Exploring optimising strategies for sampling sea-air carbon dioxide fluxes in the Southern Ocean

Nicholas Pringle

August 19, 2014

A thesis presented for the degree of
Master of Science
in the Department of Oceanography
Faculty of Science
University of Cape Town
South Africa

Abstract

A model study was undertaken to investigate the optimization of sampling strategies for returning low-uncertainty sea-air CO_2 flux measurements in the Southern Ocean.

Replicating Lenton et al. (2006) using the ORCA2/PISCES ocean biological model shows that sampling 4 times a year, every 2° in latitude and every 40° in longitude reduces the uncertainty of estimating annual CO_2 flux estimates such that sampling at a higher frequency does not reduce the total uncertainty in proportion to the increase in sampling effort. Sampling at this frequency results in a annual CO_2 uptake estimate of 0.347 PgC.yr^{-1} with a sampling error of 0.55 PgC.yr^{-1} and a total sampling uncertainty of 0.089 PgC.yr^{-1}.

Investigation of the variability of the model shows that certain regions (eg. eastern Weddel Gyre, coastal Southern Ocean) show high variability. In these regions, the contributions to variability from the annual cycle and the non-seasonal variability are approximately equal. Therefore, there are areas that should be sampled at higher frequencies than others. This knowledge suggests that sampling using objective sampling strategies would be more efficient in reducing uncertainty in flux estimates.

Employing a genetic algorithm approach to find optimized sampling strategies (not relying on a regularly spaced grid) enables the calculation of uncertainty for objective sampling strategies. optimized sampling strategies using 1000, 2000, 5000 and 7000 are tested and it was shown that the uncertainty from using optimized sampling strategies is much lower than the uncertainty from a regular gridded approach. Sampling using 1000 locations results in an annual sea-air CO_2 flux estimate of 1.81 mmol.m^{-2}.day^{-1} with a total sampling uncertainty of 0.07 mmol.m^{-2}.day^{-1} which is equivalent to sampling at the model resolution.

An optimisation method that makes use of successive Radial Basis Function (RBF) interpolations to identify sampling locations of importance is used. The resulting locations selected by this optimisation method are shown to be found in high variability locations according to the model and observational studies.

If these sampling locations were to be sampled by wave-gliders with pCO2 sensors in addition to the current ship-based sampling efforts, the uncertainty in regional CO_2 flux estimates can be reduced to 10-15% of annual flux estimates.

Declaration

I know the meaning of Plaigarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Supervisors

Dr. Pedro M.S. Monteiro
Ocean Systems and Climate Council for Scientific and Industrial Research, Stellenbosch, South Africa

Dr. Howard Waldron
Department of Oceanography University of Cape Town South Africa

Acknowledgements

I would like to thank all those who helped and supported me during this thesis. My supervisors, Dr Howard Waldron and Dr Pedro Monteiro. Pedro Monteiro for the ideas, comments, advice and patience. Schalk Kok for help with the Genetic Algorithms and RBF sections as well as general comments. The SOCCO group for their useful input. My family for their never-ending support and advice. The staff at CSIR and UCT for all their help.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background: The problem

Opportunities for oceanographic measurements in the Southern Ocean are far fewer than in the Northern Hemisphere oceans. This is especially true for measurements of CO_2 concentrations and fluxes (Lenton et al., 2006; Lenton, 2009; Monteiro et al., 2010; Wanninkhof et al., 2012). The use of underway pCO_2 measurements in commercial shipping has provided valuable and extensive data in the Northern Hemisphere, but the Southern Hemisphere lacks the intensity of commercial shipping lanes and therefore the required temporal and spatial coverage of surface ocean pCO_2 data.

The Southern Ocean and its role in the changing global carbon cycle needs to be more extensively observed and understood in order for meaningful constraints to emerge for ocean-atmosphere and climate models. The trajectory of atmospheric CO_2 and its impact on climate change will not be understood or accurately estimated without a sound knowledge of the response of Southern Ocean CO_2 concentrations and fluxes. The inter-annual change in the flux of CO_2 needs to be resolved but, due to there being too few observations, the uncertainty is too high (Takahashi et al., 2009; Wanninkhof et al., 2012).

The Southern Ocean is a sink for atmospheric carbon in that it has the ability to take up carbon from the atmosphere. As the CO_2 concentration in the ocean increases, the ability of the Southern Ocean to act as a sink becomes weakened.

"Without understanding the carbon exchange at the boundary with a high degree of certainty and being able to account for random events which may affect data gathered or phenomena being measured, it is not possible to estimate the long term impact of changing ocean CO_2 sink on the effectiveness of global emission reduction targets." - ScienceScope Volume 6 Number 1

1.2 CO_2 and Climate in the Southern Ocean

The Southern Ocean is a regulator of regional climate and global atmospheric CO_2. While the Northern Hemisphere is dominated by terrestrial landmass, the Southern Hemisphere is dominated by the circumpolar Southern Ocean. The seasonal variability of atmospheric CO_2 is dominated by

the land biosphere; whereas the Southern Ocean governs long-term trends on decadal and greater time scales.

Climate change puts the long term balance Southern Ocean-atmosphere system at risk. Changes to ocean systems are not as manageable as terrestrial and atmospheric systems and can have significant positive feedback (i.e negative impacts).

The Southern Ocean (south of 50°) holds 9% of the global oceanic anthropogenic CO_2 (Sabine et al., 2004). Two mechanisms, the solubility pump and the biological pump, mediate the storage. The solubility pump is the result of uptake due to cooling and sinking of sub-tropical waters and the biological pump is the result of phytoplankton taking up carbon and transporting it to the ocean interior (Volk and Hoffert, 1985).

While the Southern Ocean is a net sink of anthropogenic carbon, there is a large outgassing of CO_2 due to Circumpolar Deep Water upwelling south of the Polar Front (Bakker et al., 2008). This is the only place in the world where water from below 2000m with pCO_2 greater than 450 μatm is in direct contact with the atmosphere. For the Southern Ocean, the median annual sea-air CO_2 flux is calculated between -0.27 and -0.43 $PgC.yr^{-1}$ (Lenton et al., 2013). There are large differences between sources and sinks in the region which are sensitive to small adjustments in climate forcing.

1.3 Concurrent research

To determine the partial pressure of CO_2 at the ocean-atmosphere interface, ship cruise data are being used to find a numerical model to couple measurements of sea surface temperature, mixed layer depth (MLD) and chlorophyll concentration with pCO_2.

Statistical methods are also used to provide an empirical relationship between the partial pressure of CO_2 and bio-geographic factors measured by shipboard instruments.

2 Literature Review

2.1 Anthropogenic Activities and CO_2 or CO_2 and Society

Anthropogenic activities have been driving an increase in the concentration of CO_2 in the atmosphere since the Industrial Revolution in the 1850s (Keeling et al., 2011). Estimates put the pre-industrial atmospheric CO_2 concentration at 280 parts per million per volume (ppmv) which is 112 ppmv less than current day estimates of 392 ppmv (Keeling et al., 2011). In the last 420 000 years, this rate of increase of Atmospheric CO_2 concentration has not been as rapid as we now observe (Falkowski et al., 2000) since the advent of the Industrial Revolution. As CO_2 in the atmosphere acts as a greenhouse gas, an increase in atmospheric CO_2 is believed to lead to an increase in global temperature (Schimel et al., 2001).

Not all anthropogenic emissions of CO_2 remain in the atmosphere. A large percentage ($\pm$60%) is taken up by the other two reservoirs of the carbon cycle, the ocean and the terrestrial biosphere. According to Sabine et al. (2004), 48% of anthropogenic CO_2 that has been emitted has been taken up by the oceans. The increase of CO_2 in the surface oceans has resulted in a global pH decrease of 0.1 (Raven et al., 2005).

The role of the oceans in atmospheric CO_2 was not much considered until 1966 Revelle and Suess (1957) suggested that oceans were the long term regulator of atmospheric CO_2. This led to further studies into the role of the oceans in both the short term and the long term carbon cycle.

2.2 The Carbon Cycle

2.2.1 Selection

The global carbon cycle consists of different reservoirs that operate on different residence times or flux rates (Sigman and Boyle, 2000). As the smallest reservoir with the smallest residence time, the atmosphere is said to be a good indicator of the state of the global carbon cycle (Sarmiento and Gruber, 2006). Past atmospheric CO_2 concentrations have been reconstructed from air bubbles trapped in ice cores (Indermühle et al., 1999; Petit et al., 1999). Variations of CO_2 concentration from 180ppm to 280ppm were observed to occur in time with glacial-interglacial cycles with glacial maxima coinciding with CO_2 minima.

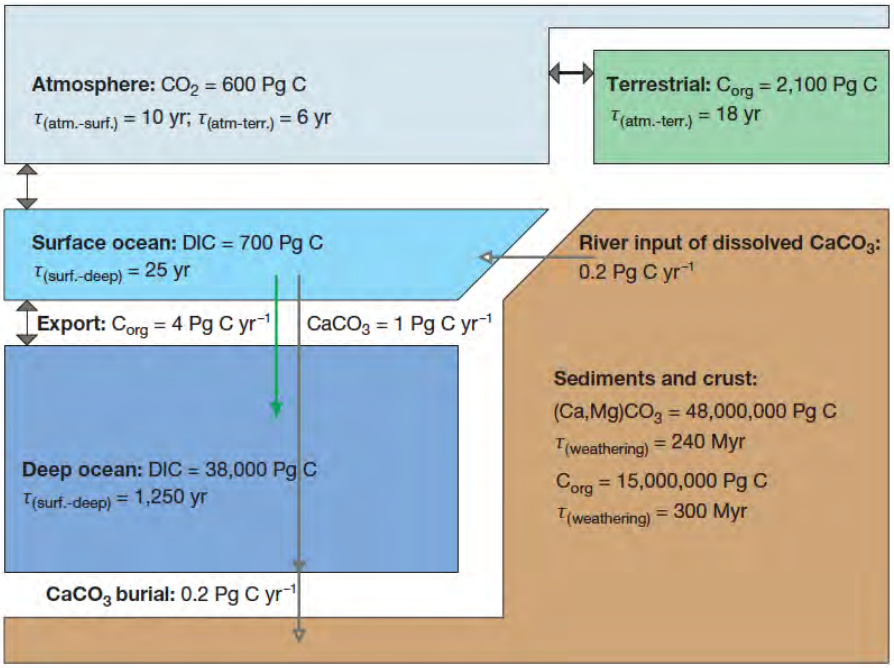

Figure 2.1: A schematic of the pre-industrial carbon cycle, including the size and residence time of each reservoir, from Sigman and Boyle (2000).

The terrestrial reservoir has a residence time of 6 to 8 years, whereas the surface ocean has a residence time in the order of tens of years. It is the deep ocean, with a residence time in the order of thousands of years, that controls the atmospheric CO_2 variations seen in the glacial-interglacial cycles (Sarmiento and Gruber, 2006).

Glacial-interglacial cycles have dominated global climate and are believed to have been a result of variations in Earth's orbit , known as Milankovitch cycles (Sigman and Boyle, 2000). The ice-core records show a close correlation between atmospheric CO_2 and temperature (Petit et al., 1999). According to Paillard and Parrenin (2004), Antarctic sea-ice extent could alter CO_2 fluxes enough to switch from one phase to another.

Deep water formation in high latitudes plays a role in transporting CO_2 from the atmosphere to the deep ocean. The deep oceans are 12% richer in inorganic carbon than the surface (Sarmiento and Bender, 1994). A mechanism transporting carbon across the concentration gradient is needed to sequester carbon in the deep ocean. Volk and Hoffert (1985) recognise three such carbon pumps: the solubility, the soft-tissue, and the carbonate pumps.

The geological reservoir has a residence time of over 500 000 years, and it is from this reservoir that anthropogenic emissions of CO_2 have primarily been sourced. As a result, the carbon cycle, once considered to have been in quasi-steady state, is now unbalanced as anthropogenic emissions of CO_2 begin to build up in the terrestrial and surface ocean reservoirs. Anthropogenic CO_2 emissions, through the carbon cycle, perturb the natural carbon fluxes at a rate greater than the ocean is able to store. The build up of carbon in reservoirs with time scales shorter than geological, such as the atmosphere and the surface ocean, has created an imbalance in the system.

