudy vs cloud free pixels**

Cloudy pixels are not retrieved by the FSI algorithm as SCIAMACHY cannot see through clouds in the wavelength range used to measure XCO<sub>2</sub>. Though it is possible to sample above cloud only CO<sub>2</sub> mixing ratios, this would ignore the lowest part of the atmosphere where SCIAMACHY is most sensitive and where the majority of carbon fluxes take place.

There are various consequences for the biosphere to atmosphere exchange when clouds are present. Firstly, clouds will lower the amount of sunlight reaching the biosphere, which will impact upon photosynthetic behaviour. Secondly, clouds have a cooling effect during the day (and a warming effect at night - though we do not sample at night), and this will impact upon the respiration behaviour of the biosphere. Thirdly, clouds often bring rain with them, which can again have an impact upon photosynthesis particularly in water stressed regions.

This sampling bias is skewed in its impact over the globe. For mid-latitude locations cloud and rain may be randomly distributed within a region though there may be some diurnal and seasonal patterns. For some tropical locations such as rainforests, rain and cloud might be expected every day at certain parts of the diurnal cycle and a SCIAMACHY overpass time of 10 am may generate a different percentage of cloud free pixels to the GOSAT overpass time of 1:30 pm and the planned OCO overpass time of 2 pm.

The cloud algorithm used to filter cloudy scenes detects clouds in the path of the light reflected from the surface. For very high latitude retrievals there may be

some situations where there is cloud present in the sun to surface path but not the surface to satellite path. This scenario is a problem for high latitudes only because of the larger angle between incident and reflected light; where the angle is small the same cloud would be present in both light paths.

#### **4.1.2 Snow/ice vs snow/ice free pixels**

Much of the higher northern latitudes is snow covered in the winter months. The FSI retrieval scheme is unable to retrieve  $\text{XCO}_2$  over snow covered pixels as snow is highly absorbing in the near infrared (unlike in the visible) and so results in a poor signal-to-noise ratio. When examining a scene average for a region such as Siberia, it is likely that the high northern latitudes will be unsampled through the winter months and the scene average will be sampled from the southern parts of the scene. One problem with this set up is the latitudinal gradient of the background  $\text{CO}_2$ . Introducing a southern sampling bias in the measurements for this scene also means introducing a clear bias from the true scene average  $\text{CO}_2$ .

A more serious problem with this set up would be to make the mistake of assuming that carbon fluxes at high latitudes are negligible in the winter months since the extreme cold will hamper respiration. Some dedicated field work in hostile conditions [Zimov et al., 1993, 1996, 1999], has shown that at very high northern latitudes the seasonal amplitude of atmospheric  $\text{CO}_2$  is very large. At the onset of winter, the deep snow in Siberia acts as a blanket protecting the soil from the extreme cold, and soil respiration provides enough heat to allow the soil to continue to respire throughout the winter.

The result of the sampling bias would appear to show negligible  $\text{CO}_2$  emission over the Siberian region when the unsampled regions are actually strongly emitting carbon. As a continuation of this project it would be interesting to examine the high northern latitudes in winter using an OCO-like instrument in glint mode.

#### **4.1.3 Ocean vs land pixels**

Two thirds of the Earth's surface is covered by water. The SCIAMACHY instrument is unable to retrieve  $\text{XCO}_2$  over ocean pixels as water is highly absorbing in the near infrared and so results in a poor signal-to-noise ratio at the detector. The carbon cycle of the unsampled ocean is better understood (air - sea exchange is a

much more gradual process with a small annual cycle). Knowledge of  $XCO_2$  over the ocean would still have its uses; for example ocean measurements could be used to provide the background concentrations.

#### **4.1.4 Time of day sample bias**

All SCIAMACHY overpass times occur at 10 am local time. The nature of the diurnal cycle for photosynthesis and for respiration is rather different. With a gradual peak in the day and a moderately smaller trough at night, respiration is a 24 hour process. This is contrasted with a very strong peak at the middle of the day and no activity in the night for photosynthesis. The net carbon exchange is much smaller than the peak of photosynthetic activity, so retrieving  $CO_2$  from any one sun-synchronous satellite will skew the understanding of the carbon cycle.

### **4.2 Scene average $CO_2$**

A like-for-like direct comparison of a single SCIAMACHY sounding with a second instrument scanning exactly the same part of the atmosphere in space and time is not possible. The ground based FTIR sites are the closest match to what SCIAMACHY measures geometrically and very few measurements would match up when SCIAMACHY is directly overhead. Some sort of averaging of both FSI and the second instrument measurements need to be made. Like-for-like column comparisons are possible with modelled atmospheric  $CO_2$  concentrations produced either from “bottom up” fluxes or from the inversion of an alternative atmospheric measurement scheme. It must be remembered that these modelled concentrations are also not a perfect representation of the true atmospheric  $CO_2$ , and so averaged results from the two could be compared to investigate potential offsets, differences in seasonality or differences in amplitude of maximum and minimum concentrations. It is also desirable to investigate some of the behaviours of the carbon cycle without comparing to a control  $CO_2$  concentration, for example to compare one year to the next.

In all of these cases, sampling biases discussed in Section 4.1 will influence the “average” concentration. The following example compares the TM3 model  $CO_2$  fields with those of retrieved FSI.

## 4.2.1 Scene average CO<sub>2</sub> biases in TM3 vs FSI comparison

### 4.2.1.1 Temporal distribution

TM3 is a global atmospheric tracer model, developed by the Max Planck Institute for Biogeochemistry (MPI-BGC), which solves the continuity equation for an arbitrary number of atmospheric tracers [Heimann and Korner, 2003]. The atmospheric transport is driven by National Center for Environmental Prediction (NCEP) meteorological fields using a model grid of  $1.8^\circ \times 1.8^\circ \times 29$  layers with 8 3-hourly time steps per day. The ocean air-sea fluxes are based on the monthly CO<sub>2</sub> climatology compiled by Takahashi et al. [2002] whilst the natural terrestrial biospheric fluxes were modeled using the BIOME-BGC model driven with daily NCEP data, using a simple diurnal cycle algorithm [Thornton et al., 2005].

Five scenes (West Europe 42-62°N by 25°W-15°E, North America 30-85°N by 170-60°W, India 4-32°N by 70-100°E, Siberia 45-85°N by 60-160°E and the Gobi Desert) were chosen for this comparison. Of interest is the seasonal cycle, and for both data types 12 monthly averaged scene CO<sub>2</sub> mixing ratios were obtained. For the TM3 this was simple, with one value per grid box the closest matching 3 hour slot (to the 10 am overpass) was taken for every day and for every grid square in the region. These were summed together and divided by the number of grid boxes to produce the mean and standard deviation of the TM3 CO<sub>2</sub> for each month.

A simple way to obtain the monthly averaged FSI CO<sub>2</sub> would be to add up all the measurements in the scene for that particular month. However, some parts of the scene are over represented with many more measurements per unit area than other parts of the scene (some parts of Scotland have very few measurements) and much of the scene is ocean with no measurements at all. This leads to a southern bias in the averaging of FSI which is not present for the TM3 average (see Section 4.2.1.3).

To remove this bias, the scene was broken up into  $1^\circ$  by  $1^\circ$  grid squares and a monthly average for each of these was obtained. Grid squares with at least one measurement in a month were considered “active” and the monthly average FSI CO<sub>2</sub> was taken from the average of all the active squares. To ensure a fair comparison, the TM3 fields were reordered into  $1^\circ$  by  $1^\circ$  grid squares and the average was taken using only the active FSI grid squares for that month (see Figure 4.1).

Note that this does not completely eradicate spatial bias. The need for at least

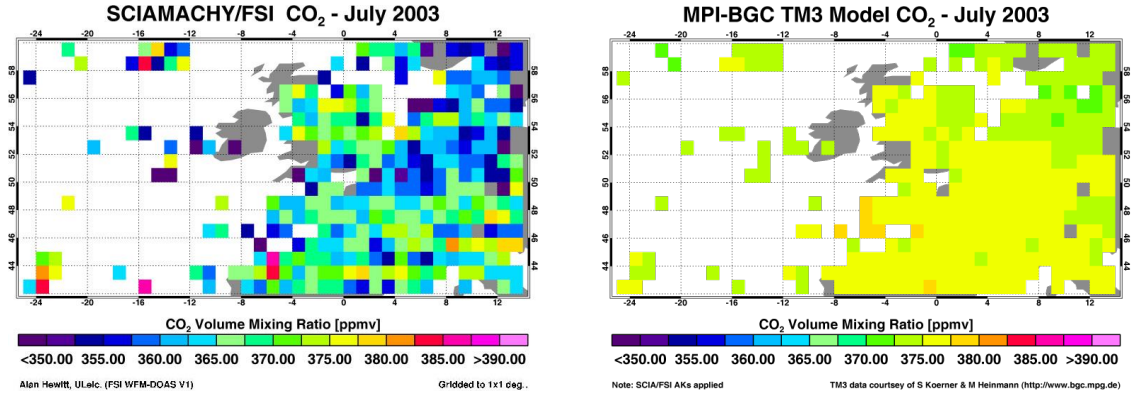


FIGURE 4.1. Collocated grid squares of the TM3 model and FSI retrieved  $\text{CO}_2$  VMRs over Western Europe. Left: The FSI retrieved  $\text{CO}_2$  VMRs over Western Europe for July 2003. Grid squares with at least one measurement in the month are active. Right: The corresponding TM3 model  $\text{CO}_2$  field.

one measurement to activate a grid square means that for some wintery scenes (e.g. Siberia and Canada) there will be a southern bias in the monthly average. This will have a knock on effect on the seasonal cycle as the average of winter grid square will be south of the average of summer grid squares. By setting a low threshold of just one measurement required to activate a grid square it is likely that this will favour the out-laying measurements and increase the standard deviation of the average compared to the standard deviation of all the measured columns in the month.

By adopting the grid square system to eradicate spatial bias, it is necessary to chose an arbitrary grid size which will affect the standard deviation on the month mean FSI  $\text{CO}_2$  (though not to the same extent that of TM3 which has a smoothing applied to it).

Similar issues are involved when considering an annual average  $\text{CO}_2$  concentration. If all the individual measurements were summed up to get an annual average  $\text{CO}_2$  concentration, there would be a spatial bias as discussed above and also a temporal bias since the number of measurements in the summer months greatly outnumber that of the winter months. The most correct method to obtain a yearly averaged  $\text{CO}_2$  would be to obtain 12 monthly averaged concentrations and then take the average of these (see Figure 4.2). As the arbitrary division into  $1^\circ$  grid squares affected the standard deviation of the monthly average, so the arbitrary division of months instead of weeks will affect the standard deviation on the yearly average.



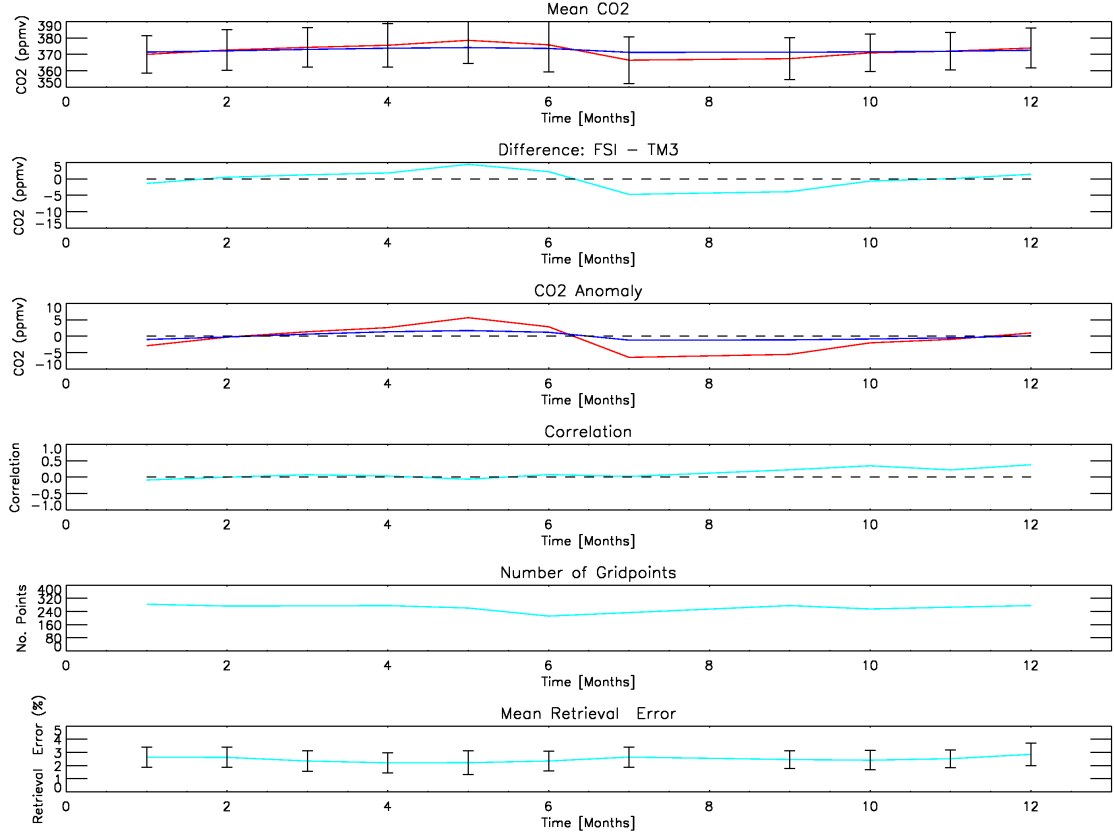


FIGURE 4.2. Comparisons between the TM3 model data (blue lines) and the FSI-retrieved  $\text{CO}_2$  VMRs (red lines) for the Indian region for the year 2003. Top Panel: The mean  $\text{CO}_2$  VMR of each scene. The error bars on the FSI data represent the  $1\sigma$  standard deviation of the mean. Second panel: The mean difference between the FSI columns and the TM3 data (equivalent to the difference between the monthly averages). The error bars represent the  $1\sigma$  standard deviation of this difference. Third Panel: The  $\text{CO}_2$  anomaly (i.e monthly averages minus the yearly mean). Fourth Panel: The correlation coefficient between the two data sets. Fifth Panel: The number of active grid squares used in the calculation of the scene means. Bottom Panel: The mean FSI retrieval error of the observed  $\text{CO}_2$  VMRs with the  $1\sigma$  standard deviation, (note that all measurements with a retrieval fitting error greater than 5% were filtered out). Measurements for August 2003 have been omitted.

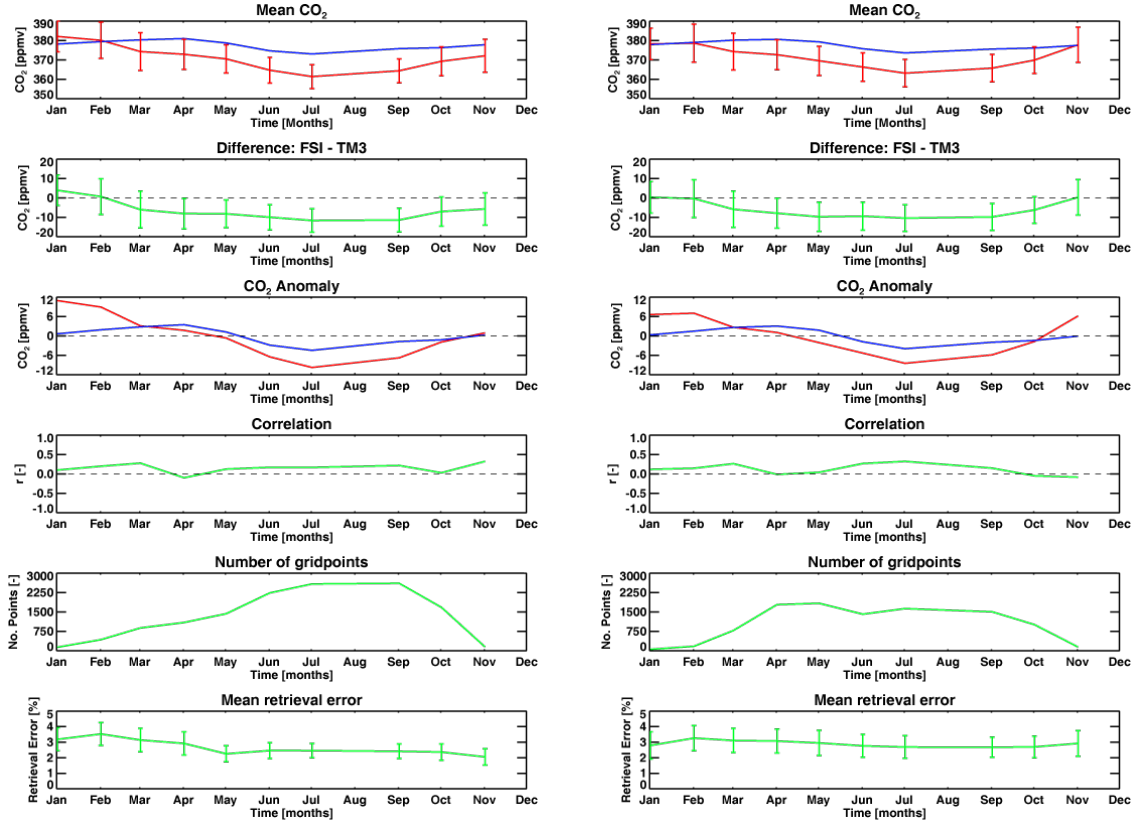


FIGURE 4.3. As Figure 4.2 but for the Siberia (left) and North America (right) regions for the year 2003 [Barkley et al., 2006c].

The TM3 model has been calibrated for an optimal match with in-situ measurements at the South Pole station. The TM3 model shows a 2% negative bias to measured columns at the Egbert, Canada FTIR site [Barkley et al., 2006c] and small differences in the seasonal cycle amplitude at this location so TM3 CO<sub>2</sub> should not be considered true.

There are considerable differences between the yearly mean CO<sub>2</sub> concentrations of TM3 and FSI for all the regions (see column 5 in Table 4.1) averaging out as a 2% negative bias. The Indian region suffers from excessive aerosol contamination in the first half of the year as the air is trapped by the Himalayas, which raises the mean concentration by such a degree as to cancel out the negative offset. India is located at a much lower latitude than the other scenes and is subject to a different background concentration (see Figure 2.1) which could explain its significantly different TM3 yearly mean CO<sub>2</sub> concentration.

The seasonal cycle amplitude of the TM3 data seems to be related to the latitude of the scene with the most northerly scenes having the greatest SCA. Ignoring India (which is heavily contaminated by aerosol) the SCA for FSI monthly averaged

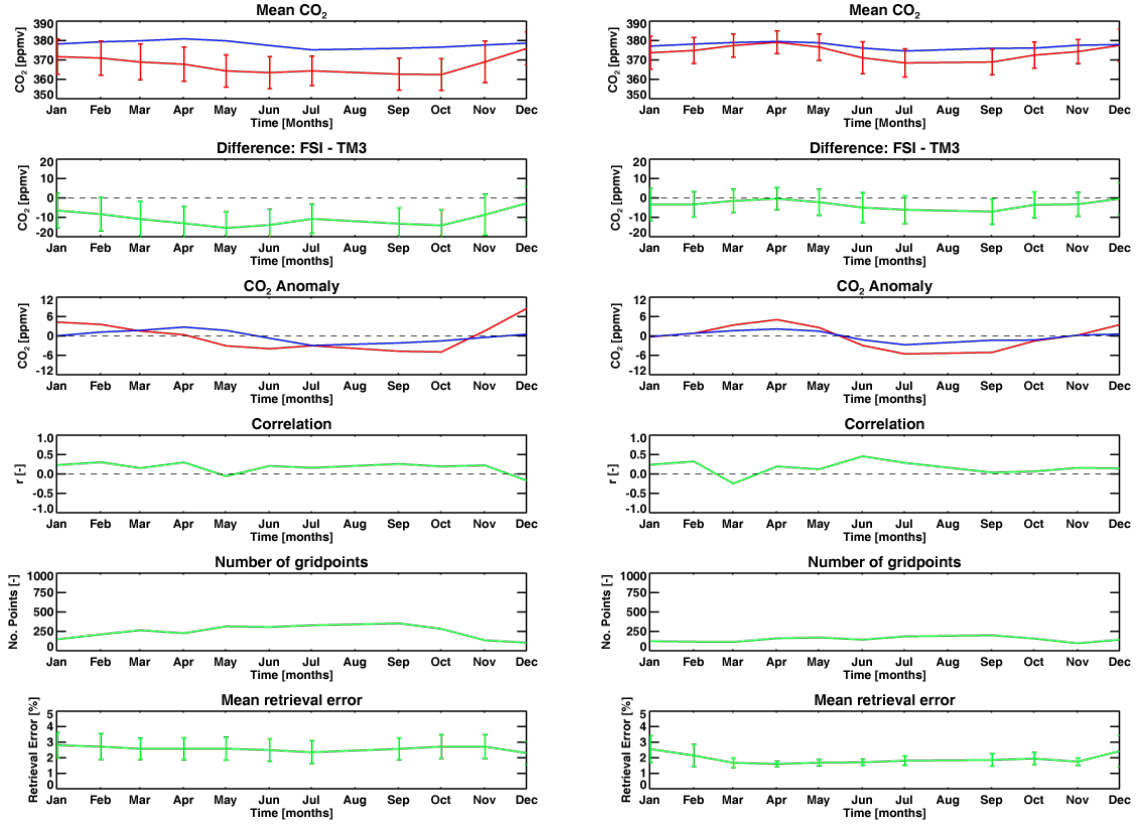


FIGURE 4.4. As Figure 4.2 but for the Western Europe (left) and Gobi desert (right) regions for the year 2003 [Barkley et al., 2006c].

CO<sub>2</sub> is 2-3 times greater than that of TM3. The minimum concentration of both SCIAMACHY and TM3 CO<sub>2</sub> for every scene occurs in July and the maximum for TM3 occurs in April. It would appear that the growing season where photosynthesis exceeds respiration is between April and June. This excess is rather steep as the maximum to minimum takes just 3 months followed by a more gentle increase over the other 9 months. Assuming that fossil fuel burning skews the biospheric seasonal cycle it is more likely that the biosphere continues take up through out July as well. The maximum monthly concentration of CO<sub>2</sub> in FSI occurs for most scenes in January, however there are issues with data availability in the winter months.

#### 4.2.1.2 Spatial distribution

Despite the differences in the magnitude of seasonal cycles and biases between the measurements, there are clear spatial similarities in the monthly averaged CO<sub>2</sub> fields of retrieved FSI and TM3 model data (see Figure 4.5). This is a very promising result as the flux values used in the TM3 model are produced with a bottom-

Table 4.1. Summary of the FSI retrievals and TM3 model comparisons. “SCA” refers to the Seasonal Cycle Amplitude and “Mean Correlation” refers to the average correlation between the monthly gridded data. Typically SCIAMACHY under estimates the yearly mean by approximately 2%, whilst the average difference between observation and model is 0-3% depending on the region.

Region	Yearly Mean		SCA		Difference		Time Series	Mean
	[ppmv]		[ppmv]		[ppmv]		Correlation	Correlation
	FSI	TM3	FSI	TM3	Mean	1 $\sigma$	[-]	[-]
Gobi Desert	374.0	377.3	10.1	4.9	3.3	6.9	0.95	0.16
North America	371.6	377.5	15.4	7.0	6.0	7.9	0.67	0.12
Siberia	371.2	377.5	20.7	7.9	7.3	7.6	0.75	0.15
Western Europe	367.5	378.1	13.5	5.7	10.7	8.6	0.47	0.16
India	372.9	372.4	12.1	2.8	0.5	8.0	0.94	0.10

up methodology. It is likely that the information used to run the TM3 model is reasonably close to the true atmospheric CO<sub>2</sub> field, therefore the spatial matching of FSI to these results indicates that the FSI retrieval is detecting real signatures of carbon fluxes.

The disparity in the seasonal cycle amplitude of the SCIAMACHY measured CO<sub>2</sub> fields to those of the TM3 model and indeed to the FTIR stations are not understood and neither is the origin of the offsets between these fields. Flux behaviour does not stay constant over the course of a month, and comparisons with shorter time frames of satellite data, such as a single overpass, would prevent a whole scene comparison and thus matching up of the spatial patterns in the data. Despite the difference in magnitude of the SCAs, the evidence from FSI vs TM3 shown in Table 4.1 suggests that SCIAMACHY/FSI is good enough to validate forward modelled CO<sub>2</sub> fields.

One of the dangers of using images such as those shown in Figure 4.5 to demonstrate relationships between two fields is that the eye wants to see patterns, sometimes even when they are not there. When considering that the correlation between active grid squares in the TM3 and FSI fields (fourth from the top Figures 4.2, 4.3 and 4.4) does not exceed 0.3 and for many months remains at zero, without also examining the plots, it would be reasonable to assume that FSI does not detect carbon fluxes. It would appear that the correlation in the large-scale features in the two fields is being undermined by the high variability of active grid squares on the 1° scale. Quantifying the large-scale patterns is difficult, it is possible to

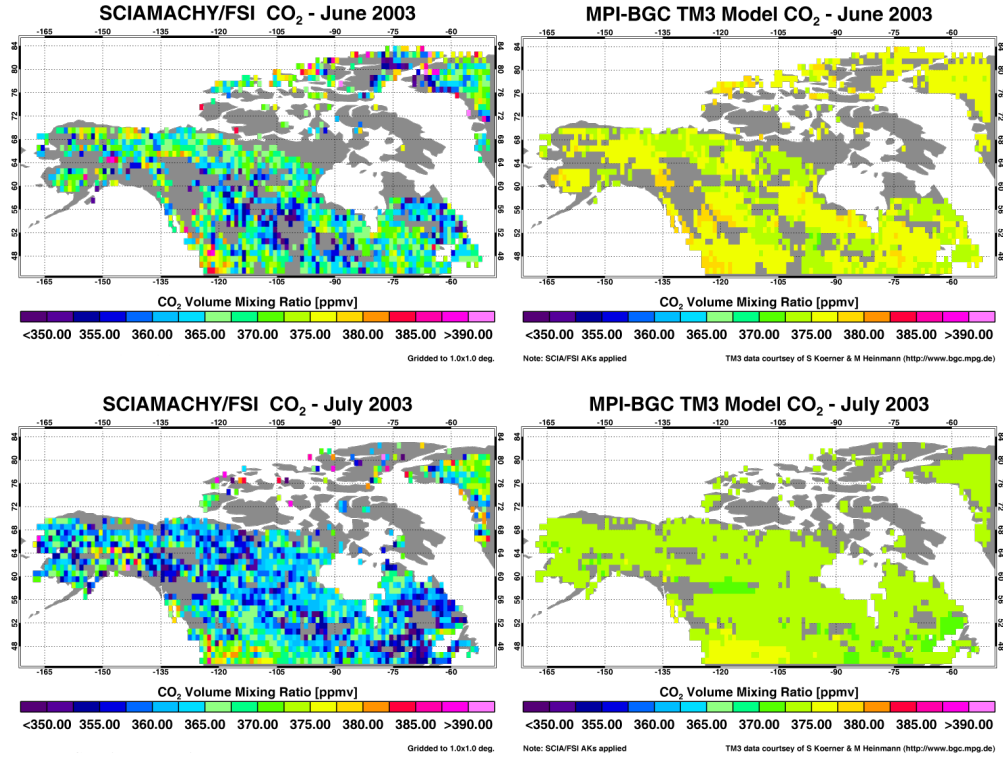


FIGURE 4.5. The SCIAMACHY/FSI monthly scene averaged CO<sub>2</sub> over North America (left) for June (top) and July 2003 (bottom) on a 1° × 1° grid are plotted next to the CO<sub>2</sub> fields produced from TM3 (right). The spatial similarities between the two fields suggest that SCIAMACHY/FSI could be used to validate forward modelled bottom-up fluxes [Barkley et al., 2006c].

apply a smoothing to the FSI field to investigate the correlation of large-scale features. However, two smoothed fields are likely to show a strong positive or negative correlation simply by chance.

#### 4.2.1.3 Different methodologies for obtaining scene monthly average

In Michael Barkley’s original comparison the monthly average FSI CO<sub>2</sub> was obtained by taking the mean of all measurements made within the scene boundary over the course of the month and the corresponding monthly average TM3 CO<sub>2</sub> was obtained by taking the mean of all the grid squares on all of the days in the month.

Two problems exist with this method, firstly that the entire scene (including over ocean) is included by the TM3 but only limited parts by FSI and secondly that there is a heavy weighting to regions of the scene that have a higher density of sampling (see Section 4.1). If the same features in the CO<sub>2</sub> field are present in

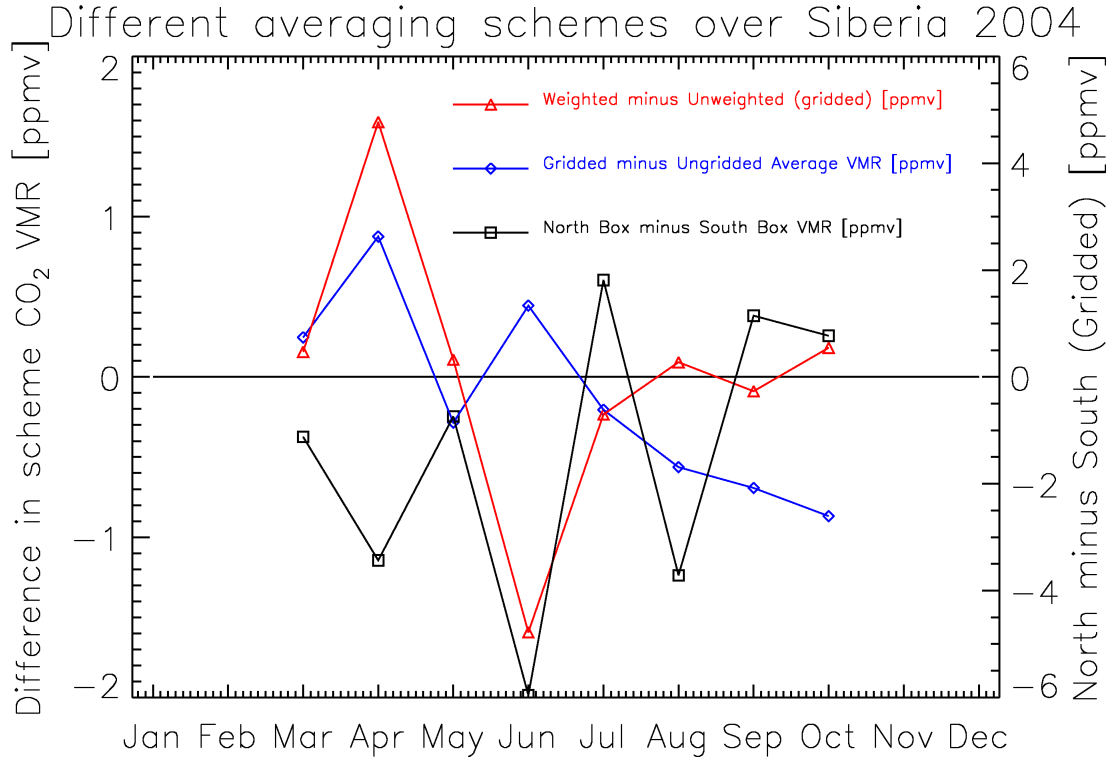


FIGURE 4.6. Different monthly averaging schemes over Siberia 2004. The blue line (use left axis) represents the difference in the monthly average between the un-gridded average (that is the mean of all the measurements in that month) and the gridded average (where the scene has been divided into grid squares, calculating the mean of each the occupied grid squares, then the mean of the grid squares is taken). The red line (use left axis) shows the difference between the weighted and un-weighted (by latitude) monthly average. The black line (use right axis) shows the difference in the gridded monthly average between the north and south boxes shown in Figure 4.7.

TM3 and FSI, it is essential that a monthly average comparison incorporates the same space.

An alternative monthly averaging system that eliminates the spatial bias and limits the weighting bias has been developed. By dividing the FSI scene into grid squares that are equivalent to the TM3 grid squares, it is possible to identify active grid squares that contain at least one FSI measurement in a month. For each active grid square the mean of FSI measured CO<sub>2</sub> is obtained and the scene monthly mean is then the mean of all the active grid squares. Only TM3 grid squares that are equivalent to FSI active squares will be averaged over for that month. In this way the TM3 and FSI CO<sub>2</sub> fields for each individual month can be considered spatially equivalent.

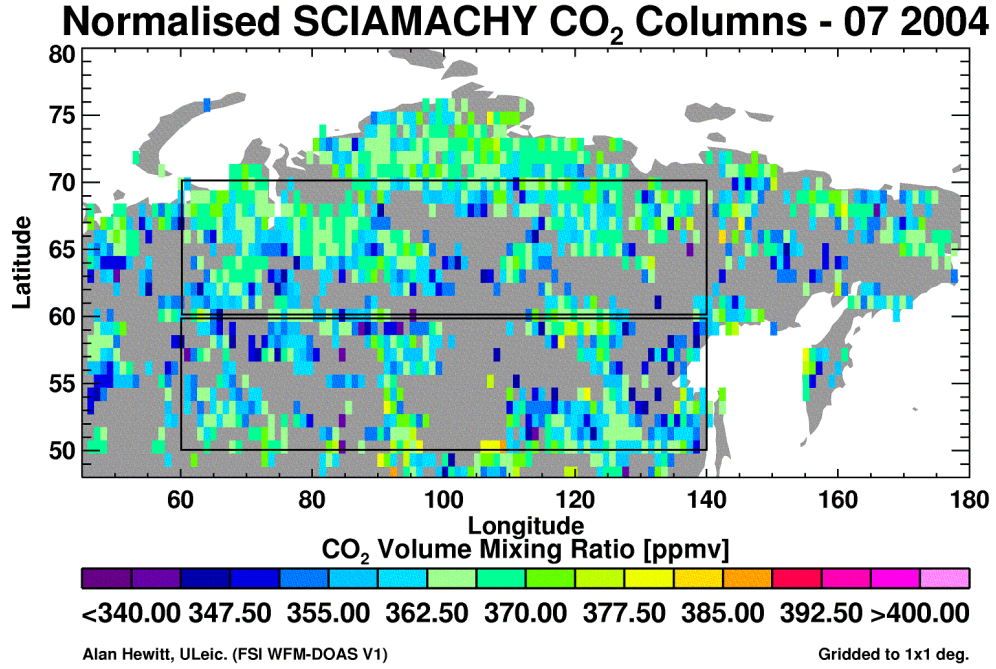


FIGURE 4.7. The map shows the north and south regions used to demonstrate how adjusting for the latitudinal distribution of measurements can produce a different monthly averaged XCO<sub>2</sub> VMR.

Figure 4.6 (blue line) shows the offset between gridded and ungridded monthly averaged FSI for the Siberian 2004 region (60° to 140° west and 50° to 70° north). The high degree of smoothing in the TM3 CO<sub>2</sub> field makes the differences in its two alternate monthly averages almost insignificant and as a result the blue line shows the difference in the averaging systems. Although this particular example does not show a definitive example of bias in the two systems, there is a negative offset in the later half of the year. The gridding method was subsequently adopted for every monthly averaging comparison ([Barkley, 2007; Barkley et al., 2006b,c, 2007]).

An alternate monthly average which has not been adopted in the analysis methods can be used to correct for differences in weighting of measurements by latitude band. In a simple example, the Siberian scene has been split into “North” (60° to 70° north) and “South” (50° to 60° north) scenes (see Figure 4.7). Using the gridding method, a monthly average is obtained for both the “North” and “South” scenes plus the number of active grid squares in both. A combined scene average is made by taking an average of the two scenes weighted according to the split in land

area of the two scenes. The difference between the combined mean (weighted) and original gridded mean (unweighted) is shown by the red line (left hand scale axis) in Figure 4.6. This difference depends on both the difference between the monthly averaged “North” and “South” scene VMR (black line using right hand scale axis) and the “North” vs “South” ratio:

$$NorthSouthRatio = \left[ \frac{ActiveNorth}{ActiveSouth} \right] \div \left[ \frac{AreaNorth}{AreaSouth} \right] \quad (4.1)$$

There is a clear difference between the monthly averaged VMR (black line), with the “South” scene typically having a higher measured CO<sub>2</sub> concentration. In the month of August the red line is unaffected as the “North” vs “South” ratio is approximately equal to one. In the month of August, there are unusually more measurements in the “North” than the “South”. The general negative trend in the red line raises questions about whether the unweighted gridding system is the most appropriate method. It has already been suggested in Section 1.4 and in Figure 2.1 that there is likely to be a significant latitudinal gradient in the CO<sub>2</sub> field. This question remains unresolved in this thesis.

#### 4.2.2 FSI vs ground stations

Ground stations in the Global View network [ESRL, 2008b] measure very high precision CO<sub>2</sub> concentrations at point locations which are often based far from local sources. Ordinarily, a greater CO<sub>2</sub> seasonal cycle amplitude would be expected for ground stations (at low altitudes) than for satellite measurements [Hewitt et al., 2006] [Barkley et al., 2007], since the flux behaviour occurs at the surface. However, the necessity to locate the stations away from sources and often on mountain tops, means they are probably more representative of the planetary boundary layer.

Obtaining data for the stations was difficult, requests had to be made for individual stations, many were not operating in the years of interest and there were gaps in some of the data. Typically, increasing the size of the latitudinal and longitudinal limits increased the correlation coefficient and also decreased the size of the seasonal cycle amplitude (see Table 4.2), both should be considered as improved FSI precision owing to the increase in measurement numbers (thus decrease in measurement noise).



Table 4.2. Summary of the in-situ ground based comparison with FSI over Western Europe and North America [Hewitt et al., 2006]. The seasonal cycle amplitudes (SCA) are given for both the ground stations and SCIAMACHY/FSI observations within 1° and 3° latitude and longitude of the respective stations. The correlation between the monthly averaged time series is also given.

Station	SCA [ppmv]			Correlation	
	ground	FSI 1°	FSI 3°	1°	3°
Begur, Spain	8.0	14.4	12.7	0.16	0.34
Monte Cimone, Italy	11.5	23.1	9.5	0.35	0.56
Plateau Rosa, Italy	9.6	28.2	13.4	0.54	0.48
Deuselbach, Germany	15.8	25.2	17.1	0.60	0.85
Neuglobsow, Germany	20.4	28.8	25.7	0.50	0.65
Zugspitze, Germany	10.4	37.0	16.3	0.30	0.61
Niwot Ridge, Colorado, USA	8.9	17.8	8.0	0.79	0.90
Southern Great Plains, Oklahoma, USA	12.8	6.3	6.4	0.64	0.56
Wendover, Utah, USA	7.3	20.5	19.8	0.77	0.75

The frequency of measurements differs from station to station, some take samples hourly others daily or weekly. The monthly average CO<sub>2</sub> was calculated for all of the measurements for that month for each of the stations. Since some stations make measurements at certain times of day whilst others are averaged over 24 hours they are representative of different parts of the diurnal cycle. For all the stations SCIAMACHY/FSI samples a lower average CO<sub>2</sub> but the magnitude of this difference varies from station to station, and the differences in diurnal sampling between the stations will play a role in this variation as well as the different latitude bands for background CO<sub>2</sub> and any regional fluxes.

Despite the problems described above, one surprising result was how well the anomaly of monthly averaged to yearly averaged FSI and ground station plots matched each other (see Figure 4.8). For Niwot Ridge, Colorado as well as other ground stations (Table 4.2) the magnitude of the seasonal cycle amplitude match very closely in the 3° limit case and timings of the minimum and maximum months also match. It would seem then, that SCIAMACHY/FSI is indeed sensitive to the near-surface variations in atmospheric CO<sub>2</sub>.

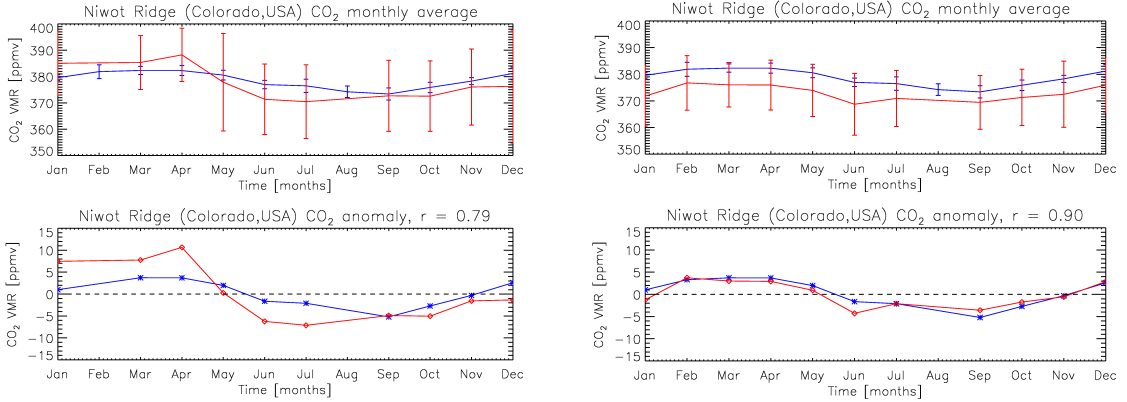


FIGURE 4.8. Time series of FSI (red) and Niwot Ridge ground station (blue) monthly averaged  $\text{CO}_2$  (top) and annual anomaly (bottom) for 2003 [Hewitt et al., 2006]. The left hand side has monthly averaged FSI  $\text{CO}_2$  from within  $1^\circ$  of the Niwot Ridge, Colorado ground station, whilst on the right monthly averaged FSI  $\text{CO}_2$  was taken from measurements within a  $3^\circ$  limit of the station. The error bars give the  $1\sigma$  standard deviation on the monthly mean.

### 4.3 Comparison with vegetation indices

The basis of this research project began with the two plots (Figure 4.9) and the question of whether the features in the atmospheric  $\text{CO}_2$  field could be explained by differences in the fluxes of the various vegetation types. “The transition from low  $\text{CO}_2$  VMRs along the Canadian Shield and the eastern coast to the higher values found over the mid-western USA, corresponds to a change in vegetation type from evergreen needle leaf, mixed and deciduous broadleaf forests to land covered by crops and large grass plains” [Barkley, 2007], is indicative of the initial interest in this problem. It seems more likely that the similarity of the two fields is an illusion as discussed in Section 4.2.1.2, and the actual cause is a gradual uptake of  $\text{CO}_2$  from west to east in the direction of the prevailing winds. The desire to relate the  $\text{CO}_2$  field to the vegetation type formed much of the early analysis of FSI, and led indirectly to the NAME inversion methods discussed in the next chapter.