2.3 Carbon Dioxide in the Oceans

The Keeling Curve (Keeling et al., 1995) (a graph that plots the ongoing change in atmospheric CO_2 concentration since 1958) shows that global atmospheric CO_2 is increasing, but this is more difficult to demonstrate in the oceans. In the 1990s, ship based pCO_2 measurements were dramatically increased through the efforts of the Joint Global Ocean Flux Study (JGOFS, 1990). The focus of the JGOFS was to quantify air-sea CO_2 exchange in the open oceans. Takahashi et al. (1997) published the first global estimate of air-sea CO_2 flux based on *in-situ* data. An ocean CO_2 uptake of 0.6 to 1.34 PgC.yr$^-$1 was estimated from 250 000 data points.

Takahashi et al. (2002), utilized 0.94 million surface water measurements of pCO_2. Takahashi et al. (2009) made use of about 3 million measurements to construct a climatological mean distribution for surface water pCO_2. Observations made in the equatorial Pacific during El Niño and observations in coastal waters were excluded. Using a mean rate of +1.5 μatm.y^{-1}, the data captured in different years was corrected by Takahashi et al. (2009) to a reference year, 2000. The increase of +1.5 μatm.y^{-1} is consistent with the decadal mean rate of increase of atmospheric pCO_2. Takahashi et al. (2009) estimated the annual mean net CO_2 uptake flux to be -1.6±0.9 PgC.yr^{-1} and the total ocean uptake flux, including anthropogenic CO_2, to be 2.0±1.0 PgC.yr^{-1}

Different methods have been used to constrain carbon dioxide (CO_2) fluxes. The Intergovernmental Panel on Climate Change (IPCC) (2001) uses atmospheric O_2/N_2 observations (Keeling and Shertz, 1992) to derive rates of anthropogenic CO_2 uptake. Inverse atmospheric models have been used to calculate land and ocean fluxes from atmospheric CO_2 data (Gurney

et al., 2002). Oceanic uptake of anthropogenic CO_2 can also be estimated from Chlorofluorocarbon (CFC) data sets of McNeil et al., 2003. Observations of reduced isotopic ratio (^{13}C) of Dissolved Inorganic Carbon (DIC) was used by Gruber and Keeling (2001). The Ocean Carbon Cycle Intercomparison Project (OCMIP-2) evaluated 19 ocean carbon cycle models with radiocarbon and CFC-11 data (Matsumoto, 2004).

Using a biogeochemical carbon cycle model (HAMOCC5) coupled online to a global ocean general circulation model (MPI-OM), Wetzel (2005) simulated and analysed the trends and variability in sea-air CO_2 flux from 1948 to 2003. The model study showed a global inter-annual variability of ± 0.50 PgC yr^{-1} dominated by ocean dynamics in the equatorial Pacific (Wetzel, 2005). Wetzel (2005) observed two patterns that emerge on a global, decadal scale. Firstly, the inter-annual variability of the equatorial Pacific decreases from ± 0.32 PgC yr^{-1} to ± 0.23 PgC yr^{-1}, and the mean outgassing of CO_2 goes from 0.70 PgC yr^{-1} to 0.58 PgC yr^{-1}. Secondly, the Southern Ocean CO_2 flux increases over the simulation period due to increasing wind velocities causing stronger upwelling and deeper mixed layers (Wetzel, 2005). Globally the estimated CO_2 flux from Wetzel (2005) was 1.49 PgC yr$^-$1 into the ocean for 1980-1989, and 1.74 PgC yr^{-1} into the ocean for 1990-1999.

The impact of mesoscale and submesoscale oceanic processes on sea surface pCO_2 is examined by Mahadevan (2004) in order to explain variability observed at length scales of order 10km. According to Mahadevan, 2004 , only small variations in surface pCO_2 are induced by submesoscale upwelling while larger variations are generated at larger scales.

A high-resolution ocean biogeochemical model, validated against shipboard and float data, was used by Resplandy et al. (2009) to estimate oceanic pCO2 and air-sea CO2 flux in the NE Atlantic. The study showed large submesoscale variability of the carbon system with gradients of 25 μatm over 20km, which is similar to the mean seasonal drawdown of pCO_2. A comparison of air-sea flux derived from their model, together with that of the floats and shipboard observations, gave a sampling error of 15 to 30%, and results from coarser model simulations showed that model resolution only accounts for 5% error. Thus, their findings indicated that the net air-sea CO_2 exchange is largely unaffected by the submesoscale variability of pCO_2.

2.3.1 Calculating Carbon Flux

The error in the annual mean CO_2 uptake flux from Takahashi et al. (2009) shows high error in the estimates. Much of the error is a result of how flux estimates are derived.

CO_2 flux (F) is calculated as:

$$F = k_w \cdot K_0 \cdot \Delta pCO_2 \tag{2.1}$$

where k_w is the gas transfer velocity, K_0 is the solubility parameter and Δ pCO_2 is $pCO_2^{sea} - pCO_2^{air}$, the difference between the partial pressure of CO_2 in the ocean and the atmosphere. The uncertainty in the gas transfer is estimated to be around 10-20% (Ho et al., 2011).

Several theoretical models have contributed to the understanding of emperically determined gas transfer velocity. The stagnant-film model (Liss and Merlivat, 1986; Broecker and Peng, 1974) assumes a thin layer (stagnant film) of water in which diffusion between the atmosphere and surface water occurs, but addresses mixing between the surface layer and the film. A dynamic interface is defined in the replacement film (Danckwerts, 1951) and the eddy impingement models (McCready and Hanratty, 1985) to address mixing between the layers. The models have the following concepts:

- The transfer velocity is dependent on the turbulence at the interface and the thickness of the film.
- Turbulence at the interface is affected by wind stress.

Nightingale (2009) refers to factors such as wave breaking, bubble formation, humidity and temperature gradients, and surfactants which may alter the gas exchange velocity.

The Schmidt number (Liss and Merlivat, 1986), Sc, is the thickness of the film and is determined by the ratio:

$$Sc = v/\epsilon \tag{2.2}$$

where v is the kinematic viscosity of water and ϵ is the molecular diffusivity of a particular gas ,therefore allowing the gas transfer velocity to be determined for different temperatures and gases.

2.4 Marine Carbonate Chemistry

CO_2 in the ocean is influenced by biology, temperature, and transport. These factors can determine whether or not an area of the ocean is a sink or a source. The reaction of CO_2 in seawater is described in the following equation:

$$\begin{aligned} CO_{2(aq)} + H_2O &\overset{K_0}{\longleftrightarrow} H_2CO_3^* \\ H_2CO_3^* &\overset{K_1}{\longleftrightarrow} H^- + HCO_3^- \\ HCO_3^- &\overset{K_2}{\longleftrightarrow} H^- + CO_3^{2-} \end{aligned} \tag{2.3}$$

Carbon in the ocean is found in three forms: dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), and particulate organic carbon (POC). Most of the CO_2 in the ocean is stored as bicarbonate (HCO_3^-).

2.5 The Southern Ocean

The Southern Ocean is a unique and important region that is strongly connected to global ocean circulation, and to climate change. It is believed to be a carbon sink. The Antarctic Circumpolar Current (ACC) connects the three major ocean basins: the Pacific, the Atlantic, and the Indian Oceans. It therefore acts as an interface for heat and fresh water transfer between the oceans. Regions of upwelling in the Southern Ocean allow for heat transfer from the deep ocean to the atmosphere. The Southern Ocean is also the only source of Antarctic Bottom Water.

Orsi et al. (1995) investigated the Antarctic Circumpolar Current and fronts in the Southern Ocean. The northenmost extent of the Sub-Antarctic waters is the Sub-Tropical Front (STF). The Sub-Antarctic Front is found further south, and the Polar Front even further south.

Thomalla et al. (2011) provide a regional characterization scheme for the Southern Ocean that improves understanding of regional differences in ecosystem sensitivity to climate forcing. The seasonal cycle is defined by four zonal regions: the Subtropical Convergence Zone (STCZ), the transition zone (TZ), the Antarctic circumpolar zone (ACZ) and the marginal ice zone (MIZ). These authors contend that the extent of inter-annual seasonal phase locking and the magnitude of integrated seasonal biomass led to additional classification of the four regions. The regions are described by high or low chlorophyll and high or low seasonality. Any combination of chlorophyll or seasonality is possible, and the combination of the two classifications

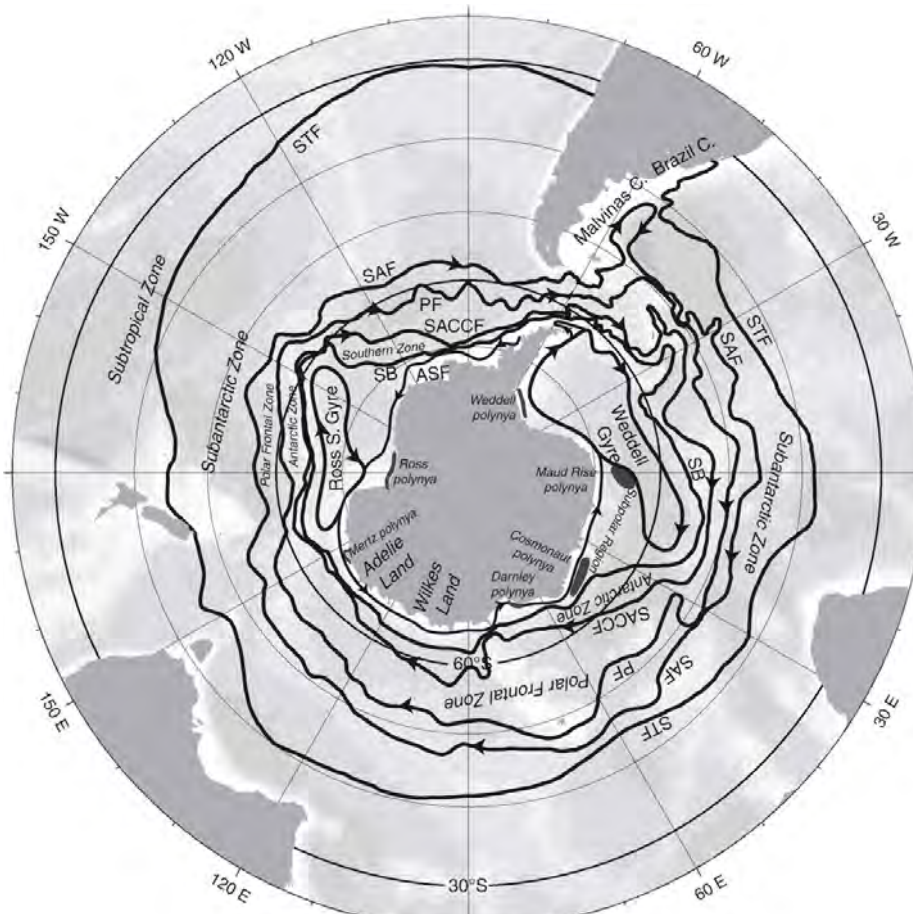

Figure 2.2: Map of the Southern Ocean including the important oceanic fronts and zones (Pickard and Emery, 1990).

can help with understanding the role of physical and biological drivers of production and CO_2 flux in the Southern Ocean.

2.6 Carbon in the Southern Ocean

According to Friedlingstein et al. (2003), model studies show a positive feedback between climate change and the carbon cycle and demonstrate that the Southern Ocean circulation, which controls geochemical uptake of CO_2, is poorly constrained in these models.

Rangama (2005) used shipboard and satellite measurements to assess small-scale variability of pCO_2 and estimate the net air-sea CO_2 flux in the Southern Ocean south of Tasmania and New Zealand. An oceanic sink of -0.08 GtC.yr^{-1} was estimated with an error of 0.03 GtC.yr^{-1}.

According to Le Quéré et al. (2007) climate change has caused a decrease in Southern Ocean CO_2 uptake. Matear and Lenton (2008) used a model study to investigate how changes in heat and freshwater fluxes as well as winds affect Southern Ocean CO_2 uptake. They show that while increase in heat and freshwater fluxes drive an increase in uptake, increase in wind stress drives an approximately equal and opposite response, resulting in a Southern Ocean response equal to that with no climate change. The positive feedback of the Southern Ocean carbon cycle on climate change will, according to Matear and Lenton (2008) become a negative feedback mechanism once the partial pressure of atmospheric CO_2 reaches that of Circumpolar Deep Water. Matear and Lenton (2008) also show that climate change could reduce the aragonite saturation state in the Southern Ocean.

The decadal variability in the fugacity of Carbon Dioxide (fCO_2) at the sea surface was analyzed by Metzl (2009) The area of focus was the south-western Indian Ocean and corresponding Antarctic sector. This study was based on seasonal cruises during the period 1991-2007. According to the shipboard observations the average annual rate of the atmospheric CO_2 increase was 1.72 ppm/yr, which is equivalent to the annual growth rate recorded at monitoring stations in the Southern Hemisphere. The oceanic fCO2 increased at a rate of 2.11$\pm$0.07 μatm.yr^{-}1 for the period 1991-2007. They found that the rate of increase is greater (0.4 μatm.yr^{-}1 faster) in the oceans than in the atmosphere. This has implications for the ocean carbon sink.

Studies have been done in specific areas of the Southern Ocean to de-

termine the sources and sinks of atmospheric carbon and how they change over time.

Bakker et al. (1997) looked at changes in carbon dioxide in surface waters during spring in the Southern Ocean. According to their study, in November 1992 the fCO_2 decreased in the Polar Frontal Zone (PFZ) from just below the atmospheric level to 50 μatm below. This was accompanied by a diatom bloom. Biological drawdown and formation of Antarctic Intermediate Water makes for combined biological and physical uptake of CO_2. According to Bakker et al (1997) the boundary between the ACC and the Weddel gyre was a significant CO_2 source as ice-cover disappeared.

Bakker et al. (2008) investigated the eastern Weddell gyre and showed how Circumpolar Deep Water (CDW) enters the Weddell gyre in the south-east and sits under the ice cover. When the sea-ice melts, the high CO_2 concentration water interacts with the atmosphere and outgassing occurs. According to Bakker et al (2008), the outgassing is counteracted by increased biological activity.

In austral autumn and early winter the southern ACC and the northern Weddel gyre were found to be supersaturated with CO_2 (Hoppema et al., 2000). The southern Weddel gyre was very undersaturated and they demonstrated that potential for autumn cooling could lead to undersaturation.. The northeastern Weddel Gyre is a CO_2 source (outgassing) due to upwelling of CO_2 rich deep water (Hoppema et al. 2000).

Metzl et al. (1999) studied the annual cycle of sub-Antarctic air-sea CO_2 flux and showed that the Sub-Antacric zone (SAZ) is a strong summer sink due to the formation of shallow seasonal mixed layer leading to biological drawdown. They found that near equilibrium values in summer account for a small CO_2 exchange and are a result of the formation of a deep winter mixed layer (Metzl et al. 1999).

Metzl et al. (2006) went on to investigate summer and winter air-sea CO_2 flux in the Seasonal Ice Zone (SIZ) and the permanent open ocean zone (POOZ). In the SIZ (south of 58°) during summer, the surface CO_2 concentrations are below equilibrium resulting in a summer sink. Metzl et al. (2006) show that in the POOZ (50-58°), surface ocean CO_2 concentrations increase from summer (sink) to winter (source). They used a model to reproduce the seasonal change in CO_2 flux and showed that summer primary production and winter vertical mixing was responsible for change.

Robertson and Watson (1995) show that observations along the ice edge in the Southern Ocean summer indicate that the region is a sink for atmospheric carbon dioxide. The strong CO_2 sink in the coastal Southern Ocean was investigated by Arrigo et al. (2008). They suggest that the estimates of large scale fluxes in the Southern Ocean do not resolve the fluxes observed in the coastal Southern Ocean. Their model study showed that Antarctic shelf waters are a very strong CO_2 sink due to high biological productivity, intense winds, efficient ventilation and extensive winter sea-ice coverage. Arrigo et al. (2008) use these results to show that highly productive waters around the antarctic continent need to be included in future budgets of anthropogenic CO_2.

Bakker et al (1997) state that a limited understanding of gas exchange mechanisms and the large spatial and temporal variability of sea-air flux hinder reliable estimates of Southern Ocean mean annual sea-air CO_2 flux.

2.7 Sampling Strategies for CO_2 flux

The uncertainty in the estimates of global and regional CO_2 uptake is a barrier to understanding the global carbon cycle and the global climate. As a result, the sampling frequencies required to accurately estimate air-sea CO_2 fluxes have been the subject of a number of studies.

Garçon et al. (1992) investigated the sampling frequency required to estimate air-sea CO_2 flux in the North Pacific. In order to reduce the uncertainty in estimating the air-sea CO_2, a high resolution time series of air-sea flux was sub-sampled to determine the optimal temporal sampling frequency. According to Garçon et al. (1992) the sampling requirements were 24 samples per year in the Pacific.

A study was done by Mémery et al. (2002) in the Mediterranean Sea in which time scales for estimating air-sea CO_2 exchange were investigated. This study also made use of a highly resolved observational time series and determined a sampling frequency of greater than 70 samples per year for the Mediterranean Sea.

In the Southern Ocean, Sweeney et al. (2002) used a summer and a winter north-south transect to determine spatial sampling required to constrain the mean ΔpCO_2 to ± 4.3 μatm. This ΔpCO_2 value can be calculated as an uptake of ± 0.1 PgC/yr using a mean Southern Ocean gas exchange coefficient. According to Sweeney at al. (2002), sampling every 5° in latitude

in summer and sampling every 10° in latitude in winter is sufficient to meet the requirements.

Takahashi and Sweeney (2002) determined the sampling required to estimate the net annual CO_2 flux by subsampling monthly flux maps of the global, gridded data set of Takahashi et al. (1997). According to Takahashi and Sweeney (2002) , sampling every 4 months, or 3 times a year is required.

Lenton et al. (2006) combined temporal and spatial sampling strategies to validate a sampling strategy that reduces Southern Ocean annual CO_2 uptake uncertainty to ± 0.1 PgC/yr. They used a time-evolving - prognostic - high-resolution - biogeochemical model to simulate fluxes. According to Lenton et al. (2006), an accurate representation of the Southern Ocean CO_2 uptake and its uncertainty provides: (1) the essential information to resolve the present mismatch between observation and model estimates of the uptake which will reduce the uncertainty in the global budget; (2) a reference against which future changes in variability can be assessed; and (3) an observational estimate to assess and validate numerical models.

In developing a sampling strategy that will resolve the temporal and spatial variability, they used Fourier Transforms with signal-to-noise ratios. This identified the dominant frequencies in time and space that control variability in the simulated air-sea CO_2 fluxes of the Southern Ocean.

Their study determined that sampling every 3 months, at 30° in longitude and 3° in latitude is sufficient to determine the net Southern Ocean CO_2 uptake. They applied their sampling strategy to the simulated air-sea fluxes in order to estimate the annual mean CO_2 uptake.

Sparse ocean sampling in the Southern Ocean especially means that the simulated data cannot be accurately assessed. Lenton et al. (2006) separated the simulated air-sea flux variability into seasonal and non-seasonal variability. They defined the seasonal variability as the daily (climatological) data and the non-seasonal variability is the variability not represented by the seasonal cycle.

Lenton et al. (2006) compared the simulated seasonal cycle to the coarse-resolution, monthly climatological maps of Takahashi et al (2002). The simulated seasonal cycle of Lenton et al. (2006) captured the variability evident in the observations in Takahashi et al (2002).

Lenton et al. (2006) calculated signal-to-noise ratios (SNR) in order to assess the air-sea flux variability. They defined the seasonal cycle to be the

signal and the nonseasonal variability to be the noise.

$$SNR = \frac{\sigma^2_{SIGNAL}}{\sigma^2_{NOISE}} = \frac{\sigma^2_{SIGNAL}}{\sigma^2_{ALLDATA} - \sigma^2_{SIGNAL}} \tag{2.4}$$

They calculated a SNR of less than 1 for most of the Southern Ocean when computing a SNR map for the daily simulated air-sea CO_2 fluxes. In those regions where the SNR is less than 1, the seasonal cycle cannot be distinguished from the non-seasonal variability. The variability of the seasonal cycle calculated in Lenton et al. (2006) was not homogenous in space; this situation is consistent with the Takahashi et al. (2002) climatology.

Lenton et al. (2006) attempted to answer the question, "[w]hat sampling was required to constrain the Southern Ocean air-sea CO_2 flux given the large variability observed in the model?". This was achieved by using two-dimensional Fourier Transforms (2D-FT) in space and time along constant sections of latitude and longitude. The plots that they produced showed that the Southern Ocean was a net uptake region; there was a more uniform spread of variability across a range of periods when the wavelength was longer (λ <5°); and the variance explained by short periods (T >20 days) and short wavelengths (λ >5°) declined rapidly.

They then derived a relationship for SNR in the frequency domain that is equal to SNR in the temporal domain.

$$SNR(f) = \frac{\sum_{n=1}^{f} H_{signal}(n)^2}{\sum_{n=1}^{f} H_{noise}(n)^2} \tag{2.5}$$

They applied this equation to the 2D-FTs that they had calculated in order to see how the SNR changed as time and space was sampled at higher temporal and spatial frequencies. Their study found that variance does not increase much once the seasonal cycle is resolved, and once the seasonal cycle is resolved the spatial sampling frequency needs to be increased in order to increase the SNR. This was consistent with the result of the SNR plots in the temporal domain being less than 1 in most regions, and confirmed that a high resolution in time or space still returns a low SNR when combined with a low resolution in space or time.

Lenton et al. (2006) showed that the statistical properties of the model and the proposed sampling strategy are consistent with the statistical properties of observations. They quantified the uncertainty of estimated mean

annual uptake by applying their sampling strategy to their simulated daily fluxes. They calculated a value of 0.6 ± 0.1 PgC/yr at the model resolution. This result was similar to Wetzel et al. (2005) which is expected as the same model was used. Applying their sampling strategy to the simulated fluxes returns a sampling uncertainty of ± 0.2 PgC/yr which translates to a sampling error of ± 0.07 PgC/yr for the 1990s. The inter-annual variability returned by their sampling strategy was ± 0.1 PgC/yr, which is the same as the value returned by sampling at the model resolution. The total uncertainty calculated (inter-annual variability and sampling error) by Lenton et al. (2006) was ± 0.1 PgC/yr giving an estimate of annual averaged uptake for the 1990s of 0.6 ± 0.1 PgC/yr. The results are summarised in Table 2.1.

Table 2.1: Comparison of the total Simulated Uptake With the Uptake from Proposed Sampling and the Sampling Error Introduced from Lenton et al. (2006).

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1990	0.67	0.63	0.21
1991	0.76	0.73	0.17
1992	0.99	0.97	0.21
1993	0.71	0.67	0.24
1994	0.56	0.52	0.18
1995	0.54	0.49	0.22
1996	0.53	0.49	0.20
1997	0.49	0.44	0.24
1998	0.38	0.32	0.26
1999	0.40	0.32	0.22
1990-1999	0.60 ± 0.12	0.56 ± 0.13	

Lenton et al. (2006) tested the present sampling strategy in order to compare the results to their proposed sampling strategy and came up with a total uncertainty value of twice that of their proposed sampling strategy. The results are summarised in Table 2.2.

Table 2.2: Comparison of the Total Simulated Uptake With the Uptake from the Current Sampling and the Sampling Error Introduced from Lenton et al. (2006).

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1990	0.67	0.67	0.37
1991	0.76	0.77	0.33
1992	0.99	1.00	0.42
1993	0.71	0.72	0.42
1994	0.56	0.55	0.38
1995	0.54	0.54	0.36
1996	0.53	0.53	0.36
1997	0.49	0.49	0.42
1998	0.38	0.38	0.37
1999	0.40	0.40	0.38
1990-1999	0.60 $\pm$ 0.12	0.61 $\pm$ 0.12	

The uncertainty calculated by subsampling the simulated data in their study was the same as that calculated by sampling at the model resolution. According to Lenton et al. (2006) this suggests: (1) that the uncertainty due to sampling frequency is small compared to the uncertainty due to inter-annual variability, and (2) sampling at higher resolution would not improve large-scale regional flux estimates,

Friedrich and Oschlies (2009) present a method for mapping surface pCO_2 on a basin scale using ARGO floats. The method was tested using an eddy-resolving biogeochemical model of the North Atlantic. They use voluntary observed ship (VOS) and ARGO float coverage to generate synthetic observations of the model, and use the observations to form a training data set for self organizing neural networks. The results, once the trained neural network is applied to the observations of SST and SSS and compared to the simulated data at model resolution, show an improvement of remote sensing based estimates with the resulting monthly mean pCO_2 maps covering 70% of the area with a RMS error of 15.9 μatm (Friedrich and Oschlies, 2009).