#### 4.3.1 Carbon Flux Calculations

It was initially assumed that the flux signatures that were observable in the FSI  $\text{CO}_2$  field were caused by very localised fluxes. To begin with, it was of interest to investigate how FSI could be used to calculate flux magnitudes if carbon fluxes and their corresponding effect on the  $\text{CO}_2$  field occurred in the same spatial region, or to put it another way if there were no wind fields and air stayed in the same place.

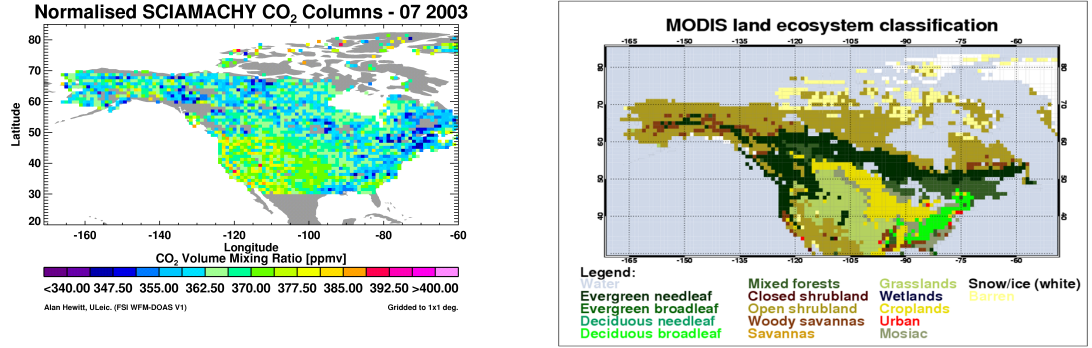


FIGURE 4.9. SCIAMACHY/FSI monthly averaged  $\text{CO}_2$   $1^\circ$  grid squares over North America for July 2003 (left) and the MODIS land cover map (right). The spatial similarities between the  $\text{CO}_2$  field and the vegetation type in these two plots may be misleading (see Section 4.2.1.2). It seems more likely that there is gradual draw-down from west to east in the direction of the prevailing winds.

Treating the “flux  $\text{CO}_2$ ” and the “transported (background)  $\text{CO}_2$ ” separately was the first step in developing the forward and inverse methods with NAME (see the giant tube thought experiment in Section 6.2.2.2).

Taking another idea from this chapter, that of active grid squares, led to a very simple method of calculating carbon fluxes. For a  $1^\circ$  by  $1^\circ$  grid square to be active, there has to be a minimum number of FSI measurements in that square for two consecutive months (or another appropriate time frame). The mass flux of carbon in each grid square can be calculated by dividing the difference of the mass of carbon in the grid square from May to June,  $\Delta C_i$ , by the area of the grid box,  $A_i$ . The area of the grid box is calculated using spherical coordinates:

$$A_i = \int_{\phi_{\min}}^{\phi_{\max}} \int_{\theta_{\min}}^{\theta_{\max}} R_E \sin \theta d\theta d\phi \quad (4.2)$$

and the mass of carbon is calculated from:

$$C_i = \text{VMR}_i \times 10^{-6} \times \left( \frac{44.0 \text{ grams mol}^{-1}}{28.96 \text{ grams mol}^{-1}} \right) \times \left( \frac{12.0 \text{ grams mol}^{-1}}{44.0 \text{ grams mol}^{-1}} \right) \times \left( \frac{P_i \times A_i}{g} \right) \quad (4.3)$$

where  $\text{VMR}_i$  is the  $\text{CO}_2$  volume mixing ratio (ppmv),  $P_i$  is the average *a priori* surface pressure (in Pascals),  $g$  is gravitational acceleration at the Earth’s surface and  $R_E$  is the radius of the Earth. The total mass flux for a scene can be calculated

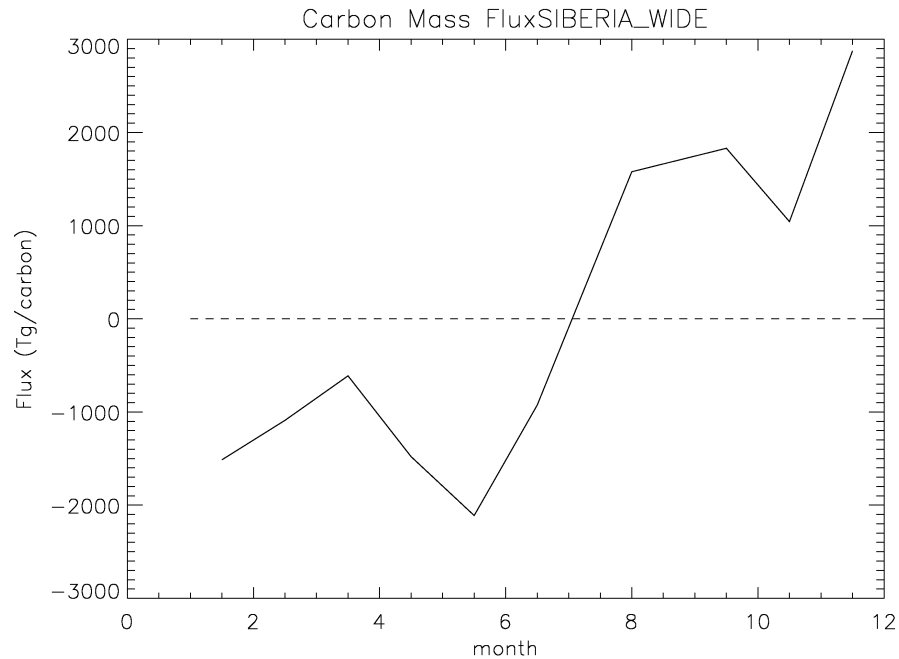


FIGURE 4.10. Simple carbon mass flux time series for Siberia 2003. Each point represents the difference in atmospheric mass of carbon (in the  $\text{CO}_2$  molecule) between consecutive months (method described in section 4.3.1)

by multiplying the flux per unit area by the total area of the scene (see Figure 4.10). With the incorrect assumption that transport plays no part in the story, it is inappropriate to sum together the whole year to obtain an annual carbon flux budget for the Siberian (or other region) in question.

Because the MODIS land classification is on the same  $1^\circ$  by  $1^\circ$  grid scale, it is possible to use the same method to break down surface fluxes by surface type. Figure 4.11 shows the Siberian scene broken down into 4 dominant land types: Mixed Forest (red), Evergreen Needleleaf (green), Cropland (yellow), and Deciduous Needleleaf (blue) plotted in units grams per square metre this time.

Both the whole scene and vegetation type carbon flux plots (Figures 4.10 and 4.11) show seasonality in the uptake of atmospheric  $\text{CO}_2$ . Since the measured changes are likely to be the result of seasonality in the background  $\text{CO}_2$  field, the results of the two plots probably say more about carbon fluxes over the northern hemisphere than they do about Siberia.

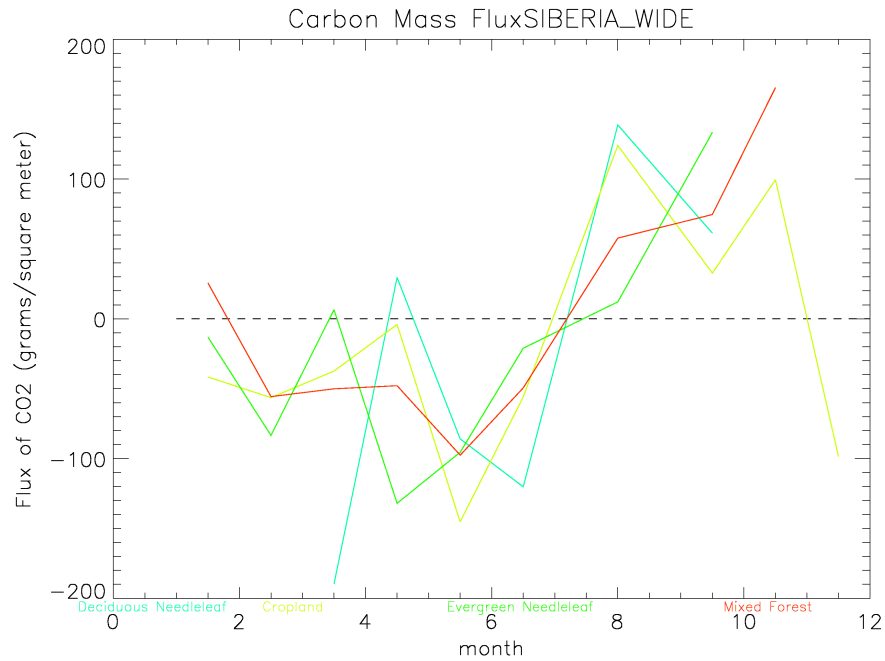


FIGURE 4.11. Simple carbon mass flux time series for Siberian land types 2003. Each point represents the difference in atmospheric mass of carbon (in the  $\text{CO}_2$  molecule) between consecutive months (method described in section 4.3.1)

### 4.3.2 Comparisons with MTCI

The question arises as to whether the variation in the atmospheric  $\text{CO}_2$  field could be related to other remotely sensed indicators of plant growth, such as vegetation greening, and whether different vegetation types are responsible for these variations. A collaborative effort was made with Jadu Dash of the School of Geography, University of Southampton, to investigate the relationship between variations in the atmospheric  $\text{CO}_2$  field and variations in the measured MTCI (MERIS terrestrial chlorophyll index [Dash and Curran, 2004]).

SCIAMACHY/FSI monthly averaged  $\text{CO}_2$  was gridded to  $1^\circ$  resolution using a maximum 3% fitting error filter. Monthly averaged MTCI was also gridded at the same resolution for 2003. A few selected latitude bands were chosen for the comparison for the different land types, depending on the availability of the land class at the particular band.

The grassland plots show a strong negative correlation between the MTCI and

the CO<sub>2</sub> for the individual latitudinal bands (see Figure 4.12). The timings of the peaks match very closely in both the northern (top) and southern (bottom) hemispheres, and there are clear differences between adjacent latitude bands which probably indicate the strong response of SCIAMACHY/FSI to near surface CO<sub>2</sub> fluxes.

The woodland plots show a strong negative correlation between the MTCI and the CO<sub>2</sub> for the individual latitudinal bands (see Figure 4.13). The timings of the peaks match very closely particularly for the northern (top) hemisphere, and there are clear differences between adjacent latitude bands which again probably indicate the strong response of SCIAMACHY/FSI to near surface CO<sub>2</sub> fluxes. Similar trends are observed for deciduous broadleaf and mixed forests (see Figure 4.14).

Next, each 1° grid square in North America was classified according to vegetation type. The maximum monthly averaged MTCI for each grid square is compared with the minimum CO<sub>2</sub> of that grid square. There is a negative relationship between maximum MTCI and minimum FSI CO<sub>2</sub>, and this relationship also depends on the type of land cover (see Figures 4.15 and 4.16). The strength of this relationship may also depend upon latitude (see Figure 4.15) something only investigated for grassland as there were not enough points to break the scene up for woodland and agricultural land. The date of the minimum FSI CO<sub>2</sub> is 1-2 months later than the date of the maximum MTCI.

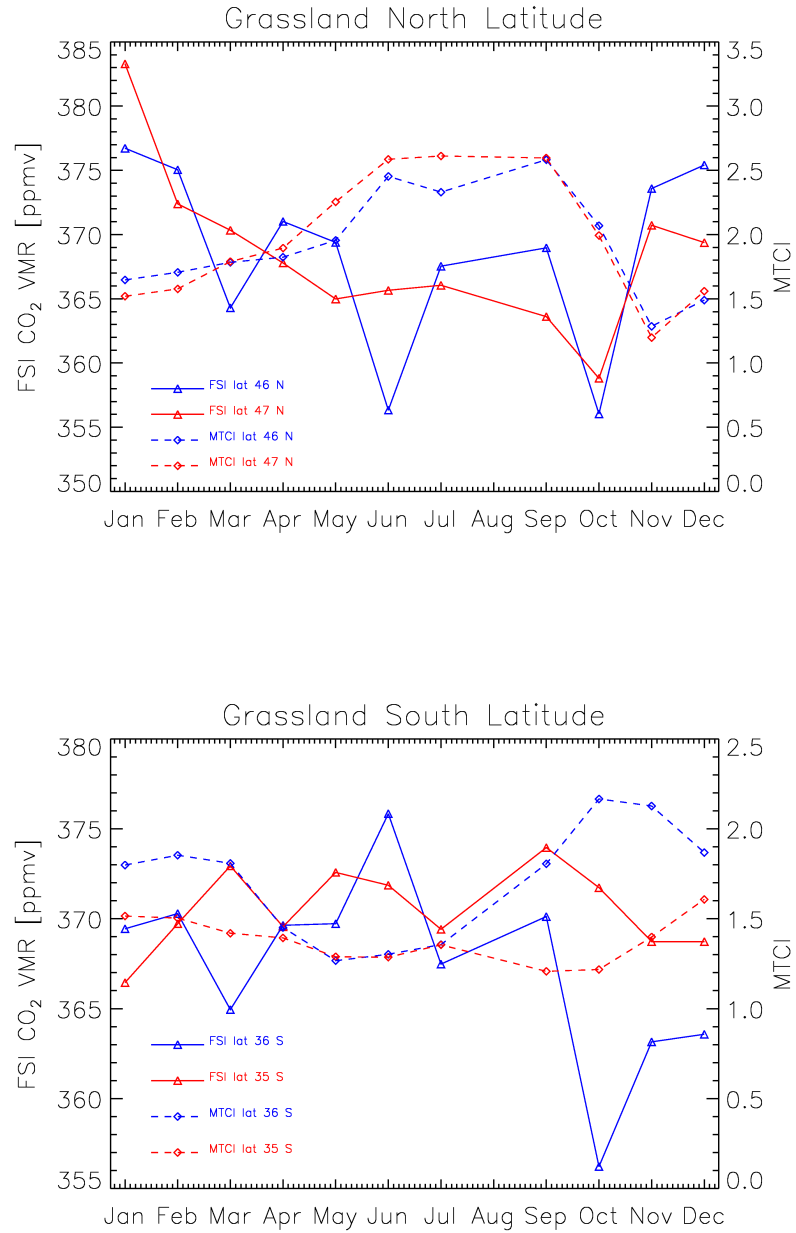


FIGURE 4.12. Monthly averaged FSI CO<sub>2</sub> versus MTCI over grasslands for selected latitude bands (across the whole globe) in 2003 [Dash and Hewitt, n.d.]. There is a strong negative correlation between the MTCI and the CO<sub>2</sub> for the latitudinal bands. The timings of the peaks match very closely in both the northern (top) and southern (bottom) hemispheres, and there are clear differences between adjacent latitude bands which probably indicates the strong response of SCIAMACHY/FSI to near surface CO<sub>2</sub> fluxes.

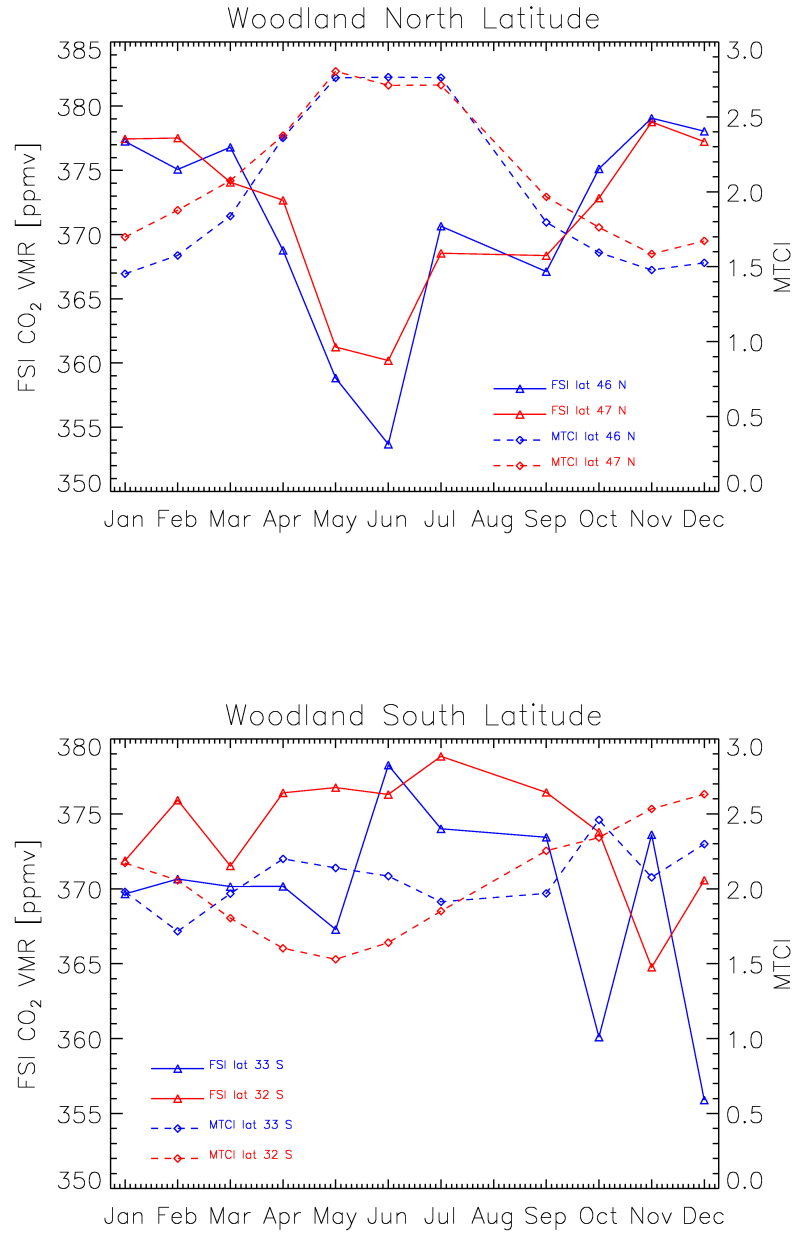


FIGURE 4.13. Monthly averaged FSI CO<sub>2</sub> versus MTCI over woodlands for selected latitude bands (across the whole globe) in 2003 [Dash and Hewitt, n.d.]. There is a strong negative correlation between the MTCI and the CO<sub>2</sub> for the latitudinal bands. The timings of the peaks match very closely particularly for the northern (top) hemisphere, and there are obvious differences between adjacent latitude bands which probably indicates the strong response of SCIAMACHY/FSI to near surface CO<sub>2</sub> fluxes.



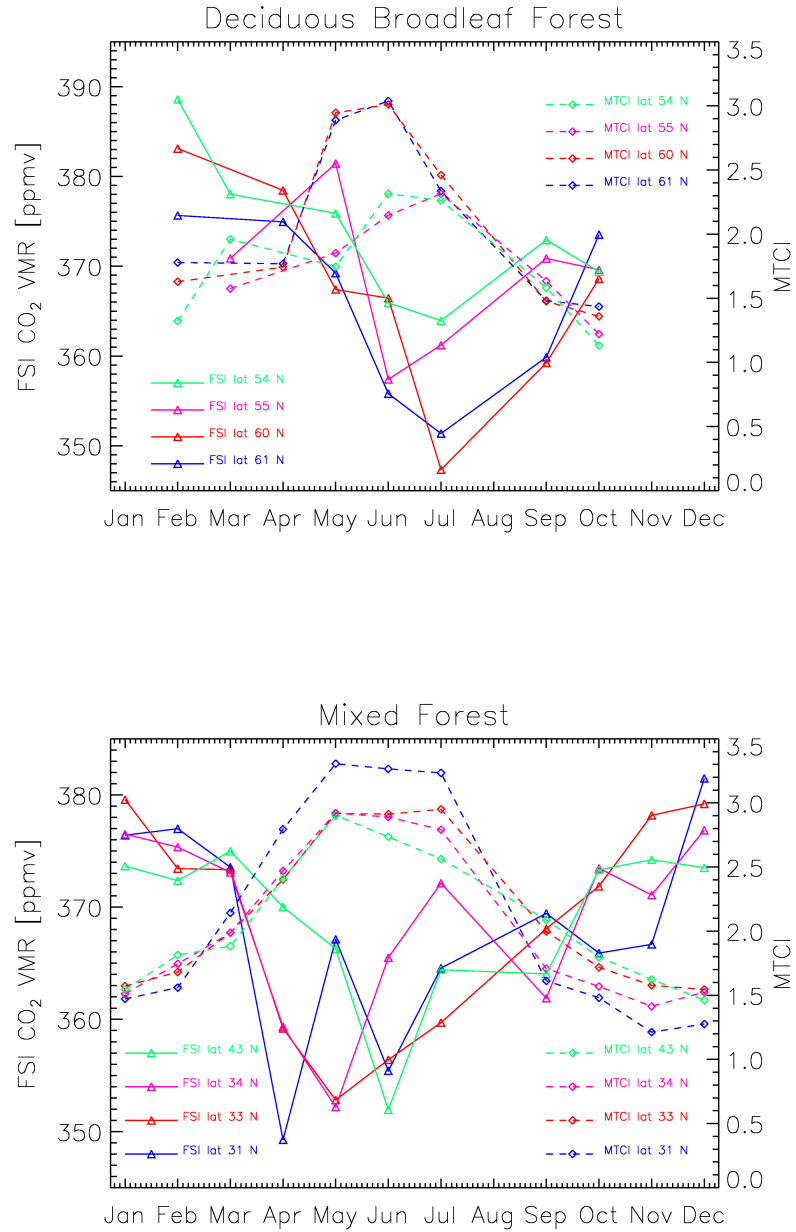


FIGURE 4.14. Monthly averaged FSI CO<sub>2</sub> versus MTCI over deciduous broadleaf (top) and mixed forests (bottom) for selected latitude bands (across the whole globe) in 2003 [Dash and Hewitt, n.d.]. Again there are strong negative correlations between the MTCI and the CO<sub>2</sub> for the latitudinal bands.

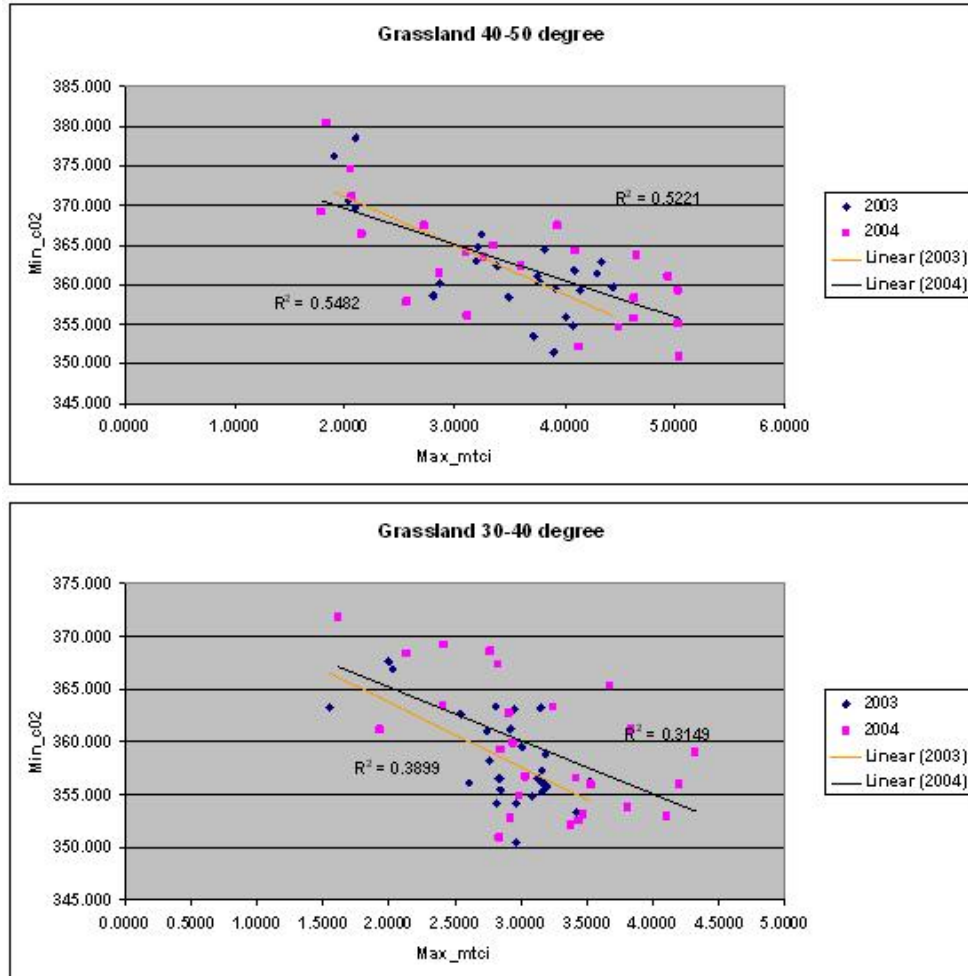


FIGURE 4.15. Maximum MTCI versus minimum FSI CO<sub>2</sub> for 1° by 1° grid squares over North American grassland in the range 40°-50° North (top) and 30°-40° North (bottom) in 2003 and 2004 [Dash and Hewitt, n.d.].

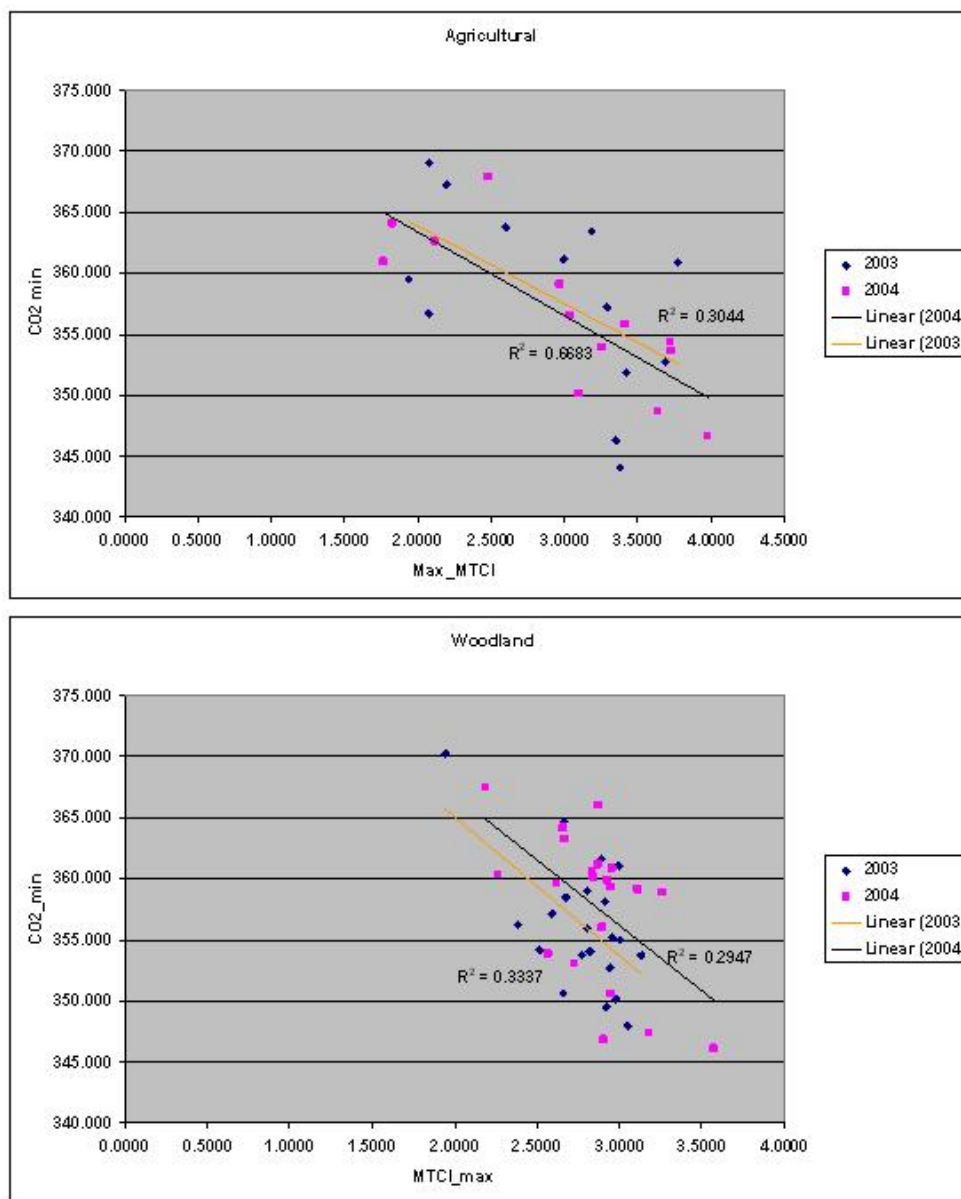


FIGURE 4.16. Maximum MTCI versus minimum FSI CO<sub>2</sub> for 1° by 1° grid squares over North American agricultural land (top) and woodland (bottom) in 2003 and 2004 [Dash and Hewitt, n.d.].

## 4.4 Summary

Miller et al. [2007] has shown that satellite retrieved atmospheric XCO<sub>2</sub> columns with bias-free precisions in the range of 1 - 10 ppmv (0.3 - 3%) will reduce uncertainties in CO<sub>2</sub> sources and sinks owing to uniform and dense global sampling. Currently, FSI retrieved XCO<sub>2</sub> columns are hovering at the upper limit of this precision range [Barkley et al., 2006a], and are subject to a (currently) unexplained negative bias in the column average.

There are some sampling biases that are unique to the SCIAMACHY instrument. Its large “footprint” size result in many more measurements being excluded as cloud contaminated pixels. The absence of a sun-glint mode results in poor quality (low signal-to-noise) retrievals over oceans and snow/ice covered regions.

There is generally good agreement between the FSI retrieved monthly averaged CO<sub>2</sub> field and the CO<sub>2</sub> fields produced by CarbonTracker and TM3, which evidence of large-scale carbon fluxes present over North America in all three sources. There are some discrepancies in the seasonal cycle amplitudes (SCAs) between FSI retrieved CO<sub>2</sub> and those of the other models, with the SCA of FSI retrievals being roughly double those of most modelled columns. This is probably a result of the near surface bias in the averaging kernel of FSI retrievals.

Despite the early promise from the monthly averaged gridded CO<sub>2</sub> maps (see Figure 4.9), FSI-retrieved CO<sub>2</sub> columns are not yet adequate for use in an inversion scheme. By applying a constant offset factor to correct for the negative bias in FSI retrieved XCO<sub>2</sub> columns, this data could already be used in a forward modelling scheme to validate modelled carbon flux magnitudes. However, it is unlikely to be used whilst the cause of the negative offset and greater seasonal cycle amplitude remain unexplained.

The differences between the SCIAMACHY/FSI retrieved CO<sub>2</sub> field and other sources of modelled or measured CO<sub>2</sub> are summarised in Table 4.3.

Table 4.3. Summary of the validation of SCIAMACHY/FSI-retrieved CO<sub>2</sub> fields against other sources of modelled and measured CO<sub>2</sub>. The second column shows the measurement uncertainty of the other data source. The third column shows the offset bias of FSI w.r.t. the other source of CO<sub>2</sub>. The fourth column shows the magnitude of the seasonal cycle amplitude of FSI w.r.t. the other source of CO<sub>2</sub>.

Validation Source	uncertainty [%]	Bias [%]	SCA factor	Comments
Single FSI CO <sub>2</sub> retrieval	2 to 3	N/A	N/A	Fitting error of individual model spectra to the instrument measured spectra [Barkley et al., 2006a].
TM3 using BIOME-BGC biospheric fluxes	N/A	-2.0	2 to 3	Similar spatial features are observed in both CO <sub>2</sub> fields. The seasonal cycle of both data sets peak at the same time, although the magnitude of the FSI SCA is much greater [Barkley et al., 2006c].
AIRS XCO <sub>2</sub> [Engelen et al., 2004]	<1	variable	3	The seasonal cycle of FSI peaks three months before that of AIRS CO <sub>2</sub> and is a factor of 3 larger (both consequences of the greater surface sensitivity of FSI). Similar spatial features are observed in both CO <sub>2</sub> fields [Barkley et al., 2006b].
FTIR Egbert Canada	0.1	-4.1	2	FSI does not detect the high concentrations of CO <sub>2</sub> detected at Egbert possibly a consequence of sampling bias (see Section 4.1) [Barkley et al., 2006c].
FTIR Park Falls	0.1	-1.1	2	Park Falls has cleaner air than Egbert which could help explain the difference in offset bias [Barkley et al., 2007].
Ground <i>in-situ</i> network [ESRL, 2008b]	negligible	-2.0	1.3	Large co-location limits are used to smooth out the noise of individual FSI retrievals. The close match in the SCA between FSI and <i>in-situ</i> measurements is evidence of the near surface sensitivity of the FSI retrieval [Hewitt et al., 2006].

# Chapter 5

## NAME background chapter

A description of the carbon cycle and the exchange of carbon between the terrestrial biosphere and the atmosphere has been described in Chapter 1. Also described earlier in this thesis is physical geometry of the column measurement of atmospheric CO<sub>2</sub> both generically and specifically related to the SCIAMACHY instrument. The limitations of FSI retrievals of atmospheric CO<sub>2</sub> concentrations have also been described. Using this information, a pre-existing Lagrangian dispersion model (NAME) was manipulated in such a way as to represent the real world situation of column measurements of atmospheric CO<sub>2</sub> being used to investigate the exchange of carbon between the terrestrial biosphere and the atmosphere.

The creation of this new tool has allowed the surface exchange of carbon to be investigated in two distinct ways. The first is a forward modelling approach in which pre-existing surface fluxes of carbon are compared with the measured atmospheric CO<sub>2</sub>, for the purpose of validating: the method, the measurement, the initialisation (or background concentration) or the pre-existing surface fluxes. The second method is an inverse modelling approach in which the atmospheric measurements are used to investigate the magnitude of the surface exchange of carbon between the atmosphere and the terrestrial biosphere.

The majority of proposed studies for investigating surface exchange of carbon using retrieved OCO-like or GOSAT CO<sub>2</sub> columns will use a Eulerian approach. The advantages of a Lagrangian model are the ability to perform regional high resolution studies, that can be tailored to the specific area of interest, such as the European heat wave in 2003 [Ciais et al., 2005]. The model can easily be adapted to work at any spatial or temporal scale. It is also easy to couple the fluxes to a Lagrangian model. When used in a forward model run, different emission invento-

ries can be used without the need to re-run the model (assuming no chemistry is performed). The problem of initialisation is also something that needs to be solved if there is any benefit from the high resolution fluxes.

There are other differences between the Eulerian and Lagrangian frameworks that are less relevant to this research project. For example, it is relatively more difficult to encompass chemical interactions between separate air parcels in the Lagrangian framework. The physically meaningful air trajectories of the Lagrangian framework are more effective in relating emissions from source to receptor. Vertical transport processes (very important in tropical regions) are better incorporated by Eulerian models.

North America has been chosen as the test case in this study, as this is probably the best understood region in terms of carbon cycle dynamics, allowing comparison with many alternative carbon flux budgets.

Chapter 5 gives a basic description of the functioning of the NAME model. Chapter 6 describes how the author adapted the NAME model to investigate the carbon cycle.

## **5.1 Numerical Atmospheric-dispersion Modelling Environment**

### **5.1.1 Origins**

As a response to the Chernobyl nuclear accident, on the 26<sup>th</sup> April 1986, the UK government sanctioned the Met Office with developing an emergency response modelling capability, that would provide detailed predictions of the transport and deposition of radioactive materials in the event of a similar incident in the future. Knowing the path of this material before it arrives would allow for infrastructure to be protected, and if necessary people to be evacuated. It could also allow for cloud seeding, to induce rain over ocean, washing out some of the radioisotopes there rather than over land.

The UK Met Office Nuclear Accident Model (NAME), an operational long-range dispersion model, was in use by 1988 [Maryon et al., 1991]. Fortunately, it has not been needed to perform its primary task, since it became operational. Before

this time, dispersion modelling was being discussed for its potential to investigate atmospheric processes, but was limited by its demands on computing resources.

As computers became faster, the potential of dispersion modelling became realised. The second generation operational dispersion NAME model, repackaged as “the Numerical Atmospheric-dispersion Modelling Environment”, is a Lagrangian particle trajectory model [Ryall and Maryon, 1998]. It’s most recent version, NAME III [Jones et al., 2004], is the version used in this study.

### **5.1.2 NAME run set up**

#### **5.1.2.1 Forward and backwards dispersion**

NAME is a Lagrangian type model, it operates by following the trajectories of a number of “tracer particles” which are meant to represent an air mass parcel. The particles are carried passively by the wind fields, plus additional transport options (such as turbulence and meander) defined by the user. NAME can operate in forward dispersion mode, which is used for studying the transport of polluted air, from a point or source region. NAME can easily be set up to run in reverse dispersion mode (where the air mass travels backwards in time) because the direction of the wind fields can be reversed, identifying the origin of the air mass. This is used to attribute the source strengths and locations of the trace gas. Either way, a run length is defined by the user, after which the model run ends and output is produced.

#### **5.1.2.2 Tracer particles**

Tracer particles represent the contribution of a “source” to the “receptor” where they are evaluated.

When run in forward dispersion mode, the particles are not meant to represent “lumps” of trace gas, but rather the trace gas density. As time progresses, it would be expected that the particles become further apart, representing the concentration of the trace gas becoming less dense. The beginning point in time of the model run, corresponds with the emission of a plume of trace gas in the real world. Likewise the beginning point in space of the real world trace gas plume is mirrored in the NAME model with the distribution of tracer particles. In the forward case, the particles would be “released” in quite a small volume. When considering very short-range



transport (city scale), the particle model does not perform very well. The third version of NAME includes a puff particle approach, not used in this study which can better resolve this.

Assuming that there was only one source of the particular trace gas of interest, a measurement of the plume concentration a couple of days downwind of the trace gas source would be made. If the trace gas is inert, it will not have decayed along the path, and its concentration will relate to the initial concentration and the amount of dispersion of the air in reaching the measurement site. NAME has some built in functions to allow it to simulate radiative decay and atmospheric chemistry if the gas is not inert. In reality the air at the measurement site would contain the same trace gas released at many different regions. This ideal scenario only exists in man made experiments using atmospherically rare trace gases [Bowne and Londergan, 1983]. This forms the basis of air quality work with the NAME model, where emission grids are used in the model.

When run in reverse dispersion mode, particles represent the density of released air as they travel in space and time. The volume of the air mass in question will contain air that has arrived from different directions (essentially the same as air being dispersed, but in reversed time). Like the situation just described above, the air will have encountered different regions of emission of the trace gas (assuming that it is a primary product), of various degrees of strength.

As the initial distribution of release particles can be quite compact in the forward model run (in the case of a chimney) or quite diffuse (in the case of a city), so too can the initial distribution of the “measurement” particles (in the case of a flask (point) or column measurement). In the model runs used in this thesis, the distribution of individual particles in the release column in the horizontal (latitude and longitude) direction is random, the vertical distribution of particles is also random but weighted with respect to the density of air as a function of height.

For short-lived atmospheric gases, the tracer particles represent the air that has been sampled, rather than the trace gas. At some point backwards in time, the air mass will have not yet experienced the trace gas source and so will be clean. For longer lived trace gases, natural or man made, there is no clean air. As the particles are followed further back in time, the diffusion is so great that it is no longer possible to identify where the air came from. The measured air will contain background concentrations, which are essentially from sources that can no longer be attributed by diffusion modelling, with the product of more recent fluxes on top.

The number of particles released is specified by the user. By releasing many particles, a Gaussian distribution of particle trajectories is obtained. If the volume of the release box (the volume that the particles are initially released from) is large or the run time is long, there is likely to be a bifurcation of the likely trajectory of air from that location. Keeping this number large will help weight the trajectories when investigating sources, or in the forward case could be used to predict the severity of a pollution event for different areas downwind. On the other hand, the model works by following each of the particles, if this number is too large the model will have to perform considerably more calculations, which will increase the amount of time needed to run the model. Experience has shown that there is little benefit in running more than a particular number of particles (this number varies with the domain size and run length) , as the result of doubling this number is practically the same. In our setup, the calculation part of the NAME run takes about the same amount of time as the reading of the meteorological files, so little can be gained by reducing the particle number further.

#### **5.1.2.3 Meteorological data**

NAME III is capable of using multiple sources of meteorological data, which would allow high resolution meteorological data to be nested onto a much larger scale region with less detailed meteorological data. This would be very useful in investigating urban pollution to the surrounding countryside, where the initial pollution is slowed down by street canyon effects. This study does not take advantage of this possibility, because high resolution data is not produced for North America. All meteorological fields are taken from the UK Met Office weather prediction model, the Unified Model (UM), which provided analysis fields with global coverage at 60 kilometre horizontal resolution and 50 vertical levels, for the period considered [Cullen, 1993].

### **5.1.3 Operation of NAME**

With the notable exception of particle mass, almost every aspect of the running of NAME functions in the same way in forwards or backwards mode.

Each particle starts with a “mass” assigned to it and coordinates within the release volume. The model run contains a (release) start time, a release end time, and a run end time. Particles can be released quickly or slowly depending on the

real life situation to be mimicked, pollution from an industrial fire may be released for many hours until it is extinguished whereas pollution from an explosion may be released almost instantly. If the start time is after the end time, the model runs backwards. A finite domain may be specified, unless the user desires to investigate transport around the globe. Certain particle behaviours are also specified before the run, these can affect the particles motion such as turbulence, or its “mass” such as chemistry.

In forward mode, the particles are much less of an abstract idea, in that they are meant to represent the trace gas density. Many trace gases (not including CO<sub>2</sub>, which reacts only with the surface) react in the atmosphere in a variety of ways that can be simulated by the model. The most obvious of these is radiative decay, rather than remove a certain number of particles for each time step that the model progresses (not a practical solution), the particle mass exponentially decays as a function of time. In this way, the idea of trace gas density is preserved.