A high-latitude North Pacific and North Atlantic annual mean air-sea

CO_2 flux sampling strategy was developed by Lenton et al. (2009) following the same methods used by Lenton et al. (2006). The study showed that a regular sampling strategy of every 6° in latitude and every 10° in longitude every three months returned the annual mean air-sea CO_2 flux to within 15% of the simulated annual mean. Key features that are highlighted in this study are that i) optimising the combination of spatial and temporal sampling results in fewer measurements than if the spatial and temporal signals are optimized independently, and ii) inter-annual variability contributes more to the uncertainty in decadal annual mean uptake than sampling error and unresolved mesoscale variability.

Sampling strategies for the ARGO array in the Indian Ocean were investigated by Schiller et al. (2004) using simulation data from an OGCM. The results show that spatial sampling of 500 km zonally and 100 km meridionally is critical for resolving intraseasonal oscillations.

In situ ocean data and remotely sensed data were used in conjunction with observations of surface-water fugacity of carbon dioxide by Chierici et al. (2009) to estimate fCO_2^{SW}. Sea-surface temperature (SST), mixed-layer depth (MLD) and chlorophyll *a* (chl *a*) contributed significantly to the fit, and two algorithms were developed depending on the presence of chl *a* data. The estimated annual CO_2 uptake was 0.0058 Gt C yr^{-1} or 0.6 mol C m^{-2} yr^{-1}.

McNeil et al. (2007) used an independent method to estimate the Southern Ocean air-sea flux of CO_2. It exploited all available surface ocean measurements for DIC and Alk beyond 1986. They estimated a Southern Ocean (<50°) CO_2 sink of 0.4±0.25 PgC.yr^{-1} and a CO_2 sink of 1.1±0.6 PgC.yr^{-1} in the sub-Antarctic zone (40° to 50° S). Standard hydrographic properties are used by to predict surface-normalised DIC and Alk, and their empirical relation is applied to World Atlas (2011) climatologies to estimate Southern Ocean air-sea CO_2 flux.

2.8 Genetic Algorithms

Genetic Algorithms (GA) draw parallels from evolutionary biology in that they use simulated evolution to search for solutions to complex problems. Genetic Algorithms were developed by John Holland and have been used in fields such as search, optimization, design and machine learning (Whitley, 2001).

The state of a GA is given by a population, and each member of the population is a complete set of parameters for the function being searched (Gershenfeld, 1999). The whole population is updated in generations with the initial generation containing a genetic pool of possible solutions. Improved solutions are obtained by using recombination by crossover as well as random alteration by mutation (Whitley, 2001). Genetic Algorithms have the following elements (Gershenfeld, 1999; Whitley, 2001):

Representation of solution: Possible solutions to a problem are represented as a binary string of a fixed length. The solution can also be represented by integers or another set ,but binary is the most common or traditional method. The solutions contain genes or characteristics which are also of fixed length.

Construction of a gene-map: A gene-map is constructed such that the gene-map associates each gene with a value. In classical genetic algorithms, each variable is represented using m bits. Therefore, a function with n variables will have a chromosome of length $m.n$ (Budin et al., 2010).

Initialisation of the population: A number of binary strings representing individual solutions are randomly created to make up the initial gene pool or population.

Once a population has been initialised a number of steps are repeated until some criteria is met which determines termination. The termination criteria should be defined in such a way to best suit the needs for the given problem.

Evaluation: Each individual solution in the population is evaluated by a fitness function to determine a fitness value.

Selection of parents: Parents are chosen from the population or solution space.

Reproduction: New offspring are created from the parents and become the new population. Operators such as cross-over, mutation and elitism are used to ensure that the offspring are different from their parents yet have the same information.

Termination criteria: A termination criterion, already declared, will break the algorithm out of the loop. This can occur, for example, when the output from the fitness function remains the same for a certain number of generations.

The diagram in Figure 2.1 shows the flow of a genetic algorithm.

2.8.1 Selection

A genetic algorithm needs to pass on information from one generation to another. Selection is the name given to the step which selects parents from the current generation whose information will be transferred into the next generation.

There are a number of selection methods that can be used in a genetic algorithm. Tournament selection and Roulette wheel selection are the most commonly used.

In Roulette wheel selection, the individuals of a generation are ranked according to their fitness. Individuals with a greater fitness have a greater probability of being chosen as parents.

Tournament selection is accomplished through randomly selecting an individual to compete against another randomly selected individual. The fittest individual is the winner and is then used as a parent for the next generation.

2.8.2 Crossover

Crossover is the process whereby the parents selected in the selection process pass on their information to children. Crossover can be done at one or more points. A random number is selected and at that point in the individual's genetic code, the string is split and the two parents exchange that part of their code. Because stronger individuals are selected in the selection stage, crossover allows for 'better' information to be passed from one generation to the next. Single point crossover is the most common type of crossover used although it is possible to use two point crossover. The probability that crossover occurs (a value that is assigned in the algorithm implementation) depends on the problem but the probability is usually quite high.

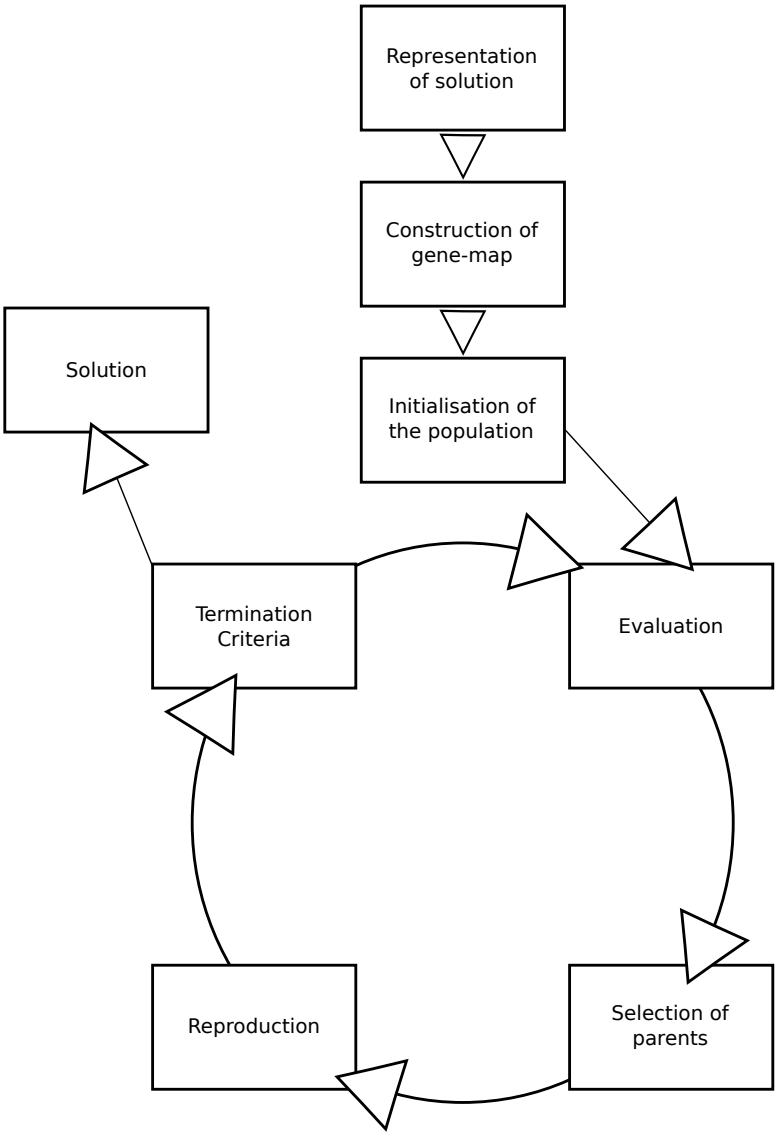

Figure 2.3: Schematic of a typical Genetic Algorithm implementation

2.8.3 Mutation

While Crossover can add genetic diversity to the population, another method of adding random diversity is often used. This method is called Mutation and involves the chance of randomly selecting a number that falls within the mutation rate which results in the possibility of the flipping of a random bit in the genetic code of the children. The chance of mutation needs to be set so as to not add too much diversity or too little. The inclusion of mutation is necessary to ensure that solutions do not simply converge to local optimum values.

2.8.4 Elitism

Elitism ensures that the information belonging to the most fit individual is passed onto the next generation. This is done by replacing a random solution from the new generation by the fittest solution from the previous solution.

2.8.5 Replacing the old generation with the new

There are many ways that one could decide to replace the old generation with the new. Most often this is done by having the new chromosomes replace the parent chromosomes.

2.8.6 Limitations of Genetic Algorithms

Genetic algorithms are not considered the first option in solving optimisation problems for a number of reasons (Whitley 2001; Goldberg 1989). These include:

- Genetic Algorithms do not perform well with expensive fitness functions. It is therefore better to use simplified fitness functions that approximate more complex fitness functions.
- Because genetic algorithms only find better solutions, it can be difficult to assign an appropriate termination criterion.
- Genetic algorithms tend to converge to local optima and rely on random mutations to find global optima.

2.8.7 Use of Genetic Algorithms in Oceanography

The problem of how to optimally deploy a suite of sensors to estimate the oceanographic environment is addressed by Heaney et al (2007). An optimal way to estimate and predict the ocean environment is to assimilate measurements from dynamic and uncertain regions into a dynamic ocean model. In order to determine the sensor deployment strategy that optimally samples the region's uncertainty, Heaney et al. (2007) present a Genetic Algorithm approach. The scalar cost function used by Heaney et al. (2007) is defined as a weighted combination of a sensor suite's sampling of the ocean variability, ocean dynamics, transmission loss sensitivity and modelled temperature uncertainty. The advantage of the GA approach is that the user can determine "optimal" *via* a weighting of constituent cost functions, which include ocean dynamics, acoustics, cost, and time. A numerical example with three gliders, two powered AUVs and three moorings is presented by in their study to illustrate the optimization approach in the complex shelfbreak region south of New England.

Models can better predict the ocean state with the inclusion of field data, and the use of models can improve mapped fields generated from measured data. The challenge is limited resources. According to Heaney et al. (2007), an integrated optimization-assimilation-modeling system is being developed to perform optimal ocean sampling and ocean prediction. The system involves the combination of an ocean model with data assimilation capabilities, an ensemble of measurement platforms, and a non-linear constrained global optimization subsystem.

2.9 Questions and Aims

- Is there a temporal and spatial sampling frequency that can achieve a significantly lower sampling error without a significant expense in sampling?
- How can we decide what constitutes an efficient trade-off between sampling error and sampling effort?
- If we attempt to abandon the use of a regular grid to sample, can the sampling effort be reduced?
- Does the optimal sampling strategy of the decadal mean sample the

simulated annual means with sufficient accuracy?

- Does the the optimal sampling strategy of the mean seasonal or annual cycle sample each simulated year with sufficient accuracy?
- Which regions in the Southern Ocean need to be sampled at the highest frequency to reduce uncertainty in the annual mean CO_2 uptake.
- Can a new sampling strategy be proposed that optimizes for a multi-platform approach (i.e. Ships. gliders, and moorings?

3 Data and Methods

3.1 Observational data

Observational data from Takahashi et al. (2009) was used to assess and validate the simulated model data. The data is based on 3 million measurements of surface water pCO_2 obtained between 1970 and 2007 and produces a climatological mean distribution of CO_2 flux with a resolution of 4° latitude by 5° longitude. The measurements represent open ocean observations and measurements made during El Niño are excluded.

3.2 Model Data

The simulated model data was obtained from ORCA2-LIM, an ocean model that is based on the ORCA2 global configuration of OPA version 8.2 (Aumont and Bopp, 2006 from Madec et al. 1998) coupled with the dynamic-thermodynamic ice model developed at Louvain-La Neuve (Timmerman et al. 2003 in Aumont and Bopp, 2006). The mean horizontal resolution of the ocean model is 2° by 2° $\cos(\Phi)$ (where Φ is latitude). There are 30 vertical levels; twelve of which are located in the top 125 m. The parameterization of Gent and McWilliams (1990) is used for the effects of unresolved mesoscale eddies poleward of 10° latitude. Lateral mixing, both on tracers and momentum, is performed along isopycnal surfaces *sensu* Lengaigne et al. (2003). The flow of deep water over bathymetry is represented using the bottom boundary layer (BBL) proposed by Beckmann and Doscher (1997). The Blanke and Delecluse (1993) prognostic model of turbulent kinetic energy (TKE) is used to compute vertical eddy and viscosity coefficients. Climatological atmospheric forcing is constructed from various data sets consisting of daily NCEP/NCAR 2m atmospheric temperature averaged over 1948-2003 (Kalnay et al., 1996), monthly relative humidity (Trenberth et al., 1989), monthly ISCCP total cloudiness averaged over 1983-2001 (Rossow and Schiffer, 1999), monthly precipitation averaged over 1979-2001 (Xin and Arkin, 1997), and weekly wind stress based on ERS satellite product and TAO observations (Menkes et al., 1998). Surface heat fluxes and evaporation are computed using empirical bulk formulas described by Goose (1997). Modelled sea surface salinity is restored to the monthly WOA01 data set (Conkright et al., 2002) with a timescale of 40 days.

The ocean biogeochemical model used is the Pelagic Interaction Scheme

for Carbon and Ecosystem Studies (PISCES) derived from the Hamburg Model of Carbon Cycle version 5 (HAMOCC5) (Aumont and Bopp, 2006; Aumont et al., 2003). The model has 24 compartments. Phytoplankton growth can be limited by nitrate, phosphate, ammonium, silicate and iron (Aumont and Bopp, 2006). There are two phytoplankton size classes and two zooplankton size classes (Aumont and Bopp, 2006).

Simulated model data from an ocean biogeochemical model are used as an idealised world where the results from different sampling strategies applied to this idealised world can be compared, In this way, the simulated model data provides a theoretical "perfect" knowledge.

The data used in this study comes from two different model runs using the ORCA2 ocean model coupled with PISCES, a 5-day averaged dataset from 2000-2009, and a daily resolution dataset from 1998-2007. The daily resolution model output is used to explore the optimisation of sampling strategies that use a regular grid (simulating ship-based sampling). The methods explored in Lenton et al. (2006) are repeated on this daily resolution model output from the ORCA2/PISCES model. In order to replicate Lenton et al. (2006), daily model output is required, and the Southern Ocean includes all locations south of 40°S. The 5-day mean model output is used to explore the more computationally expensive methods of optimisation sampling strategies that use Genetic Algorithms (GA) and Radial Basis Functions (RBF). There is an error in the 5-day model simulation that does not affect its use in this study. This error is due to an error in the parameterisation of mixed layer depth dynamics. As the exploration of optimized sampling strategies treats the model output as theoretical perfect knowledge, errors in the model output do not affect the results of the optimisation methods.

3.3 Flux calculations

The model data set used in this study has two variables: Dissolved Inorganic Carbon (DIC) flux and ΔpCO_2. The model makes use of the equation derived by Wanninkhof (1992) to calculate sea-air carbon flux. In this study, negative sign in flux represents flux into the ocean. To calculate pCO_2 from ΔpCO_2, the annual average of atmospheric CO_2 is removed from the ΔpCO_2 for each year in the dataset.

3.4 Data Analysis

Data analysis was done using Python, and all scripts for the data analysis are available in the appendix. Python packages, Numpy, Scipy and Matplotlib are the most commonly used scientific packages.

3.5 Position of Ocean Fronts

Ocean fronts in the Southern Ocean were calculated using the Mean Dynamic Topography (MDT) product from (Rio et al., 2011).

The Subtropical front was plotted using the 0.35 contour. The middle of the Subantarctic front was plotted using the 0.030 contour line. The middle of the Polar front was plotted using the -0.48 contour. The north of the Southern ACC front was plotted using the -0.943 contour. The Southern Boundary of the ACC was plotted using the -1.244 contour line. The position of these fronts are summarised in Table 3.1.

Table 3.1: Position of ocean fronts in the Southern Ocean.

Southern Ocean Front	Contour from Rio et al., 2009
Sub Tropical Front	0.35
Subantarctic front	0.030
Polar front	-0.48
Southern ACC front	-0.943
Southern Boundary	-1.244

3.6 Signal to Noise Ratios

Signal-to-noise ratio was calculated according to Lenton et al. (2006) in order to assess the importance of the climatological seasonal cycle and the non-seasonal variability in explaining the overall variance. The standard deviation of all the data was calculated as the standard deviation of the time series at each location. The inter-annual standard deviation was calculated as the standard deviation over the annual means for each year in the data set. The seasonal cycle (the mean annual cycle) was calculated as the mean climatological value for each day in the year. The standard deviation of the

seasonal cycle is the standard deviation in the seasonal cycle or mean annual cycle for each location. The non-seasonal data was defined as the difference between the time series for all the data and the seasonal cycle (repeated over 10 years) at each location. The standard deviation was calculated for the non-seasonal data.

Lenton et al. (2006) defined the seasonal cycle to be the signal and the nonseasonal variability to be the noise and their definition is used in this study. The signal-to-noise ratio was calculated by dividing the standard deviation of the mean annual cycle with the standard deviation of the non-seasonal data.

$$SNR = \frac{\sigma^2_{SIGNAL}}{\sigma^2_{NOISE}} = \frac{\sigma^2_{SIGNAL}}{\sigma^2_{ALLDATA} - \sigma^2_{SIGNAL}} \tag{3.1}$$

This method was used to define SNR for each location in the Southern Ocean.

3.7 Seasonality of CO_2 flux in the Southern Ocean

Seasonality, or the part of the overall variance that is explained by the seasonal cycle, was calculated in the same manner as Thomalla et al. (2011). Variability in the seasonal cycle was computed as the variance explained by the regression of CO_2 flux onto the climatological mean seasonal cycle. Values closer to 0 show less reproducibility in the mean seasonal cycle and values closer to 1 indicate that the time series reproduces the climatological mean seasonal cycle.

The seasonal cycle was defined as the mean value for each time step in the year over the 10 year model period. There is no phase or amplitude modulation.

3.8 Fourier Transforms

For a function $f(x)$ the Fourier transform (Bracewell and Bracewell, 1986) is:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\ e^{-2\pi i x \xi}\, dx, \tag{3.2}$$

For a function $f(x, y)$ the two-dimensional Fourier transform is:

$$\hat{f}(\xi_x, \xi_y) = \iint f(x, y) e^{-2\pi i(\xi_x x + \xi_y y)} \, dxdy \tag{3.3}$$

The basis functions are a product of generalised sinusoids with a frequency u in the x direction and generalised sinusoids with a frequency v in the y direction:

$$b(u, v) = e^{-2\pi i(\xi_x x + \xi_y y)} \tag{3.4}$$

So, the point (u, v) in the frequency domain corresponds to the basis function with frequency u in x and frequency v in y.

The Discrete Fourier Transform (DFT) is a unitary transformation in that it preserves the norm of a vector with the Fast Fourier Transform (FFT) being an important version of the DFT that can be solved much faster than the DFT (Gershenfeld, 1999). The Fourier transform is based on the Fourier series in which periodic time series data can be represented by a sum of sinusoidal components. The Fourier transform represents a timeseries in a frequency domain.

3.9 Testing all the possible regular grid strategies

Testing each possible sampling strategy that makes use of a regular grid can be done with time, depth, latitude and longitude dimensions. For the purpose of this investigation, the depth dimension was not used as the study involves surface measurements. The latitude data was assumed to be collected at the same resolution as the model grid. In this case, the model grid used had a N-S resolution of 2 $^\circ$ cos(Φ).

The method that was used is an exhaustive search method that aimed to test all regular grid sampling possibilities. It followed the following steps:

1. A sampling strategy is selected (spatio-temporal frequency).
2. The mean is calculated for each possible permutation of a sampling strategy.
3. The standard deviation of these means is calculated as the sampling error.
4. A new sampling strategy is selected.
5. This is repeated until all sampling strategies are selected.

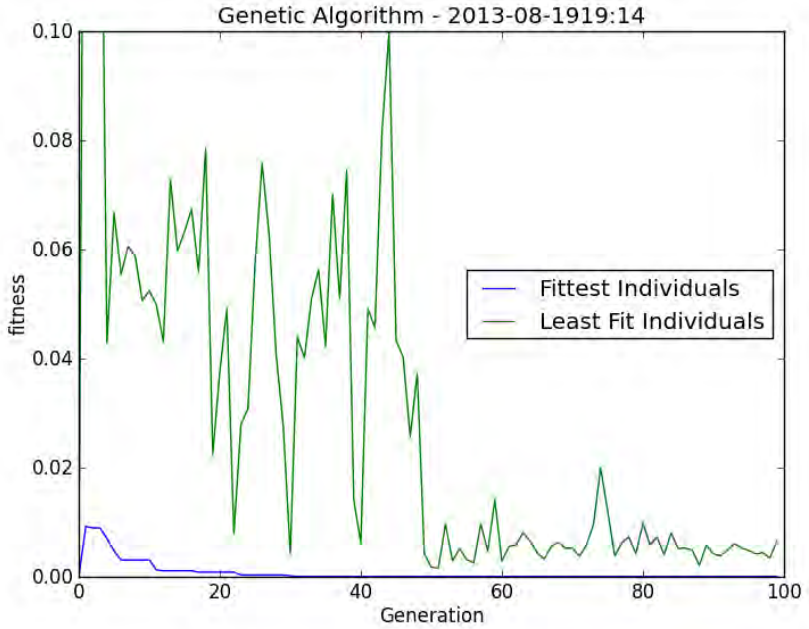

Figure 3.1: An example of the history of a genetic algorithm run showing the fitness value for the fittest and least fit individual in each successive generation

3.10 Genetic Algorithm

The genetic algorithm can be used to find a subset of points from the data set that has a minimal error in predicting the pCO_2 for the entire data set. There are two problem specific aspects of the GA, the encoding of the information and the fitness function which allows evaluation of the possible solutions. The outputs (consider optimized sampling strategies) from the genetic algorithm are tested in a similar manner to the regular grid strategies. Sampling with a certain number of optimized locations is considered a single sampling strategy. For example; the genetic algorithm can be used to find a set of 1000 locations. The genetic algorithm can be repeated to obtain a number of solutions for sampling with 1000 locations. This is done for 2000, 5000, and 7000 locations.

An example showing how the fitness is minimised, and how the least fittest individual converges is shown in Figure 3.1.

3.10.1 Implementation of Genetic Algorithm

Representation of solution and construction of gene-map: The solutions or chromosomes are represented as binary strings. The length of this string is dependent on the number of locations available to sample. Each location in the search space is represented as a unique binary string. The binary strings are stored as keys in a dictionary or associative array. The values associated with these keys are also dictionaries. These dictionaries contain the key's 'location' and 'value' and their associated location stored as a tuple (immutable list) and value stored as a float. Initially the GA implementation included all the locations in the climatological seasonal cycle of the decadal dataset. The gene-map not only contains valid sampling locations but it must also hold genes that do not necessarily represent valid sampling locations. These currently include the genes that are necessary for the 'padding' of the binary strings. In order to enable the genetic algorithm to search through solutions as fast as possible, changes were made to the data set. The 73 day by 40° by 180° data is regridded onto a 64 day by 64° by 128° grid and the last data point is removed in order for there to be 2^n - 1 sampling locations. This meant that only valid sampling locations would be represented in the binary hash table. The locations that are non-oceanic remain unmasked and have a value of zero which means they are considered valid sampling locations and can be included in a solution. The same sampling locations can be selected more than once on one solution as no additional cost is assigned to this.

Initialisation of the population: The population was initialised with 20 chromosomes. This meant that the initial population contains 20 random solutions. Each chromosome has a length of n times the length of the binary string representing the gene. n is set at different values according to how many locations are allowed in the sampling strategy solution. This effectively means that each solution chromosome contains the location of 1000, 2000, 5000, and 7000 possible sampling locations. It is important to note that the solution may not contain n unique solutions. It is also important to note that the genes that are used to 'pad' the tail of the gene-map can be selected. The length of

the binary string representing the gene is determined by the number of available locations that can be sampled. The GA implementation using the regridded data removes the possibility of invalid locations being selected.