Particles can also react chemically, this is because NAME allows many different chemical particles to be released at around the same time and not necessarily from the same place. Around 40 predominant atmospheric species can react together in NAME, based on the reaction rates of the Met Office STOCHEM model [Collins et al., 1997]. Trace gases can physically be removed from the atmosphere by wet deposition (e.g. SO<sub>2</sub> into acid rain) or dry deposition, which again involves a change in particle mass. Particles can carry masses of multiple emitted species, thus secondary products gain mass as the primary mass falls. In backwards time, the result of any of these behaviours could potentially be modelled, however for some cases this is more difficult than forward modelling.

The particles move (and change mass) discreetly, the model run is divided into a number of time steps. Shorter time steps are needed if very short range transport is being investigated, however this adds to the run time of the model. For longer range, 15-minute (900 second) time steps are more appropriate. Because of the modular nature of the NAME code, the particle trajectories can be obtained using either a homogeneous diffusion scheme for long-range transport or a complicated (and computationally expensive) inhomogeneous velocity memory scheme for near source applications.

Particles are advected each time step using a random walk scheme:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + [\mathbf{u}(\mathbf{x}_t) + \mathbf{u}'(\mathbf{x}_t) + \mathbf{u}_l'(\mathbf{x}_t)] \Delta t \quad (5.1)$$

where  $\mathbf{x}_t$  is the particle position vector at time  $t$ ,  $\mathbf{u}(\mathbf{x}_t)$  the mean wind velocity vector,  $\mathbf{u}'(\mathbf{x}_t)$  the turbulent velocity vector for small scale turbulent motions,  $\mathbf{u}_l'(\mathbf{x}_t)$  the velocity vector for low-frequency horizontal meandering, and  $\Delta t$  the time step [Ryall and Maryon, 1998]. Ryall and Maryon describe in further detail the additional transport options built into NAME. Convective processes are the most difficult to model correctly, which is unfortunate as this is the dominant part of circulation in tropical regions, where satellites have perhaps the most to tell about the carbon cycle. The bolted-on convection model in NAME (which was left off in this research project) is activated when a convective cloud is present with a depth greater than  $\eta = 0.1$  (approximately 100 millibars) and a base below  $\eta = 0.8$  (approximately 800 millibars). In this situation, all particles between the ground and the cloud top are considered to either be in updraught or downdraught, with the proportion of updraught and downdraught particles and their vertical velocities being determined by their altitude and the convective cloud fraction,  $C_{mod}$ , taken from the unified model (further details are given by Maryon et al. [1999]).

Particles are followed until they leave the domain, the model will run until all the particles have left the domain or until the run time ends. For source estimation purposes, it is desirable for particles to leave the domain and so a suitably long run time should be chosen. The situation of all particles leaving the domain is incredibly rare, so it should be expected that the computer will need to run for at least as long as it takes to open and close all of the meteorological files. A balance needs to be struck between the two, clearly a larger domain will need a longer run time (as the particle has further to travel to reach the domain boundary). As a rule of thumb, it is assumed that when particles are released near the west boundary, they will be blown to the east boundary, take a typical free troposphere speed and the distance to the boundary to get a time; double this time to be safe to obtain a suitable maximum run time.

#### 5.1.4 Output of NAME runs

In principle, the output could contain the mass and coordinates of every released particle for every time step. However, this would be a very large output file. A library of common output parameters for the data exists, that can be specified in the NAME run file. Unusual output parameters could be specified with a modification of the NAME code.

The output parameters requested need to be chosen carefully, they should be

tailored to the way that the experiment works in real life. The model runs in this thesis are geared towards investigating surface fluxes of carbon dioxide. To investigate fluxes of other atmospheric gases consideration needs to be made to describing the measurement, the background and where the sources/sinks happen.

Typically, it would be useful to have a four-dimensional (latitude, longitude, height, time) grid, so that for every time step in the model, the number or mass of particles inside each grid box is added. The resolution of the four dimensions can be modified and it may only require one height or time (see Figure 5.1).

Knowledge about deposition of the trace gas may be desired, it may be useful to output meteorological parameters. Knowledge of where and when particles leave the scene will allow characterisation of the background.

### 5.1.5 Validation of NAME

The Kincaid field experiment [Bowne and Londergan, 1983], provided a perfect testing of the ability of the NAME model to represent short-range transport of atmospheric gases. Plumes of  $\text{SF}_6$  were released from the power plant stack, and this gas was measured downstream at distances of 500 metres to 50 kilometres away. In the study by Jones et al. [2004], it was shown that the improvements to NAME III have led to improved modelling of short-range transport. Although far from perfect, the correlation coefficient of modelled and measured concentrations is 0.47 for NAME III, and 75 % of all modelled concentrations are within a factor of two of the measurements.

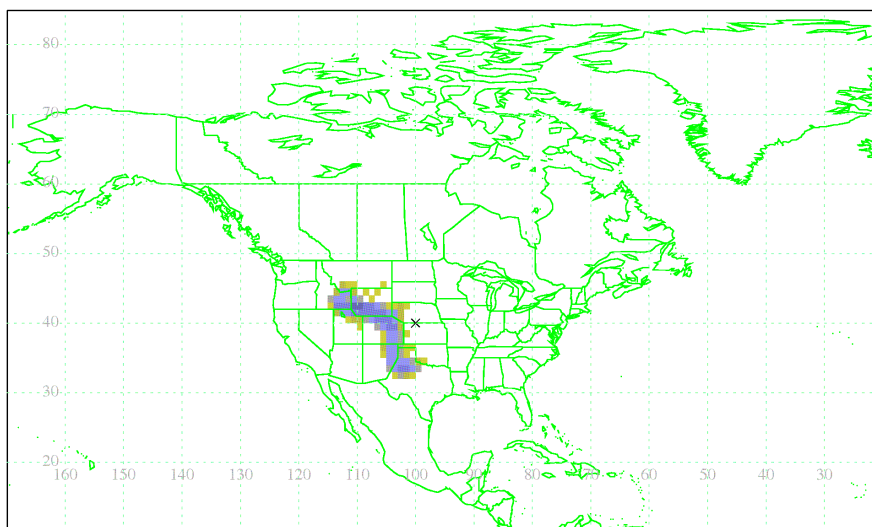
For longer range transport, the European Tracer Experiment (ETEX) carried out in 1994, provided real information that could be used to compare actual measurements with modelled concentrations. In their findings [Ryall and Maryon, 1998], the correlation coefficient of the long-range measured versus modelled concentrations are around 0.6 and vary a little depending on which of the extra functions of NAME (e.g K-distribution) were turned on. As with most other models, NAME overestimated the concentrations of the trace gas for the ETEX campaign.

# NAME III (version 4.3)

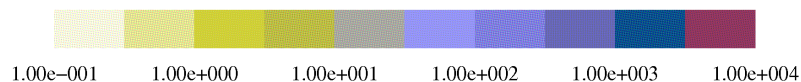
# Particles

From 0 – 1000m agl

0900Z 01/08/2003 to 1200Z 01/08/2003



Maximum value = 4.25e+002



Start of release: 1200UTC 01/08/2003  
End of release: 0900UTC 01/08/2003  
Release rate: 1.000000g/s  
Release location: 100.0000W 40.0000N  
Release height: 5000.000m agl +/- 5000.000m

Pollutant:  
Met data: NWP Flow.Global  
Run time: 1608UTC 21/02/2009  
Model run in backward mode

Met Office (GMR) Crown copyright

FIGURE 5.1. Output from NAME. The plot shows the density of particles within one kilometre of the surface one day after being released from a 1° by 1° column centered on the cross.

## Chapter 6

# Development of a method to investigate surface fluxes of carbon dioxide

Chapter 6 describes the adaptation of the NAME model (so it can be coupled with satellite-retrieved atmospheric CO<sub>2</sub> columns and carbon flux models). This setup was designed to include a parameter to assign the surface flux influence to the measurement (or particle release) column. The NAME model was set up to output the origin of the release particles in the domain, in order to initialise the background atmospheric CO<sub>2</sub> concentration. Two offline methods of investigating regional scale carbon fluxes were then developed to take advantage of the adapted NAME model. The forward modelling method can be used to compare the measured change in atmospheric CO<sub>2</sub> mass over the domain with the change in CO<sub>2</sub> mass resulting from coupling the NAME output with a carbon flux model. The linear regression inversion method uses satellite-retrieved atmospheric CO<sub>2</sub> columns and NAME output to obtain carbon flux magnitudes for the ecoregions present in the domain. These methodologies are synthetically tested in Chapter 7 and tested with real SCIAMACHY/FSI-retrieved XCO<sub>2</sub> columns in Chapter 8.

### 6.1 Author Developments

Incidental to the original goal of using SCIAMACHY/FSI retrieved CO<sub>2</sub> to investigate carbon fluxes, the author has made significant technical developments in the

application of Lagrangian modelling to CO<sub>2</sub> inversion. By adapting the NAME model to be compatible with whole column measurements of trace gas concentrations, the author has opened the door to others, who are now using these NAME specifications with satellite measurements of other gases (both long and short lived). This work was also the first to use any Lagrangian model to invert satellite measurements of atmospheric CO<sub>2</sub> and many original ideas described in this chapter were developed and applied by the author to be able to do this.

Because of the pioneering nature of the work, much of the technical description of the setup has been retained in this thesis for future reference. It has not been possible to separate the technical developments from the purely scientific developments without interrupting the logical flow, from the original NAME model to the inverse (and forward) modelling of satellite retrieved XCO<sub>2</sub>. The following sections will summarise the authors technical and scientific developments.

### 6.1.1 Technical developments

In Section 5.1.2.2, the author established the physical basis of the tracer particles. The physical basis is subtly different when the model is run forwards or backwards in time (in this research project all NAME runs were run backwards in time). The release particles represent the column of air measured by the satellite, not the particles of trace gas. This ensures continuity of the air parcel and thus surface influence for inversion or forward modelling purposes. It also allows negative carbon fluxes to be obtained.

In Section 6.2.2.1, the idea of a “surface influence height”, a technique for modelling the amount of time (and where) that a column of air had previously spent at the surface is developed. The author has also shown that the optimum extent of the surface influence is between the surface and the lower extent of the PBL.

The author gives simple reasons for running the NAME model in reverse time mode, for the purposes of forward and inverse modelling, in Section 6.2.2.3. The author also discusses the horizontal limits of the model release column and identifies that this should ideally be at least as large as the horizontal limits of the measurement ensemble, though not much greater, as the release column is more likely to bifurcate in its origin.

In Section 6.2.3, the author rules out the possibility of using a simple “source



to receptor” method of investigating carbon fluxes, showing that it is not possible to initialise the background. The author identifies the optimal set up as having a horizontal domain boundary greater than the horizontal extent of the continental region, with the boundary edges over ocean.

It was established that it would not be possible to use SCIAMACHY/FSI to initialise the background concentration of atmospheric  $\text{CO}_2$ , owing to the inability of SCIAMACHY to retrieve  $\text{CO}_2$  over ocean. The author suggests a method in which a single satellite instrument can obtain both the initial column and background concentrations of atmospheric  $\text{CO}_2$ , thus limiting the offset bias (see Figure 6.6).

In Section 6.2.4.1, the author developed a grid based method, “pseudo time”, to include an ensemble of satellite retrieved  $\text{XCO}_2$  columns into a forward or inverse modelling scheme. The author also describes an improved satellite overpass based method, “satellite time”, to include ensemble measurements in Section 6.2.4.2. These developments have allowed efficient running of NAME using HPC.

In Section 6.3.2.1, each of the calculations required to obtain the change in mass of atmospheric  $\text{CO}_2$  through the inversion of SCIAMACHY/FSI are shown. The calculations for forward modelling are similar. In Section 6.3.3.1 the author developed a method for testing the “freeze” initialisation method and, in Section 6.3.3.2, a method for testing the “edge box” initialisation method.

## 6.1.2 Scientific developments

The idea of investigating carbon fluxes using a measured change in the mass of satellite retrieved  $\text{XCO}_2$  has already been established in Section 4.3.1. In Section 6.2.1, it was decided to use NAME to identify where and how long this air mass had been in contact with the surface (and thus influenced by surface fluxes); as well as investigating the links between land cover types suggested by Barkley et al. [2006a]. The idea of dividing the surface contact into homogenous flux ecoregions is developed.

In Section 6.2.2.2, the author uses a thought experiment to show that it is not possible to obtain diurnal carbon flux information by inverting trace gas measured concentrations from a single sun-synchronous satellite and that only a diurnal average flux for each homogeneous ecoregion can be obtained.

In Section 6.2.3, the author describes the hypothetical “freeze” method for identifying the origin of background air into the scene, which could be used to initialise the background concentration of atmospheric CO<sub>2</sub>. It was established that modelled concentrations of atmospheric CO<sub>2</sub> from CarbonTracker could be used to initialise the background for the purposes of inverse modelling of SCIAMACHY/FSI XCO<sub>2</sub>, though this may need to correct for the offset bias between the two data sets.

In Section 6.2.3.1, the author develops the “edge box” method for identifying the origin of background air into the scene, which was used to initialise the background concentration of atmospheric CO<sub>2</sub>, owing to the technical challenges required to develop the superior “freeze” method. However, a test of a ‘freeze-like’ initialisation method was carried out in Section 7.1.

In Section 6.3.1, the author describes how an ensemble of measurements could be used to overcome the limitations of individual retrievals of SCIAMACHY/FSI XCO<sub>2</sub>.

In Section 6.3.2, the author identifies the strengths and weaknesses of using homogenous flux ecoregions in an inversion scheme. The length of the inversion period is also discussed. A suitable inversion method (based on linear regression) was designed which could incorporate the various factors (this was done offline).

## 6.2 Development of NAME method

The NAME model was manipulated so that it could be used with satellite column measurements to investigate surface fluxes of carbon. The information required from the NAME model output is the surface contact time and information about where particles leave the domain (to be used to initialise the background). This section details the method development, starting with a very simple description of carbon cycle behaviour then adding levels of complexity, to produce a tool capable of investigating the surface exchange of carbon.

### 6.2.1 The basics

In the FSI analysis chapter (Section 4.3.1), a very crude method for investigating fluxes is described in which the measured monthly mean mass of carbon dioxide in the air over a certain region is subtracted from the monthly mean mass over the

same region but for the previous month. The change in the mass of CO<sub>2</sub> in the column could tell us about the exchange with the surface below, except that the air (which measured in the previous month) will have moved.

The basic format for investigating the fluxes is similar, the satellite measures the mass of carbon dioxide in a column of air and then again after a period of time. The flux is worked out as **mass change** = **flux** × **time** × **area**, where the area is the amount of the column touching the ground (this can change significantly in unusual atmospheric conditions such as convective systems, see Section 6.2.2.4) and the time is the length of time that the column air touches the surface. It is difficult at first to separate out the flux signal from the tendency of the column of air to disperse. A tool used in some of the thought experiment cases (a giant hollow tube that prevents horizontal movement of air in the initial column) allowed separation of these behaviours.

The area of the column is proportional to the mass change in the column. It is clear then that SCIAMACHY/FSI should tell us about the mass of carbon dioxide in a column and NAME should tell us about the amount of time that column of air spends in contact with the surface. Flux across the domain varies both spatially and temporally. The “two measurement” approach works well for eddy flux measurement sites because the domain is small; only one flux value is needed for the entirety of the domain and the time between the two measurements is small compared to the diurnal flux cycle [Baldocchi et al., 1988].

This method will not work in the case of SCIAMACHY, because of its orbit. The minimum amount of time between two measurements of the same air is 24 hours. Even if the air has not moved, it will have encountered a significant range of temporal surface fluxes. It may have encountered some rather different spatial fluxes as well. The only exception are measurements very close to the North Pole (there is no vegetation at the South Pole to investigate), where there is a little overlap between the SCIAMACHY swaths which are 100 minutes apart. A geostationary satellite or a chain of instruments following in sequence would allow this to happen.

For a single polar orbiting satellite, many pairs of measurements will be used to investigate the fluxes. It is then assumed that the domain can be divided spatially and temporally into homogenous flux ecoregions (see Figure 6.1, [Hargrove and Hoffman, 2005]).

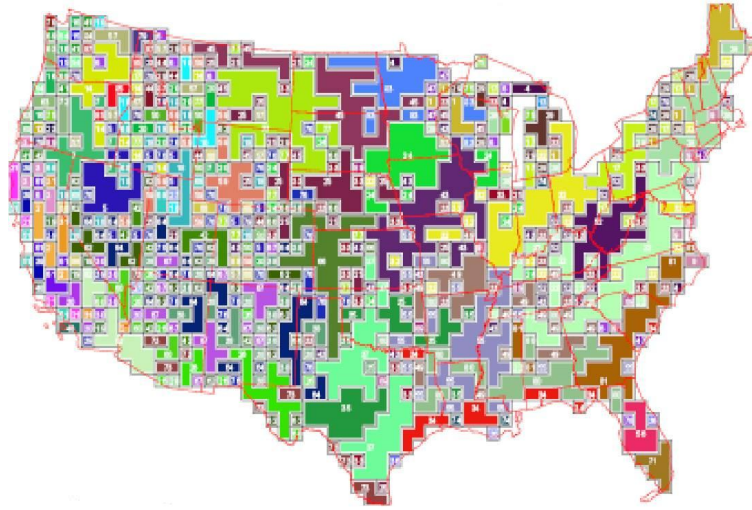


FIGURE 6.1. Using a multivariate geographic clustering technique [Hargrove and Hoffman, 2005], the North American region was divided into a number of quantitative ecoregions (30 are shown here gridded to a  $1^\circ$  by  $1^\circ$  map, each ecoregion being identified by an arbitrary colour).

## 6.2.2 Thought experiments

The following thought experiments will examine measurement pairs first. A method will be found that allows the NAME model to replicate the “real world” pair of measurements. Each pair at this point will have a mass difference and a time for each ecoregion. From here, there are two methods to study the carbon cycle.

The first (forward) method involves taking a known flux value for the ecoregions from CarbonTracker [ESRL, 2008a], multiply by the surface residence time of the particles, sum all the regions together to get the total  $\text{CO}_2$  mass flux and compare it with the SCIAMACHY  $\text{CO}_2$  mass flux. The second (inverse) method will involve combining all the column pairs, and finding an optimal flux magnitude for each ecoregion, that best agrees with the pairs.

### 6.2.2.1 Thought experiment Case 1: Giant tube and constant fluxes

In this case the giant tube prevents particles from leaving the initial release column. The flux at the surface is treated to be the same everywhere at all times. The tube starts at a latitude and longitude position under a SCIAMACHY overpass, and moves horizontally to a position where SCIAMACHY will measure the following day (see Figure 6.2).

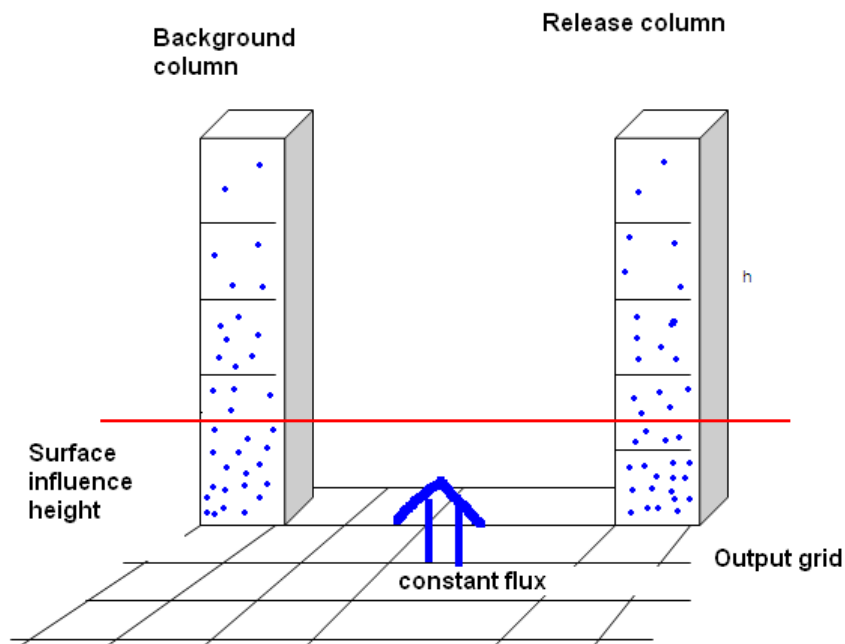


FIGURE 6.2. Release particles are horizontally constrained within the tube. The contact area of the particles within the tube remains constant. The surface flux is constant in this example therefore the contribution of the surface flux to the column is not influenced by the tube moving across the domain.

The measurement footprint of the satellite is not necessarily the same size as the tube cross-section area, but does this matter? In Chapter 4,  $\text{CO}_2$  concentrations are described in terms of the volume mixing ratio. However, this is converted for ease of comparison of all the FSI measurements. The instrument really detects the number of molecules per cross-sectional area, as the intensity of absorption depends on the number of  $\text{CO}_2$  molecules that light interacts with on its way to the instrument. The mixing ratio comes from this number divided by the total number of air molecules in the same column (calculated from the surface pressure). Assuming that the  $\text{CO}_2$  in the giant tube is well mixed at the initial measurement, an instrument with a large footprint (SCIAMACHY) or small footprint (OCO-like) should give the same value of  $\text{CO}_2$  molecules per cross-sectional area. The initial tube can be sub-divided into narrower tubes, the mass of  $\text{CO}_2$  in the narrow tubes is proportional to the cross-sectional area. The mass difference between the two measurements will also be proportional to the cross-sectional area. However, the flux is the same for both the narrow and larger tubes. This has to be considered in the inversion.

In the real world scenario, air would be free to move around up and down in the tube. Individual particles in a NAME representation of this would also be

free to move vertically. The cross-section area at the ground would remain the same throughout the time period between the two measurements. During this time, the real world air would constantly be in contact with the surface. However, individual air molecules might collide with the surface only occasionally. Because of the discrete time steps used in NAME, it is not possible to count the number of collisions that particles made with the surface. Instead, some other way of indicating the amount of time that released particles (measured column air) were influenced by surface fluxes was needed.

The output from NAME will describe the amount of time particles spent in a 2-dimensional latitude longitude grid (adding 900 seconds of time for each particle in each grid box for every time step). A third dimension can be added to the output grid to describe particles up to a certain height above the surface being under the influence of surface fluxes. The maximum height could be the top of the troposphere (basically the 2-dimensional case), although this does not represent certain weather conditions very well. From a physical point of view the height to the planetary boundary layer (PBL) would make sense, as this air gets mixed quite readily and particles anywhere here could be said to be influenced by surface fluxes. A drawback here is that the PBL is not a fixed height, and varies greatly between day and night, which could introduce a diurnal bias. A fixed height could be chosen that is fairly close to the surface instead, although there would be no physical basis for this arbitrary height.

Assuming that a fixed height is used, the output from NAME should match the real world situation. In the real world case with the giant tube, the amount of time that particles are in contact with the surface remains the same, air is in contact with the surface across the whole area of the tube for the entirety of the run. Therefore, particles in the NAME model should have a constant value of surface influence in this situation. The PBL is a height at which the air is quite turbulent, allowing mixing of air in this height range. The vertical extent of this mixing height changes between day and night. It is easy to see that if PBL were taken for the surface influence height, that the surface influence would not be constant (see Figure 6.3). At the release time, particles are distributed randomly in the horizontal direction, and randomly with a pressure weighting in the vertical direction. Although individual particles are free to move vertically in the column, the physics of the model will keep the statistical height distribution of particles the same for any point in the future (or past if the model runs in reverse mode, in this case there is no particular reason to run NAME in forward or backward time).

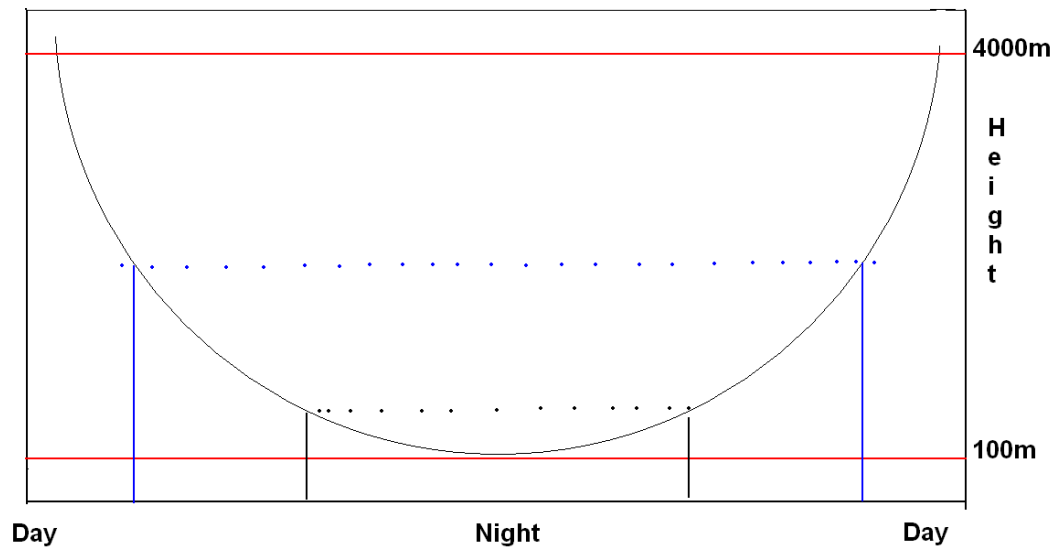


FIGURE 6.3. The NAME model was run with a defined maximum and minimum planetary boundary layer (PBL) height. The curved line represents the PBL height as it varies between day and night. Particles below this line will have greater vertical speeds and will be in frequent contact with the surface whereas particles above this line have slow vertical speeds and have no contact with the surface. As day turns to night, the PBL height becomes lower and most particles will leave the layer (the black particle remains in the PBL until much later in the evening than the blue particle). As the amount of air in the PBL becomes less in the night, the influence upon particles remaining in the PBL becomes stronger, which needs to be represented in the method. Many different surface influence heights have been tried during this work, although it would seem that using the minimum height of the PBL provides the best physical comparison to the real atmosphere.

NAME could be used to realise the tube scenario if only one particle is released, with the latitude and longitude coordinates representing the centre of the tube. One obvious problem is that 24 hours after the release of the particle, there is no guarantee that the particle will be underneath the satellite to make the second measurement. It is not unreasonable to expect air from a compact region to stay together for a 24 hour period in some atmospheric conditions, as the particles would follow pressure contours. However, this is not likely in a column, as wind speed and direction varies with altitude. Methods of obtaining a background concentration are discussed in Section 6.2.3.

### 6.2.2.2 Thought experiment Case 2: Giant tube and variable fluxes

The release column of measured air will typically have a similar surface influence area at any later or earlier time, because the statistical vertical distribution in the atmosphere of release particles depends on the rate at which pressure changes with height. The horizontal diffusion of particles makes it harder to recognise that the vertical distribution of the column is the same, so the giant tube idea is used here to make this point. Case 4 gives examples where this assumption breaks down.

In this case, the surface flux of carbon varies both in time and space across the domain. To keep it simple, there are 4 spatially different flux areas in the domain, and each of these four has a different flux for day and night (see Figure 6.4). The giant tube will be under the influence of 8 possible flux magnitudes between the two measurements. NAME will record the amount of time that particles spend over these 8 domains (so will have a latitude, a longitude, and a time dimension to the output array - in this case the height dimension is one dimensional since the air is always in contact with the ground in the giant tube case). The inversion problem will be of the form:

$$\frac{\text{mass difference}}{\text{surface area}} = \sum_i \mathbf{f}_i \times \mathbf{t}_i \quad (6.1)$$

where  $\mathbf{f}_i$  is the flux magnitude for area  $i$ , and  $\mathbf{t}_i$  is the surface residence time of air over area  $i$ . There is not enough information in a single measurement pair (release column and background concentration) to separate out the 8 individual flux magnitudes, which is why many measurement pairs need to be considered. It is quite clear that the model will be able to cope very well with the spatial variation of the fluxes, as both the measurements and the NAME particles can be made over the entirety of the domain.

The NAME release particles will provide temporal information about the residence times. However, potential temporal information about fluxes may be difficult to obtain because both measurement pairs will be made at the same part of the diurnal cycle (SCIAMACHY has a local overpass time of 10 am). It was decided that any inversion set up with measurements only at a single point in the diurnal cycle, would be unable to represent the diurnal fluxes, and therefore that only a daily averaged flux for each “homogenous ecoregion” could be investigated. The photosynthetic and respiratory components of the surface exchange cannot be separated with measurements from a single sun-synchronous satellite. However, the diurnal



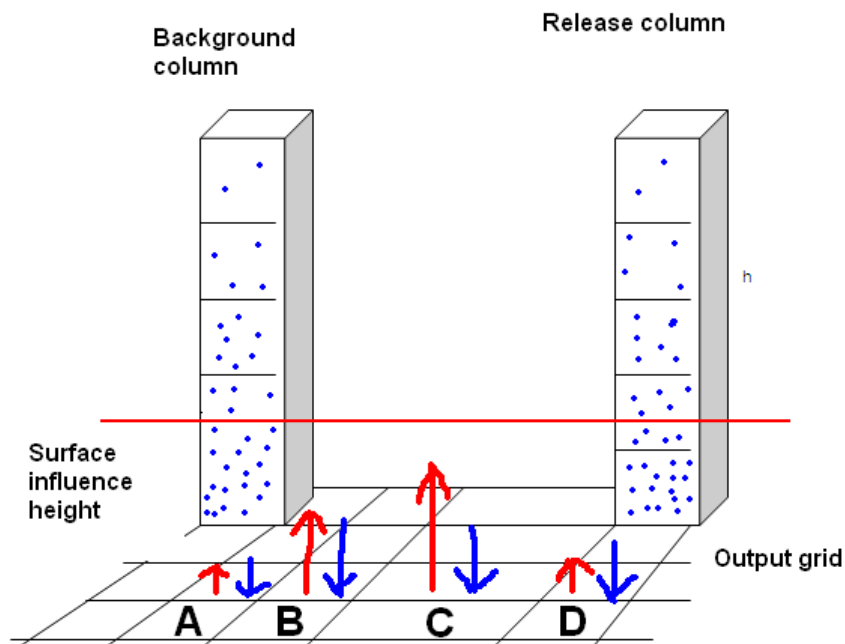


FIGURE 6.4. Similar to Figure 6.2 shown in Case 1. The magnitude of fluxes are represented by the size of the arrows (red for night blue for day). The mass change in the column will depend on the length of time spent over each region, and the flux magnitude over that region.

average flux is still a very useful result. It would allow the capacity of biospheric uptake, under different climatic conditions, to be investigated. Such information could be used to project the future uptake of carbon by the terrestrial biosphere, one of the key questions of carbon cycle research.

In the forward modelling case, the effect of the diurnal cycle can still be investigated, as CarbonTracker fluxes have 3-hourly outputs. The 3-hourly case can be used to examine whether an individual pair of measurements (from CarbonTracker columns) produces the same mass difference as the result of multiplying the carbon tracker fluxes by the surface residence time. This is effectively a test of our initialisation of the model, i.e. how well can the background  $\text{CO}_2$  be constrained? This question is examined in Section 7.1.

### 6.2.2.3 Thought experiment Case 3: Horizontal dispersion

If the particles are free to move in the model, as they are in the atmosphere, it is clear that the two measurement column situation is going to break down. If the release column is horizontally large (compared to say weather map isobars), the release particles could no longer be said to represent those of the column.

The air would quickly bifurcate further and would represent distinct trajectories and therefore different fluxes and backgrounds. The release column area can still be larger than a measurement footprint as it is desirable to have an ensemble measurement. The break down of the two measurement column approach to obtain the mass flux suggests that a more sophisticated way of obtaining the background CO<sub>2</sub> concentration is needed.

For a Lagrangian scheme, the most appropriate way of initialising the model run would be to weight the background measurement based on the number of particles that come from various directions. A background measurement value then needs to be defined for each direction that the air arrives from. This turned out to be very difficult. Two approaches have been studied. One involved running NAME for a single day and finding where the particles were after a complete diurnal cycle (see Section 7.1.2). The model has a gridded output and the NAME trajectories would be used to weight the background. Of course the difficulty here is to have measurements at the same time and place as all the grid squares. The other main idea (see Section 7.1.1) involves running the model for as much time as needed for the particles to leave the domain, the domain edge could either define the edge of the land area or could go further out into sea (this may not be appropriate for the SCIAMACHY instrument as it cannot measure over the ocean). Because there is not expected to be any significant fluxes between the air and sea, air at these locations will have less influence from local sinks and sources. It is possible therefore to use a measurement from a previous day, rather than the day that particles arrive at the edge.

In the giant tube case, it did not matter whether the model was run in forwards or backwards mode. However, when the particles are free to disperse, the surface area of interest should be near the centre of the domain. To investigate in reverse mode, only a few release domains around and over this area are needed. To investigate the same area forwards, particles must be released everywhere in the domain (which is impractical). Backwards time releases also gave the option of defining the release source by the satellite overpasses.

#### **6.2.2.4 Thought experiment Case 4: Vertical and horizontal dispersion**

There are some unusual cases where the initial vertical distribution of the particles (from the release location) is not the same as that at the end. For example, there may be convective behaviour at the location where particles are released from, and

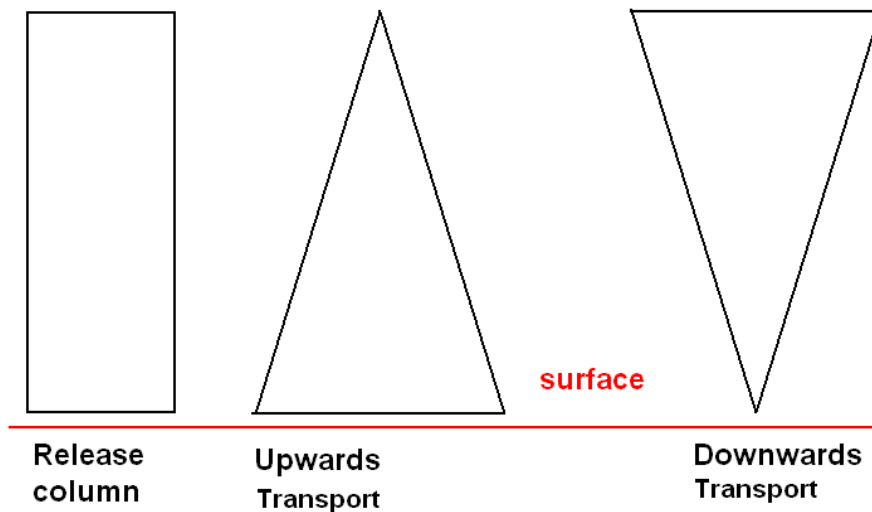


FIGURE 6.5. The image on the left represents the shape of the column at the moment it is being measured (thus the shape of the release particle box). This shape represents the exponential decrease of particles/pressure with height. Imagine that just before the measurement was made, there was upwards convection around the location of the measurement. Many more of the release particles would have been distributed in the lower parts of the atmosphere; if it was possible to see the shape of the release column at this time, it may look like the middle image. In this case the total surface influence time is greater than the run period. In the example of mass downwards movement of release particles (right) it is possible that none of the release particles are within the surface influence height, and the total surface influence time in this case is much less than the run period. Away from the tropics, convective forcings becomes less significant, therefore the total surface influence time should be similar to the run period.

just a few hours earlier most of the particles could have been near the ground (see Figure 6.5). If the model is run for a very long time the vertical distribution of particles will become random based on pressure levels again, but on the time scale of the run there will be a significant increase in surface residence time of this air. This is why it is important that the output grid contains vertical information and not just output for latitude and longitude.

Another unusual case would be for air having two preferred directions, so air low in the column is from the east and higher in the column is from the south. For the initialisation measurement, the column to the south has a high concentration of  $\text{CO}_2$ . However, most of the excess  $\text{CO}_2$  is in the lowest part of the column, giving the high column value where the air that arrives in the release column is much closer to the typical background for this time of year. This would modify the mass

difference in the model, without having a basis in the physical world. There is a case here for having an extra dimension to the output grid, so that the model can recognise particles released at different heights in the column. It would be difficult to use this information in a practical way, in principle the column measurement value could be used to project a likely column profile. This would also mean that the output grid is increased in size (having an extra dimension). This is less of a problem where particles are followed to the domain boundary, as the length of time spent over the minimal flux ocean would smooth out the column gradient.

### 6.2.3 Initialisation

Cases 3 and 4 have shown that there is going to be some difficulty in initialising the background value of the release column air. The problem was first approached by having as small a time period between the initialisation and the release measurement as possible. This would mean that the surface area over which a flux value is obtained will be fairly small, and assuming that both measurements have a high precision, there would be a fairly detailed flux map of the area. This would be the best possible flux inversion method that could be used for the SCIAMACHY retrieved columns.

However, it is clear that air will disperse in different directions requiring column measurements over several parts of the domain. In the day prior to the release, SCIAMACHY will not measure over all parts of this domain. The level of detail that can be achieved in this way is subject to the condition that SCIAMACHY can measure all these areas. One possible solution was to consider measurements that were made in the same NAME grid box, but on the nearest earlier or later day. Because SCIAMACHY has global coverage every six days, the initialisation measurement for every NAME grid square would be within three days of the actual air mass. However, that air mass could have had a very different trajectory and experienced very different flux conditions, to the one on the day the air mass was actually present at the initialisation grid square. It would be unreliable to use this measurement as an initialisation.

It was not possible to obtain a reliable initialisation value that was spatially the same but temporally different from the actual initialisation. This effectively ruled out the possibility of a one day NAME run. The only other reliable way of initialising the release column air would be to measure air at the same time and near to the location where the initialisation would have been made, had there been a

satellite overpass. This would be reasonable if the surface between the measurement grid square and the NAME grid square that particles are said to be in at the end of the run, has only a minimal carbon exchange and it could be presumed that the particles would pass very soon through the measurement grid square.

The first idea in this case is to set up the domain boundaries just past the west and east coasts of the scene. Although SCIAMACHY cannot measure over the ocean, it is able to measure columns that are just inside the coastal boundary. With an absence of terrain for air to maneuver around, air over the ocean is likely to travel along the same latitudinal band, and over the short distance between the measurement and the domain boundary, it would be reasonable to assign the measured value at the coastal location to the NAME grid square at the boundary.

The NAME output would now consist of two files for each run. The first is the surface residence time grid with fine latitudinal and longitudinal resolution, vertical resolution of several possible surface influence heights, no temporal resolution (since diurnal cycle behaviour cannot be distinguished) and an extra dimension in the form of distinguishing the initial height that particles were released at (NAME allows many particle types to be released, particles released between zero and one kilometres are labeled inert1 in these runs). This file records the time in each grid square for each time step (i.e. 900 seconds multiplied by the number of particles in the grid box). The second file records the place where particles leave the scene, the latitude and longitude resolution of this grid is the same as the other, except that only the very outside grid squares are recorded (making a much smaller output file). The grid has a vertical resolution of one kilometre (which would allow a mixing ratio value at that height, if it is possible to project a column profile, as suggested in Case 4). It has a temporal resolution of once per day, since there will be a different column mixing ratio value for the background for each day. This file records the last grid box each particle was in immediately before leaving the domain.

One of the things working in favour of such a method is that the path of a polar orbiting satellite from north to south, means that there will be measurements more or less parallel to the east and west boundary edges for every single day. It is only a minor problem that the same coverage does not exist along the north and south boundaries, because almost all particles arrive from the west boundary. Releases with any significant contribution to the background from the north or south can be filtered out of the inversion step. The problems for SCIAMACHY in this situation are:

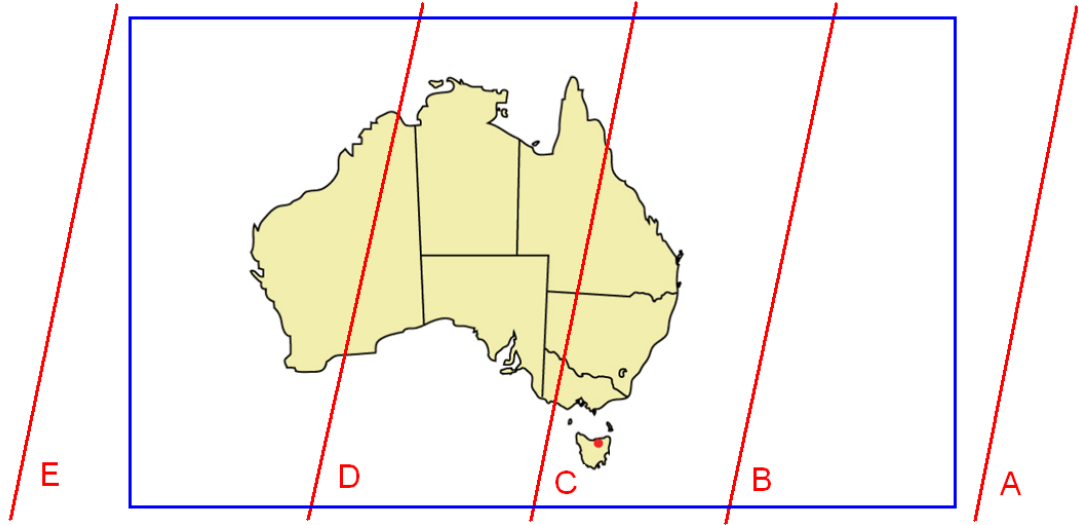


FIGURE 6.6. Initialising the background using a CO<sub>2</sub> retrieving satellite in sun-glint mode. The closest sun-glint swath (red lines) to the western and eastern boundaries (blue rectangle) are used to define the background concentration at that boundary (in this case swaths E and A). The swath and boundary are divided into discreet latitudinal sections that are large enough to contain an ensemble of CO<sub>2</sub> measurements yet not so large as to average over the latitudinal CO<sub>2</sub> gradient discussed in Section 2.1. This method could be tested using an OCO-like instrument.

- the gap created along the orbit by the alternating nadir and limb modes.
- the scene to be investigated (North America) does not have a convenient rectangular shape that would fit with the domain edge.
- the scene to be investigated (North America) has a mountainous western edge, that SCIAMACHY seems unable to reliably retrieve over.
- cloud contamination.

The second and third itemised problems can be removed if the domain boundary is further out in the ocean, and the measurements are made out there also (see Figure 6.6). Unfortunately, SCIAMACHY is unable to measure out over the ocean. However, this method could work with an instrument measuring in continuous glint mode (also resolving the first problem), such as an OCO-like instrument.