Evaluation: There are two fitness functions available in the algorithm that can determine the fitness value for a chromosome solution:

- The first is a combination of the difference between the population mean and sample mean, and the difference between the population standard deviation and the sample standard deviation with an added function that multiplies the count of non-unique sampling locations used as well as assigning non-valid locations a value of $1e20$ which would increase the cost. The regridded data GA implementation did away with the padded and non-valid locations and did not penalise for non-unique locations being chosen. The inclusion of the standard deviation of the data was considered sufficient to ensure a good distribution of sampling locations. The fitness function is defined as:

$$f = (\bar{x}_{population} - \bar{x}_{sample}) + (\sigma^2_{population} - \sigma^2_{sample}) \tag{3.5}$$

- The second makes use of a Radial Basis Function which interpolates the sampled data points. The difference between the actual data and the interpolated data is squared at each node/pixel/point and the sum of these values is the fitness value. The RBF interpolation could be applied to the annual mean (a 45x180 grid) but not on the 73x45x180 grid and was therefore discarded as a sensible fitness function.

Selection of parents: Both tournament selection and roulette wheel selection can be used to select parents for the following generations.

Reproduction: Single-point cross-over, mutation and elitism are all used in this stage of the algorithm. The single-point cross-over can occur at any bit in the binary string. The mutation rate is set to a very low number so that only one in every 40 solutions experiences mutation.

Termination criteria: The algorithm terminates after a defined number of iterations. 10, 100, 1000, 10000 iterations have been tried. In

order to find the optimized sampling strategies using the regridded GA implementation, it was decided that 100 iterations would be sufficient.

This implementation of a genetic algorithm addressed the limitations of genetic algorithms in that a simplified fitness function is used. The computation time was therefore not significant. The tendency for genetic algorithms to converge to local optima was not considered a limitation in this implementation as this problem requires that local optima are found and not global optima. Furthermore, the use of the genetic algorithm to find local optima makes the the choice of appropriate termination criteria simpler.

3.11 Radial Basis Function interpolation

The genetic algorithm method was computationally too expensive to use in conjunction with multiple radial basis function calls per generation. For this reason, a new method was proposed to return a sampling solution.

The 5-daily averaged CO_2 data was used. The algorithm used a Wendland C2 compact support Radial Basis Function (Floater and Iske, 1996). The Cost function used to evaluate the interpolated sample is:

$$1/m \sum_{i=0}^{m} (sample(x^i, y^i) - population(x^i, y^i))^2 \tag{3.6}$$

The goal was to minimize this cost function and therefore return an interpolated sample with the least difference from the population data. This method finds a sample that, when interpolated, has the least error between sample and population. Since it is computationally too demanding to solve this problem using a formal optimization technique, it was assumed that successively selecting sampling locations at the point of greatest error will produce a reasonable solution. The location with the largest error was chosen and placed in the sample. The steps required to find the sampling locations are:

1. Randomly select 4 sampling locations to use in the sample data.
2. Use a radial basis function to interpolate the sample data.
3. Compare the interpolated sample data to the population data.

4. Select the location where the error is greatest and add this location to the list of sampling locations.

5. Repeat until N locations have been found

The locations were saved at intervals of fifty in order for the comparisons to be made between different sampling frequencies.

The algorithm used the annual cycle as the population data set from which the sampling locations are decided. Once the sampling locations are chosen, the sampling strategy can be tested on each year within the decadal data set. The expectation is that with a sufficiently large sampling size, the mean and the standard deviation of the sampled data will closely resemble that of the population data.

The implementation is summarised in the schematic in Figure 3.2.

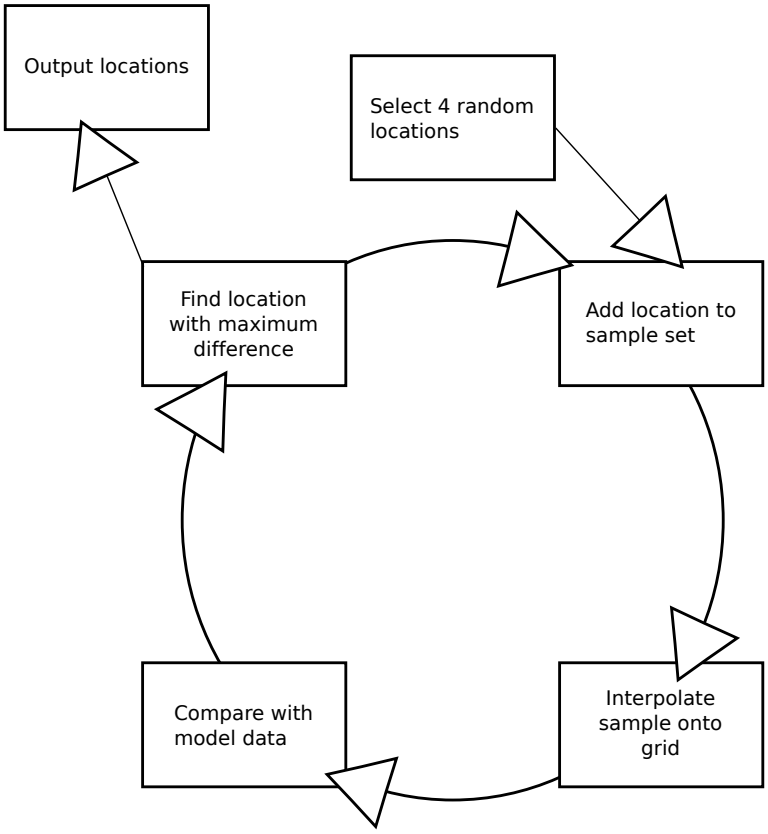

Figure 3.2: Schematic of Radial Basis Function implementation used in this study to identify optimal sampling locations

4 Results

4.1 Introduction

Two approaches were used to explore sampling strategy optimisation. The first, relied on regular grid sampling strategies and made use of daily resolution model output. This was used as a basis to explore further optimisation methods. The second approach used Genetic Algorithms (Section 3.10) to identify a group of optimized sampling strategies which can be compared in a similar way to the regular gridded approach. A radial basis function optimisation method was used to identify a set of possible locations by successively minimising the root mean squared error (Section 3.11).

4.2 Model Data Analysis - characteristics of variability

In this section, the simulated model data from the ORCA2/PISCES model was analysed. The data from these model runs were used in subsequent sampling strategy experiments. The daily resolution model data output was used to investigate sampling strategies that relied on a regular grid, replicating Lenton et al. (2006). The 5-day mean resolution model data output was used in the investigation of sampling strategies that did not rely on regular grids (using Genetic Algorithms and Radial Basis Functions). The simulated model data was compared to Takahashi et al. (2009) and the characteristics of variability of the data was investigated. A summary of the comparison is presented in Table 4.1.

In order to validate the simulated model data, the data was compared to Takahashi et al. (2009) observation data. The global mean sea-air CO_2 flux in mol C per m^2 per month is -0.038 for the observations (Takahashi et al. 2009). The 5 day averaged 2000 to 2009 simulated data global mean CO_2 flux was -0.048 mol C per m^2 per month and the 1998 to 2007 daily resolution model data global mean was -0.028 mol C per m^2 per month ($molC.m^{-2}.month^{-1}$). Negative sea-air fluxes represent uptake of CO_2 by the ocean. The Southern Ocean (south of 40°) mean CO_2 flux from observations was -0.046 mol C per m^2 per month (Takahashi et al. 2009). For the 5 day and daily resolution data the mean values were -0.09 and -0.029 mol C per m^2 per month respectively.

Table 4.1: Comparison of the simulated annual mean CO_2 fluxes with the annual mean from observations for the Southern Ocean South of 40° in molC.m^{-2}.month^{-1}. The data is also separated into the Atlantic (70°W - 20°E), Indian (20°E - 140°E), and Pacific (140°E - 70°W) sections of the Southern Oceans.

Region	Takahashi et al. (2009)	5-Day Mean Resolution	Daily Mean Resolution
Global	-0.038	-0.048	-0.028
Southern Ocean	-0.046	-0.09	-0.029
Atlantic	-0.055	-0.092	-0.011
Indian	-0.060	-0.92	-0.036
Pacific	-0.032	-0.068	-0.028

The CO_2 flux from Takahashi et al. (2009) ranged from values of ±-6.0 teragrams per year to ±2.0 teragrams per year in the Southern Ocean. Positive fluxes (outgassing) are evident in regions south of 50° and negative fluxes to the north. Figure 4.1 shows the mean CO_2 flux from Takahashi et al. (2009) for the Southern Ocean. The Air-Sea CO_2 flux ranged from 0.16 to -0.26 Mole Carbon per m^2 per month for the mean data. The values ranged from 0.50 to -0.81 Mole Carbon per m^2 per month for the overall data.

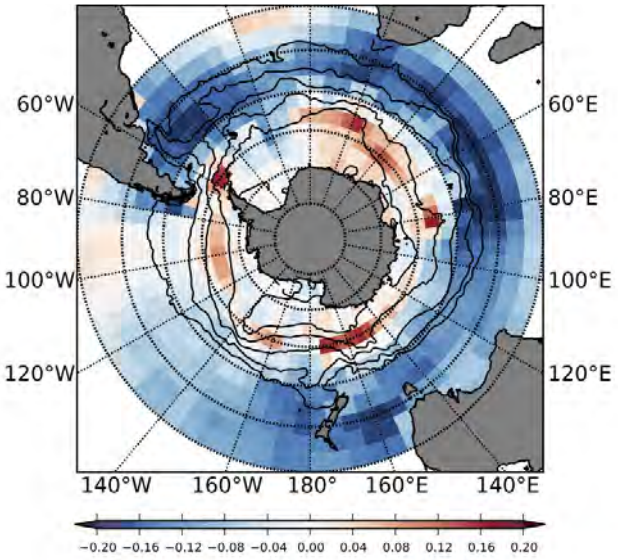

Figure 4.1: Mean CO_2 flux for a reference year 2000 from Takahashi et al. (2009) for the Southern Ocean South of 30° in molC.m^{-2}.month^{-1} with the positions of the Southern Ocean fronts overlaid.

Figure 4.2 shows the distribution of model derived CO_2 flux in the Southern Ocean over the period 2000 to 2009 taken from the 5 day mean resolution model data. The data shows that the decadal mean of the 5 day resolution model data ranged between 0.06 mol C per m^2 per month and -0.44 mol C per m^2 per month. This range of 0.50 mol C per m^2 per month is very similar to the observational data from Takahashi et al. (2009) which has a range of 0.42 mol C per m^2 per month. However, the 5-day resolution model data has a very low maximum outgassing and a larger maximum uptake. The 5 day resolution model data overestimated uptake showing almost no outgassing (Figure 4.2). The range over the entire (global) 5 day mean resolution model data set was 2.99 mol C per m^2 per month. Values greater than the observations are mostly due to the finer spatial and temporal resolution of the model. The maximum outagassing value for air-sea carbon dioxide flux was 0.93 mol C per m^2 per month and the maximum uptake was -2.06 mol C per m^2 per month.

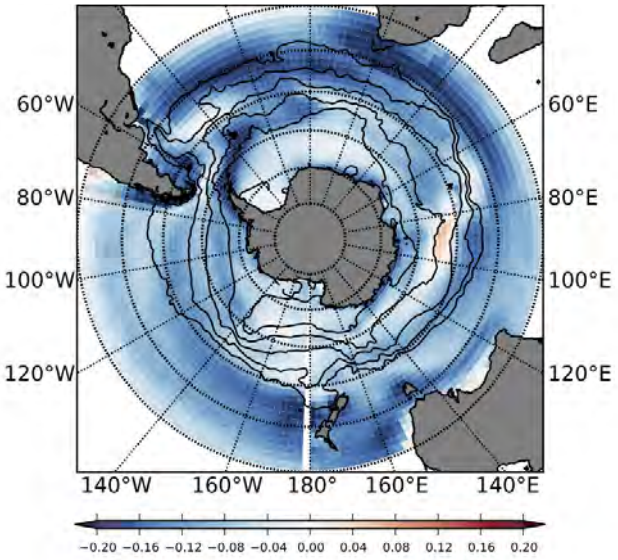

Figure 4.2: Decadal mean CO_2 flux for 2000-2009 (5-day mean model output) for the Southern Ocean South of 30° in molC.m^{-2}.month^{-1} with the positions of the Southern Ocean fronts overlaid.

The only outgassing depicted in FIgure 4.2 is between 80°E and 100°E, just south of the Polar Front. Strong uptake is depicted north of the subtropical front.

Figure 4.3 shows the distribution of model derived sea-air CO_2 flux in the Southern Ocean over the period 1998 - 2007 taken from the daily resolution model data. In Figure 4.3 there is a positive flux (outgassing) in most regions south of 50° which was more similar to the values seen in the observations than the 5 day mean data. The most striking differences were south of 50°. Strong outgassing can be observed east of the Antarctic Peninsula around the Weddel and Scotia seas. Strong outgassing is also depicted in the eastern Weddell gyre, south of the Polar Front. North of 50° there is more similarity to the 5-day resolution model output. The range in the decadal mean of the daily resolution model data was 0.62 mol C per m^2 per month. This was greater than both the observational data and the 5 day resolution model data. The maximum sea-air CO_2 flux was 0.28 mol C per m^2 per month and the minimum CO_2 flux was -0.34 mol C per m^2 per month. The range in all of the data of the daily resolution model data was 9.25 mol C per m^2 per month. This was significantly higher than both the observational data and the 5 day resolution model data. This was expected as the monthly

resolution of the Takahashi et al (2009) observations and the 5 day resolution of the other model data would average out and smooth over the extremes. The maximum sea-air CO_2 flux is 4.37 mol C per m^2 per month and the minimum CO_2 flux is -4.88 mol C per m^2 per month.

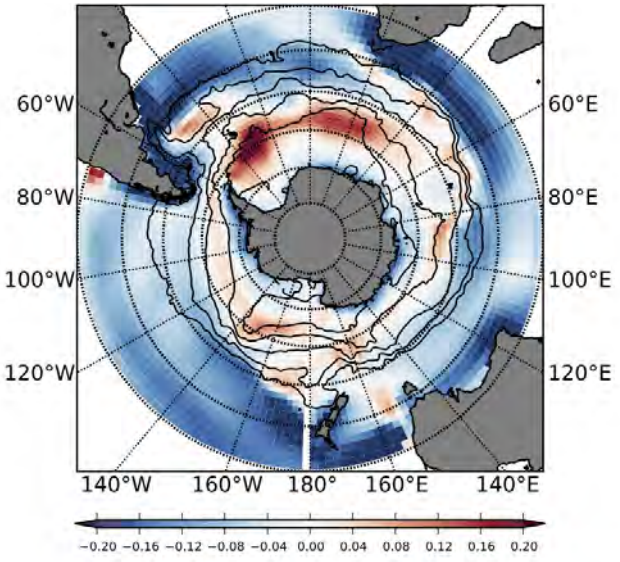

Figure 4.3: Decadal mean CO_2 flux for 1998 - 2007 (daily model output) for the Southern Ocean South of 30° in molC.m^{-2}.$month^{-1}$ with the positions of the Southern Ocean fronts overlaid.

Table 4.1 shows that, for the observational data and 5-day resolution model output, the Indian and Atlantic Ocean sections of the Southern Ocean were the ocean basins with the highest mean uptake. In contrast, for the daily resolution model output, the Atlantic Ocean had the lower mean uptake due to outgassing in the Scotia Sea and Weddell gyre (Figure 4.3).

Figure 4.4 shows the decadal mean for ΔpCO_2 (sea-air) for the daily resolution model data. As expected, the regions of high positive ΔpCO_2 closely resemble those of a strong outgassing and regions of negative ΔpCO_2 resemble regions of uptake.

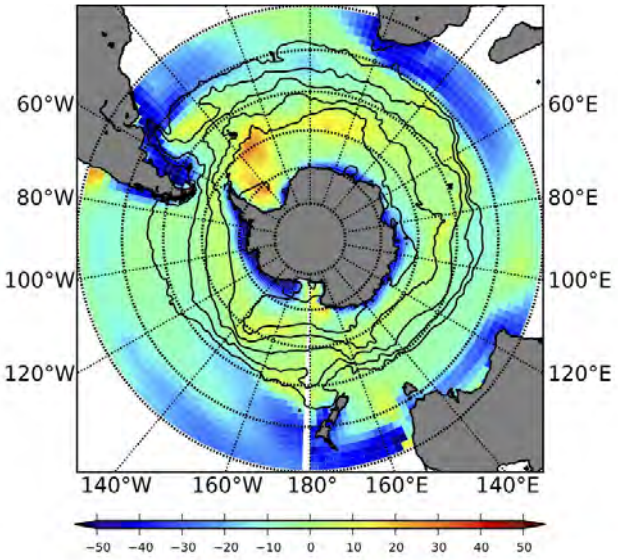

Figure 4.4: Decadal mean ΔpCO_2 for the Southern Ocean for the daily resolution 1998 - 2007 data for the Southern Ocean South of 30° in μatm with the positions of the Southern Ocean fronts overlaid.

The mean decadal ΔpCO_2 distribution exhibits a similar pattern to the sea-air CO_2 flux. On longer timescales CO_2 flux is driven by ΔpCO_2. The strong positive uptake around the Antarctic continent is driven by the large ΔpCO_2. In other regions, the positive uptake is found along western boundary currents that bring tropical and subtropical water south into the cooler temperate mid-latitudes.

Figure 4.5. shows the time series of the seasonal cycle of sea-air CO_2 flux for observations from Takahashi et al. (2009), 5 day resolution simulated model data, and daily resolution data. The maximum CO_2 flux value from Takahashi et al. (2009) was -0.016 and the minimum is -0.079 Mole Carbon per m^2 per month. The highest average CO_2 flux was in September and the lowest was in February. The maximum CO_2 flux value from the 5 day resolution simulated model data was -0.06 and the minimum was -0.13 Mole Carbon per m^2 per month. The highest average CO_2 flux was in September and the lowest was in February. The maximum CO_2 flux value from daily resolution data was -0.02 and the minimum was -0.07 Mole Carbon per m^2 per month. The highest average CO_2 flux was in February and the lowest was in December.

The daily resolution and the 5 day resolution data differed greatly. This was due to an error surface boundary layer dynamics in the model used to create the 5-day averaged data. The daily resolution data fits within the range of the observation data, but shows relatively higher values in summer. The 5 day resolution model data displays a stronger seasonal cycle, but with lower values year round. Both the 5 day and daily resolution data show large variability in summer months.

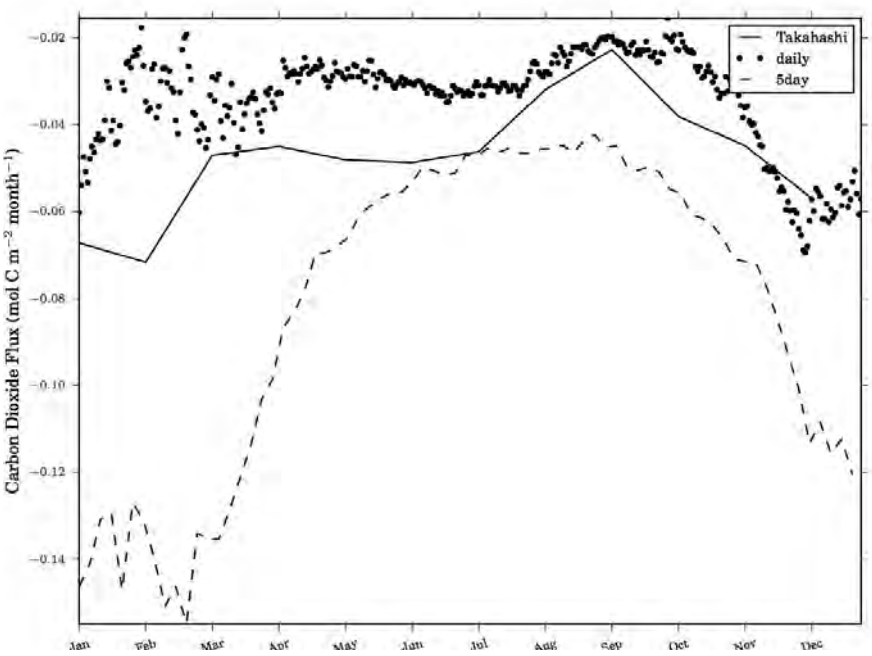

Figure 4.5: Climatological seasonal cycle for observed CO_2 flux (solid line) for a reference year 2000 from Takahashi et al (2009), 5 day resolution model data (dashed line) and daily resolution model data (dotted line) in molC.m^{-2}.month^{-1} for the Southern Ocean South of 30°

Figure 4.6a provides the February (summer) map of the observations from Takahashi et al. (2009) and Figure 4.6b shows the August (winter) observations. From the summer observations it appears that much of the flux was neutral with some outgassing but the majority of the Southern Ocean exhibits strong uptake especially to the west of Patagonia. In winter the negative flux regions move northwards and most of the Southern Ocean exhibits neutral flux with the exception of the ice margin where there is very strong outgassing. The maximum value for February was 0.13 Mole Carbon per m^2 per month and the minimum value was -0.67 Mole Carbon per m^2

per month. For August the maximum was 0.43 Mole Carbon per m^2 per month and the minimum was -0.46 Mole Carbon per m^2 per month.

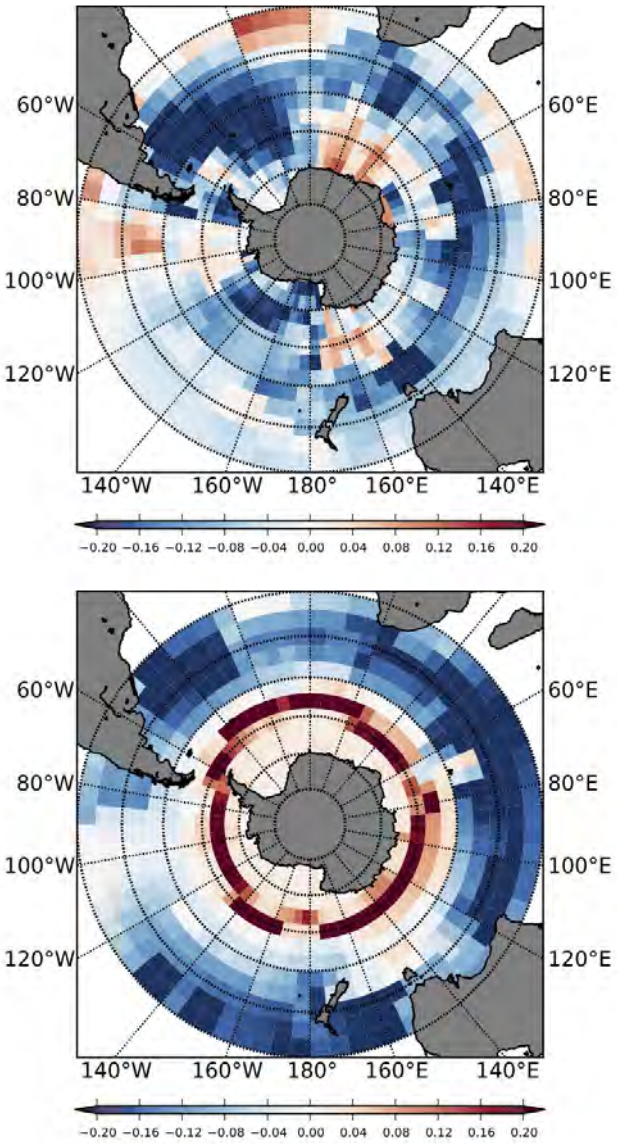

Figure 4.6: a) Mean February CO_2 flux, and b) Mean August CO_2 flux for a reference year 2000 from Takahashi et al (2009) in molC.m^{-2}.$month^{-1}$ for the Southern Ocean South of 30°

To assess and compare the regional contrasts in seasonal extremes of the in model data and observations, the climatological Summer (January, February, March) and Winter(July, August, September) model data were

compared to the most recent data from Takahashi et al. (2009). The summer and winter climatological means were calculated by taking the mean of the average daily values for Summer (January, February, March) and Winter (June, July, August).