It became clear that it would not be possible to use SCIAMACHY measurements for the initialisation. Although it is expected that the variation of the column CO<sub>2</sub> is much smaller out over the ocean, it would be the first choice to use the same instrument for measuring the initial and final column concentrations. It is expected

that this would cancel out many systematic uncertainties in the retrieved columns. The most appropriate alternative would be modelled columns, and CarbonTracker was chosen because it had data available for the desired dates and times and it was based on measured CO<sub>2</sub> concentrations. Using a different data set for initialisation is likely to introduce biases in fluxes, as SCIAMACHY/FSI data is likely to be offset against CarbonTracker. Some comparisons have been made between the two data sources, which should give some indication as to how much the mass difference is affected by using different data for initialisation and final measurements (see Section 2.2.1).

In-house analysis of CarbonTracker model and FSI/SCIAMACHY CO<sub>2</sub> columns suggest a mean offset bias of around 1.5% for the North American region, though this offset varies somewhat with both latitude band and with season [Comyn-Platt and Boesch, 2009]. Results shown in Section 8.1 also indicate a mean bias of 1.5% between the CarbonTracker weather product and the FSI/SCIAMACHY columns, again with seasonal differences. It is quite clear from the plots (see Figure 8.1) that there is a lot more complexity than just a simple offset between the two column values.

Though the variability will get filtered out when considering many pairs of background and foreground CO<sub>2</sub> columns, the biases introduced by the different seasonal cycle amplitudes will not. No attempt has been made to correct for this, since tweaking this value will directly change the magnitude of the inversion results. It is critical to the integrity of the method that the inversion results are not made to fit observed quantities. The forward run (see Section 8.1) and inversion (Section 8.2) were run both with and without the 1.015 scaling factor suggested by Schneising et al. [2008].

### 6.2.3.1 “Edge box” initialisation method

Extracting information about where and when particles leave the domain boundary requires some significant changes to the way that the NAME model outputs information. Rather than spending a significant amount of time setting this up straight away, an alternative method for initialising the background was used in the NAME setup. Rather than recording the time and place where particles leave the domain, the amount of time particles spend in the “edge” grid boxes (the 1° by 1° grid boxes adjacent to the domain edge) was recorded. Because output is taken for every 15 minute time step, it is very likely that particles will be recorded for

several time steps before they leave the domain.

There are some drawbacks from this method, for example particles traveling to two different domain edges at different speeds would skew the weighting of particles in favour of the slower moving (thus longer resident) particles. When particles all travel to the same boundary, this will have much less effect on the precision of the initialisation. The advantage of this method was that it could be used straight away.

After testing both initialisation methods (see Section 7.1), it is clear that there is a significant improvement in using the “particle leaving the scene” or “freeze” initialisation method, and that modifying the NAME model to output this information would be worthwhile.

#### **6.2.4 NAME model setup**

Ideally, the centre of the NAME release square lines up with the centre of the ensemble of satellite measurements. However, it is essential that the inversion method employed to investigate fluxes is capable of identifying the ensemble in an automated way. This required the development of a suitable labelling system for measurement ensembles. The author was unable to identify a suitable labelling system centring on the satellite swath and instead developed a grid based system (see Section 6.2.4.1) for identifying satellite measurement ensembles. Difficulties occur when ensembles fall across two adjacent grid squares, which can be seen later in Figure 8.8, . Some grid boxes (such as Box AA) may contain very few measurements and may therefore have a very high standard deviation. Nevertheless, the grid labeling system was used throughout this research project. The author retrospectively developed a labelling system that could align NAME release squares with satellite overpasses (see Section 6.2.4.2). However, owing to time pressures this system has yet to be employed.

##### **6.2.4.1 Original NAME setup**

NAME can be set up so that the release coordinates are specified by a central source point, plus the width of the latitudinal source, the length of the longitudinal source, the height of the vertical source and the length of release time. The release time may as well be instant for our purposes, and the vertical coordinates can be set by height above ground level, which would be the same for every release. Ideally,



the latitude and longitude dimensions are the same as those of the nadir ensemble footprint (about  $3^\circ$  in both directions) and centred upon the ensemble.

Originally, the NAME output was intended to be used with a simulated annealing method, that has previously been used to investigate trace gas fluxes [Manning et al., 2003], and it was assumed that this method would be adopted to investigate the carbon fluxes. At the start of the project was a rather trivial matter of associating the measured mass flux ( $\Delta\text{CO}_2$ ) with the NAME time output. The method for associating the two involved describing the time of the measurement and the release time. In the case of satellite measurements, there are typically many measurements made at roughly the same time, but over many locations. This is quite different to the measurement input that was typically included from point measurements (one location and measurements separated in time). The time information had no physical significance in the simulated annealing method, it is only used to match the measurement with the release of particles. In order to identify different release and measurement locations on the same day, the latitudinal and longitudinal locations were converted into a “pseudo-time” which could be understood by the simulated annealing method.

The requirement of “pseudo-time” constrained the way that the model could be set up. It was disappointing therefore that it was not possible to obtain fluxes from the simulated annealing method. Future use of NAME to investigate the carbon cycle with GOSAT or OCO-like measurements would probably benefit from swapping “pseudo time” with “satellite time”. In Section 6.2.4.2 a method of applying this to an OCO-like instrument is suggested.

The first step of the inversion method requires reading the NAME output files, which had a filename with either 8 or 12 number characters, depending on whether there was only one measurement in a day (so that 1<sup>st</sup> April 2003 has filename NAME01042003.txt) or whether there were several (so that the other four characters represented the 24 hour clock). The simulated annealing inversion scheme expected a regular sequence of measurements (perhaps once an hour) and the pseudo time was an integer number representing the number of hours since a start time (usually the start of the year). The simulated annealing inversion scheme had no problem with measurement gaps.

Keeping the date information preserved in the filename was convenient, because both the processed SCIAMACHY grids and the CarbonTracker initialisation grids would also have the same date in the filename. The simulated annealing scheme

does not deal with time in a discrete way, a central time and time resolution is needed for each measurement. By allowing one minute each side of the central time, the maximum number of latitude and longitude release squares allowed in the domain is 24 (from 24 hours) by 30 (30 lots of 2 minutes per hour). Release locations were thus gridded, with the longitude represented by the hours and the latitude by the minutes. This limit was sufficient for 23 longitudinal and 13 latitudinal  $3^\circ$  by  $3^\circ$  to be gridded together across the landmass in the North American domain. The domain boundary was defined as  $170^\circ$  to  $20^\circ$  west and  $10^\circ$  to  $85^\circ$  north, particles were only released within  $130^\circ$  to  $61^\circ$  west and  $30^\circ$  to  $69^\circ$  north.

The problem with this set up is that the centre of the release grid does not match up with the centre of the SCIAMACHY ensemble. This is not a great source of error, as only SCIAMACHY footprints inside the grid box divide need to be used to give an ensemble mean. The initial experiment used even larger grid boxes at  $10^\circ$  by  $10^\circ$  and required less than 24 time slots, meaning that minutes could be dropped from the pseudo time. Although this was not the best set up for the model, it was chosen because it was simple to set up and would produce almost identical results to a set up with matched release square centres. Running the NAME model to get the right filename (hence pseudo-time) required a different directory for each release location. In the gridded case, the directory name would include the pseudo time, although the actual directory names used letters instead of numbers (allowing scripted changes to be made).

#### **6.2.4.2 Re-examining the NAME setup**

In Section 6.2.4.1, the requirement of “pseudo-time” by the simulated annealing method added constraints to the way that the NAME setup could be manipulated to investigate the carbon cycle. Suggested below is a satellite centric labeling system for OCO-like retrievals (since OCO had a failed launch in 2009). The details can be adjusted for other sun-synchronous  $\text{CO}_2$  measuring instruments (GOSAT or even SCIAMACHY).

The planned ensemble size of OCO was much smaller (than SCIAMACHY), OCO would have imaged eight footprints (only four of which are available) across track with a total width of just over 10 kilometres. This means that there would have been a lot more to gain by centring the release box over the ensemble. It would have been useful to have around 20 or more footprints in the ensemble, and if 24% of retrievals pass the cloud filter [Miller et al., 2007], it would have been

useful to take at least 20 rows of pixels along track (there was no oscillating mode in OCO, so the track was continuous), and so would have been appropriate for the latitude distribution of release to be every half degree along the track. Cloud cover is not random however, some regions and times would have been much less cloudy, and the size of the release square would have needed to be modified to reflect this.

By freeing the constraint of needing a pseudo time to identify the measurement and the release time set, it becomes clear that the directory labeling of the NAME runs should be set up to match the satellite orbit. There is a 16-day repeat cycle for the satellite orbit or 233 individual orbits, for 16 days the instrument operates in nadir mode and the next 16 days it operates in glint mode. The NAME run directories would be identified with an N (for nadir or G for glint) followed by three numbers to identify the orbit and then some more numbers to identify the latitude of the ensemble. In this way it is known which day the orbit takes place, there is no need to run the model for the wrong days reducing run time by a factor of 16. For the nadir runs, the latitude and longitude coordinates are the same for every repeat orbit. For glint runs, the longitude will be in a different place each time because the position of the sun in the sky changes through the seasons.

It would be of great benefit to the inversion scheme if the background values of  $\text{XCO}_2$  were retrieved by the same instrument as those at the release location (see Figure 6.6). It is advantageous to measure over ocean in order to get a background value, since this will be less affected by local fluxes. The OCO instrument was designed to operate for 16 days in glint mode, followed by 16 days in nadir mode. If for a large scene (such as North America) a ten day run was required, this meant that there was only a small window in which all ten days could have oceanic background measurements. Running for shorter periods would have increased the amount of days that there were all glint mode available background measurements, but it would have been less than half of the 32-day cycle. If inversion results were desired for any release day, then the nearest nadir (or glint) measurements inside the coastal boundary would have been used. This is not ideal, as there would certainly have been a significant amount of surface residence time between air passing over the coastal boundary and then being measured by the instrument.

## 6.3 Offline method developments

### 6.3.1 FSI measurement conditions

The final inversion scheme needs to consider the fact that some measurements are more reliable than others. Each mass difference in the column  $\text{XCO}_2$  has a corresponding error value to indicate how reliable the measurements were. The earlier analysis of SCIAMACHY retrievals suggests that the FSI algorithm performs badly in certain situations. Measurements over hilly terrain seem to vary wildly, despite the fact that they have a very small column fit error. Aerosol contamination is also not dealt with very well. It is also likely that the FSI cloud filter is too lenient with partial cloud conditions.

As a result the  $\text{XCO}_2$  column fit error is a poor indicator of how far from the truth the single  $\text{XCO}_2$  measurement is likely to be. In the case of an OCO-like instrument, the ensemble column fit error would be useful in the inversion scheme, and would allow even a single footprint  $\text{XCO}_2$  measurement within the ensemble to be used. The mass difference error should encompass both the error at the release square and also the error of the background measurement.

This is not possible in the case of FSI  $\text{XCO}_2$  measurements. It was reasoned that the ensemble air would encounter very similar source conditions and would quite likely have come from the same background. For this reason the variability of individual measurements in the ensemble should be fairly small. Large variations would be brought about by aerosol or cloud contamination in the background air. A filter was applied so that for any day a minimum of 20 footprints needed to fall with the  $3^\circ$  by  $3^\circ$  release grid square for it to be considered in the inversion (the number 20 was chosen arbitrarily). The standard deviation of the measured  $\text{XCO}_2$  columns was then used to represent the measurement error. This meant that having an increased number of measurements in an ensemble did not (in the FSI case) improve the ensemble precision error. Ensembles with a standard deviation of more than 10 ppmv were filtered out, most fell in the range 6 to 12 ppmv. Validation work by Peters et al. [2007] in the North American region suggest a precision of the CarbonTracker  $\text{CO}_2$  column of  $\Delta = 0.5 \pm 0.9 \text{ ppmv}$  (see Section 2.2.1). The mass flux precision error for FSI measurements was based only upon the ensemble standard deviation.

For individual FSI  $\text{XCO}_2$  measurements to be included in the ensemble they must pass a number of filters.

- The measurement must pass the SCIAMACHY PMD Identification of Clouds and Ice/snow (SPICI) [Krijger et al., 2005].
- The measurement must be a forward scan pixel.
- The measurement SZA must be less than  $75^\circ$ .
- The measurement must have been made over land.
- The column  $\text{XCO}_2$  fitting error must be less than 3%.

No aerosol filtering is used, although the Bremen group are looking at using the Total Ozone Monitoring Spectrometer (TOMS) to identify regional episodes of high aerosol (such as saharan dust events), so that measurements at these times are excluded [Schneising et al., 2008]. This method is not suitable for individual footprints, as the TOMS instrument does not fly onboard ENVISAT with SCIAMACHY.

### 6.3.2 Development of an inverse modelling technique to investigate $\text{CO}_2$ fluxes

The NAME model and  $\text{XCO}_2$  measurements have been conjoined to examine individual measurement flux behaviours, and can now be combined together in an inversion scheme. Since it has proved successful before with NAME inversions [Manning et al., 2003], albeit for point sampling locations, the simulated annealing method was initially considered for use in the inversion scheme.

The problem that needs to be solved is one of optimisation. In this case there will be a mass difference and a surface residence time array for every release. The task is to find a flux value for each element of the surface array, so that the difference between the observed mass difference and the mass difference from multiplying the residence times by the flux values is minimised.

The array in question (in the North American example) is made up of 150 latitudinal by 75 longitudinal,  $1^\circ$  by  $1^\circ$  ground pixels. The peak of the number of measurements that pass through the filters occurs in the summer, and consists of several hundred measurements per month, see Figure 6.7. Minimising the problem becomes very difficult if the number of fluxes to be obtained is greater than the number of measurements made. An assumption was made, that the flux magnitude

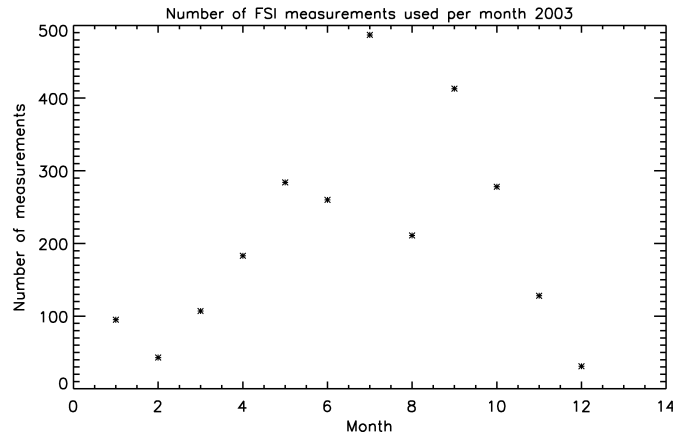


FIGURE 6.7. Number of FSI measurements used in NAME inversion per month.

of groups of pixels are the same, based upon the idea of homogeneous ecoregions. The total time over each of these areas is then considered and there should be far fewer ecoregions than there are measurements.

The assumption of a homogeneous ecoregion is a reasonable way of examining regional scale fluxes of carbon. Satellite remote sensing will not allow for individual carbon budgeting for small scale ecosystems, but it is reasonable for many similar ecosystems experiencing similar weather conditions to have similar photosynthetic and respiratory activity. There is another subtle assumption being made, that the flux magnitudes of the ecoregions stay the same throughout the period over which particles are released. To make sure that there were sufficient measurements, this timescale meant releases over a month plus ten days backwards in time and so was around 40 days.

It would be advantageous to shorten this time, certainly it would be interesting to compare an early 10-day release period in the month to a late 10-day period, to examine whether the flux inversion scheme returns the same values for both time periods. If the daily averaged flux magnitudes of a particular ecoregion are wildly different at the beginning and end of the run then the results will not be particularly meaningful. Clearly, a year is much too long a time period to consider, given that vegetation has an annual cycle of growth. A month is probably too long a time period as well, as the physical processes which affect the carbon exchange in the biosphere are primarily based on the weather, which is quite variable over a month long timescale.

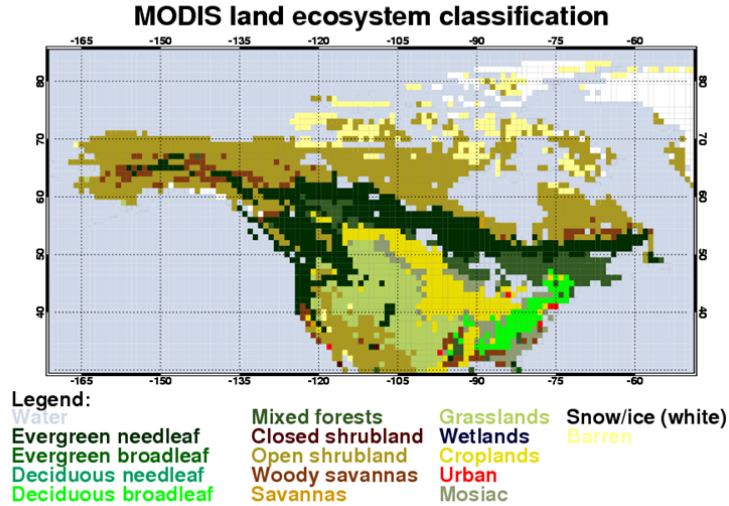


FIGURE 6.8. MODIS land classification gridded to a  $1^\circ$  by  $1^\circ$  map.

It is clear that for each inversion the balance between running for long enough to get the necessary amount of measurements and not so long as to have very different flux conditions needs to be met. Certainly, reducing the length of time the model runs backwards will help the latter whilst not affecting the former, though it would suggest having a much smaller domain.

Identifying the homogeneous ecoregions can be done with a combination of a priori information and other remote sensing information. Ideally, the flux values obtained in the inversion scheme would be compared against bottom up information that has been scaled up to regional levels based on the same ecoregion map (the idea of homogeneous ecoregions would be essential in the scaling up process). Such a data set does not exist at the time of writing, though it is likely to be developed in the near future. Many biospheric models use weather conditions to define the flux behaviour across a domain, though they do not need to consider homogeneous ecoregions. These models are not ideal for comparison as they do not prove actual measurement data. The ecoregion does not need to contain the same pixels for each time period in question, and the way the NAME information is outputted allows this to be changed very easily (basically define all the pixels in the surface array that belong to that ecoregion and sum up the residence time).

Defining the ecoregion map would require more specialised knowledge of vegetation behaviour. It would make sense to collaborate in future with a group looking at scaling up the localised eddy fluxes to a regional scale. For the purposes of investigating the SCIAMACHY/FSI inversions, the MODIS global land cover product,

MOD12Q1, was used [Friedl et al., 2002]. This data set has been gridded onto a 1° by 1° map of the world, with the majority of a certain vegetation being identified as that pixel's land classification (see Figure 6.8 and also Section 6.3.2.2 where the consequences of applying the majority rule are discussed further). There are 17 different vegetation types identified by the MOD12Q1 product, which have been reclassified into eight possible types in this work. This was done either to group together the presumed negligible flux surface types, or to ignore vegetation that was either very rare or absent from the North American scene. The eight vegetation types considered, and their MODIS land classification equivalents are as follows:

- Grass
  - Grassland: Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
- Savanna
  - Woody Savanna: Lands with herbaceous and other understorey systems, and with forest canopy cover between 30% and 60%.
  - Savannas: Lands with herbaceous and other understorey systems, and with forest canopy cover between 10% and 30%.
- Crop
  - Croplands: Lands covered with temporary crops followed by harvest and a bare soil period.
  - Crop Mosaic: Lands with a mosaic of croplands, forests, scrubland and grasslands in which no one component comprises of more than 60 %.
- Scrubland
  - Open Scrubland: Lands with woody vegetation less than 2 metres tall, and with shrub canopy cover between 10% and 60%.
  - Closed Scrubland: Lands with woody vegetation less than 2 metres tall, and with shrub canopy cover greater than 60%.
- Mixed Forest
  - Mixed forests: Land dominated by trees with a percent cover greater than 60% and height exceeding 2 metres. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types.



- Deciduous
  - Deciduous needleleaf forests: Land dominated by needleleaf woody vegetation with a percentage cover greater than 60% and height exceeding 2 metres. Consists of seasonal tree communities with an annual cycle of leaf-on and leaf off periods.
  - Deciduous broadleaf forests: Land dominated by broadleaf woody vegetation with a percentage cover greater than 60% and height exceeding 2 metres. Consists of seasonal tree communities with an annual cycle of leaf-on and leaf off periods.
- Evergreen
  - Evergreen needleleaf forests: Land dominated by needleleaf woody vegetation with a percentage cover greater than 60% and height exceeding 2 metres. Almost all trees remain green all year, canopy is never without green foliage.
  - Evergreen broadleaf forests: Land dominated by broadleaf woody vegetation with a percentage cover greater than 60% and height exceeding 2 metres. Almost all trees remain green all year, canopy is never without green foliage.
- Negligible flux regions
  - Oceans, Desert and areas with permanent snow or ice cover.

In the case of having  $M$  homogeneous ecoregions (eight have been identified for this trial run) and  $N$  observations, the inversion scheme needs to solve the matrix equation  $\tau \cdot \underline{\mathbf{F}} = \underline{\mathbf{O}}$ , where  $\tau$  is the matrix of all the contribution times that are multiplied by the net land use  $\text{CO}_2$  fluxes  $\underline{\mathbf{F}}$  to obtain the observed carbon mass fluxes  $\underline{\mathbf{O}}$ .

$$\begin{pmatrix} \tau_{11} & \tau_{12} & \tau_{13} & \dots & \tau_{1N} \\ \tau_{21} & \tau_{22} & \tau_{23} & \dots & \tau_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ \tau_{M1} & \tau_{M2} & \tau_{M3} & \dots & \tau_{MN} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \dots \\ F_M \end{pmatrix} = \begin{pmatrix} O_1 & O_2 & \dots & O_N \end{pmatrix} \quad (6.2)$$

In principle any inversion scheme that solves Equation 6.2 could be used, without having to re-run the NAME model. Some time was spent modifying the simulated annealing scheme, so that it would be capable of inverting satellite retrieved XCO<sub>2</sub>. However, some settings were not capable of being modified, for example measurement precision can be stated but the same value is applied for every measurement (which is appropriate for single point measurement sites). Although the simulated annealing scheme was capable of producing inversion results, these did not deviate from the *a priori* flux values used to start the model. The most likely explanation for this non-result, is that the high *a priori* measurement uncertainty, being relatively large compared to the actual variation between the FSI measured CO<sub>2</sub> ensembles, made it unlikely that measurements would nudge the model away from the *a priori* fluxes.

A linear regression method was used in this study, which was simple to manipulate. There are many possible solutions to the inversion problem, from which it is desirable to select the best possible solution, to which an error estimate will be assigned. From Equation 6.2, it is possible to obtain the maximum *a posteriori* solution [Rodgers, 2000]. The mean state averaged over the probability density function (PDF),  $\hat{\mathbf{x}}$ , or the *a posteriori* solution (and its corresponding *a posteriori* error covariance,  $\hat{\mathbf{S}}$ ) can be obtained from:

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{y} + \mathbf{S}_a^{-1} \mathbf{x}_a) \quad (6.3)$$

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \quad (6.4)$$

where  $\mathbf{K}$  is the weighting function matrix,  $\mathbf{S}_\epsilon$  is the measurement error covariance matrix,  $\mathbf{S}_a$  is the *a priori* covariance matrix,  $\mathbf{y}$  is the measurement vector,  $\mathbf{x}_a$  is the *a priori* flux vector and  $\hat{\mathbf{S}}$  is the error covariance matrix.

Each of these elements must now be related to one of the elements already discussed in this chapter.  $\mathbf{K}$ , the weighting function matrix, has dimensions of number of measurements  $\times$  number of states (the number of homogeneous ecoregions). This array will contain information about the residence times of NAME particles over the ecoregions for each NAME release. The measurement vector,  $\mathbf{y}$ , will contain information about the change in carbon dioxide mass between the measured column and the background. The measurement error covariance matrix,  $\mathbf{S}_\epsilon$ , is a diagonal matrix with the dimensions: number of measurements  $\times$  number of measurements.

The precision of the mass change in carbon dioxide for each measurement is used to fill the diagonal. The state vector,  $\hat{\mathbf{x}}$ , has as many elements as there are ecoregions, and suggests the most likely daily averaged flux magnitude for each of those regions. The corresponding error covariance matrix,  $\hat{\mathbf{S}}$ , suggests a precision value for each element of the state vector (see Figure 6.9 for a diagrammatic overview of the inversion method).

There are corresponding *a priori* fluxes,  $\mathbf{x}_a$ , and covariance,  $\mathbf{S}_a$ , though these are only needed to start the model. By keeping the diagonal elements of the *a priori* covariance large, the *a posteriori* state vector fluxes should be independent of the *a priori* fluxes. There may be occasions where it is favourable to fix the fluxes over a certain type (i.e. minimal over the ocean), in which case a small *a priori* covariance would be used. Since most of the surface influence time in the trial run was spent over the ocean, this function will prove to be useful, particularly in cases where the foreground is always larger or always smaller than the background.

### 6.3.2.1 Units

Two final tasks remain, deciding on the units for Equation 6.4, and converting to those units. There is a certain amount of flexibility with all the elements of the equation, though the flux is the least flexible. The units must either be in number of molecules of carbon per unit area per unit time, or mass of carbon (or carbon dioxide) per unit area per unit time. Mass of carbon dioxide per square metre per day was chosen as the units, because these are the same units used in the vegetation community.

The units of  $\mathbf{y}$  and  $\mathbf{K}$  need to be made to balance the units of the equation. Since the units of  $\mathbf{K}$  are in time, it would make sense to record the mass flux in units grams per unit area. This is ideal since the satellite measures vertical column density (VCD), in units molecules per unit area. The FSI retrieval and the CarbonTracker values are outputted as volume mixing ratio values, which are easily changed to mass per unit area.

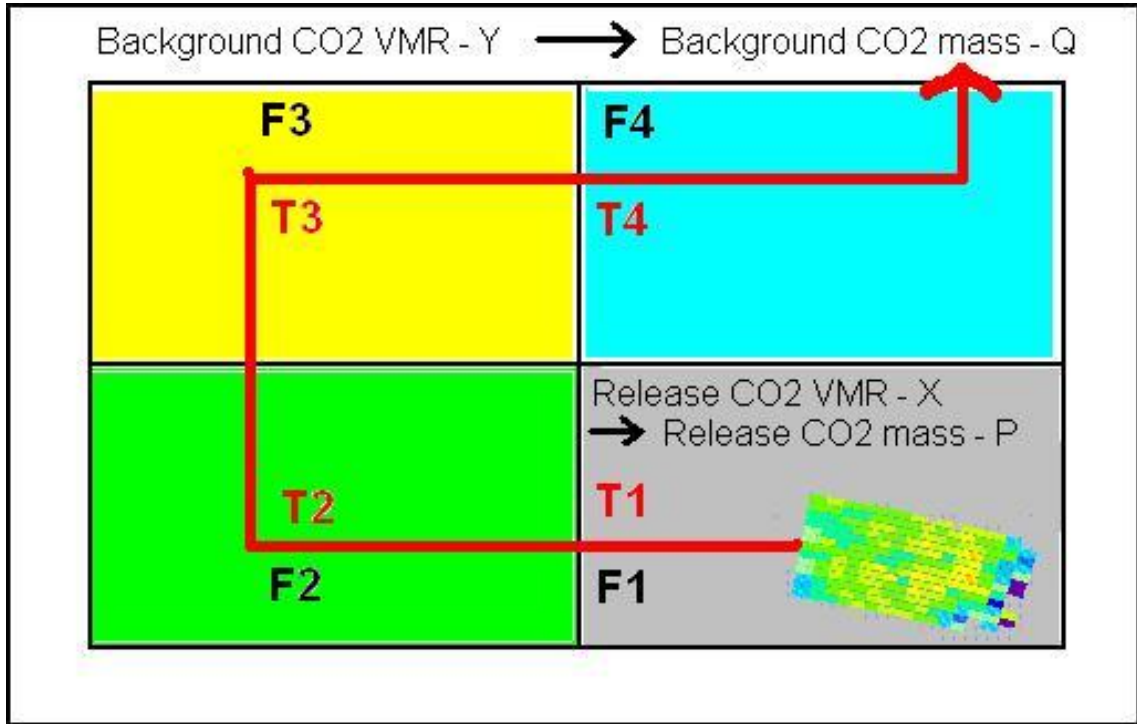


FIGURE 6.9. Overview of the inversion scheme used to investigate CO<sub>2</sub> fluxes. For each ensemble measurement of XCO<sub>2</sub> (labeled X on this diagram), there is corresponding NAME output for the surface influence time and “edge box” origin of the air-mass. The “edge box” origin information is weighted so that it sums to unity. A modelled CO<sub>2</sub> field, such as CarbonTracker (see Section 2.2.1), is multiplied by this weighting to give the background atmospheric CO<sub>2</sub> concentration of the satellite measured air (labeled Y on this diagram). Both the satellite measured and background CO<sub>2</sub> concentrations were converted into a total column mass of CO<sub>2</sub> (labeled P and Q respectively). For each of the “N” ensemble measurements and NAME releases (see Equation 6.2) Q is subtracted from P to give the observed carbon flux, “O”. Meanwhile, the surface influence time (labeled T1, T2, T3, T4 with the red arrow on this diagram) of each NAME release is partitioned by the “M” homogeneous flux ecoregions (labeled F1, F2, F3, F4 with the four coloured regions on this diagram). Assuming that NAME perfectly captures the atmospheric transport, that the “M” flux regions are constant and that FSI and CarbonTracker both capture very close to the true CO<sub>2</sub> column values then solving the matrix equation (from Equation 6.2) could identify the magnitudes of the homogeneous flux ecoregions. Owing to measurement uncertainty, this becomes a many solution problem and a linear regression method (Equation 6.3) was used in the inversion scheme to identify the best solution to the “M” flux magnitudes.

To get the weighting function matrix units correctly, it is useful to reconsider the giant tube approach. If none of the particles leave the scene, then the total surface influence time should be on average the same as the total run time (see Sections 6.2.2.2 and 6.2.2.4). For the trial run, the maximum height at which particles can be considered to be in contact with the surface is 4000 metres. If initially the release of  $\mathbf{N}_p$  particles is vertically random, weighted by pressure, then the surface contact time is:

$$\mathbf{t}_c = \mathbf{t}_r \times (\mathbf{N}_p)^{-1} \times \frac{\mathbf{P}_s}{(\mathbf{P}_s - \mathbf{P}_{sih})} \quad (6.5)$$

where  $\mathbf{t}_c$  is the surface contact time,  $\mathbf{t}_r$  is the model run time, and the fraction is used to adjust for the variation of residence time to the maximum surface influence height ( $\mathbf{P}_{sih}$  is pressure at the maximum surface influence height and  $\mathbf{P}_s$  is surface pressure). The maximum surface contact height is just a tool to allow the model to work out surface residence time, and the actual contact time should not differ if this is raised or lowered. The same is true of the number of particles. The output from NAME must be divided by the number of particles (the pressure fraction is actually removed from the  $\mathbf{y}$  element of Equation 6.4, it is discussed here as this is where it is physically relevant). Output from NAME is in multiples of 900, the number of seconds in each time step, which means that no factor is needed to correct this. Because the final output is desired in days, the elements are also divided by the number of seconds in a day, before the total time for each surface time for that run is summed up. The weighting function matrix,  $\mathbf{K}$ , is then obtained by running the same calculation for every NAME run with a corresponding measurement.

If desired, the release column can be weighted to the SCIAMACHY averaging kernel. This can be done by releasing particles at ten different heights with 1000 metre intervals from the surface up to 10 kilometres (it would not be possible for particles above 10 kilometres to touch the surface, so these are ignored). Pressures taken for every kilometre from the ECMWF 6 hourly 1.125° by 1.125° data set can be used to weight the number of particles released at vertical heights, which can be weighted again by the SCIAMACHY averaging kernel. The model is already heavily weighted towards the lowest part of the column (because there are more particles originating here and because those particles are more likely to be inside the maximum surface height), and so weighting with respect to the averaging kernel will have little effect on the distribution of surface residence time.

Getting the measurement vector,  $\mathbf{y}$ , in the right units is a much more compli-

cated procedure. The first stage is to have one mixing ratio value for the release square and one for the background. The release square value will be the mean of the retrieved mixing ratio values of the SCIAMACHY footprints that pass all of the filters (assuming there is a minimum of 10 of them), and the measurement error covariance (which for SCIAMACHY foreground and CarbonTracker background, is going to only have contributions from SCIAMACHY) is the standard deviation of the individual footprints (also in ppmv).

The background VMR was obtained from the weighted distribution of the global ( $6^\circ$  by  $4^\circ$ ) CarbonTracker weather grid squares around the outside of the domain. This was coupled offline with the ten day average weighting of the  $1^\circ$  by  $1^\circ$  “edge box” particles, obtained using the methodology described in Section 6.2.3.1 (see Figure 6.10).

For every day in 2003 plus the last ten days of 2002, the mixing ratios at the 25 height levels of the CarbonTracker  $\text{CO}_2$  weather product for each of the grid boxes on the east, west, north and south boundaries of the domain are converted into a single vertical mixing ratio column with the SCIAMACHY averaging kernel applied. Then for every day in 2003, each grid box on the boundary consists of the mean of the vertical mixing ratio of the same column for that day and for each of the ten previous days (as particles could have left the scene on any of those days). Finally the weighting of the NAME domain edge  $1^\circ$  by  $1^\circ$  grid boxes was used to define the weighting of the larger CarbonTracker boxes, and this was used to find an overall background  $\text{XCO}_2$  column. There are eight CarbonTracker  $\text{CO}_2$  weather fields for each day, and the daily background was defined by the 9 am to midday product only.

A filter is applied so that only FSI footprint ensembles with a standard deviation of less than 10 ppmv are included in the inversion scheme. A second filter excludes NAME releases where the surface pressure (from ECMWF) is greater than 950 millibars, which is intended to discard measurements made over hilly terrain (which SCIAMACHY does not deal with very well).

The next stage is to calculate the mass of carbon dioxide in the foreground and background columns (plus the measurement covariance mass, since this uses the same equation). The mass of the foreground column can be worked out from the mass of air in the column. The magnitude of the weighting function matrix,  $\mathbf{K}$ , varies with surface influence height (set at 4000 metres in this test, but in future should be more appropriate at the minimum planetary boundary

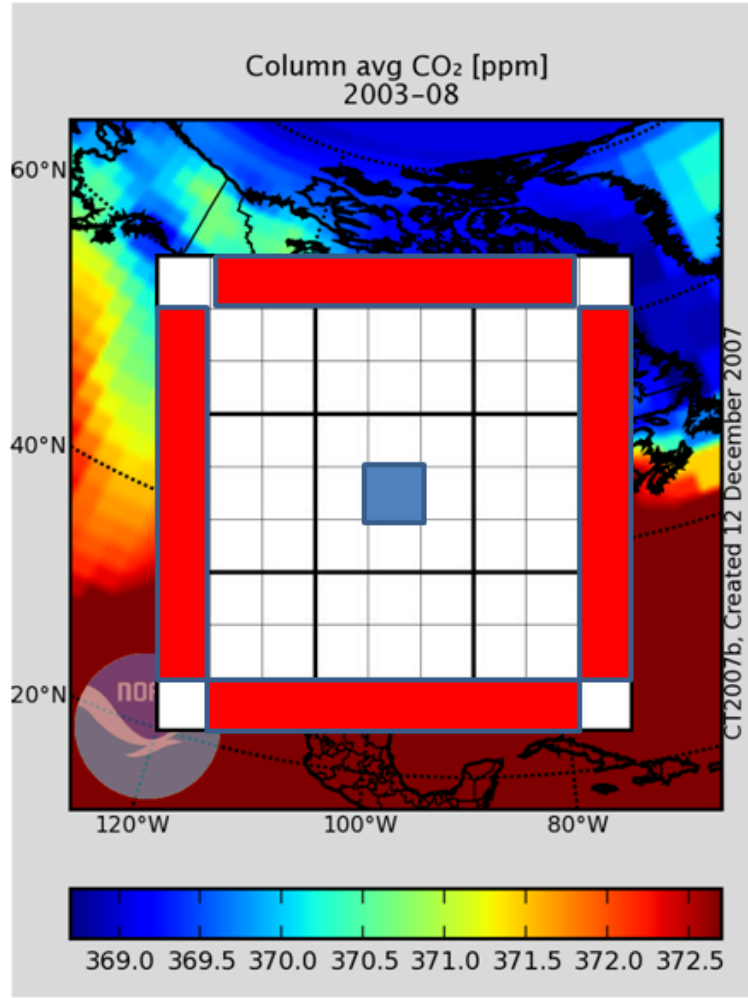


FIGURE 6.10. Particles travel backwards in time after being released (blue square). The time spent in the grid squares adjacent to the domain edge (shown in red) is used to weight the background when it is coupled to the CarbonTracker weather product (underneath). This coupling of the background  $\text{CO}_2$  concentration is carried out offline. This figure courtesy of NOAA has been modified by the author.

layer height of 100 metres, see Figure 6.3). It is more convenient to compensate for the variability of the surface influence height in the values of the measurement vector,  $\mathbf{y}$ , and the error covariance matrix,  $\hat{\mathbf{S}}$ . First subtract the surface pressure (in millibars) from the pressure at 4000 metres, see equation 6.7. This pressure difference can be converted to grams per square metre through:

$$1mb = 100 \frac{N}{m^2} = 1 \times 10^5 \text{grams} \times g \quad (6.6)$$

The gravity,  $g$ , was calculated for the specific latitude and altitude. To obtain the mass of carbon dioxide in the measurement column (in grams per square metre),

multiply the mass of air between the ground and 4000 metres by the ratio of CO<sub>2</sub> mass to air mass,  $\frac{44.0}{28.8}$ , then multiply by the measured volume mixing ratio at the release square. The same calculation was made to obtain the background CO<sub>2</sub> column mass and the measurement error CO<sub>2</sub> column mass. The measurement vector,  $\mathbf{y}$ , was obtained by subtracting the mass in the measurement column with the mass in the background column for each of the measurement pairs in that month. The error covariance matrix,  $\hat{\mathbf{S}}$ , was made by inserting the elements of the mass error column into a diagonal matrix.

The surface pressure can be easily obtained from the ECMWF 6 hourly, 1.125° by 1.125° data set. Because the pressure levels are not regularly spaced, obtaining pressure at 4000 metres requires a few steps. The closest ECMWF grid square to the centre of the release grid square is used to obtain this pressure. The 60 temperature, humidity and pressure values for that square plus the surface elevation are used to identify the heights of each level. The heights are found by incrementing the hydrostatic equilibrium equation:

$$dz = - \left( \frac{dP}{P} \right) \times \left( \frac{RT}{Mg} \right) \quad (6.7)$$

Gravity,  $g$ , can again be calculated for the specific latitude and altitude. To get the pressure at exactly 4000 metres, the 60 pressures are converted to log pressures which are then interpolated to obtain the log pressure at 4000 metres.

### 6.3.2.2 Majority rule for MODIS land product

Where the MODIS land classification [Friedl et al., 2002] was used in this thesis for forward and inverse modelling, the author has employed a majority rule for labeling 1° by 1° grid squares. The finest resolution 5km MODIS product was unsuitable for inverse modelling, as not only would the NAME output be increased by a factor of approximately  $20 \times 20$ , but it would also mean that many more release particles would be needed to produce a gaussian distribution in the surface contact time. The 1° by 1° grid resolution meant using either a majority system, where the mode land classification type is used to label each grid square or alternatively each grid square is labeled by the fraction of each land classification type. The MODIS product has been used for testing the inverse and forward modelling techniques developed in this thesis, since it is the intension of the author that these tools incorporate



Table 6.1. The different weighting of selected North American land classifications (shown in Figure 6.8) of the MODIS land product for the finest scale and for the 1° majority rule. The factor that using the majority rule changes this weighting is also shown.

land classification	weighting [%]		factor change
	fine scale	1° majority	
water	47.8	47.3	0.99
evergreen needleleaf forest	8.21	10.6	1.29
deciduous broadleaf forest	1.61	1.70	1.06
mixed forest	4.15	3.66	0.88
open scrubland	16.6	20.0	1.20
woody savanna	4.43	2.50	0.56
grasslands	6.25	5.59	0.89
croplands	4.28	4.50	1.05
crop/veg mosaic	2.42	2.00	0.83
barren	2.48	1.45	0.58
savanna	0.66	0.068	0.10

homogeneous flux ecoregions to identify fluxes, consequently the simpler majority rule was used in testing.

Comparing the total weighting of the North American scene (10° to 85° north and 170° to 20° west) between the fine scale and the 1° majority grid squares has shown that the weighting of certain land classifications are significantly distorted by applying the majority rule. Water is a special case in this scene (containing the Atlantic and Pacific Oceans), since almost every box that contains water is entirely water. Relatively few water containing grid squares are coastal grid squares - among the few locations where the majority system could push it into either water or another land classification (small inland lakes are not likely to make a majority water grid square). As a result, there is very little difference between the two systems.

There is an obvious bias in using the majority system, in that it over represents the land types which are most dominant in the scene. The two most common land types, after water, are open scrubland (over weighted by a factor of 1.2) and evergreen needleleaf forest (over weighted by a factor of 1.29). This is best understood when considering that many grid squares could be 60 to 90% of a

common land type interspersed with patches of much rarer land types. As a result, the rarer land types are significantly under represented using the majority rule (see Table 6.1).

### **6.3.3 Development of a forward modelling technique to investigate CO<sub>2</sub> fluxes**

Although this technique is a great deal simpler than the inversion, it can be adapted to investigate many different questions. In each example, the output will be the mass change per unit area of carbon dioxide in the release column. There are two methods that can achieve this output (see Sections 6.3.3.1 and 6.3.3.2) and it is desirable to compare these results.

The reader should not be confused between the “forward” modelling technique described in this section, and the fact that the NAME model is run in reverse time to derive the dilution matrix (or surface residence time). See Figure 6.11 for an overview diagram of the forward modelling method developed in this research project.

#### **6.3.3.1 Mass change from coupling NAME with a pre-existing flux product**

Before particles are released from the “measurement” column, the output grid is set up to have the same resolution as the pre-existing flux product. The domain boundary defined in the NAME run should not be greater than the domain of the flux product. If the flux product has a time resolution (CarbonTracker has 3 hourly fluxes), the NAME model can also be set with the same time resolution if desired (although it may be of interest to obtain a 12 hourly NAME dilution matrix).

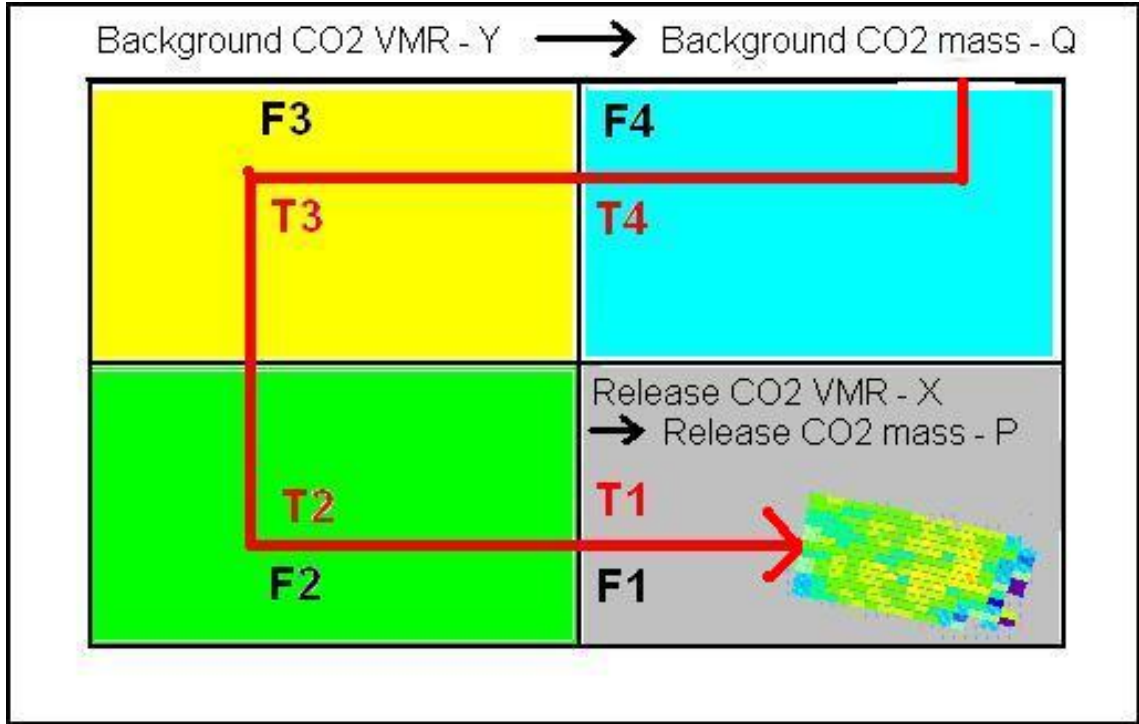


FIGURE 6.11. Overview of the forward modelling scheme used to investigate CO<sub>2</sub> fluxes. It may be the intension of the forward modeler to use SCIA-MACHY/FSI CO<sub>2</sub> to validate the carbon flux magnitudes of a model such as CarbonTracker (see Section 2.2.1) or it may be be desirable to perform a synthetic test to validate the initialisation method (as described in Section 7.1); in either case the basic methodology is the same (the diagram shows the former example). NAME is run backwards in time from the receptor location producing surface influence and background origin output. The origin information is weighted so that it sums to unity. A modelled CO<sub>2</sub> field, such as CarbonTracker, is multiplied by this weighting to give the background atmospheric CO<sub>2</sub> concentration of the receptor location air (labeled Y on this diagram). The mass change of the receptor air can be calculated by multiplying the flux magnitudes (labeled F on this diagram) of each grid square by the surface influence time (labeled T on this diagram), and summing the total (in North America this would be on a 1° by 1° resolution, for simplicity the diagram contains just four flux regions). The mass change over the scene plus the mass of the CO<sub>2</sub> in the background column (labeled Q on this diagram, this is calculated from Y) should equal the mass of CO<sub>2</sub> in the receptor column (labeled P on this diagram). Since there will be many releases, it is possible to plot a graph of  $Q + (F \times T)$  vs P (for example the synthetic test of “edge box” initialisation test shown in Figure 7.2).

Particles are released and followed in reverse time for a set release period. The dilution matrix time from NAME is first divided by the number of seconds in a day (assuming that the model fluxes are in units of  $\text{grams}_{CO_2} \text{ metre}^{-2} \text{ day}^{-1}$ ), then treated in the same way as in Equation 6.5 to obtain the surface influence time array. If particles are released from only a part of the column, the equation looks like this:

$$\mathbf{t}_c = \mathbf{t}_r \times (\mathbf{N}_p)^{-1} \times \frac{(\mathbf{P}_{low} - \mathbf{P}_{high})}{(\mathbf{P}_s - \mathbf{P}_{sih})} \quad (6.8)$$

where  $\mathbf{t}_c$  is the surface contact time,  $\mathbf{t}_r$  is the model run time,  $\mathbf{P}_{low}$  is the pressure at the lowest part of the release column and  $\mathbf{P}_{high}$  is the pressure at the highest part of the release column. Multiplying the surface influence time array by the flux array and summing up for all locations and times produces the total mass flux in the column (or part of column) in units  $\text{grams}_{CO_2} \text{ metre}^{-2} \text{ day}^{-1}$ .

### 6.3.3.2 Mass difference between release column and background

If particles are released in only part of the column, the mass of air can be obtained from Equation 6.6 by taking the pressure difference between the lower and upper boundary of the release box. For a whole column release the lower limit is the surface pressure and the upper limit is zero pressure.

Take the volume mixing ratio in the release column (from a model or a measurement) in ppmv and multiply by one million to get the fraction of  $\text{CO}_2$  molecules in the air. Multiply by the mass of air in the column and multiply by the ratio of  $\text{CO}_2$  mass to air mass,  $\frac{44.0}{28.8}$ , to get the mass of carbon dioxide in the (part of the) column. Obtain the background concentration (by for example the “edge box” method in Section 6.2.3.1), and do the same. Subtract the background carbon dioxide mass from the release box mass to obtain the total mass difference in the column (or part of column) in units  $\text{grams}_{CO_2} \text{ metre}^{-2} \text{ day}^{-1}$ .

This set up has been used in a synthetic test of the proposed “freeze” initialisation method (see Section 7.1.2).

## 6.4 Summary

In this chapter, a method was developed in which level 2 satellite CO<sub>2</sub> data (the individual satellite footprint data), specifically from SCIAMACHY but more generally for sun-synchronous satellite instruments, can be used with a Lagrangian dispersion model to obtain information about the surface exchange of carbon between the terrestrial biosphere and the atmosphere.

In Sections 6.2.2.1, 6.2.2.2, 6.2.2.3 and 6.2.2.4 it was shown how information could be obtained about where and when column air (represented in the model by particles) spent time in contact with the surface (thus experienced the surface fluxes), prior to being measured by the satellite.

In Sections 6.2.3 and 6.2.3.1 it was shown how to initialise the background concentration of the released air, thus obtaining a CO<sub>2</sub> mass difference in the domain (from Section 6.3.3.2), with which to compare to the surface fluxes.

After manipulating the NAME model to output the information above, two distinct methods for investigating the surface fluxes were developed. The inverse method (Section 6.3.2), makes use of the assumption of homogeneous flux regions to obtain flux information from level 2 satellite CO<sub>2</sub> data. The forward model (Section 6.3.3) can be used in many different ways (discussed in Chapter 8), and involves coupling the NAME surface influence time output with a pre-existing surface flux model (see Section 6.3.3.1).

In Chapter 7, the forward modelling and inversion techniques will be tested using synthetic satellite measurements.

# Chapter 7

## Synthetic testing of forward and inversion method

In Section 6.3.2, an inverse method was developed for investigating the surface exchange of carbon between the terrestrial biosphere and the atmosphere. In this chapter, a number of idealised synthetic tests are employed to validate the tools developed in this research project. Initialisation is the most challenging area of Lagrangian modelling; in Section 7.1 synthetic testing is performed on the “edge box” and “freeze” methods used in both forward and inverse modelling.

A method for testing the inversion technique is demonstrated in Section 7.2. This idealised experiment is used to show how varying the satellite instrument precision (i.e. using different satellites such as SCIAMACHY, GOSAT, OCO-like) would affect the *a posteriori* carbon flux error covariance,  $\hat{\mathbf{S}}$  (see Equation 6.4).

### 7.1 Validation of initialisation method using CarbonTracker

The most difficult challenge to overcome in the inversion of a Lagrangian model is that of initialisation. Any improvements to the precision of the measurement at the release location is meaningless if the background concentration cannot be characterised. For a gas such as carbon dioxide, whose variation in the atmosphere is much smaller than its mean concentration, this is particularly important; how can the mass flux between the background and the measurement be measured if the background concentration is unknown?

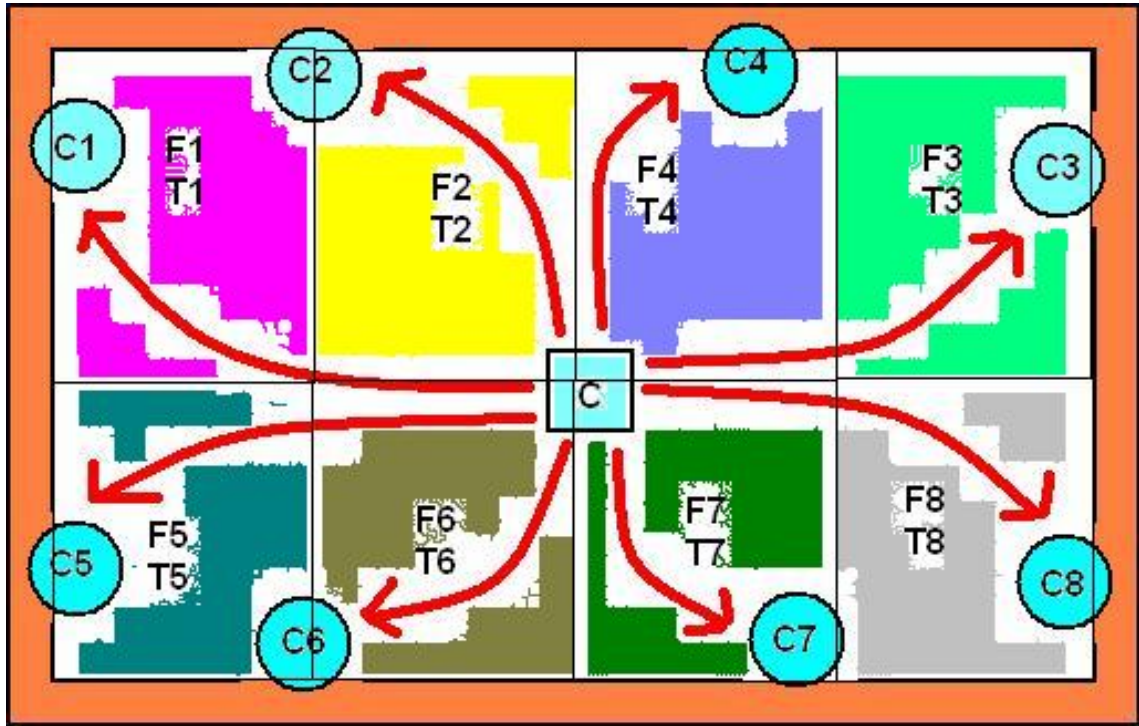


FIGURE 7.1. Overview of the “edge box” and “freeze” initialisation methods. In the validation of both the “edge box” and “freeze” initialisation methods, the release domain contains 64000 (20 latitudinal  $\times$  40 longitudinal  $\times$  80 temporal) surface influence output times for each of which exists a CarbonTracker carbon flux. This diagram simplifies the NAME output (shown as red arrows) as containing 8 spatial regions without temporal resolution. Multiplying the output residence times (T1, T2, etc) by the corresponding CarbonTracker flux product (F1, F2, etc) and then taking the sum of all the boxes, produces the CO<sub>2</sub> mass change in the domain (the vertical axis of Figures 7.2, 7.3 and 7.4). In the “edge box” method, the residence time of particles in the whole column (i.e. with no vertical or time resolution) in the very outside grid squares of the domain (shown in peach colour) was recorded. The residence time for each edge-box was weighted to the total edge-box residence time. A 10-day average of CarbonTracker weather XCO<sub>2</sub> (mixing ratio in the whole column) was associated with each edge-box (shown as C1, C2, etc) and this was multiplied by the weighted time. The total of all the boxes gives the background CO<sub>2</sub> mass and this is subtracted from the release CO<sub>2</sub> mass (C) to give the CO<sub>2</sub> mass change (the horizontal axis of Figure 7.2). The “freeze” method is similar but particles do not leave the scene; after every three hours the grid box of the release particle (including vertical resolution) is used to weight the background CarbonTracker weather concentration and thus obtain the CO<sub>2</sub> mass change (Figures 7.3 and 7.4). For the next three hour step, the previous background concentration (the weighted sum of C1, C2, etc) becomes the release CO<sub>2</sub> mass (C).

### 7.1.1 Validation of “edge box” initialisation method

The first test concerns the “edge box” initialisation method that was used in the inversion of SCIAMACHY CO<sub>2</sub> (see Section 6.2.3.1). Here, a weighted background was obtained by coupling the CarbonTracker weather 1° by 1° product with the surface influence height output from NAME.

The following describes the online NAME setup. Particles left the domain if they moved outside the range 30° - 50° north and 80° - 120° west. Particles were randomly released (over a 1 hour period) in a 1° by 1° box centred on 40° north and 100° west, covering a height range of 0 to 10 kilometres (the vertical release was randomly distributed, weighted by the pressure altitudes). As with every run, the planetary boundary layer height was allowed to vary between 100 and 4000 metres. The model was run for 31 days in August 2003 using a release time of around 10 am local time. The maximum run length was 10 days backwards or until every particle left the domain. Two output types were produced, the first was surface influence time for every three hours for each of the one degree squares in the domain. The second output gave the number of particles between 0 and 10 kilometres for every three hourly time step for the grid squares on the very edge of the domain (80° - 81° west, 119° - 120° west, 30° - 31° north, 49° - 50° north).

The following describes the “edge box” initialisation method. The NAME model setup outputted residence time of particles in the edge 1° boxes for each of the 80 3-hour time periods and this was weighted by the total “edge box” residence time. Concurrently, the equivalent CarbonTracker weather XCO<sub>2</sub> columns were obtained for each of the 1° by 1° boxes for each of the 80 time periods by pressure weighting the 25 CO<sub>2</sub> levels. The background CO<sub>2</sub> value was then obtained by coupling the weighted NAME output with CarbonTracker weather. The foreground CO<sub>2</sub> concentration was simply obtained from the pressure weighted concentration of the 25 levels at the time and place of the particle release. Figure 7.1 provides a conceptual overview of this validation method.

A possible source of error will come from the fact that some particles will spend many of the 15 minute NAME time steps in the edge grid squares, whilst others will spend fewer. Any errors resulting from treating CarbonTracker levels as layers are likely to be small in this case owing to the pressure weighting and as other than the first height, the variability between the levels is small.



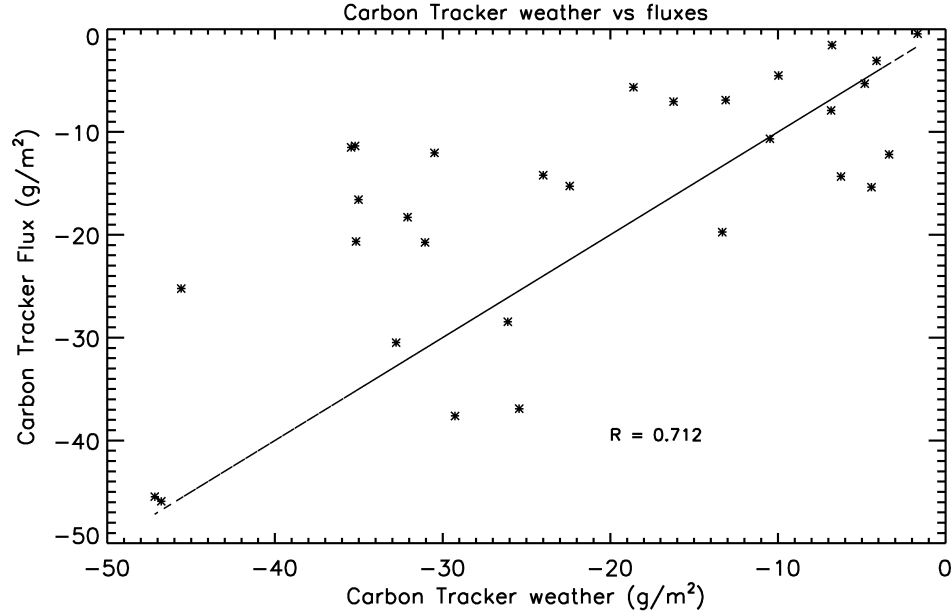


FIGURE 7.2. Initialisation test using edge method.

The CarbonTracker weather product was produced by adding the result of the CarbonTracker flux product to the previous CarbonTracker weather modified by the wind field (this is described in Section 2.2.1). Although the CarbonTracker wind fields are not based on the UM, they should be very similar. Consequently, it would be reasonable to assume a one to one relationship between the mass change owing to CarbonTracker fluxes multiplied by the NAME surface influence time (see Section 6.3.3.1) and the mass change between the foreground and background columns (see Section 6.3.3.2).

Figure 7.2 shows the 31 release days from the month of August 2003, both in units of grams per square metre. A one to one line has been included as a guide, although the difference in the weather product does seem greater than that of the flux product. The correlation coefficient  $R^2 = 0.71$  indicated that the edge square method of initialising the background could be improved upon. In this example the maximum surface influence height was set as 100 metres, only a negligible difference was seen when this was set as 4000 metres.

### 7.1.2 Validation of “freeze” initialisation method

The precision of the background initialisation could be improved upon by producing a NAME output format which could specify the location of each particle in time and space just before it leaves the domain (as described in Section 6.2.3). In this way, it would be possible to “freeze” the CO<sub>2</sub> concentration of each particle as it exits the domain and average out over all the particles to get the background. However, this is a rather complicated output, requiring changes to the operation of NAME, which have yet to be implemented. A variation on the proposed “freeze” method was devised to investigate how much improvement could be made by adopting a method that records when particles leave the domain.

The following describes the online NAME setup. Instead of letting the particles leave the domain (as in Section 7.1.1), particles that initially started at the release box were followed backwards in time and a mass flux was calculated between each 3-hour time step. This is possible whilst none of the particles leave the domain. Additionally, in this case particles were only released from the lowest CarbonTracker layer (between 0 and 80 metres). To work out the concentration for the two time steps, NAME outputs were generated for every three hours for every 1° by 1° box and for each of the CarbonTracker layers in the troposphere. This should result in a greater spread of mass fluxes, since the mass change between 10 am and 1 pm is likely to be strongly negative and between 10 pm and 1 am it should be positive.

The first time step (between 7 am and 10 am on the day of release) for each of the 13 runs produced a very large mass change in the part of the column (between 0 - 80 metres) where the particles were released (mass changes are thus calculated using Equation 6.8). Because the bottom CarbonTracker weather level varies the most during each 3 hour period the first time step has a mass change greatly outside the order shown in Figures 7.3 and 7.4. For each of the 13 days used in this example the ratio of the mass change for the first period is not one to one; this is very likely to be a result of treating CarbonTracker levels as layers. This does not seem to affect any of the other time steps. Ignoring the first time step, the next 23 mass changes (three days backwards in time) for each of the 13 release days are shown in Figures 7.3 and 7.4.

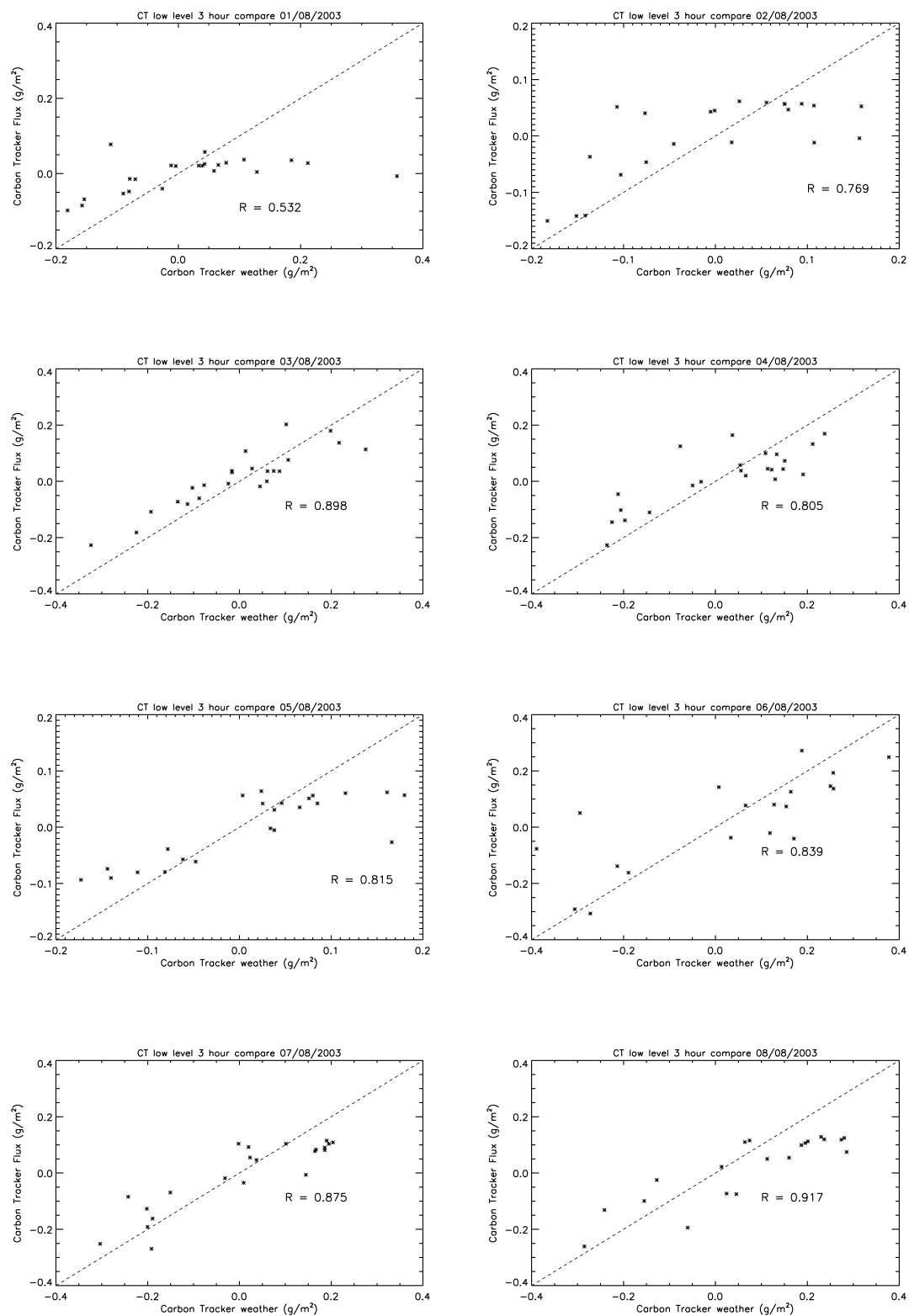


FIGURE 7.3. Initialisation test using “freeze” method days 1 - 8.

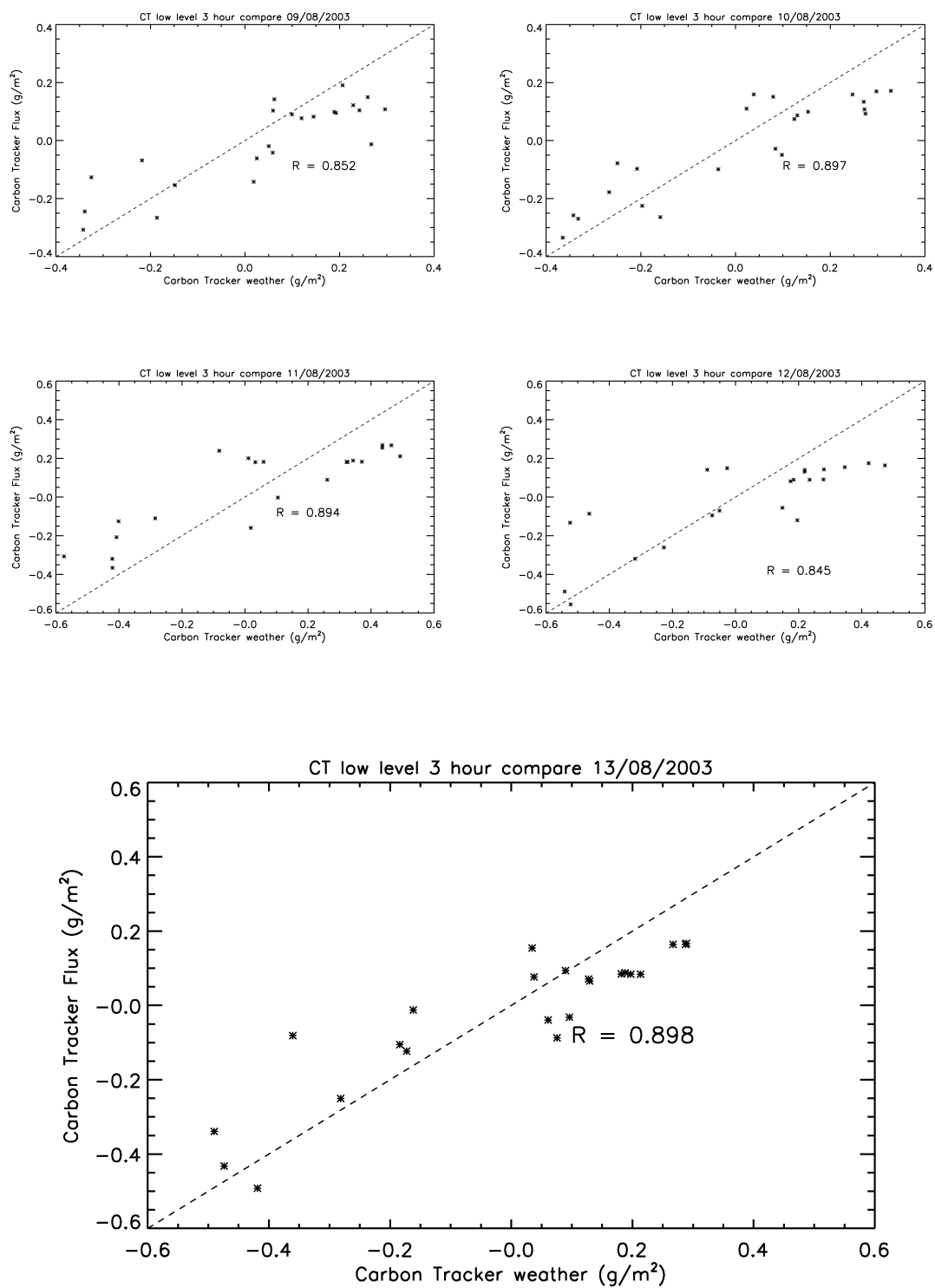


FIGURE 7.4. Initialisation test using “freeze” method days 9 - 13.

There is a significant improvement in the correlation coefficient, between the “edge box” initialisation method ( $R^2 = 0.71$  shown in Figure 7.2) and the “freeze” initialisation method ( $R^2$  typically around 0.85, shown in Figures 7.3 and 7.4). The two days where the correlation coefficient is lower than the edge box method (the 1<sup>st</sup> and 2<sup>nd</sup> of August), were days when the effect of the CarbonTracker fluxes to the mass change were very small.

The strength of the correlations shown in Figures 7.3 and 7.4 indicate that it is possible to obtain precise background values of carbon dioxide using NAME with either measured or modelled concentrations.

On the 1<sup>st</sup>, 2<sup>nd</sup> and 5<sup>th</sup> of August, the magnitude of the mass change owing to the change in the CarbonTracker weather product is greater than that of the residence time multiplied by the CarbonTracker mass flux. This should not be the case, since the weather product is a direct consequence of the flux product. It is possible that this bias is an artifact of the different wind fields used in the two models or the height at which the fluxes are emitted into the CarbonTracker weather product. In this test a maximum surface influence height of 100 metres was specified.

## 7.2 Validation of inversion method

With the question of initialisation solved (see Section 7.1), it is possible to invert the atmospheric carbon dioxide concentrations. An idealised experiment was set up in order to test the performance of the inversion scheme developed in this thesis (Section 6.3.2, see method overview Figures 7.5 and 7.5). As well as demonstrating that the method developed does work, it also tested the relationship between measurement precision and the precision of the inversion.

The following describes the online NAME setup. 211 NAME output files from August 2003 (where  $\geq 10$  FSI-retrieved columns satisfied the criteria described in Section 6.3.1) were used in the inversion. Particles were released from 3° by 3° grid boxes between 0 and 10 kilometres. The domain covers the area 170° - 20° west and 10° - 85° north. Particles were followed backwards in time for 10 days or until they exit the domain. The maximum surface influence height was defined as 4000 metres (the same height was used in the forward and inverse modelling stages). The NAME output recorded the surface influence time for each of the 150 by 75 grid squares (without any time of day information).

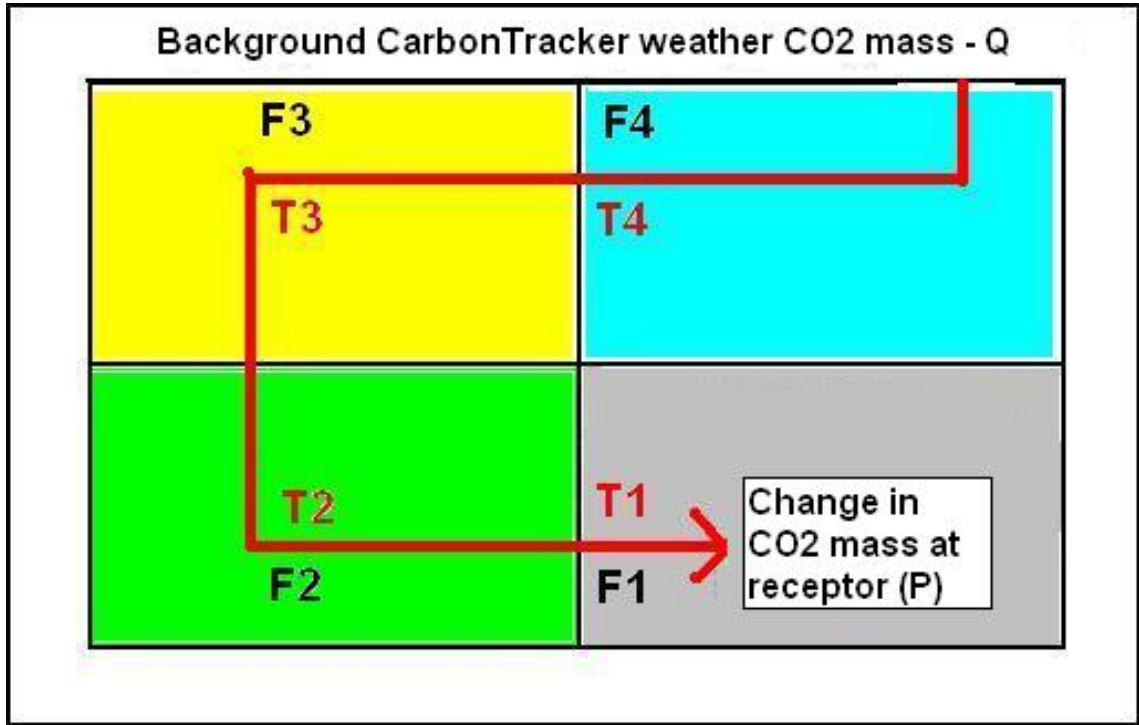


FIGURE 7.5. Overview of the synthetic test used to validate the linear regression inversion scheme (developed in Section 6.3.2). The method used the modelled NAME output from all 211 releases from August 2003 (those with a corresponding SCIAMACHY/FSI ensemble measurement). The NAME output was coupled with seven synthetic homogeneous carbon-flux ecoregions (located as shown in Figure 7.7), of various sizes (therefore varying surface influence times, it is expected that *a posteriori* carbon flux error covariance would be inversely related to surface influence time). The forward modelling approach (shown here, see Section 6.3.3) was used to calculate the mass change (shown as  $P$ ) at the receptor for each of the 211 NAME model runs. One out of four possible simulated “satellite ensemble measurement errors” was then added:  $\pm 0.1$  ppmv (a nearly perfect test),  $\pm 1.0$  ppmv (equivalent to a GOSAT or OCO-like ensemble),  $\pm 3.0$  ppmv (equivalent to a SCIAMACHY/FSI ensemble) or  $\pm 6.0$  ppmv (a very large measurement error), obtaining a mass change at the receptor ( $P + dP$ ). This overview is continued in Figure 7.6.

The following describes the offline forward method. The surface was defined to have seven synthetic homogeneous emission regions, so that each of the 150 by 75 grid squares of that type emitted at the same rate (regardless of time of day). In order to only validate the inversion part of the code (not the initialisation), the background concentration for every release was set to be 370 ppmv of  $\text{CO}_2$ . The surface influence time was coupled with the fluxes and summing over the 150 by 75 grid squares produced the mass change in the release column (from Section 6.3.3.1). This was converted into a concentration (the reverse of Equation 6.6)

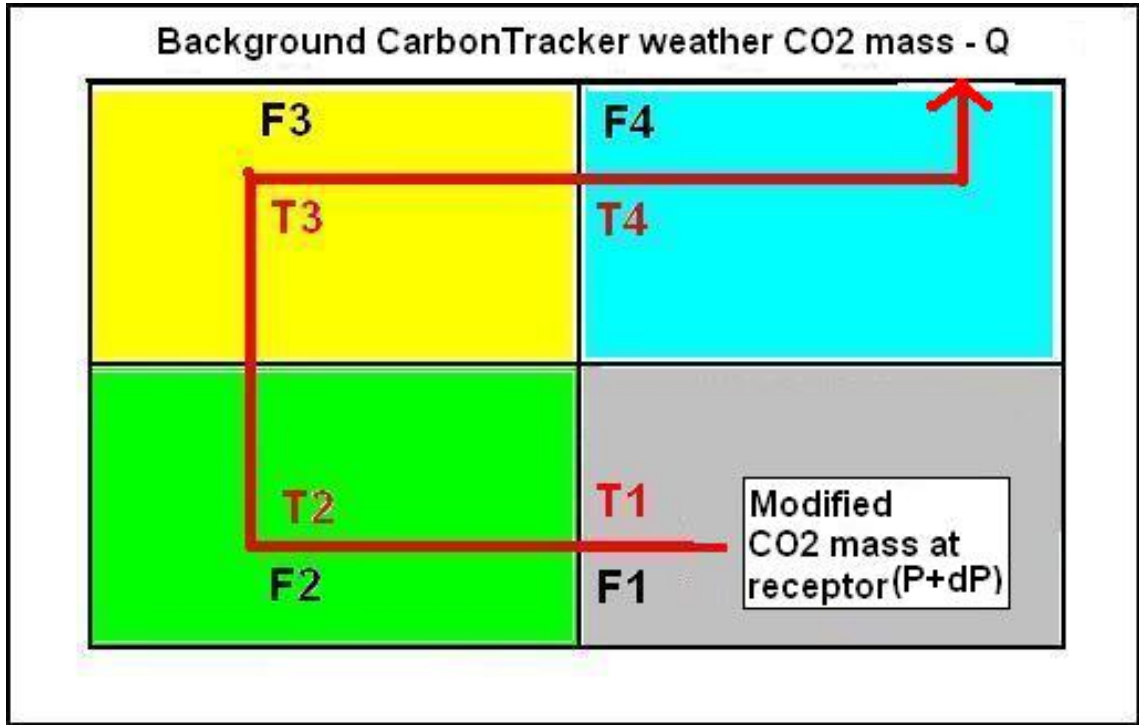


FIGURE 7.6. Overview of the synthetic test used to validate the linear regression inversion scheme. Continuing the method overview from Figure 7.5, the mass change over the scene ( $P + dP - Q$ ) was used in the linear regression inversion scheme (shown here, see Equation 6.3). The results of the inversion scheme validation are shown in Table 7.1.

and was added to the background to produce 211 simulated concentrations at the “measurement” columns.

Rather than use the pre-existing MODIS land classification map, it was decided to create one for this experiment (see Figure 7.7). The first land type was the equivalent of the ocean for the North American domain, making up roughly two-thirds of grid squares but concentrated on the domain boundary. The other six land types were set in the middle and account for ever decreasing numbers of grid squares. This was intended to show that the *a posteriori* error covariance,  $\hat{\mathbf{S}}$ , (i.e. the precision of the inversion result) is not the same for each land classification and would be reduced as the amount of surface influence time is increased (see Equation 6.4). It turned out that the surface influence time for the second land type was longer than that of the first, simply because particles did not spend as much time near the domain edges.

The following describes the offline inversion method. A simulated measurement error was added to each of the 211 measurement columns (it was not necessary to demonstrate the working of this method to add measurement error on to the

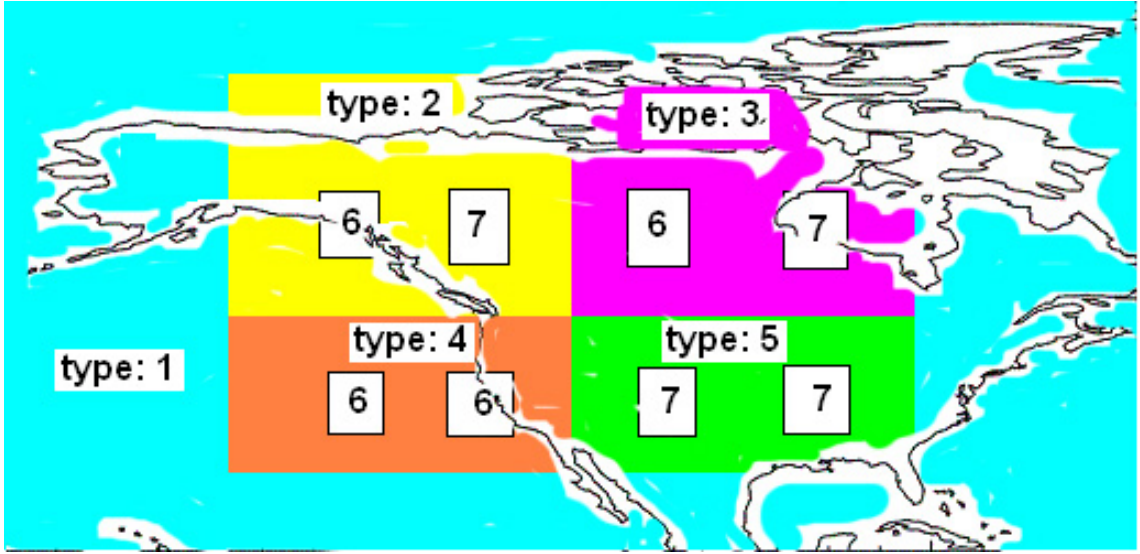


FIGURE 7.7. Spatial diagram showing the arbitrary layout of the synthetic homogeneous carbon flux ecoregions used in the linear regression inversion method test (described in Section 7.2). The first synthetic ecoregion approximately represents the very weak emitting background region to the scene (it is intended that the domain edge of any future scene be over ocean). The weighting of the scene by surface area does not necessarily reflect the weighting of surface influence time (for a poorly sampled ecoregion it is expected that the *a posteriori* carbon flux error covariance would also be poor). Surface influence times of the seven regions are shown in Table 7.1.

background also). The measurement error was chosen to be a random noise added to each of the 211 measurements that was not greater than a set threshold. The four settings presented here are  $\pm 0.1$  ppmv,  $\pm 1.0$  ppmv,  $\pm 3.0$  ppmv and  $\pm 6.0$  ppmv. A linear regression inversion was then run with the new values of simulated  $\text{CO}_2$  columns (see Equation 6.3) with no *a priori* constraints (all *a priori* fluxes set to 0 grams per square metre per day, and *a priori* covariance set to an arbitrarily large number).

Taking the case of nominal (0.1 ppmv) measurement error, it is clear that the inversion scheme works (see Table 7.1). Even for the very rare land classification types, the method was able to return the original flux magnitudes. For the largest simulated error, the flux for (heavily sampled) land types 1 and 2 were overestimated, resulting in very underestimated fluxes for the less sampled land types. For comparison, GOSAT and OCO-like instruments would have an ensemble measurement uncertainty,  $\mathbf{S}_e$ , of 1-2 ppmv, with SCIMACHY/FSI around 3-6 ppmv.



Table 7.1. The *a posteriori* fluxes and error covariance from the linear regression inversion of synthetic CO<sub>2</sub> columns. Using the methodology described in Section 7.2 (see overview Figures 7.5 and 7.6), a synthetic validation of the linear regression inversion scheme described in Section 6.3.2 was performed. The region was split into seven homogeneous flux ecoregions (shown in Figure 7.7). Equation 6.3 was used to obtain the best solution for the seven homogeneous flux ecoregions on four separate occasions: using a negligible synthetic measurement error of  $\pm 0.1$  ppmv, using a synthetic measurement error of  $\pm 1.0$  ppmv (similar to an ensemble of GOSAT or OCO-like measurements), using a synthetic measurement error of  $\pm 3.0$  ppmv (similar to an ensemble of SCIAMACHY/FSI measurements) and using a very large synthetic measurement error of  $\pm 6.0$  ppmv. The near perfect ( $\pm 0.1$  ppmv) measurement error case returned almost exactly the forward modelled fluxes for even the least sampled regions, whereas the very large ( $\pm 6.0$  ppmv) measurement error case struggled to return the forward modelled fluxes even for heavily sampled regions. The *a posteriori* flux covariance,  $\hat{\mathbf{S}}$ , appears to scale linearly to the *a priori* measurement error covariance,  $\mathbf{S}_\epsilon$ . The *a posteriori* flux covariance,  $\hat{\mathbf{S}}$ , also appears to be inversely proportional to the total surface influence time (of all 211 10-day NAME runs).

	simulated land classification						
	1	2	3	4	5	6	7
total surface influence time (days)	305.9	331.9	117.5	77.32	74.93	51.53	18.16
Original flux ( $\text{g}_{\text{CO}_2} \text{ m}^{-2} \text{ day}^{-1}$ )	0.000	-7.000	3.000	2.000	-4.000	5.000	1.000
measurement error covariance, $\mathbf{S}_\epsilon$	<i>a posteriori</i> fluxes, $\hat{\mathbf{S}}$						
0.1 ppmv	0.017	-7.008	3.006	1.946	-3.947	4.960	0.965
1.0 ppmv	0.062	-7.047	2.986	1.902	-3.779	4.852	0.310
3.0 ppmv	-0.559	-7.029	2.426	1.448	-1.919	4.873	4.997
6.0 ppmv	1.270	-6.400	1.570	-0.468	-4.302	0.400	-9.236
measurement error covariance, $\mathbf{S}_\epsilon$	<i>a posteriori</i> flux covariance, $\hat{\mathbf{S}}$						
0.1 ppmv	0.024	0.018	0.057	0.050	0.103	0.117	0.215
1.0 ppmv	0.241	0.178	0.569	0.506	1.028	1.164	2.135
3.0 ppmv	0.719	0.530	1.687	1.507	3.048	3.444	6.130
6.0 ppmv	1.410	1.045	3.258	2.954	5.869	6.583	10.82

The diurnal variability of the homogeneous ecoregions was ignored in this inversion. If several time periods per grid square were to be used in the inversion scheme (assuming that measurements are only taken at one point in the diurnal cycle) it is essential that there are differences in the residence times for the different parts of the diurnal cycle. It is expected that real homogeneous ecoregions will be blocked together over vast areas, and that released particles from a particular column that spend day time over a particular ecoregion are likely to spend night time there as well; the consequences of this are that it is not possible to separate out the different diurnal elements in the inversion scheme. It would be possible to artificially create a very mixed landscape in order to separate out the diurnal elements, but this defeats the purpose of simulating the results and would not be any different from the results shown in Table 7.1.

Another major drawback of including the diurnal cycle in the inversion scheme is that it dramatically increases the number of different ecoregions. One consequence of this is that the total surface residence time is reduced, which has been shown to increase the *a posteriori* covariance (in other words reduce the precision of the inversion results). The other main consequence is that the inversion scheme will, at some point switch from being an over-defined to an under-defined system. If there are 7 ecoregions  $\times$  8 parts of the diurnal cycle, there would be 56 solutions, with just 211 measurements (or even fewer if inversion period is reduced from a calendar month to a week).

In this idealised experiment there seems to be a linear relationship between measurement error and *a posteriori* covariance error. The covariance error is a better guide to the performance of the retrieval than the fluxes; for example the *a posteriori* flux for land type 5 is actually closer to the real solution for 6 ppmv measurement error than for 3 ppmv, which is the result of chance. It would not be unreasonable for an ensemble of OCO-like or GOSAT CO<sub>2</sub> column retrievals to achieve a measurement precision of 1 ppmv. At this precision it would seem possible to obtain reasonable flux magnitudes, except for land types with little surface residence time. Without a strict cloud filter, SCIAMACHY column measurements are not likely to be improved by taking an ensemble, and would seem to be closer to the 6 ppmv measurement error or 2%. Inversion of SCIAMACHY retrievals are likely to produce much less reliable flux information for common land types and will not be able to provide meaningful information about the rarer land types.

## 7.3 Summary

It has been shown that the adoption of the “freeze” initialisation method into the NAME model would provide a level of improvement over the “grid box” method that would justify the considerable amount of work needed to achieve this.

Chevallier et al. [2007] suggested a similar validation approach (to that applied in section 7.2) for testing their Bayesian inference inversion scheme. Rather helpfully for this comparison, Chevallier et al. [2007] quotes monthly averaged rather than annual flux error covariance. They suggested an OCO measurement covariance of 2 ppmv and also perturbed both the wind fields and the forward modelled fluxes (rather than keep them homogeneous). They used the Terrestrial Uptake and Release of Carbon (TURC) model [Lafont et al., 2002] to simulate the biosphere to atmosphere exchange of CO<sub>2</sub>. The linear regression inversion method results from Table 7.1 shows an improvement on the TURC modelled monthly averaged carbon flux uncertainties ( $\pm 4g_C \text{ m}^{-2} \text{ day}^{-1}$ ) and possibly over the Bayesian inference inversion of OCO-like XCO<sub>2</sub> (which reduced the TURC flux uncertainties by 60% [Chevallier et al., 2007]). The linear regression inversion method also performs well compared with CarbonTracker monthly averaged flux uncertainties over North America (approximately  $\pm 4g_{CO_2} \text{ m}^{-2} \text{ day}^{-1}$  [ESRL, 2008a]). The ensemble Kalman Filter inversion developed by Feng et al. [2009] will also provide separate carbon fluxes for different land classifications. However, a comparison with the inversion of synthetic OCO measurements was hampered by the complicated units given by Feng et al. [2009].

Overall the synthetic test of the inversion method demonstrates the improvements that GOSAT or an OCO-like instrument could provide to the understanding of regional scale carbon fluxes.

In Chapter 8, the forward and inverse modelling techniques are tested using real SCIAMACHY/FSI XCO<sub>2</sub> data.

## Chapter 8

# Analysis of forward and inverse modelling of FSI

The synthetic testing performed in Chapter 7 suggests that the forward and inverse modelling methodologies (developed in Sections 6.3.3 and 6.3.2 respectively) would prove a useful additional tool in the investigation of biosphere to atmosphere carbon fluxes. As already discussed in Section 4.4, the satellite retrieved CO<sub>2</sub> columns from SCIAMACHY/FSI do not meet the requirements (for inverse modelling of satellite XCO<sub>2</sub> measurements) to improve upon the carbon fluxes provided by the inversion of ground network measurements [Miller et al., 2007]. Nevertheless, SCIAMACHY/FSI does provide a useful first test of real data on the forward and inverse methods and could provide superior carbon flux information than the ground network in regions where the sampling network is very sparse [Houweling et al., 2004].

The test of the forward modelling scheme in Section 8.1 makes use of the CarbonTracker weather product to initialise the background. As previously discussed in Section 2.2.1, the high spatial and temporal frequency of point location measurements over North America result in a very precise product (equivalent to satellite measurement precision  $\Delta = 0.05 \pm 2.7 \text{ ppmv}$  and  $R^2 = 0.83$ ) to initialise the background [Peters et al., 2007]. The offset-bias of approximately 1.5% of the column concentration between SCIAMACHY/FSI and CarbonTracker weather is an additional complication to the forward modelling test. The carbon flux information is also provided from CarbonTracker.

The test of the inverse modelling scheme in Section 8.2 also uses the CarbonTracker weather product to initialise the background. The linear regression inversion assumes a SCIAMACHY/FSI measurement error covariance,  $\mathbf{S}_\epsilon$ , of 6 ppmv (see

Equation 6.3). The scheme used the eight ecoregions (adapted from the MODIS land classification system) described in Section 6.3.2.

## 8.1 Analysis of FSI forward modelling results

The methodology used in the analysis of the forward modelling scheme which couples NAME output with SCIAMACHY/FSI-retrieved XCO<sub>2</sub> and the CarbonTracker weather and flux products was described in Section 6.3.3.1 (see Figure 6.11 for a diagrammatic overview). The mass change of CO<sub>2</sub> in the domain is obtained in two ways: by coupling the NAME surface residence time output with the CarbonTracker flux product and by subtracting the satellite-retrieved CO<sub>2</sub> mass at the receptor from that at the background (through coupling of the “edge box” residence time with the CarbonTracker weather product). This technique was used to validate the CarbonTracker flux product. Applying this scheme was challenging, some of the difficulties faced are described below. The results from comparing individual mass changes (Figures 8.1 to 8.7) were obscured by the high level of noise in the SCIAMACHY/FSI-retrieved XCO<sub>2</sub> columns (see Section 8.1.1). These results were reanalysed using the clustering method, described in Section 8.1.2, which reduced the noisiness of individual SCIAMACHY/FSI XCO<sub>2</sub> retrievals.

### 8.1.1 Comparison with individual releases

The following describes the online NAME setup. Particles were released from the 3° by 3° release box between 0 and 10 kilometres. Particles were followed backwards in time for 10 days (or until they left the domain limits of 10° to 85° north and 170° to 60° west) in 15 minute time steps. The total amount of time spent within the maximum surface influence height (4000 metres) in the 150 longitudinal by 75 latitudinal, 1° by 1° grid boxes was recorded in the NAME output. Also recorded was the total time spent between 0 and 10 kilometres in the “edge boxes”, as this information would be used to initialise the background concentration of the measured column (see Section 6.2.3.1).

The following describes the conditions for coupling NAME releases with SCIAMACHY/FSI ensemble measurements. The 3° by 3° NAME release boxes were gridded over the area 30° to 69° north and 130° to 61° west. To be active on any particular day, the NAME release box must contain a minimum of 10 SCIAMACHY/FSI-

retrieved XCO<sub>2</sub> columns that meet the criteria set out in Section 6.3.1, with an ensemble standard deviation less than 10 ppmv and surface pressure greater than 950 millibars. The measurement vector,  $\mathbf{y}$ , was obtained from the mean of the ensemble footprints and the measurement error covariance matrix,  $\mathbf{S}_e$ , was obtained from the standard deviation (as discussed in Section 6.3.1, using an ensemble of SCIAMACHY/FSI measurements does not greatly improve the precision owing to its lax aerosol filtering).

The first carbon mass flux (represented on the vertical axis of Figures 8.1 to 8.7) was obtained (for each release) by subtracting the SCIAMACHY/FSI-retrieved CO<sub>2</sub> mass at the release column from the weighted CO<sub>2</sub> background (CarbonTracker weather coupled with the “edge box” NAME output). In an attempt to remove the diurnal cycle signature from the inversion, only CarbonTracker data from the same part of the diurnal cycle (9 am to midday) was used.

The initialisation would have been improved if the edge box time was recorded in 24-hour periods rather than the ten day run. This was not used in this case, because it should be used in conjunction with a true “particle leaving the scene” initialisation method. Edge box height information was not taken in this case, this would have resulted in a massive increase in the output file sizes (this problem has been solved by the satellite centric releases described in Section 6.2.4.2 which vastly reduces the number of unnecessary releases).

The second carbon mass flux (represented on the horizontal axis of Figures 8.1 to 8.7) was obtained (for each release) by coupling NAME surface influence time with the CarbonTracker flux product. Starting at 9 am on the date of the release, a 10-day average flux value was obtained for every 1° by 1° grid box in the domain, by taking the mean of the 80 flux time periods (8 3-hour time slots for 10 days).

The Figures 8.1 through to 8.7 contain information about the relationship between the CarbonTracker model and the SCIAMACHY/FSI retrieval scheme. Unlike in Section 7.1.2, a linear one to one relationship should not be expected. Potential sources of uncertainty include: SCIAMACHY/FSI measurement error, uncertainties in the “edge box” initialisation, uncertainties in the surface influence time, uncertainties in the CarbonTracker fluxes and uncertainties in the CarbonTracker weather.

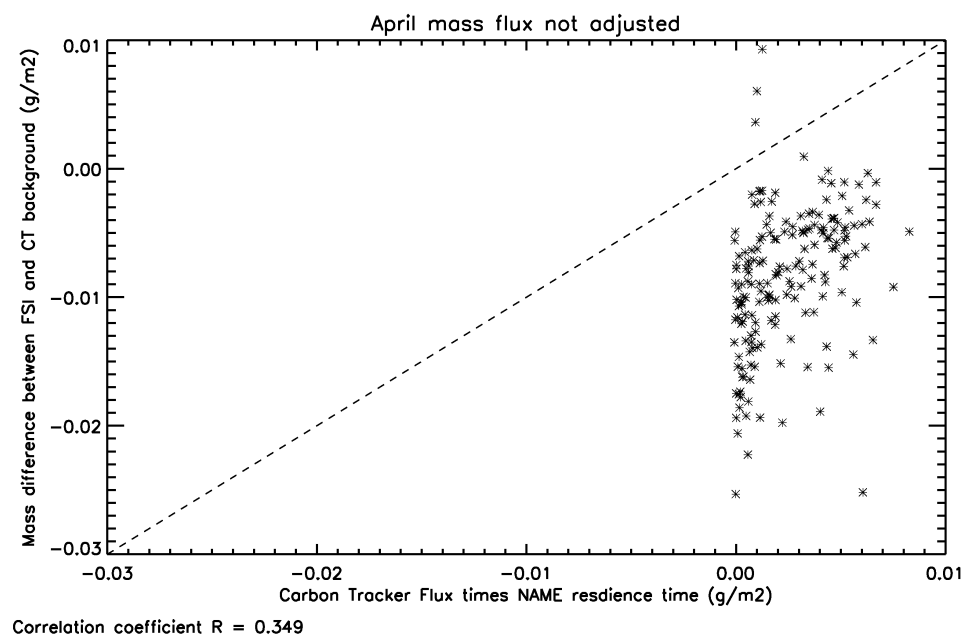
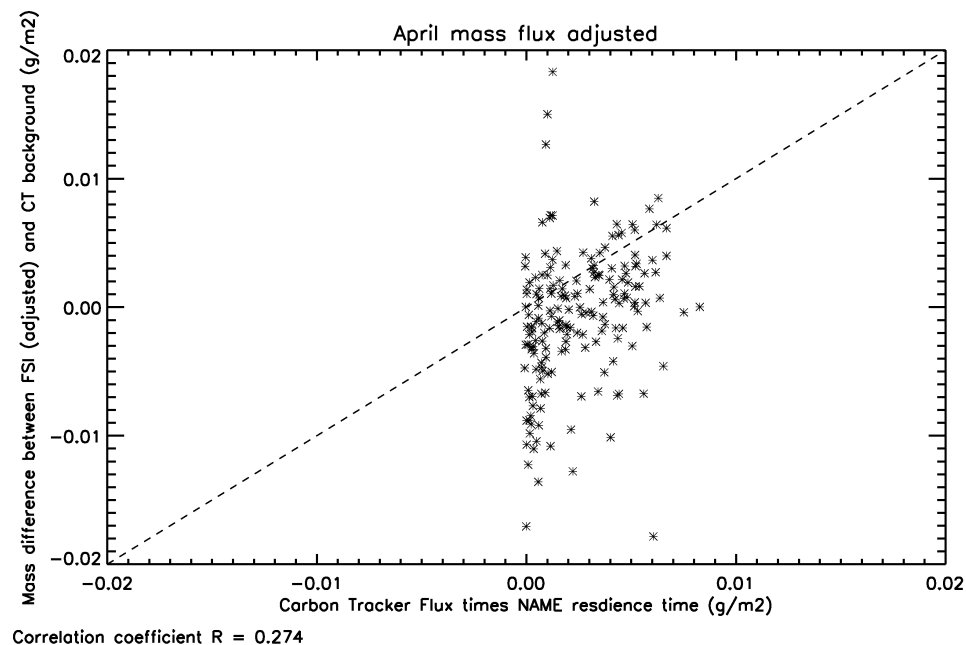


FIGURE 8.1. Mass flux plots for April 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

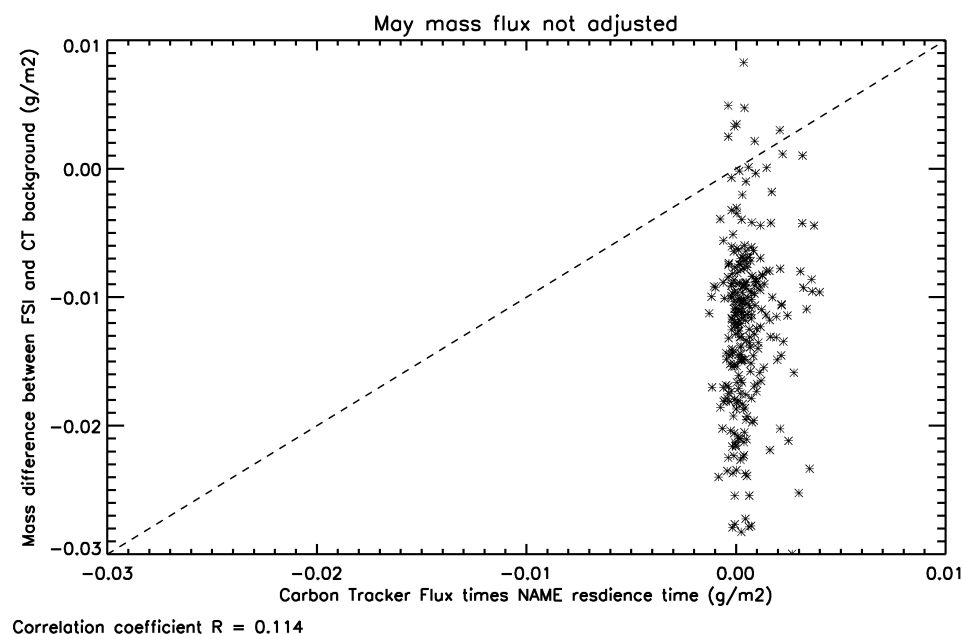
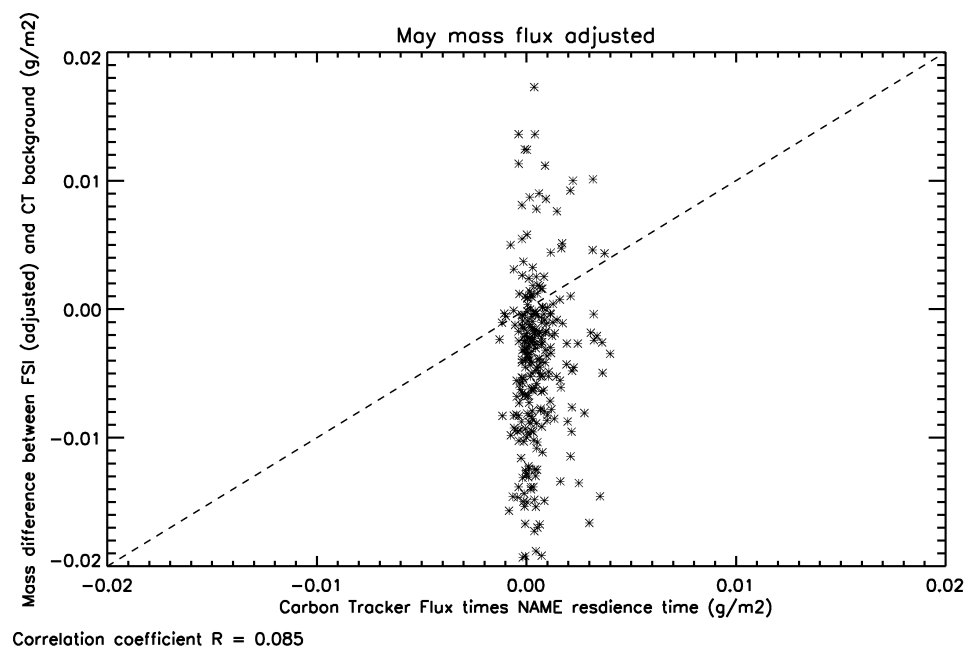


FIGURE 8.2. Mass flux plots for May 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.



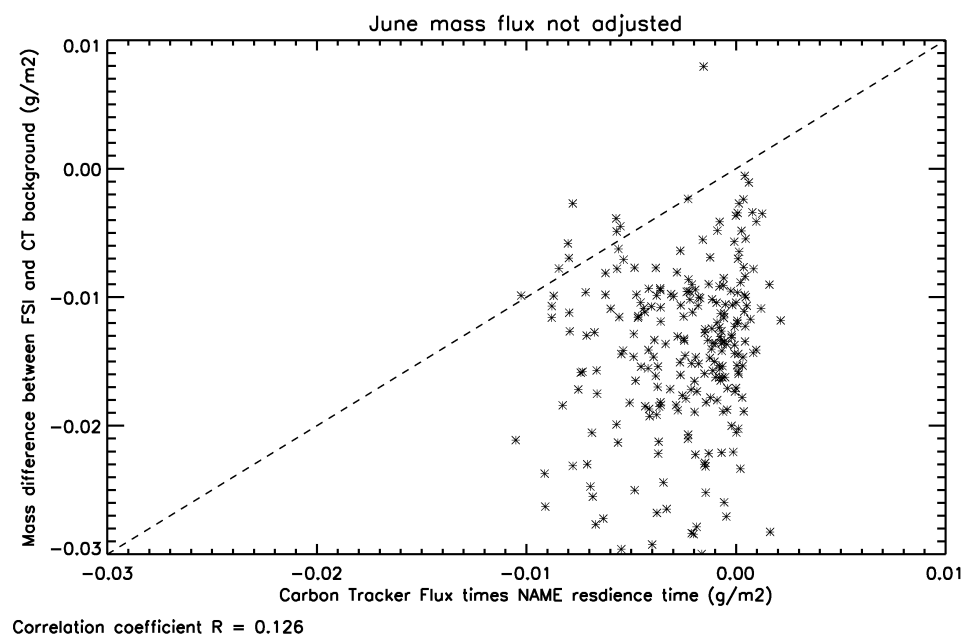
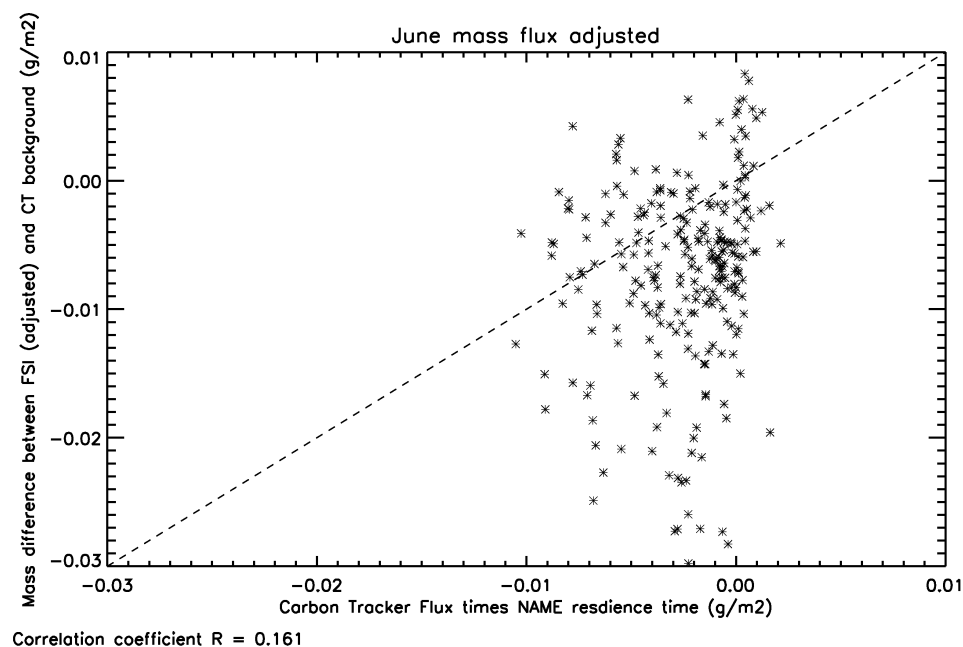


FIGURE 8.3. Mass flux plots for June 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

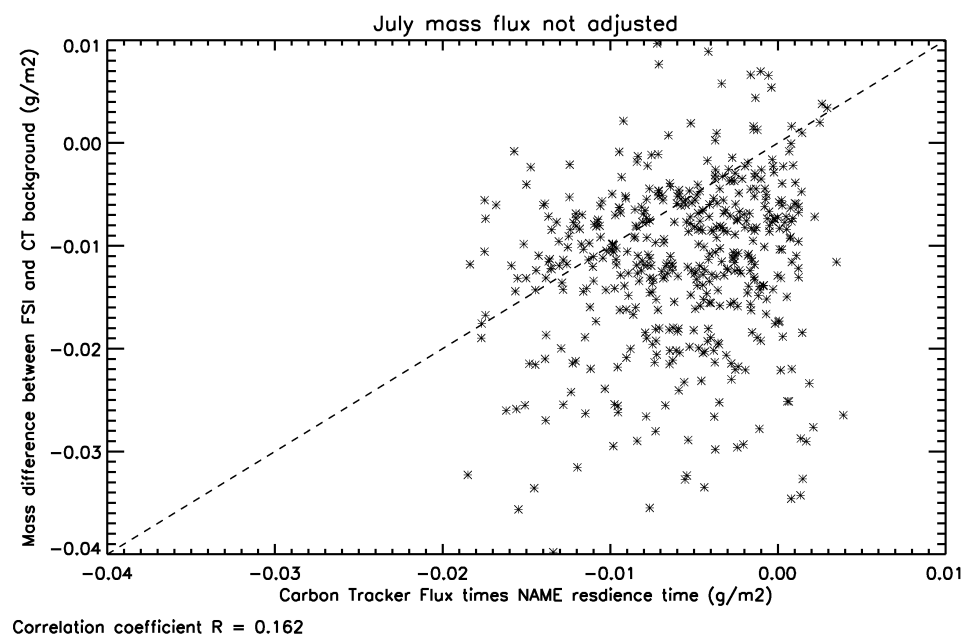
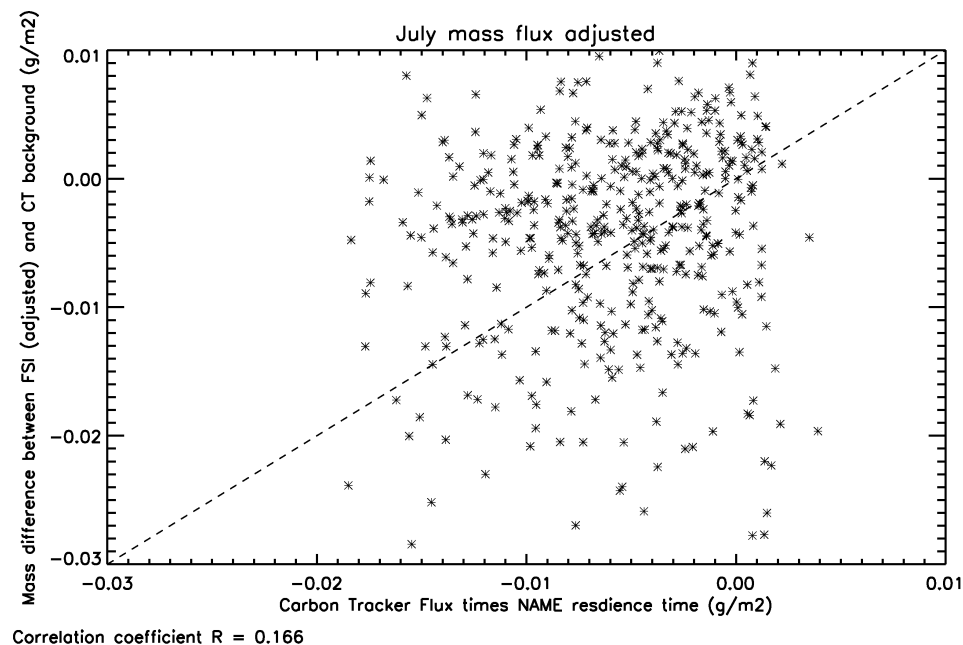


FIGURE 8.4. Mass flux plots for July 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

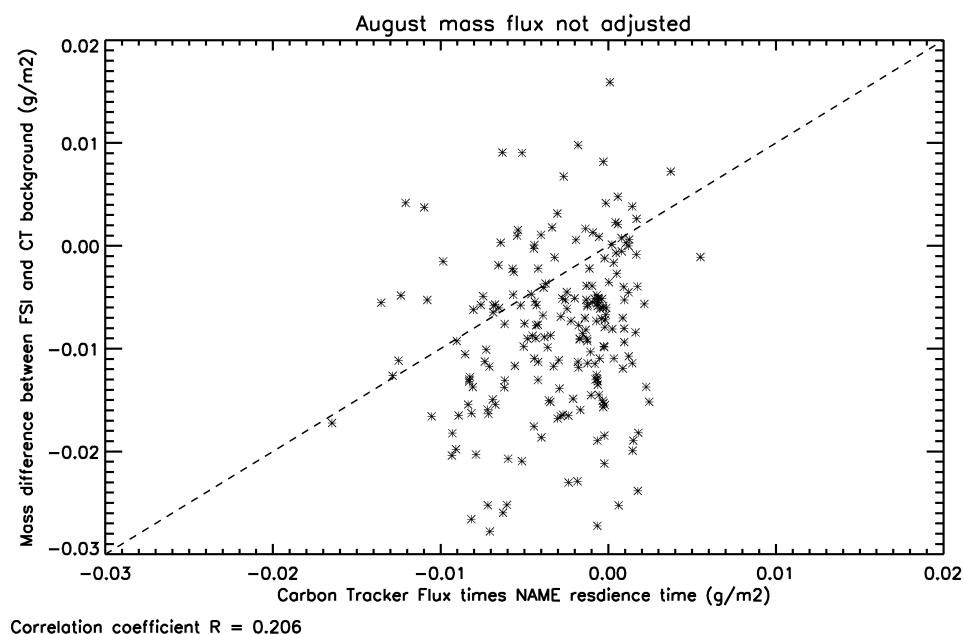
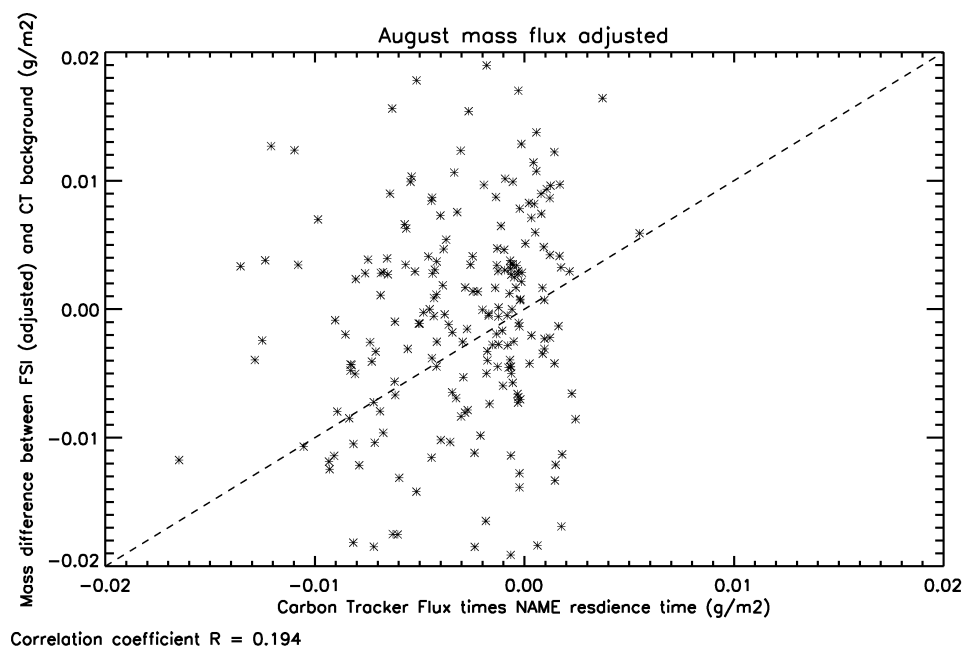


FIGURE 8.5. Mass flux plots for August 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

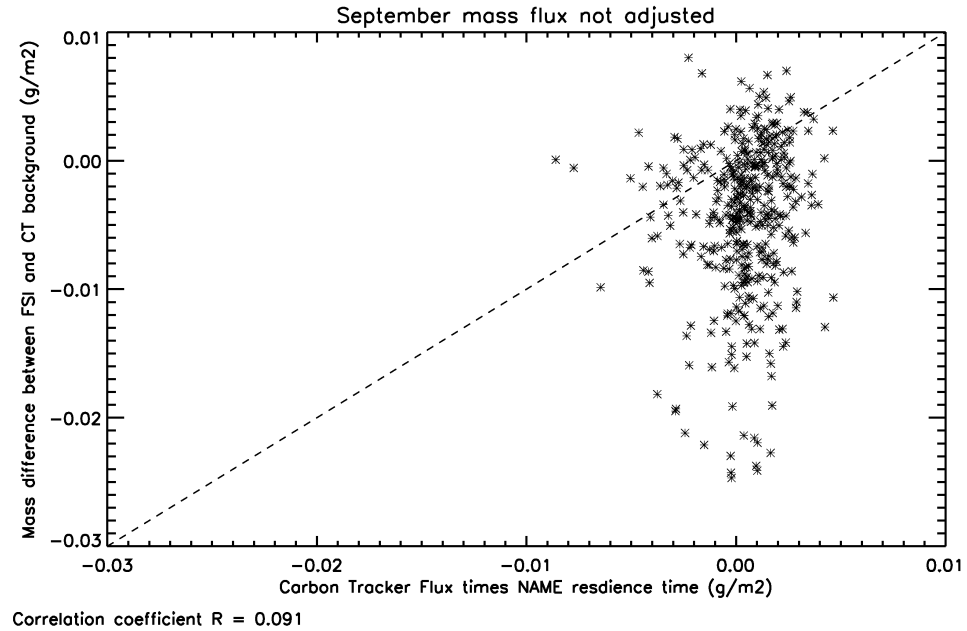
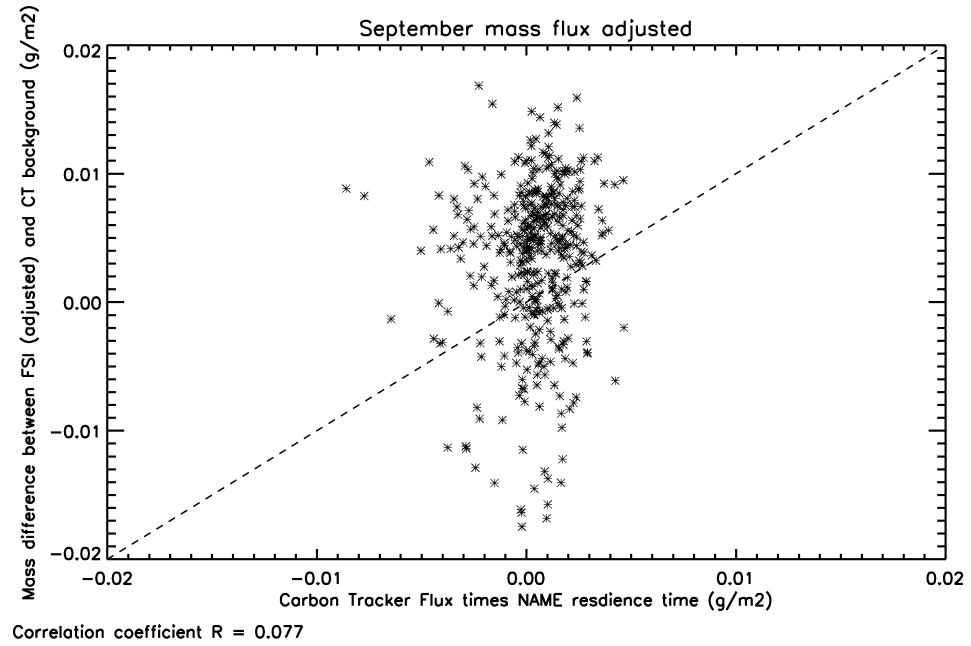


FIGURE 8.6. Mass flux plots for September 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

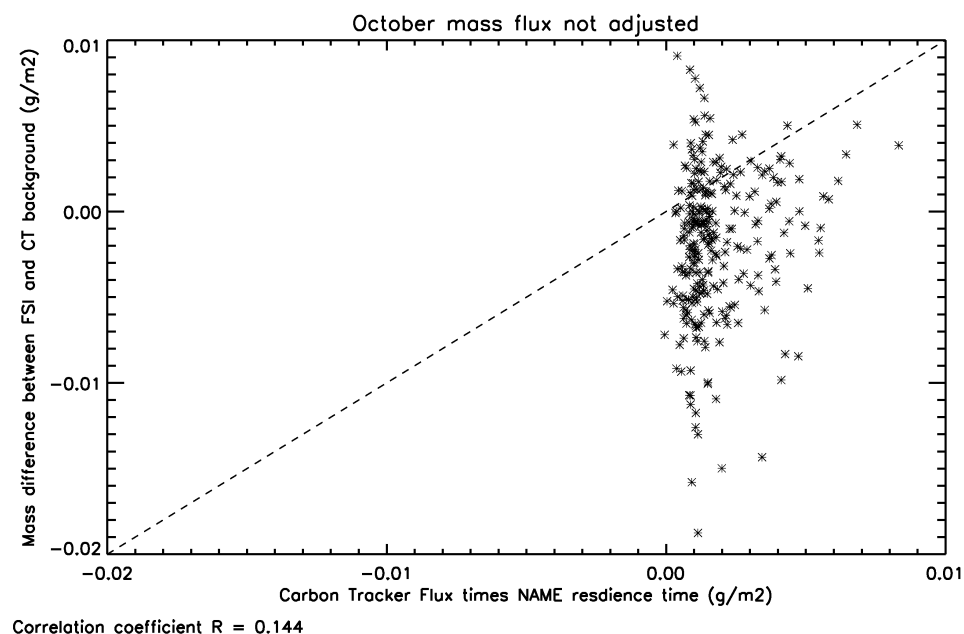
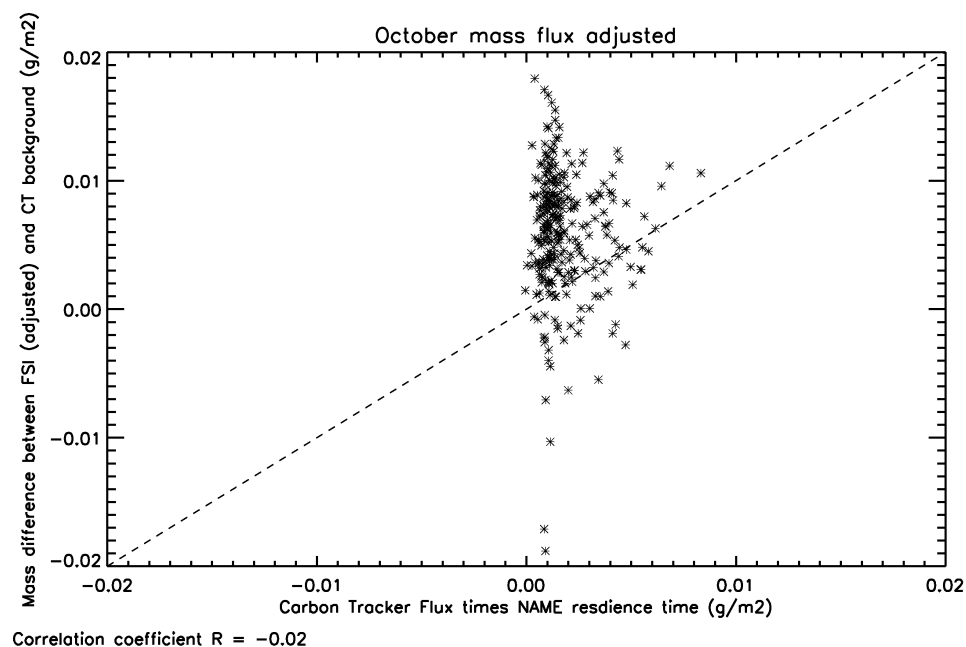


FIGURE 8.7. Mass flux plots for October 2003. The horizontal axis shows mass flux obtained by coupling the NAME output with CarbonTracker flux product, whilst the vertical axis shows the mass difference by subtracting the release concentration from the background. SCIAMACHY/FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and SCIAMACHY/FSI in the top plot whilst no offset was made in the bottom plot.

In April 2003 (Figure 8.1), there is only a very slight correlation between the two methods of obtaining mass difference in the column. In this month, North America appears to have been a mild source of  $\text{CO}_2$  to the atmosphere (judging from the horizontal axis). It would appear that the offset-bias between SCIAMACHY/FSI and CarbonTracker is greater than 1.5% this month, as the centre of the scatter in the upper image is just below the one to one line. The distribution of points in the vertical direction is a little greater than that in the horizontal direction, a sign perhaps that SCIAMACHY/FSI is noisier than CarbonTracker.

In May 2003 (Figure 8.2), there is no correlation between the two methods used to obtain the column mass flux. The offset-bias between FSI and CarbonTracker is significantly greater than 1.5% in this month. The CarbonTracker flux product would suggest that North America was a net neutral producer of  $\text{CO}_2$  to the atmosphere in the month of May. The large distribution of points in the vertical direction this month, suggests either the FSI scheme is not very reliable this month and/or that the initialisation method is not performing well this month.

In June 2003 (Figure 8.3), there is no correlation between the two methods of obtaining mass flux. The CarbonTracker flux product suggests that North America was a moderate sink of atmospheric  $\text{CO}_2$  this month. The rounder “shotgun blast” shape of these plots (the variation in both directions being the same), suggests that the measurements are quite close to the model this month. The offset-bias between SCIAMACHY/FSI and CarbonTracker is a little greater than 1.5% this month.

July 2003 (Figure 8.4), is similar to June. It would seem that FSI/SCIAMACHY increased by 1.5% is a close match to the CarbonTracker product. North America appears to have been a strong sink of atmospheric  $\text{CO}_2$  this month. The amount of scatter is about the same in both axes, and is considerably greater than the other months. Again there is little correlation between the two methods of obtaining mass flux.

In August 2003 (Figure 8.5), there appears to be no correlation between the two methods of obtaining mass flux. North America was a moderate sink of atmospheric  $\text{CO}_2$  this month. There is a greater distribution of points in the vertical axis. Increasing FSI/SCIAMACHY by 1.5% nullifies the mean offset-bias with the CarbonTracker weather product this month.

In September 2003 (Figure 8.6), there appears to be no correlation between the two methods of obtaining mass flux. North America was a net neutral source/sink of atmospheric  $\text{CO}_2$  this month. There is a greater distribution of points in the

vertical axis. The offset-bias between FSI/SCIAMACHY and CarbonTracker is smaller than 1.5%.

In October 2003 (Figure 8.7), there appears to be no correlation between the two methods of obtaining mass flux. North America was a mild source of atmospheric CO<sub>2</sub> this month. There is a greater distribution of points in the vertical axis. The offset-bias between FSI/SCIAMACHY and CarbonTracker is smaller than 1.5%.

The consistent factor in Figures 8.1 through to 8.7 is that there is a lot of noise in the individual mass flux calculations. This is a result of the precision of CarbonTracker, the precision of SCIAMACHY/FSI and the initialisation method. The spread of points in the vertical axis is a little bigger than the spread in the horizontal axis, and suggests the combined precision of the mass change is of the same order as the mass change. Figures 8.1 through to 8.7 also demonstrate that the offset between SCIAMACHY/FSI mean column and CarbonTracker weather mean column is not a constant 1.5%, although monthly offsets seem to lay within the range  $-1.5 \pm 0.5\%$ . There may be a seasonality about this offset, although a second years analysis would be required to investigate this.

Only very limited information was obtained in this first analysis. There are a lot of similarities between the monthly mean 1° by 1° gridded FSI/SCIAMACHY and CarbonTracker weather products, by averaging the results it will be possible to demonstrate the similarities between the two sources.

### 8.1.2 Comparison with clustered releases

In order to reduce the impact of FSI measurement uncertainty, the forward scheme was reanalysed by averaging the CO<sub>2</sub> mass change over spatially clustered regions (see Figure 8.8). It is likely that in a month where the biosphere is a net absorber of carbon from the atmosphere, receptor locations with greater contact time would see greater uptake. It is also quite likely that particles released near to the domain edges would see less surface influence time and in such a month would see a smaller uptake of carbon (hence the decision to cluster geographically). This links to the observation that in months of strong uptake there is a general trend in the atmospheric CO<sub>2</sub> concentration to decrease from west to east (see Section 4.3).

The following changes were made to the offline forward modelling scheme. The releases were clustered into eight locations of area 9° by 9° (the same area as nine of the original 3° by 3° release squares) for each calender month (see Figure 8.8).

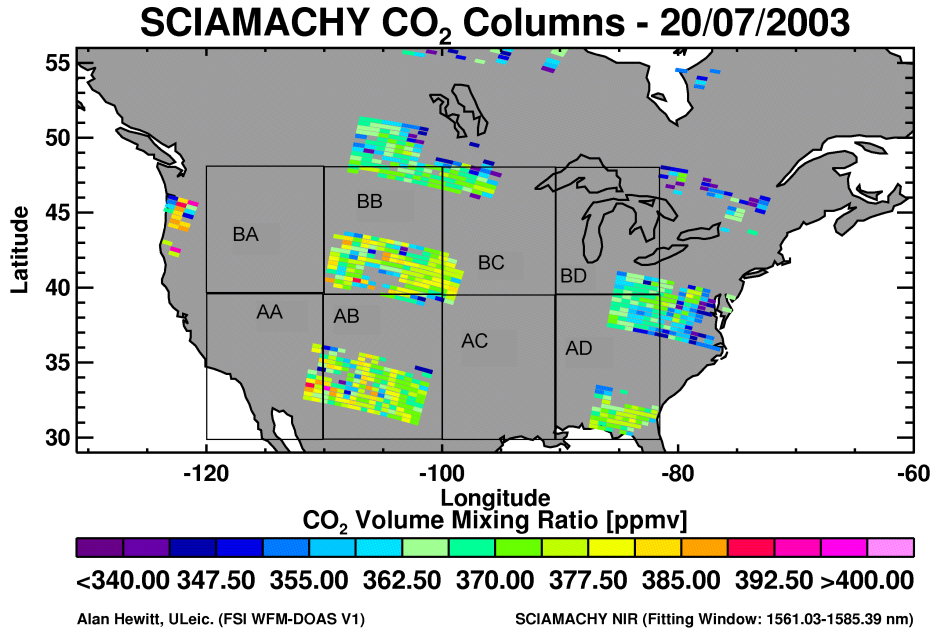


FIGURE 8.8. Location of North American Release regions.

These areas were labeled AA (30°-39° north and 118°-109° west), AB (30°-39° north and 109°-100° west), AC (30°-39° north and 100°-91° west), AD (30°-39° north and 91°-82° west), BA (39°-48° north and 118°-109° west), BB (39°-48° north and 109°-100° west), BC (39°-48° north and 100°-91° west), BD (39°-48° north and 91°-82° west), releases outside of one of these areas were ignored. If ten or more releases occur in one of these boxes in a calendar month, that area will be considered in the reanalysis.

This method is rather crude, it works because some release areas will have a longer surface contact time than others and thus are likely to experience greater fluxes. Averaging of the individual mass fluxes should smooth out some of the noise seen in Figures 8.1 through to 8.7. The locations AD, BC and BD should have experienced the most surface contact time, assuming that the general wind direction is from the west, and are expected to have greater negative flux in months where North America acts as a sink of atmospheric CO<sub>2</sub>, and have a greater positive flux when North America acts as a source of atmospheric CO<sub>2</sub>.

There is a vast improvement on the quality of information available in the reanalysis Figures 8.9 through to 8.16. Each month is also distinctive in these Figures, unlike the “shotgun blast” figures from the individual mass flux analysis.



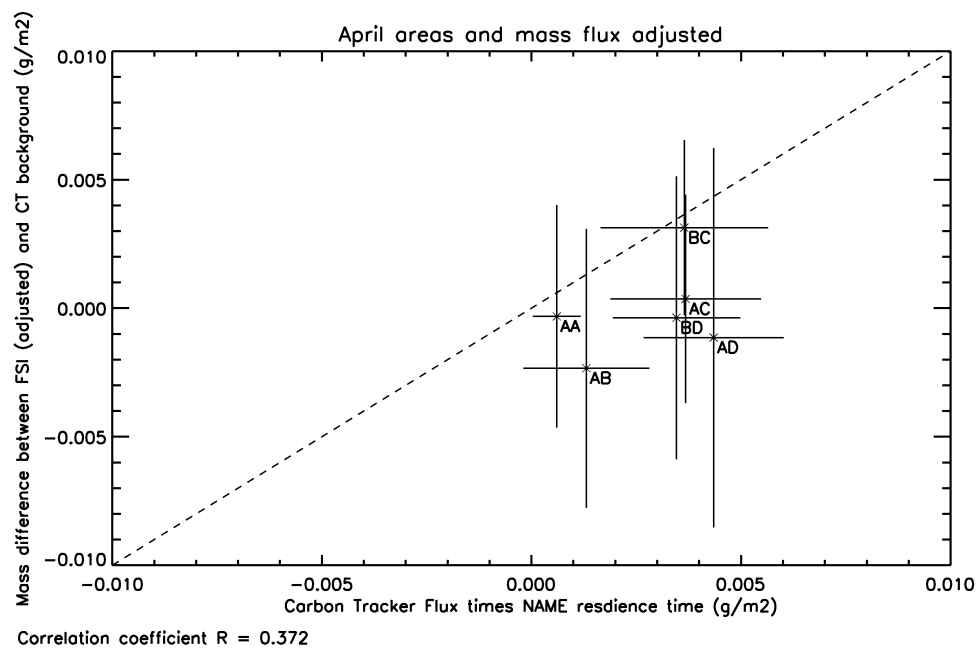


FIGURE 8.9. Area mass flux plots for April 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

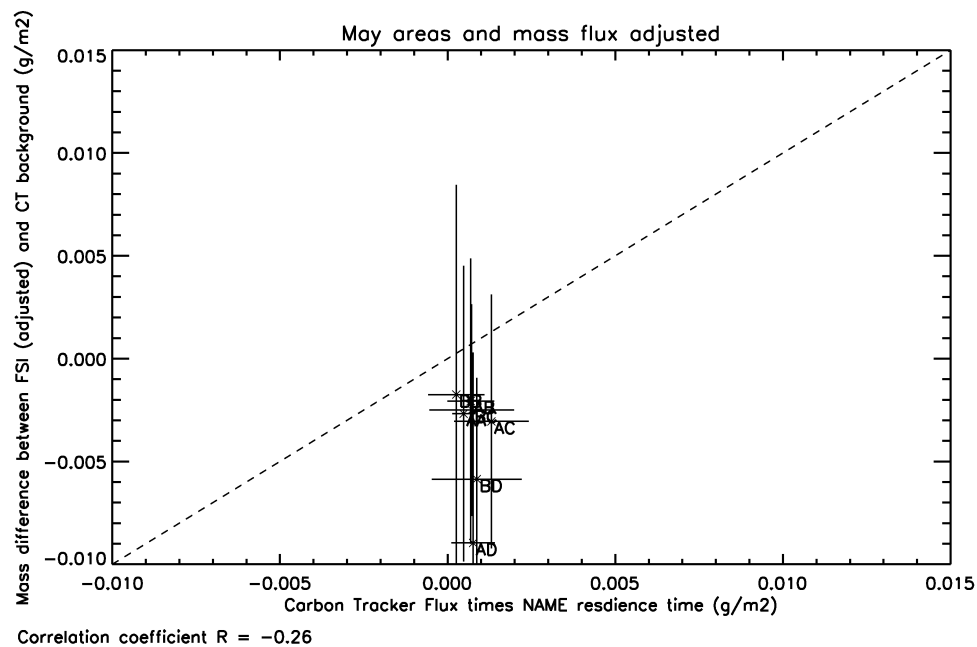


FIGURE 8.10. Area mass flux plots for May 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

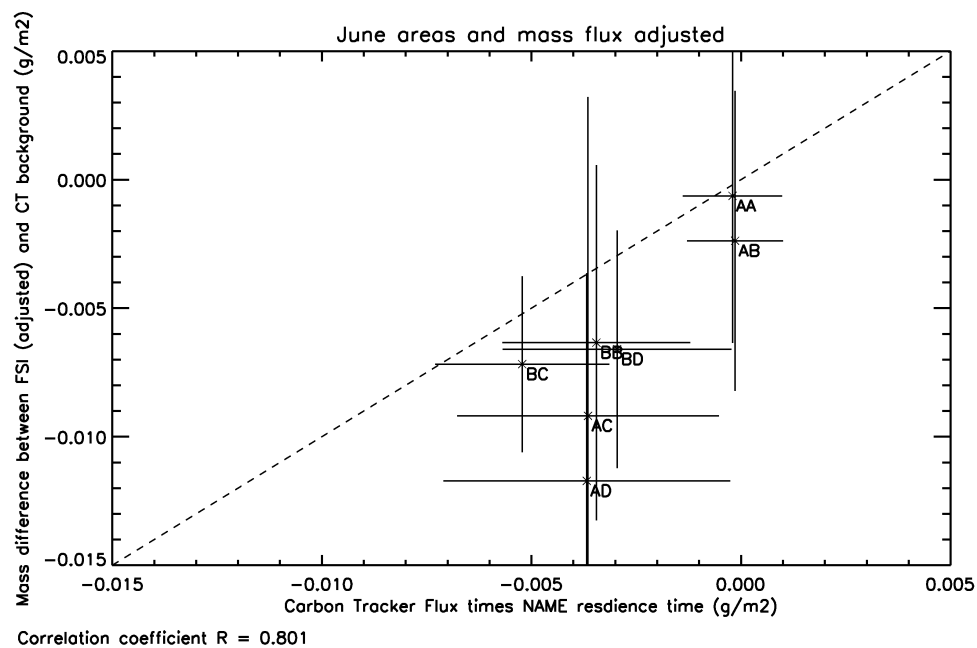


FIGURE 8.11. Area mass flux plots for June 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

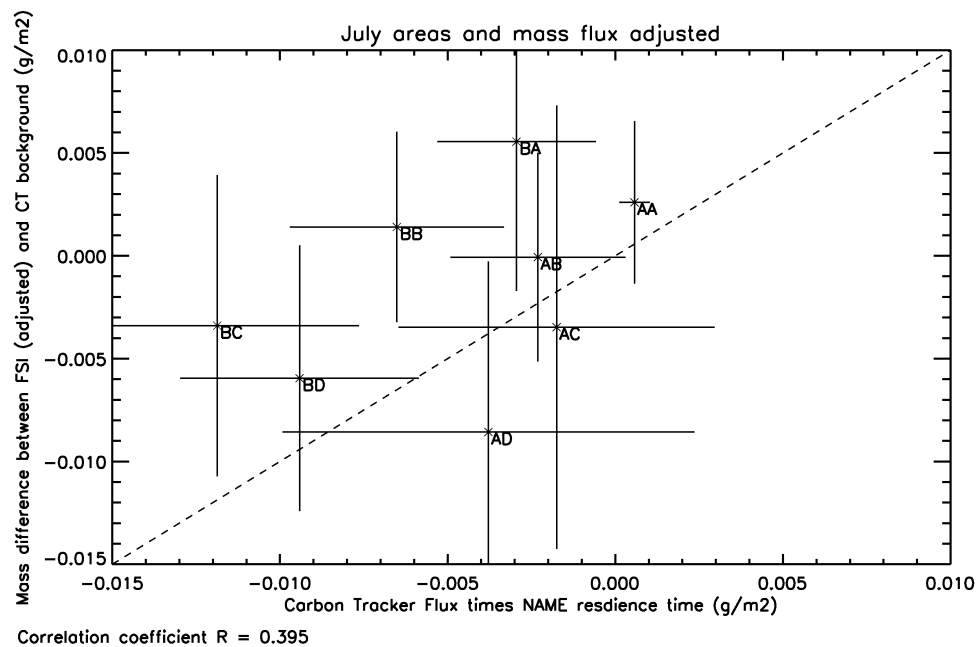


FIGURE 8.12. Area mass flux plots for July 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

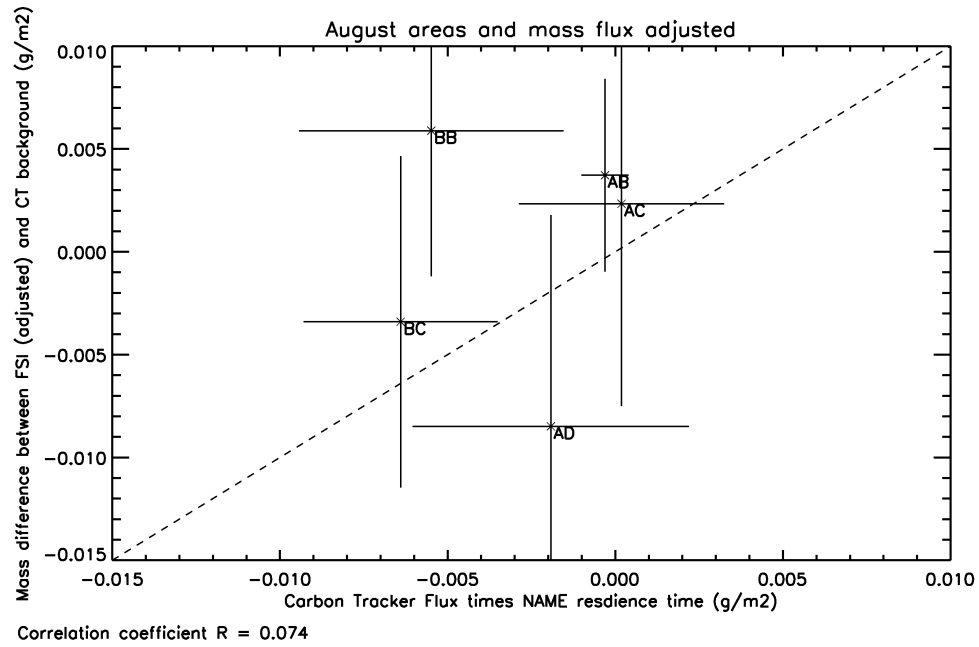


FIGURE 8.13. Area mass flux plots for August 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

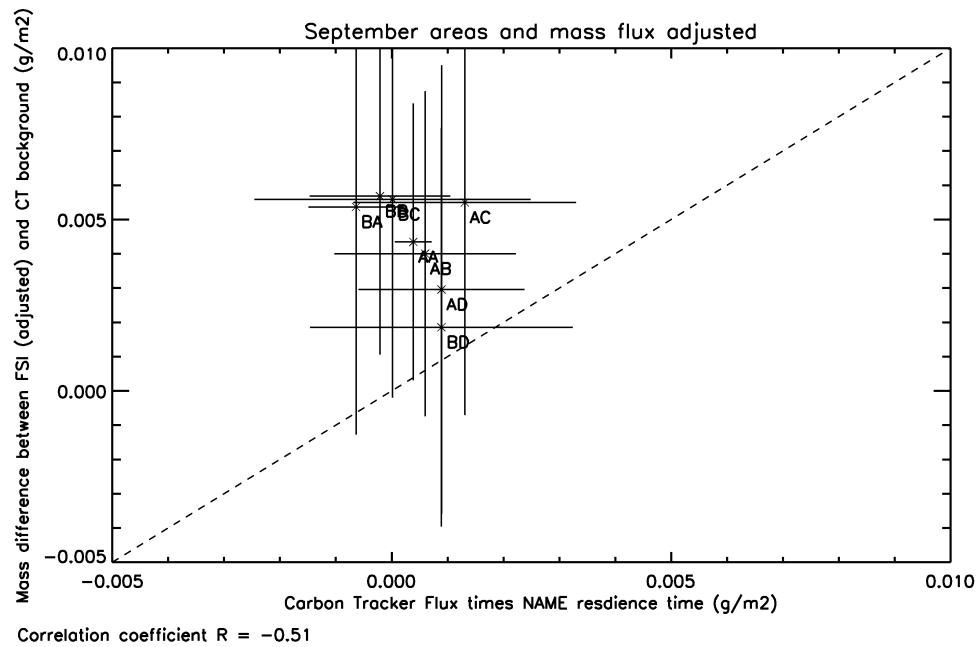


FIGURE 8.14. Area mass flux plots for September 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

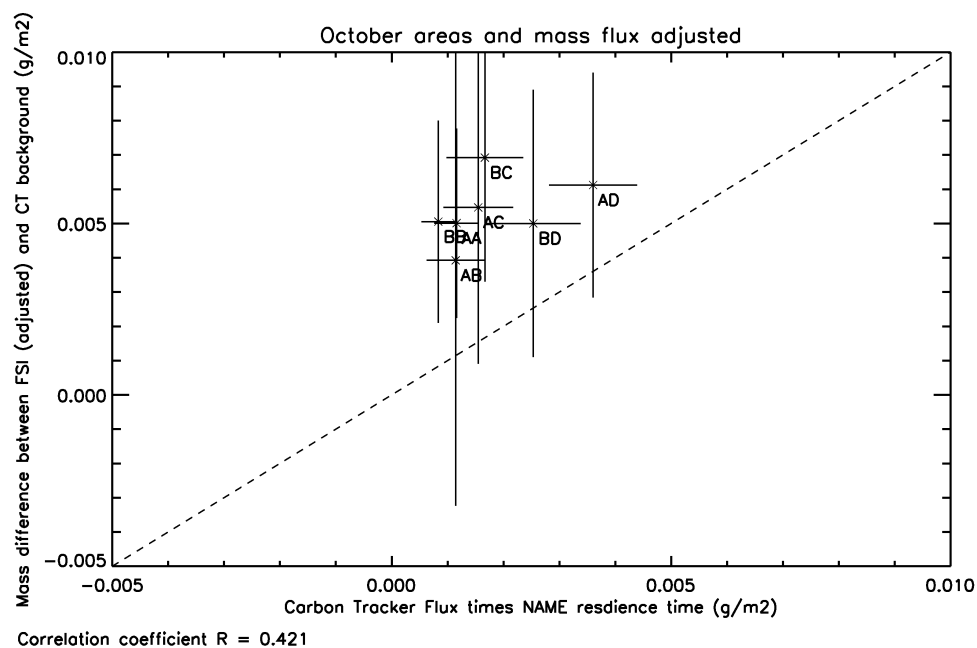


FIGURE 8.15. Area mass flux plots for October 2003. FSI was multiplied by 1.015 to adjust for the mean offset between CarbonTracker and FSI.

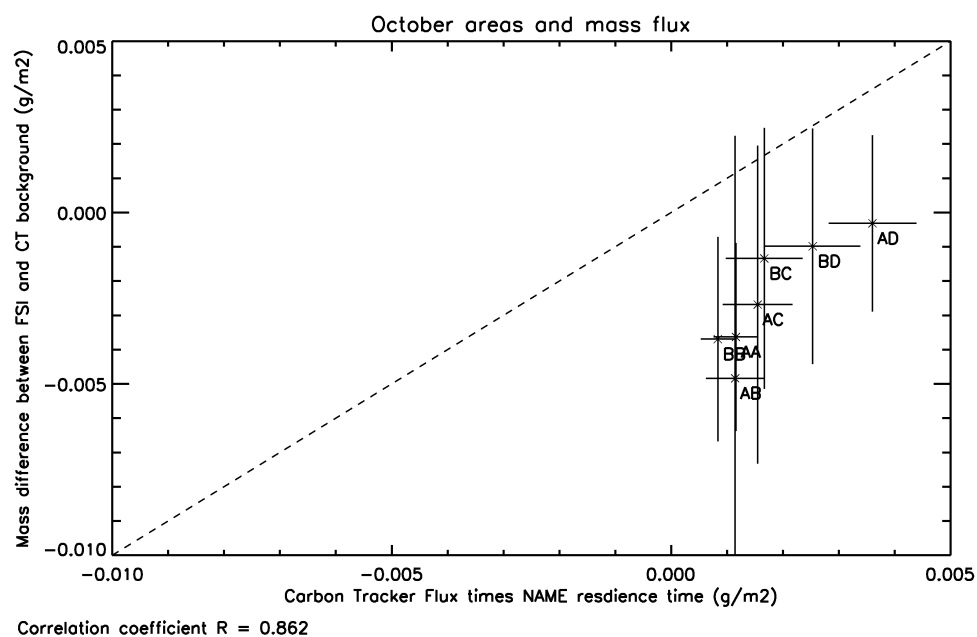


FIGURE 8.16. Area mass flux plots for October 2003. No offset was made between CarbonTracker and SCIAMACHY/FSI.

The month of April (Figure 8.9), has all the points below the dashed line suggesting that the offset between FSI and CarbonTracker is greater than 1.5%. North America is a mild source of atmospheric CO<sub>2</sub> this month, it is encouraging to see that areas AA and AB which should have had little surface contact time have only a small flux contribution in both axes whilst area BC has a more significant positive flux contribution in both the axes.

The month of May (Figure 8.10), seems to have a larger offset than 1.5% between FSI and CarbonTracker (as seen in Figure 8.2). CarbonTracker fluxes suggest that North America was flux neutral this month, and six of the areas are bunched together near zero flux in both axes in the figure. It is quite possible that the low concentration air in box AD has been brought in from the Caribbean rather than the west and the discrepancy comes from inadequate initialisation.

The month of June (Figure 8.11), has an offset between FSI and CarbonTracker just a little greater than 1.5%. North America was a moderate sink of atmospheric CO<sub>2</sub> this month, and the relationship between area (thus surface contact time) and mass flux (in both axes) is very strong in this plot. Areas BC, AC and AD have the largest mass change in both axes, suggesting that most of the background air is from the west with mass difference linked to surface residence time.

The month of July is quite similar to June with stronger uptake of atmospheric CO<sub>2</sub>. Areas BC, BD and AD have the largest mass change in both axes (Figure 8.12), again suggesting that most of the background air is from the west with mass difference linked to surface residence time.

The month of August had far less SCIAMACHY coverage, as a result only five of the areas had ten or more ensembles pass the filters.

The month of September (Figure 8.14), has an offset of just less than 1.5% between FSI and CarbonTracker. North America was a net neutral carbon emitter this month, according to the CarbonTracker fluxes. The small spread in mass flux calculated from subtracting the background agrees with this.

The month of October has been shown twice, with an offset of 1.5% in Figure 8.15 and with no offset in Figure 8.16. The largest mass changes in the domain are seen in areas AD, BC and BD in a month where North America is a mild source of atmospheric CO<sub>2</sub>.

The initial analysis and the reanalysis both clearly indicate that in their present forms CarbonTracker and FSI are not compatible for investigating the surface ex-

change of carbon. Despite having no time resolution, the initialisation method used in these tests is not much worse than that shown in Figure 7.2, where there is a clear linear relationship between the two axes. Because the cause of the offset-bias between SCIAMACHY/FSI and modelled atmospheric CO<sub>2</sub> concentrations is not understood, it is difficult to correct for. It would be possible to run the inversion, correcting for the offset-bias based on the results in this section, but this would essentially be fixing the results of the inversion to match the CarbonTracker flux products.

In order to only investigate the daily averaged flux magnitudes and not the diurnal cycle effects, only CarbonTracker background columns at the same part of the diurnal cycle as the measurement were used. This was a little difficult, as the measurements were always made at a local time, whereas the CarbonTracker product produced a value at the centre of a 3-hour time window based on GMT. This created a number of difficulties, especially as the domain covered a large range of time zones. A different CarbonTracker weather time zone was used for the east and the west domain edges, obtaining the correct 3-hour period for the south and north edges required binning longitudes into the correct time zone. This meant that some background values were  $\pm 90$  minutes away from the same part of the diurnal cycle as the satellite overpass time. Because CarbonTracker fluxes are very small over the oceans (where the background concentrations were defined), this effect was not particularly significant and a similar experiment where all 8 time zones were used to define the background produced almost the same result.

Although it is possible to use the SCIAMACHY averaging kernel to weight the CarbonTracker background columns, this would in turn introduce a bias with the CarbonTracker flux product which assumes a averaging kernel of 1 throughout the column. It is not possible to nudge the FSI retrieved column to represent a measured column without the averaging kernel biases at certain parts of the column.

The incompatibility of the satellite measurements with the modelled background and the lack of satellite background measurements over the ocean have ruled out the possibility of further investigation of the carbon cycle using FSI measurements with the forward approach. The possibility of overcoming these problems with GOSAT or OCO-like measurements will be discussed in Chapter 9.

## 8.2 Demonstration of FSI inverse modelling

The methodology used in the linear regression inversion of SCIAMACHY/FSI-retrieved XCO<sub>2</sub> was described in Section 6.3.2 (see Figure 6.9 for a diagrammatic overview). Carbon fluxes were obtained for the eight MODIS (majority) ecoregions (defined in Section 6.3.2; only the five most sampled ecoregions are shown in Figures 8.19 and 8.20). These carbon fluxes were obtained for the calendar months April through to October by finding the *a posteriori* solution,  $\hat{\mathbf{x}}$ , from Equation 6.3 and the corresponding *a posteriori* covariance,  $\hat{\mathbf{S}}$ , obtained from Equation 6.4. This demonstration of the linear regression inversion scheme represents the first time that carbon fluxes have been obtained through the inversion of satellite retrieved XCO<sub>2</sub> columns. This methodology can be improved in future by defining genuine homogeneous carbon flux ecoregions (only ocean is expected to act as a homogeneous flux ecoregion as it has a negligible carbon flux), using a method such as multivariate clustering [Hargrove and Hoffman, 2005] and through a more precise instrument such as GOSAT.

The online NAME settings were set thus: maximum surface influence height of 100 metres, 150 longitude by 75 latitude 1° by 1° grid boxes in the domain, edge box initialisation (the results in Figures 8.17 and 8.19 were corrected for a 1.5% offset-bias between SCIAMACHY/FSI and CarbonTracker weather, whereas the results in Figures 8.18 and 8.20 were not corrected). NAME releases were only made for SCIAMACHY/FSI ensembles that satisfy the conditions outlined in Section 6.3.1, consequently there were not enough of these releases to include the winter months in the inversion.

When the SCIAMACHY/FSI XCO<sub>2</sub> measurements are adjusted by the offset factor of 1.5% (Figure 8.17), the resulting Ocean fluxes obtained through the inversion method are not far from the expected zero. When considering the results of the inversion scheme for all of the MODIS land types (Figure 8.19), it is promising to see that Evergreen Forest types have the largest seasonal cycle amplitude and the magnitude of this  $\pm 5$  grams metre<sup>-2</sup> day<sup>-1</sup> is in the same range as most bottom up estimates [Hollinger et al., 2004].

The shape of the seasonal cycle amplitude as well as the magnitudes are significantly altered when the adjustment factor of SCIAMACHY/FSI to the carbon tracker background is changed. However, these early inversion results indicate that satellite products have the potential to contribute to the understanding of the carbon cycle.

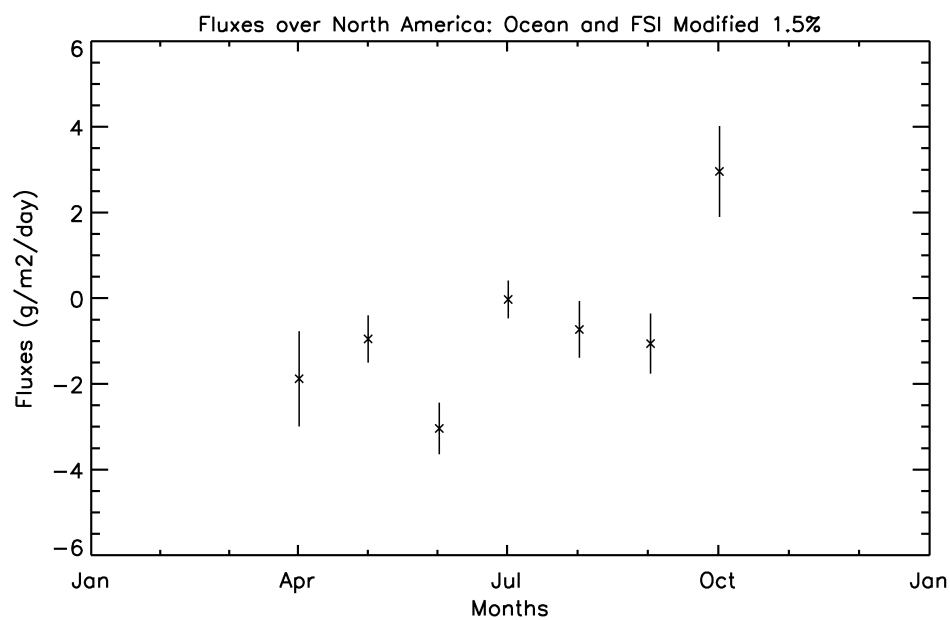


FIGURE 8.17. Monthly mass flux for ocean ecoregion 2003. FSI was multiplied by 101.5% to adjust for the mean offset between CarbonTracker and FSI.

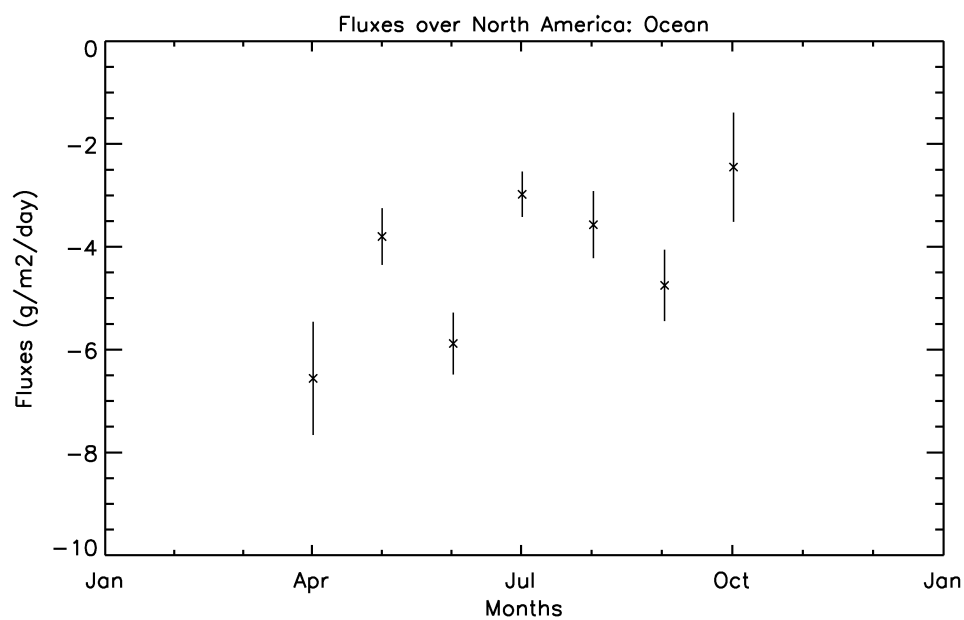


FIGURE 8.18. Monthly mass flux for ocean ecoregion 2003. No offset was made between CarbonTracker and SCIAMACHY/FSI.



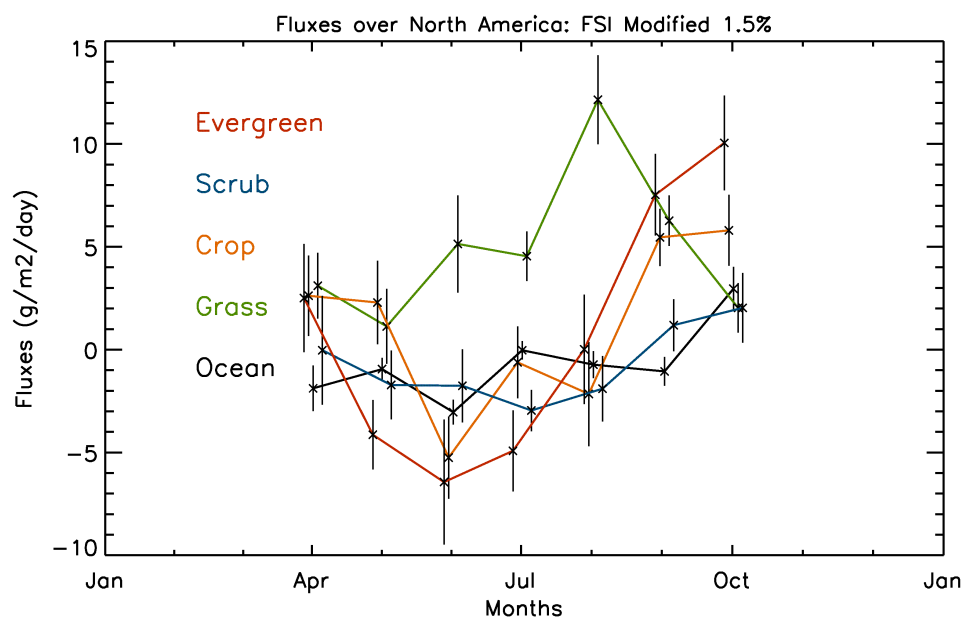


FIGURE 8.19. Monthly mass flux for 5 ecoregions 2003. FSI was multiplied by 101.5% to adjust for the mean offset between CarbonTracker and FSI.

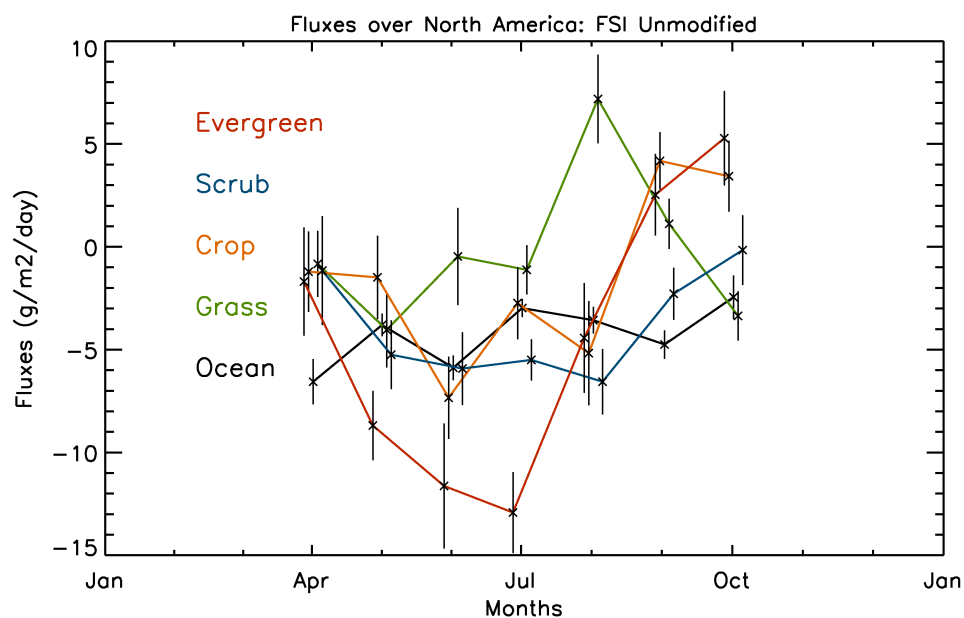


FIGURE 8.20. Monthly mass flux for 5 ecoregions 2003. No offset was made between CarbonTracker and SCIAMACHY/FSI.

## 8.3 Summary

The connection between carbon fluxes and changes in the atmospheric CO<sub>2</sub> field were not immediately obvious as there was a lot of random noise in the satellite retrievals. The reanalysis of the forward test using the area method (seen in Figures 8.9 to 8.15) filters out most of the noise and shows a strong link between CarbonTracker fluxes and SCIAMACHY/FSI measurements of the atmospheric CO<sub>2</sub> field.

The inversion results shown in Section 8.2 represents the first time that satellite retrieved CO<sub>2</sub> has been inverted to produce biosphere-to-atmosphere carbon fluxes (including using a Eulerian scheme). This represents a significant development in the human understanding of the carbon cycle. It is the intension of the author that more precise satellite retrieved CO<sub>2</sub> columns and a genuine homogeneous carbon flux ecoregion map (such as that provided by multivariate clustering [Hargrove and Hoffman, 2005]) be used in this inversion scheme in order to investigate regional scale carbon fluxes in regions poorly sampled by the ground network.

At the time of publication, there are no other satellite inverted carbon fluxes to compare with the results shown in Section 8.2. It does not seem appropriate to compare these preliminary results (using an unsuitable instrument [Miller et al., 2007]) with established bottom-up and top-down carbon fluxes. This test should be carried out with OCO-like or GOSAT retrieved CO<sub>2</sub> columns.

Miller et al. [2007] performed a synthetic test to investigate the effects of a systematic bias (such as the offset bias between SCIAMACHY/FSI and CarbonTracker) on CO<sub>2</sub> flux inversions. Miller et al. [2007] concluded that applying a simple factor to correct for a constant offset would have a minimal effect on the inversion result. Seasonal and spatial biases (both of which apply to SCIAMACHY/FSI vs CarbonTracker) complicate the correction factors. An offset-bias smaller than 1.5% may be more difficult to detect, though it would have a smaller effect on the inverted fluxes.

Chapter 9 will examine further applications of the techniques discussed throughout the thesis.

# Chapter 9

## Conclusions and Future Directions

### 9.1 Assessment of FSI WFM-DOAS retrievals

The original objective of this research project was to investigate whether the existing generation (i.e. SCIAMACHY) of satellite instruments (with the ability to measure atmospheric CO<sub>2</sub> concentrations) were capable of improving on the current understanding of the carbon cycle. A theoretical study by Rayner and O'Brien [2001] demonstrates that satellite retrievals with a precision of 2.5 ppmv (around 1%) or better are needed by an inversion scheme to improve upon the inverted global carbon flux magnitudes obtained from the ground network CO<sub>2</sub> measurements. However, the ground network is not evenly distributed and Miller et al. [2007] has shown that satellite retrievals can be used to improve various regional carbon flux estimates if they have a precision free of offset-bias in the range of 1 - 10 ppmv (0.3 - 3%).

In Chapter 4, the FSI-WFM-DOAS XCO<sub>2</sub> retrieval algorithm was validated against other sources of CO<sub>2</sub> field data (both measured and modelled), which tend to indicate a negative offset between the retrieved CO<sub>2</sub> vertical mixing ratio and those retrieved at FTIR locations (which are believed to be closer to the “true” vertical mixing ratio). The measurement precision of individual column XCO<sub>2</sub> retrievals is roughly similar to the fitting error between the radiative transfer model (SCIATRAN) and the retrieved SCIAMACHY spectra, which are in the range 2-3% (i.e. within the range specified by Miller et al. [2007]).

The cause of the offset between SCIAMACHY/FSI and other CO<sub>2</sub> fields is not understood, although it is believed to be mainly the result of uncertainty in

the radiative transfer model and spectroscopy and is possibly also affected by the sampled vs unsampled biases as discussed in Chapter 4.

Another possible use of SCIAMACHY/FSI retrieved  $\text{XCO}_2$  fields would be to validate top-down or bottom-up modelled carbon flux estimates. In this forward modelling case, the negative offset-bias and the seasonal cycle amplitude bias are much less of a problem. Since they are predictable, it is not difficult to apply an offset correction to the SCIAMACHY/FSI retrieved  $\text{CO}_2$  field.

## 9.2 Assessment of NAME method

A more general objective of this research project was to develop a Lagrangian method to investigate the surface exchange of carbon between the terrestrial biosphere and the atmosphere. Such a method would have advantages over global transport models (such as the ability to perform regional high resolution studies tailored to the specific area of interest and it being easier to model offline), although it would make initialisation of the background  $\text{CO}_2$  concentration more difficult.

The developed model contains features essential to investigating the carbon cycle and allows interchange of several types of information. Although specifically designed with SCIAMACHY/FSI retrieved  $\text{CO}_2$  columns in mind, both GOSAT and OCO-like  $\text{CO}_2$  measurements could be used with a relatively small change to the model set up. The release of tracer particles is made in a column to reflect the distribution of the satellite retrieved  $\text{XCO}_2$ . The two main model outputs are the surface residence time of the tracer particles (used to constrain flux magnitudes) and the origin of the tracer particles (used to initialise the background  $\text{CO}_2$  concentration).

Two distinct methods for investigating the surface fluxes were developed: the inverse method (Section 6.3.2) makes use of the assumption of homogeneous flux regions to obtain flux information from satellite retrieved  $\text{XCO}_2$ , whilst the forward method (Section 6.3.3) was designed to be used with satellite retrieved  $\text{CO}_2$  fields to validate modelled carbon fluxes.

Two techniques were developed to initialise the background concentration, the “edge box” method (Section 7.1.1) and the “freeze” method (Section 7.1.2). The freeze method has yet to be adopted, as it would require significant changes to the way that the NAME model outputs information. An experiment was set up to

demonstrate the advantages of the freeze method and both initialisation systems were tested using CarbonTracker data. The current edge box method performs reasonably well with a coefficient of determination,  $R^2 = 0.5$ , whereas the freeze method improves significantly upon this,  $R^2 = 0.8$ , suggesting that it would be worthwhile making the changes to the NAME output.

A synthetic test was made on the inversion method, in an otherwise idealised experiment by adding a measurement error to the true synthetic CO<sub>2</sub> concentration. The carbon flux uncertainty was strongly correlated to both the synthetic measurement uncertainty and also to the residence time of each homogeneous ecoregion (see Table 7.1). An alternative model (discussed in Section 7.2), included both a day and night ecoregion for each vegetation type. This would have the effect of doubling the number of ecoregions and would make finding the optimal carbon flux magnitudes more challenging.

Section 8.1 investigated using SCIAMACHY/FSI retrieved CO<sub>2</sub> fields to validate the CarbonTracker carbon fluxes. Since there were significant differences between the two CO<sub>2</sub> fields (both in terms of the negative offset and the larger seasonal cycle amplitude of SCIAMACHY/FSI as well as the large scatter in measured columns) this proved a rather difficult comparison. The reanalysis of the forward test using the area method (seen in Figures 8.9 to 8.15) filters out most of the noise, subsequently the correlation improved between the CarbonTracker and SCIAMACHY/FSI data sets.

The inversion results shown in Section 8.2 represents the first time that satellite retrieved CO<sub>2</sub> has been inverted to produce biosphere to atmosphere carbon fluxes (including using a Eulerian scheme). At the time of publication, there are no other satellite inverted carbon fluxes to compare with the results shown in Section 8.2. These results were not compared with established bottom-up and top-down carbon fluxes, though the methodology has been proven to be sound (as shown by the synthetic testing in Section 7.2) the potential of the inversion method would best be demonstrated once the following are adopted: GOSAT or OCO-like satellite retrieved CO<sub>2</sub>, a system for identifying the homogeneous flux ecoregions (such as that shown in Figure 1.7) and a “freeze” method for initialising the background atmospheric CO<sub>2</sub> field. However, the flux magnitudes (shown in figure 8.19) were of the correct order of magnitude that would be expected for the vegetation types; the seasonal cycle of Evergreen is the largest and for Ocean it is very small (see Figure 8.17).

## 9.3 Outlook

### 9.3.1 NAME model development

As well as the scientific advancements described in this thesis, the author has made an number of technical developments to the UK Met Office NAME model (see Section 6.1.1). The most important of these is adapting the model so that the receptor location becomes a vertical column (rather than a point release location) and can thus be coupled with satellite-retrieved trace gas concentrations. Whilst the setup in this thesis is tailored to the investigation of surface exchange of atmospheric CO<sub>2</sub>, the developments made by this author are now being used by the Leicester group, the Met Office and other groups to investigate other satellite-retrieved atmospheric trace gases.

The author has described the functioning of the “freeze” initialisation method (see Section 7.1.2), where the model outputs the time and location of each tracer particle as they exit the domain. Time constraints prevented such a change being made by the author, although developing this method should not be technically challenging.

In Section 6.2.4.2 a method of initialising the background using the same instrument as that used to obtain the receptor CO<sub>2</sub> concentration was described. This method was not adopted for SCIAMACHY/FSI since it requires use of the a satellites sun-glint mode in order to retrieve over ocean. This improvement could be tied to another suggested change, the use of “satellite time”, which could be used to centre satellite ensemble measurements onto the NAME receptor location when the measurement and model are coupled (see Section 6.2.4.2).

### 9.3.2 Investigation of high northern latitude carbon fluxes in winter

One area of interest that has been discussed previously in this thesis is the origin of high concentrations of atmospheric CO<sub>2</sub> in the high northern latitudes during the winter months [Zimov et al., 1993, 1996, 1999]. The measurement of atmospheric concentrations by SCIAMACHY/FSI in nadir mode is limited by the extremely large solar zenith angle (in some cases the sun may be below the horizon at 10 am, resulting in no solar reflected NIR radiation being detected) and also the fact

that these areas are usually snow covered resulting in a poor signal-to-noise ratio. Although a thermal infrared instrument (such as AIRS) could measure XCO<sub>2</sub> in this region it is not sensitive to the surface where the fluxes are taking place.

An alternative method for detecting trace gases in the atmosphere is using solar occultation (see Section 2.4.2.1). These measurements have a much higher signal-to-noise ratio, as the solar radiation is detected directly and so will be very precise. Solar occultation measurements can also provide vertical resolution information about the atmospheric CO<sub>2</sub>. Another useful feature is that the satellite crosses the polar region just over 14 times in 24 hours, so there is a much higher frequency of satellite measurements. One disadvantage is the very low spatial resolution of solar occultation, as a single line of sight may stretch over 1000 kilometres in the troposphere, although this line of sight sees roughly along the longitude band so the Siberian and Canadian regions could still be distinguished. Another problem with this method is that the long line of sight dramatically increases the likelihood of cloud contamination when measurements are made in the troposphere, which is where it is most desirable to monitor the changes in the CO<sub>2</sub> concentration.

GOSAT does not have a solar occultation mode. However, SCIAMACHY does operate in solar occultation and the Bremen group are currently working on the retrieval of CO<sub>2</sub> using this measurement mode for high latitude studies.

### 9.3.3 Inversion of retrieved CO<sub>2</sub> columns

SCIAMACHY/FSI retrieved XCO<sub>2</sub> was at the lower precision end of the range needed to improve upon the current understanding of carbon fluxes [Miller et al., 2007]. For some regions, the inversion of SCIAMACHY/FSI retrieved XCO<sub>2</sub> would already improve upon both the *a priori* and ground networks over some select regions (such as Africa and South America shown in Figure 2.4), were it not the case that the retrievals are subject to a negative offset-bias of unknown origin compared to the CarbonTracker CO<sub>2</sub> field needed to initialise the background concentration.

The failed launch of OCO was a setback to the future use of the NAME inversion method developed in this research project. It is not yet clear whether NASA will launch an OCO replica instrument or develop a completely new satellite instrument to monitor changes in the atmospheric CO<sub>2</sub> field. The measurement precision of GOSAT retrieved XCO<sub>2</sub> is close to 1%, making it suitable for use with the NAME inversion scheme. It is expected that the North American region would be used

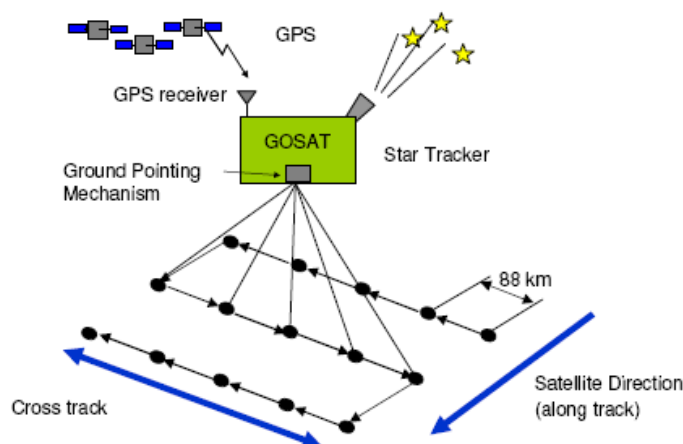


FIGURE 9.1. The across track viewing geometry of GOSAT (image courtesy of JAXA).

as a demonstration of the quality of the inversion scheme as it could be compared with other bottom-up and top-down carbon flux information. Assuming that it agrees well with other carbon flux models, the NAME inversion scheme could be used with GOSAT retrieved  $\text{XCO}_2$  to reduce carbon flux uncertainty over other continental regions (such as Africa and South America).

The identification of homogeneous ecoregions is likely to be more difficult in less inhabited parts of the world where there may be little or no directly measured information on vegetation and soil types. For these regions the classification of ecoregions would probably be done remotely (using weather products and chlorophyll indices).

### 9.3.3.1 GOSAT

The Greenhouse gases Observing SATellite (GOSAT) was successfully launched in early 2009 by the Japan Aerospace Exploration Agency (JAXA). Its mission objectives are to retrieve both  $\text{XCO}_2$  and  $\text{XCH}_4$  at a precisions of 1% and 2% respectively and to reduce sub-continental carbon flux uncertainties by a half. It was intended that the five year GOSAT mission would run in parallel with the failed OCO mission, since the two instruments would give credibility to each other assuming that they both agree.

The GOSAT instrument is a Fourier transform spectrometer designed to measure scattered and reflected sunlight in the NIR and thermal IR spectral regions at a spectral resolution of  $0.2\text{-}0.5\text{ cm}^{-1}$ . It has three NIR channels ( $0.75\text{-}0.78\text{ }\mu\text{m}$ ,



1.56-1.72  $\mu\text{m}$  and 1.92-2.08  $\mu\text{m}$ ) and one channel in the thermal IR (5.5–14.3  $\mu\text{m}$ ). The optical path length will be determined the from oxygen A-band (0.76  $\mu\text{m}$ ). GOSAT is also fitted with a cloud and aerosol detector to determine aerosol optical thickness and the presence of clouds (in order for filtering to take place).

GOSAT, like SCIAMACHY, has a low altitude and sun-synchronous polar orbit. GOSAT has an equator crossing time of 1:30 pm putting its measurements near the peak of the diurnal photosynthetic cycle. GOSAT has a wide across track swath width, very much like SCIAMACHY, which lends itself to the measurement ensemble release approach discussed in Chapter 6. The actual across track swath of GOSAT comprises of five across track “footprints” which unlike their SCIAMACHY equivalent do not join together (see Figure 9.1). The spatial resolution of the individual footprints are 10 kilometres (the resolution of the cloud and aerosol detector is 3 kilometres). Since the footprints do not overlap, the GOSAT coverage is not continuous. However, the wide swath of the instrument means that the coverage of GOSAT is effectively global.

Unlike the OCO instrument, GOSAT does not operate in sun-glint mode. This does limit the option of using the same instrument to initialise the background concentration over the ocean and also measure at the release location as has been discussed in Section 6.2.4.2. The added value of measuring both the initial and final concentrations with one instrument stems from the fact that there are offsets in the measured  $\text{CO}_2$  field between instruments and also with modelled  $\text{CO}_2$  fields.

### 9.3.4 Final Comments

The work undertaken in this thesis forms an integral part of the current research into carbon cycle processes. With the right set up, Lagrangian modelling offers a credible alternative to data assimilation in the determination of regional scale carbon fluxes. The higher precision  $\text{XCO}_2$  measurements offered by GOSAT and a future replacement OCO mission, are capable of enhancing the understanding of carbon flux magnitudes produced solely by the *in-situ* network. Both remotely sensed and *in-situ* sources of  $\text{CO}_2$  field information should be complementary rather than competitive and it is important that a modelling framework be developed that can make use of both.

# Bibliography

- Alvarez, L. W., Alvarez, W., Asaro, F. and Michel, H. V. [1980], ‘Extraterrestrial cause for the Cretaceous-Tertiary extinction’, *Science* **208**(4448), 1095–1108.
- Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P. J., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B. C., Taguchi, S. and Zhu, Z. [2006], ‘Transcom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO<sub>2</sub> fluxes 1988-2003’, *Global Biogeochemical Cycles* **20**(1).
- Baldocchi, D. D., Hicks, B. B. and Meyers, T. P. [1988], ‘Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods’, *Ecology* **69**(5), 1331–1340.
- Barkley, M. P. [2007], Measuring atmospheric carbon dioxide from space, PhD thesis, University of Leicester.
- Barkley, M. P., Friess, U. and Monks, P. S. [2006a], ‘Measuring atmospheric CO<sub>2</sub> from space using Full Spectral Initiation (FSI) WFM-DOAS’, *Atmospheric Chemistry and Physics* **6**(11), 3517–3534.
- Barkley, M. P., Monks, P. S. and Engelen, R. J. [2006b], ‘Comparison of SCIAMACHY and AIRS CO<sub>2</sub> measurements over north America during the summer and autumn of 2003’, *Geophysical Research Letters* **33**(20).
- Barkley, M. P., Monks, P. S., Friess, U., Mittermeier, R. L., Fast, H., Korner, S. and Heimann, M. [2006c], ‘Comparisons between SCIAMACHY atmospheric CO<sub>2</sub> retrieved using (FSI) WFM DOAS to ground based FTIR data and the TM3 chemistry transport model’, *Atmospheric Chemistry and Physics* **6**, 4483–4498.
- Barkley, M. P., Monks, P. S., Hewitt, A. J., MacHida, T., Desai, A., Vinnichenko, N., Nakazawa, T., Arshinov, M. Y., Fedoseev, N. and Watai, T. [2007], ‘Assessing

- the near surface sensitivity of SCIAMACHY atmospheric CO<sub>2</sub> retrieved using (FSI) WFM-DOAS’, *Atmospheric Chemistry and Physics* **7**(13), 3597–3619.
- Blackman, F. F. and Matthaei, G. L. C. [1901], ‘On the reaction. of leaves to traumatic stimulation.’, *Ann. Bot.* **15**, 553.
- Boesch, H., Toon, G. C., Sen, B., Washenfelder, R. A., Wennberg, P. O., Buchwitz, M., de Beek, R., Burrows, J. P., Crisp, D., Christi, M., Connor, B. J., Natraj, V. and Yung, Y. L. [2006], ‘Space-based near-infrared CO<sub>2</sub> measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin’, *Journal of Geophysical Research D: Atmospheres* **111**(23).
- Boucher, O. and Haywood, J. [2001], ‘On summing the components of radiative forcing of climate change’, *Climate Dynamics* **18**(3-4), 297–302.
- Bovensmann, H. and Burrows, J. P. [1998], SCIAMACHY scientific requirements document, Technical report, DLR.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S. and Rozanov, V. V. [1999], ‘SCIAMACHY: Mission objectives and measurement modes’, *Journal of the Atmospheric Sciences* **56**(2), 127–150.
- Bowne, N. E. and Londergan, R. J. [1983], Overview, results, and conclusions for the EPRI plume model validation and development project: Plains site, Technical report, TRC.
- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. and Jones, P. D. [2006], ‘Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850’, *Journal of Geophysical Review* **111**, D12106.
- Buchwitz, M. and Burrows, J. P. [2004], Retrieval of CH<sub>4</sub> CO and CO<sub>2</sub> total column amounts from SCIAMACHY near-infrared nadir spectra retrieval algorithm and first results, *in* ‘Remote Sensing of Clouds and the Atmosphere VIII Proceedings of SPIE’.
- Buchwitz, M., Rozanov, V. V. and Burrows, J. P. [2000a], ‘A near-infrared optimized DOAS method for the fast global retrieval of atmospheric CH<sub>4</sub>, CO, CO<sub>2</sub>, H<sub>2</sub>O, and N<sub>2</sub>O total column amounts from SCIAMACHY ENVISAT-1 nadir radiances’, *Journal of Geophysical Research D: Atmospheres* **105**(D12), 15231–15245.

- Buchwitz, M., Rozanov, V. V. and Burrows, J. P. [2000b], ‘A correlated-k distribution scheme for overlapping gases suitable for retrieval of atmospheric constituents from moderate resolution radiance measurements in the visible/near-infrared spectral region’, *Journal of Geophysical Research D: Atmospheres* **105**(D12), 15247–15261.
- Burrows, J. P., Chance, K. V., Crutzen, P. J., van Dop, H., Geary, J. C., Johnson, T. J., Harris, G. W., Isaksen, I. S. A., Moortgat, G. K., Muller, C., Perner, D., Platt, U., Pommereau, J. P., Rodhe, H., Roeckner, E., Schneider, W., Simon, P., Sundqvist, H. and Vercheval, J. [1988], *SCIAMACHY a european proposal for atmospheric remote sensing from the ESA polar platform*. published by Max-Planck-Institut fr Chemie, Mainz, Germany.
- Calvin, M. and Benson, A. A. [1948], ‘The path of carbon in photosynthesis’, *Science* **107**(2784), 476–480.
- Canadell, J. G., Quere, C. L., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A. and Marland, G. [2007], ‘Contributions to accelerating atmospheric CO<sub>2</sub> growth from economic activity carbon intensity and efficiency of natural sinks’, *Proceedings of the National Academy of Sciences of the United States of America* **104**(47), 18866–18870.
- Chan, D., Ishizawa, M., Higuchi, K., Maksyutov, S. and Chen, J. [2008], ‘Seasonal CO<sub>2</sub> rectifier effect and large-scale extratropical atmospheric transport’, *Journal of Geophysical Research D: Atmospheres* **113**(17).
- Chevallier, F., Breon, F. and Rayner, P. J. [2007], ‘Contribution of the orbiting carbon observatory to the estimation of CO<sub>2</sub> sources and sinks: Theoretical study in a variational data assimilation framework’, *Journal of Geophysical Research D: Atmospheres* **112**(9).
- Chevallier, F., Engelen, R. J., Carouge, C., Conway, T. J., Peylin, P., Pickett-Heaps, C., Ramonet, M., Rayner, P. J. and Xueref-Remy, I. [2009], ‘AIRS-based versus flask-based estimation of carbon surface fluxes’, *Journal of Geophysical Research D: Atmospheres* **114**(20).
- Chiti, T., Certini, G., Grieco, E. and Valentini, R. [2010], ‘The role of soil in storing carbon in tropical rainforests: The case of ankasa park ghana’, *Plant and Soil* **331**(1).

- Ciais, P., Denning, A. S., Tans, P. P., Berry, J. A., Randall, D. A., Collatz, G. J., Sellers, P. J., White, J. W. C., Trolier, M., Meijer, H. A. J., Francey, R. J., Monfray, P. and Heimann, M. [1997], ‘A three-dimensional synthesis study of d18O in atmospheric CO<sub>2</sub> 1. surface fluxes’, *Journal of Geophysical Research D: Atmospheres* **102**(5), 5857–5872.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., Noblet, N. D., Friend, A. D., Friedlingstein, P., Grunwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T. and Valentini, R. [2005], ‘Europe-wide reduction in primary productivity caused by the heat and drought in 2003’, *Nature* **437**(7058), 529–533.
- Collins, W. J., Stevenson, D. S., Johnson, C. E. and Derwent, R. G. [1997], ‘Tropospheric ozone in a global-scale three-dimensional Lagrangian model and its response to NO(x) emission controls’, *Journal of Atmospheric Chemistry* **26**(3), 223–274.
- Comyn-Platt, E. M. and Boesch, H. [2009], Validation of carbon dioxide measurement made by SCIAMACHY. Undergraduate Project Report.
- Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C. and Crisp, D. [2008], ‘Orbiting Carbon Observatory: Inverse method and prospective error analysis’, *Journal of Geophysical Research D: Atmospheres* **113**(5).
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H. and Rowntree, P. R. [1999], ‘The impact of new land surface physics on the GCM simulation of climate and climate sensitivity’, *Climate Dynamics* **15**, 183–203.
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. and Totterdell, I. J. [2000], ‘Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model’, *Nature* **408**(6809), 184–187.
- Cullen, M. J. P. [1993], ‘The unified forecast/climate model’, *Meteorological Magazine* **122**(1449), 81–94.
- Dash, J. and Curran, P. J. [2004], ‘The MERIS terrestrial chlorophyll index’, *International Journal of Remote Sensing* **25**(23), 5403–5413.

- Dash, J. and Hewitt, A. J. [n.d.], MTCI vs FSI-WFM-DOAS retrieved CO<sub>2</sub>. To be submitted for publication.
- Davidson, E. A. and Janssens, I. A. [2006], ‘Temperature sensitivity of soil carbon decomposition and feedbacks to climate change’, *Nature* **440**(27081).
- Dilling, L., Doney, S. C., Edmonds, J., Gurney, K. R., Harriss, R., Schimel, D., Stephens, B. and Stokes, G. [2003], *The role of carbon cycle observations and knowledge in carbon management*, Vol. 28.
- Dils, B., Maziere, M. D., Muller, J. F., Blumenstock, T., Buchwitz, M., de Beek, R., Demoulin, P., Duchatelet, P., Fast, H., Frankenberg, C., Gloudemans, A., Griffith, D., Jones, N., Kerzenmacher, T., Kramer, I., Mahieu, E., Mellqvist, J., Mittermeier, R. L., Notholt, J., Rinsland, C. P., Schrijver, H., Smale, D., Strandberg, A., Straume, A. G., Stremme, W., Strong, K., Sussmann, R., Taylor, J., van den Broek, M., Velasco, V., Wagner, T., Warneke, T., Wiacek, A. and Wood, S. [2006], ‘Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O’, *Atmospheric Chemistry and Physics* **6**(7), 1953–1976.
- Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B. and Gatti, L. V. [2009], ‘Observational constraints on recent increases in the atmospheric CH<sub>4</sub> burden’, *Geophysical Research Letters* **36**(18).
- Dlugokencky, E. J., Houweling, S., Bruhwiler, L., Masarie, K. A., Lang, P. M., Miller, J. B. and Tans, P. P. [2003], ‘Atmospheric methane levels off: Temporary pause or a new steady-state?’, *Geophysical Research Letters* **30**(19), ASC 5–1 – ASC 5–4.
- Dufour, E. and Breon, F. [2003], ‘Spaceborne estimate of atmospheric CO<sub>2</sub> column by use of the differential absorption method: Error analysis’, *Applied Optics* **42**(18), 3595–3609.
- Eisinger, M. and Burrows, J. P. [1998], ‘Tropospheric sulfur dioxide observed by the ERS-2 GOME instrument’, *Geophysical Research Letters* **25**(22), 4177–4180.
- Engelen, R. J., Andersson, E., Chevallier, F., Hollingsworth, A., Matricardi, M., McNally, A. P., Thepaut, J. N. and Watts, P. D. [2004], ‘Estimating atmospheric CO<sub>2</sub> from advanced infrared satellite radiances within an operational 4D-Var

- data assimilation system: Methodology and first results', *Journal of Geophysical Research D: Atmospheres* **109**(19), D19309 1–9.
- Engelen, R. J., Serrar, S. and Chevallier, F. [2009], 'Four-dimensional data assimilation of atmospheric CO<sub>2</sub> using airs observations', *Journal of Geophysical Research D: Atmospheres* **114**(3).
- ESRL, N. [2008a], 'CarbonTracker'.  
**URL:** <http://carbontracker.noaa.gov>
- ESRL, N. [2008b], 'GLOBALVIEW CO<sub>2</sub>'.  
**URL:** <http://www.esrl.noaa.gov>
- Farman, J. C., Gardiner, B. G. and Shanklin, J. D. [1985], 'Large losses of total ozone in antarctica reveal seasonal ClO<sub>x</sub>/NO<sub>x</sub> interaction', *Nature* **315**(6016), 207–210.
- Feng, L., Palmer, P. I., Boesch, H. and Dance, S. [2009], 'Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman Filter', *Atmospheric Chemistry and Physics* **9**, 2619–2633.
- Frankenberg, C., Platt, U. and Wagner, T. [2005], 'Retrieval of CO from SCIAMACHY onboard ENVISAT: detection of strongly polluted areas and seasonal patterns in global CO abundances', *Atmospheric Chemistry and Physics* **5**, 1639–1644.
- Frerick, J. [2006], The sciallc commandline tool, ESA technical note PO-TN-ESA-GS-2307, Technical report, ESA.
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C. [2002], 'Global land cover mapping from MODIS: Algorithms and early results', *Remote Sensing of Environment* **83**(1-2), 287–302.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C. and Zeng, N. [2006], 'Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison', *Journal of Climate* **19**(14), 3337–3353.

- Friess, U. [2001], Spectroscopic measurements of atmospheric trace gases at Nueymayer Station, Antarctica, PhD thesis, University of Heidelberg.
- Fung, I. Y., Doney, S. C., Lindsay, K. and John, J. [2005], ‘Evolution of carbon sinks in a changing climate’, *Proceedings of the National Academy of Sciences of the United States of America* **102**(32), 11201–11206.
- Gazetteer, W. [2003], ‘World gazetteer’.  
**URL:** <http://www.world-gazetteer.com/>
- Gottwald, M., ed. [1987], *SCIAMACHY, Monitoring the changing Earth’s atmosphere*, DLR.
- Gruber, N., Gloor, M., Fletcher, S. E. M., Doney, S. C., Dutkiewicz, S., Follows, M. J., Gerber, M., Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Muller, S. A., Sarmiento, J. L. and Takahashi, T. [2009], ‘Oceanic sources, sinks, and transport of atmospheric CO<sub>2</sub>’, *Global Biogeochemical Cycles* **23**(1).
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T. and Yuen, C. [2003], ‘Transcom 3 CO<sub>2</sub> inversion intercomparison: 1. annual mean control results and sensitivity to transport and prior flux information’, *Tellus, Series B: Chemical and Physical Meteorology* **55**(2), 555–579.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T. and Yuen, C. [2002], ‘Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models’, *Nature* **415**, 626630.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M. and Taguchi, S. [2004], ‘Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks’, *Global Biogeochemical Cycles* **18**(1), GB10101–18.



- Hansen, J. E., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., Peterson, T. and Karl, T. [2001], ‘A closer look at united states and global surface temperature change’, *Journal of Geophysical Review* **106**, 23947–23963.
- Hansen, J. and Sato, M. [2004], ‘Greenhouse gas growth rates’, *Proceedings of the National Academy of Sciences of the United States of America* **101**(46), 16109–16114.
- Hargrove, W. W. and Hoffman, F. M. [2005], ‘Potential of multivariate quantitative methods for delineation and visualization of ecoregions’, *Environmental Management* **34**(1 SUPPL.), S39–S60.
- Hatch, M. D. and Slack, C. R. [1966], ‘Photosynthesis by sugar-cane leaves. a new carboxylation reaction and the pathway of sugar formation.’, *Biochemical Journal* **101**(1), 103–111.
- Haug, G. H., Gunther, D., Peterson, L. C., Sigman, D. M., Hughen, K. A. and Aeschlimann, B. [2003], ‘Climate and the collapse of Maya civilization’, *Science* **299**(5613), 1731–1735.
- Heimann, M. and Korner, S. [2003], The global atmospheric tracer model TM3 model description and users manual, Technical report, Max Planck Institute for Biogeochemistry (MPI-BGC) Jena Germany.
- Heimann, M. and Reichstein, M. [2008], ‘Terrestrial ecosystem carbon dynamics and climate feedbacks’, *Nature* **451**(7176), 289–292.
- Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C. and Celarier, E. [1997], ‘Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data’, *Journal of Geophysical Review* **102**, 1691116922.
- Hewitt, A. J., Barkley, M. P. and Monks, P. S. [2006], Retrieval of CH<sub>4</sub> CO and CO<sub>2</sub> total column amounts from SCIAMACHY near-infrared nadir spectra retrieval algorithm and first results, in ‘3rd Workshop on the Atmospheric Chemistry of ENVISAT (ACVE-3)’.
- Hollinger, D. Y., Aber, J., Dail, B., Davidson, E. A., Goltz, S. M., Hughes, H., Leclerc, M. Y., Lee, J. T., Richardson, A. D., Rodrigues, C., Scott, N. A., Achuatavarier, D. and Walsh, J. [2004], ‘Spatial and temporal variability in forest-atmosphere CO<sub>2</sub> exchange’, *Global Change Biology* **10**(10), 1689–1706.

- Houweling, S., Breon, F. M., Aben, I., Rodenbeck, C., Gloor, M., Heimann, M. and Ciais, P. [2004], ‘Inverse modeling of CO<sub>2</sub> sources and sinks using satellite data: A synthetic inter-comparison of measurement techniques and their performance as a function of space and time’, *Atmospheric Chemistry and Physics* **4**(2), 523–538.
- Houweling, S., Hartmann, W., Aben, I., Schrijver, H., Skidmore, J., Roelofs, G. J. and Breon, F. M. [2005], ‘Evidence of systematic errors in SCIAMACHY observed CO<sub>2</sub> due to aerosols’, *Atmospheric Chemistry and Physics* **5**(11), 3003–3013.
- IPCC [2001], *Climate Change 2001: Synthesis Report: Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York.
- IPCC [2007], *Climate Change 2007: Synthesis Report: Forth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York.
- Jones, A., Thomson, D., Hort, M. and Devenish, B. [2004], The UK Met Office next generation atmospheric dispersion model, NAME III, Technical report, UK Met Office.
- Keeling, C. D. [1960], ‘The concentration and isotopic abundances of carbon dioxide in the atmosphere’, *Tellus* **12**.
- Kiehl, J. T. [1986], ‘Changes in the radiative balance of the atmosphere due to increases in CO<sub>2</sub> and trace gases’, *Advances in Space Research* **6**(10), 55–60.
- Kiehl, J. T. and Trenberth, K. E. [1997], ‘Earth’s annual global mean energy budget’, *Bulletin of the American Meteorological Society* **78**(2), 197–208.
- Kleipool, Q. [2003a], Algorithm specification for dark signal determination, Technical report, SRON.
- Kleipool, Q. [2003b], Recalculation of OPTEC5 non-linearity, report containing the NL correction to be implemented in the data processor, Technical report, SRON.
- Kleipool, Q. [2004a], SCIAMACHY: Evolution of dead and bad pixel mask, Technical report, SRON.
- Kleipool, Q. L. [2004b], SCIAMACHY: Orbital variation of dark signal, Technical report, SRON.

- Kneizys, F. X., Abreu, L. W., Anderson, G. P., Shettle, E. P., Chetwynd, J. H., Shettle, E. P., Berk, A., Bernstein, L., Robertson, D., Acharya, P., Rothman, L., Selby, J. E. A., Allery, W. O. and Clough, S. A. [1996], The MODTRAN 2/3 report and LOWTRAN 7 model, Technical report.
- Krijger, J. M., Aben, I. and Schrijver, H. [2005], ‘Distinction between clouds and ice/snow covered surfaces in the identification of cloud-free observations using SCIAMACHY PMDs’, *Atmospheric Chemistry and Physics* **5**(10), 2729–2738.
- Lafont, S., Kergoat, L., Dedieu, G., Chevillard, A., Karstens, U. and Kolle, O. [2002], ‘Spatial and temporal variability of land CO<sub>2</sub> fluxes estimated with remote sensing and analysis data over western Eurasia’, *Tellus, Series B: Chemical and Physical Meteorology* **54**(5), 820–833.
- Law, R. M., Chen, Y. and Gurney, K. R. [2003], ‘Transcom 3 CO<sub>2</sub> inversion inter-comparison: 2. sensitivity of annual mean results to data choices’, *Tellus, Series B: Chemical and Physical Meteorology* **55**(2), 2580.
- Lewis, S. L., Lopez-Gonzalez, G., Sonke, B., Affum-Baffoe, K., Baker, T. R., Ojo, L. O., Phillips, O. L., Reitsma, J. M., White, L., Comiskey, J. A., Djuikouo, M. N. K., Ewango, C. E. N., Feldpausch, T. R., Hamilton, A. C., Gloor, M., Hart, T., Hladik, A., Lloyd, J., Lovett, J. C., Makana, J., Malhi, Y., Mbago, F. M., Ndangalasi, H. J., Peacock, J., Peh, K. S., Sheil, D., Sunderland, T., Swaine, M. D., Taplin, J., Taylor, D., Thomas, S. C., Votere, R. and Woll, H. [2009], ‘Increasing carbon storage in intact African tropical forests’, *Nature* **457**(7232), 1003–1006.
- Lichtenberg, G., Kleipool, Q., Krijger, J. M., Soest, G. V., Hees, R. V., Tilstra, L. G., Acarreta, J. R., Aben, I., Ahlers, B., Bovensmann, H., Chance, K., Gloudemans, A. M. S., Hoogeveen, R. W. M., Jongma, R. T. N., Noel, S., Pitters, A., Schrijver, H., Schrijvers, C., Sioris, C. E., Skupin, J., Slijkhuis, S., Stammes, P. and Wuttke, M. [2006], ‘SCIAMACHY level 1 data: Calibration concept and in-flight calibration’, *Atmospheric Chemistry and Physics* **6**(12), 5347–5367.
- Lichtenberg, G., Kleipool, Q., Krijger, J. M., van Soest, G., van Hees, R., Tilstra, L. G., Acarreta, J. R., Aben, I., Ahlers, B., Bovensmann, H., Chance, K., Gloudemans, A. M. S., Hoogeveen, R. W. M., Jongma, R., Noel, S., Pitters, A., Schrijver, H., Schrijvers, C., Sioris, C. E., Skupin, J., Slijkhuis, S., S. P. and Wuttke, M. [2005], ‘SCIAMACHY level1 data: Calibration concept and in-flight calibration’, *Atmos. Chem. Phys. Discuss.* **5**, 8925–8977.

- Lin, B., Chambers, L., Jr., P. S., Wielicki, B., Hu, Y., Minnis, P., Loeb, N., Sun, W., Potter, G., Min, Q., Schuster, G. and Fan, T. [2010], ‘Estimations of climate sensitivity based on top-of-atmosphere radiation imbalance’, *Atmospheric Chemistry and Physics* **10**, 1923–1930.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Chow, V. Y., Gottlieb, E., Daube, B. C. and Matross, D. M. [2007], ‘Designing Lagrangian experiments to measure regional-scale trace gas fluxes’, *Journal of Geophysical Research D: Atmospheres* **112**(13).
- Lugina, K., Groisman, P., Vinnikov, K., Koknaeva, V. and Speranskaya, N. [2005], Monthly surface air temperature time series area-averaged over the 30-degree latitudinal belts of the globe, 1881-2005, Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy.
- Mann, M. E., Bradley, R. S. and Hughes, M. K. [1999], ‘Northern hemisphere temperatures during the past millennium: Inferences uncertainties and limitations’, *Geophysical Research Letters* **26**(6), 759–762.
- Manne, A. S. and Richels, R. G. [2001], ‘An alternative approach to establishing trade-offs among greenhouse gases’, *Nature* **410**(6829), 675–677.
- Manning, A. J., Ryall, D. B., Derwent, R. G., Simmonds, P. G. and O’Doherty, S. [2003], ‘Estimating European emissions of ozone-depleting and greenhouse gases using observations and a modeling back-attribution technique’, *Journal of Geophysical Research D: Atmospheres* **108**(14), ACH 3–1 – ACH 3–12.
- Maryon, R. H., Ryall, D. B. and Malcolm, A. L. [1999], The NAME 4 dispersion model: scientific documentation., Technical Report Turbulence and Diffusion note 262, UK Met Office.
- Maryon, R. H., Smith, F. B., Conway, B. J. and Goddard, D. M. [1991], ‘The U.K. nuclear accident model’, *Progress in Nuclear Energy* **26**(2), 85–104.
- Masarie, K. and Tans, P. [1995], ‘Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record’, *Journal of Geophysical Research D: Atmospheres* **100**(D6), 11593–11610.
- Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H.,

- Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L. and Law, R. M. [2007], 'Precision requirements for space-based XCO<sub>2</sub> data', *Journal of Geophysical Research D: Atmospheres* **112**(10).
- Miller, S. D., Marandino, C. and Saltzman, E. S. [2010], 'Ship-based measurement of air-sea CO<sub>2</sub> exchange by eddy covariance', *Journal of Geophysical Research D: Atmospheres* **115**.
- Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. and New, M. [2004], A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901 to 2000) and 16 scenarios (2001 to 2100)., Technical report, Tyndall Centre for Climate Change Research.
- Molina, M. J. and Rowland, F. S. [1974], 'Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone', *Nature* **249**, 810–812.
- Montenegro, A., Brovkin, V., Eby, M., Archer, D. and Weaver, A. J. [2007], 'Long term fate of anthropogenic carbon', *Geophysical Research Letters* **34**(19).
- Nakicenovic, N., Victor, N. and Morita, T. [1998], 'Emissions scenarios database and review of scenarios', *Mitigation and Adaptation Strategies for Global Change* **3**(2-4), 95–120.
- Olsen, S. C. and Randerson, J. T. [2004], 'Differences between surface and column atmospheric CO<sub>2</sub> and implications for carbon cycle research', *Journal of Geophysical Research D: Atmospheres* **109**(2), D02301 1–11.
- Palmer, P. I. [2008], 'Quantifying sources and sinks of trace gases using space-borne measurements: Current and future science', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**(1885), 4509–4528.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C. and Tans, P. P. [2007], 'An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker', *Proceedings of the National Academy of Sciences of the United States of America* **104**(48), 18925–18930.

- Philipona, R., Durr, B., Ohmura, A. and Ruckstuhl, C. [2005], ‘Anthropogenic greenhouse forcing and strong water vapor feedback increase temperature in europe’, *Geophysical Research Letters* **32**(19), 1–4.
- Platt, U., Perner, D. and Patz, H. [1979], ‘Simultaneous measurement of atmospheric CH<sub>2</sub>O, O<sub>3</sub> and NO<sub>2</sub> by differential optical absorption in the near UV’, *Journal of Geophysical Review* **84**, 6329–6335.
- Platt, U. and Stutz, J., eds [2008], *Differential Optical Absorption Spectroscopy*, Springer.
- Press, W. H., ed. [1992], *Numerical Recipies in Fortran: The art of scientific computing, second edition.*, Cambridge University Press.
- Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. J. and Tett, S. F. B. [2006], ‘Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset’, *Journal of Climate* **19**, 446–469.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C. and Kaplan, A. [2003], ‘Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century’, *Journal of Geophysical Research* **108**(D14), 4407.
- Rayner, P. and O’Brien, D. [2001], ‘The utility of remotely sensed CO<sub>2</sub> concentration data in surface source inversions’, *Geophysical Research Letters* **28**(1), 175–178.
- Remedios, J. J., Parker, R. J., Panchal, M., Leigh, R. J. and Corlett, G. [2006], Signatures of atmospheric and surface climate variables through analyses of infrared spectra SATSCAN-IR, in ‘Proceedings of the first EPS/METOP RAO Workshop, ESRIN’.
- Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H. and Burrows, J. P. [2010], ‘A method for improved SCIAMACHY CO<sub>2</sub> retrieval in the presence of optically thin clouds’, *Atmospheric Measurement Techniques* **3**, 209–232.
- Rodenbeck, C., Gerbig, C., Trusilova, K. and Heimann, M. [2009], ‘A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models’, *Atmospheric Chemistry and Physics* **9**, 53315342.

- Rodgers, C. D., ed. [2000], *Inverse methods for atmospheric sounding: Theory and practice*, World Scientific.
- Rothman, L., Jacquemart, D., Barbe, A., Benner, C. D., Birk, M., Brown, L. R., Carleer, M. R., Chackerian Jr., C., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Flaud, J.-M., Gamache, R. R., Goldman, A., Hartmann, J.-M., Jucks, J. W., Maki, A. G., Mandin, J.-Y., Massie, S. T., Orphal, J., Perrin, A., Rinsland, C. P., Smith, M., Tennyson, J., Tolchenov, R. N., Toth, R. A., Vander Auwera, J., Varanasi, P. and Wagner, G. [2005], ‘The HITRAN 2004 molecular spectroscopic database’, *J. Quant. Spectrosc. Radiat. Transfer* **96**, 193–204.
- Rozanov, V. V., Buckwitz, M., Eichmann, K. U., de Beek, R. and Burrows, J. P. [2002], ‘SCIATRAN a new radiative transference model for geophysical applications in the 240-2400nm spectral range’, *Adv. Space Res.* **29**, 1831–1835.
- Ryall, D. B. and Maryon, R. H. [1998], ‘Validation of the UK Met Office NAME model against the ETEX dataset’, *Atmospheric Environment* **32**(24), 4265–4276.
- Sarmiento, J. L. and Gruber, N. [2002], ‘Sinks for anthropogenic carbon’, *Physics Today* **55**(8), 30.
- Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macatangay, R. and Warneke, T. [2008], ‘Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - part 1: Carbon dioxide’, *Atmospheric Chemistry and Physics* **8**(14), 3827–3853.
- Smith, P., Smith, J., Wattenbach, M., Meyer, J., Lindner, M., Zaehle, S., Niederer, R., Jones, R. J. A., Montanarella, L., Rounsevell, M., Reginster, I. and Kankaanpää, S. [2006], ‘Projected changes in mineral soil carbon of European forests 1990 to 2100’, *Canadian Journal of Soil Science* **86**(2), 159–169.
- Smith, T. M. and Reynolds, R. [2005], ‘A global merged land/sea surface temperature reconstruction based on historical observations (1880 to 1997).’.
- Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J. and Nojiri, Y. [2002], ‘Global sea-air CO<sub>2</sub> flux based on climatological surface ocean CO<sub>2</sub> and seasonal biological and temperature effects’, *Deep Sea Research II* **49**, 1601–1622.

- Thornton, P. E., Running, S. W. and Hunt, E. R. [2005], Biome-BGC: Terrestrial ecosystem process model, version 4.1.1, Technical report, DAAC Oak Ridge National Laboratory.
- Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. and Chapin, F. S. [2006], ‘Methane bubbling from siberian thaw lakes as a positive feedback to climate warming’, *Nature* **443**(7107), 71–75.
- WMO [2005], *WMO WDCGG Data Summary WDGCC No.29 GAW Data Volume IV-Greenhouse Gases and Other Atmospheric Gases*.
- WMO [2006], *Scientific Assessment of Ozone Depletion*.
- Wuttke, M. W., Noel, S., Skupin, J., Gerilowski, K., Bovensmann, H. and Burrows, J. P. [2004], SCIAMACHY on ENVISAT: Instrument monitoring and calibration two years after launch, Vol. 5570, pp. 391–400.
- Yi, C., Li, R., Bakwin, P. S., Desai, A., Ricciuto, D. M., Burns, S. P., Turnipseed, A. A., Wofsy, S. C., Munger, J. W., Wilson, K. and Monson, R. K. [2004], ‘A nonparametric method for separating photosynthesis and respiration components in CO<sub>2</sub> flux measurements’, *Geophysical Research Letters* **31**(17), L17107 1–5.
- Zimov, S. A., Davidov, S. P., Voropaev, Y. V., Prosiannikov, S. F., Semiletov, I. P., Chapin, M. C. and Chapin, F. S. [1996], ‘Siberian CO<sub>2</sub> efflux in winter as a CO<sub>2</sub> source and cause of seasonality in atmospheric CO<sub>2</sub>’, *Climatic Change* **33**(1), 111–120.
- Zimov, S. A., Davidov, S. P., Zimova, G. M., Davidova, A. I., III, F. S. C., Chapin, M. C. and Reynolds, J. [1999], ‘Contribution of disturbance to high-latitude amplification of atmospheric CO<sub>2</sub>’, *Science* **284**, 1973–1976.
- Zimov, S. A., Zimova, G. M., Daviodov, S. P., Daviodova, A. I., Voropaev, Y. V., Voropaeva, Z. V., Prosiannikov, S. F., Prosiannikova, O. V., Semiletova, I. V. and Semiletov, L. P. [1993], ‘Winter biotic activity and production of CO<sub>2</sub> in siberian soils a factor in the greenhouse effect’, *Journal of Geophysical Research* **98**(D3), 50175023.