Figures 4.7a and 4.7b show the summer (January, February, March) and Winter (July, August, September) simulated CO_2 data from the 5-day averaged data for 2000 to 2009 and Figures 4.8a and 4.8b show the summer (January, February, March) and Winter (July, August, September) simulated CO_2 data from the daily resolution data for 1998 - 2007. The daily resolution model output seems to have captured the outgassing at the ice margins more closely than the 5-day averaged data. The daily CO_2 data better represented the temporal and the spatial variability that can be seen in the observations than the 5 day mean data.

Both the 5-day and the daily resolution data captured the range of the observation data, however, the daily resolution model output shows more widespread outgassing in summer than the observational data but is similar to the winter observational data.

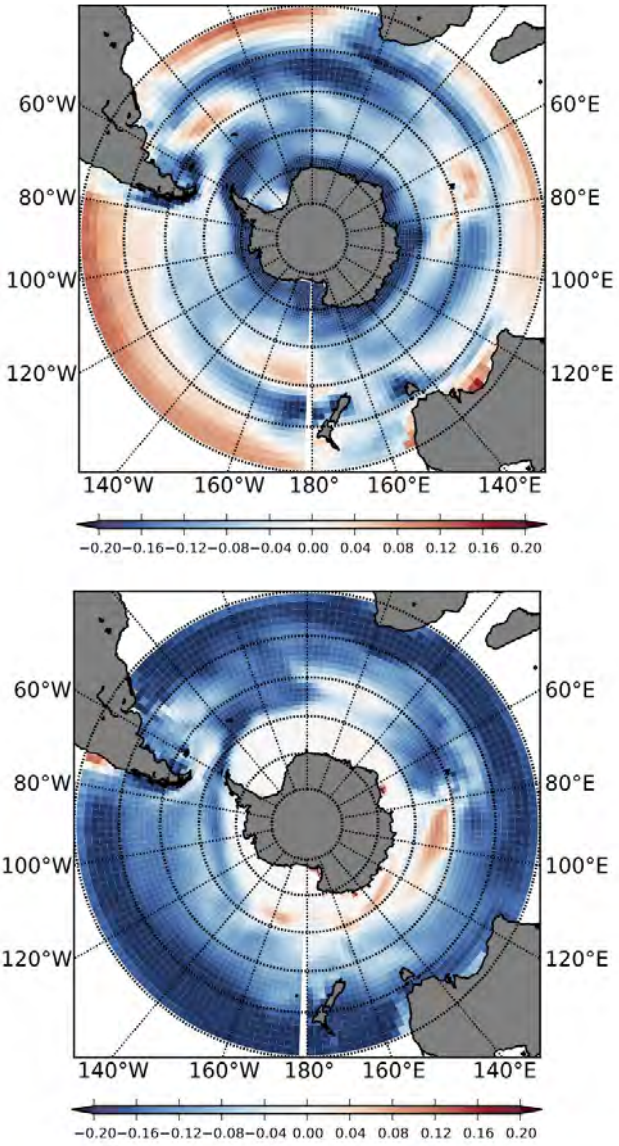

Figure 4.7: a) Mean climatological summer (top) and b) Winter (bottom) CO_2 flux for 2000 - 2009 (5-Day mean model output) for the Southern Ocean South of 30°

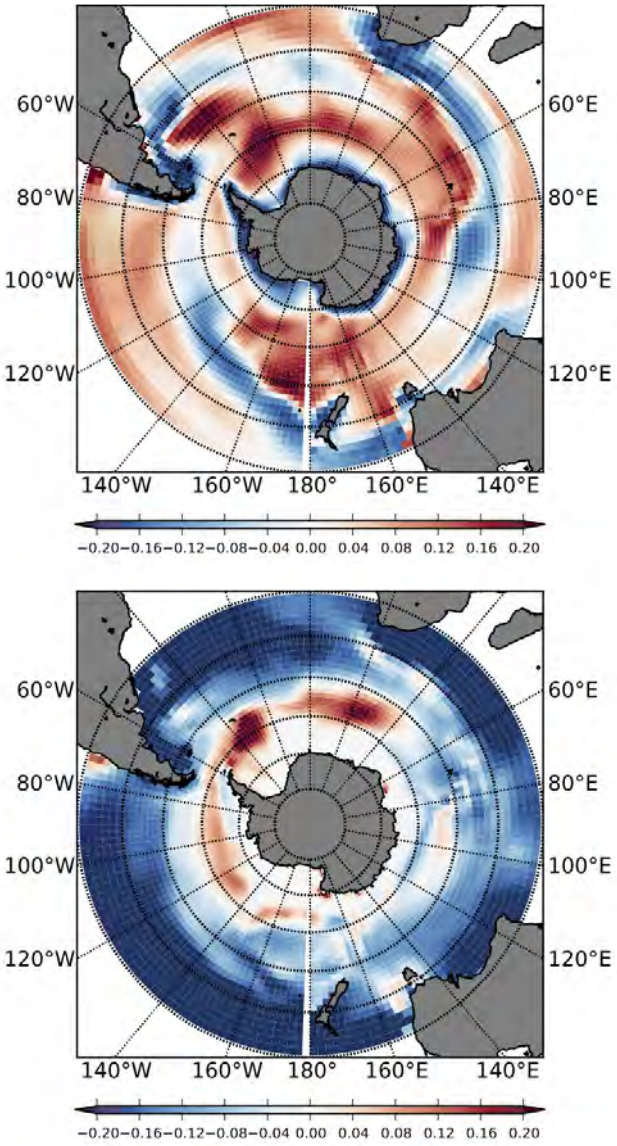

Figure 4.8: a) Mean climatological summer (top) and b) Winter (bottom) CO_2 flux for 1998 - 2007 (daily model output) for the Southern Ocean South of 30°

The summer mean had a very strong negative flux (CO_2 uptake) along the edge of the Antarctic continent which became a zero flux over winter. A strong negative flux can be seen between 30° and 40° S in winter. The seasonal shift from North to South can be seen when comparing summer and winter flux. The sea-ice in winter resulted in a zero CO_2 flux. The

outgassing in the Weddell and Scotia seas was stronger in summer but was present in winter whereas in most other regions the outgassing was squeezed out by the colder waters in winter. In winter the outgassing was reduced in the Weddell sea and predominantly occurred in the Scotia Sea.

Table 4.2: Comparison of the summer and winter simulated mean CO_2 fluxes from the daily model output and the 5-day mean model output with the annual mean from observations.

Season	Takahashi et al. (2009)	5-Day Mean Resolution	Daily Mean Resolution
Summer	-0.012	-0.06	-0.037
Winter	-0.052	-0.11	-0.001

The summer and winter Takahashi et al. (2009) observational data, 5-day mean model data, and daily resolution model data are summarised in Table 4.2. The difference between summer and winter sea-air CO_2 flux is important in order to understand what is driving variability.

To investigate sampling strategies it is important to determine which regions experience strong seasonality and in which regions the non-seasonal variability contributes the most to the to the overall variability. In order to determine this two methods were used: the signal-to-noise ratio (see Section 3.6) was calculated for each pixel and the reproducibility of the seasonal cycle (see Section 3.7) was calculated.

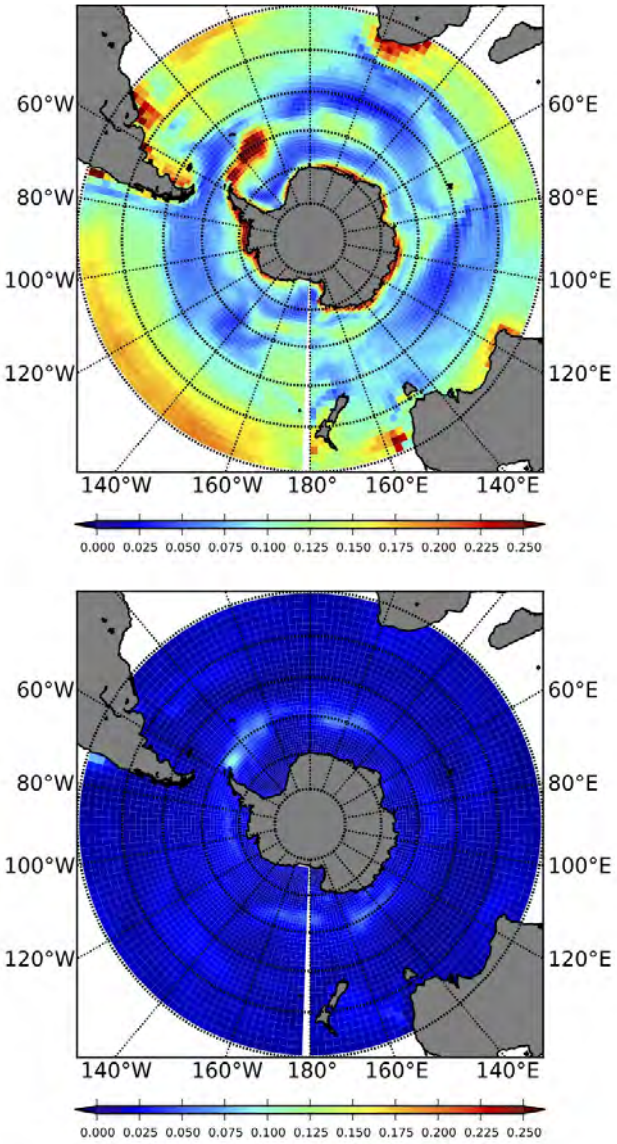

Figure 4.9: Standard deviations of Southern Ocean CO_2 flux from 1998-2007 (daily model output) for a) all the data, and b) inter-annual (standard deviation across annual means), for the Southern Ocean South of 30°

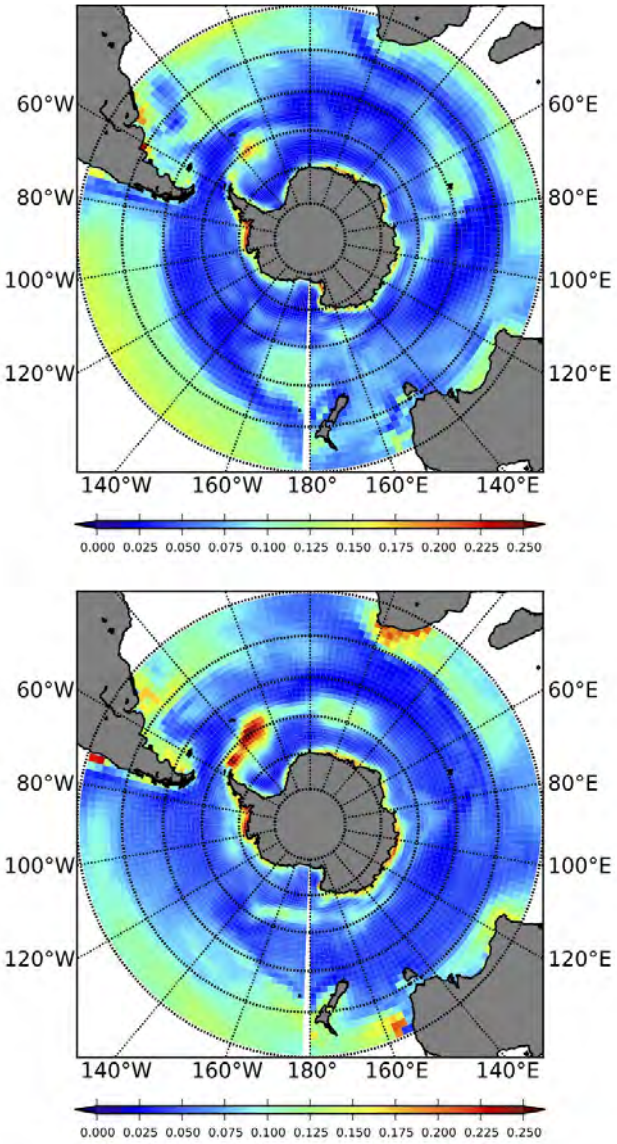

Figure 4.10: Standard deviations of Southern Ocean CO_2 flux from 1998-2007 for a) the seasonal (mean annual) cycle, and b) the non-seasonal (sub-seasonal and inter-annual), for the Southern Ocean South of 30°

The inter-annual standard deviation is shown in Figure 4.9b. When compared to the standard deviation of all the data (Figure 4.9a), standard deviation of the seasonal cycle (Figure 4.10a), and the standard deviation of the non-seasonal data (Figure 4.10b), it can be seen that inter-annual variability contributed a small amount to the variability in the model data,

especially around 60°S. The regions with elevated inter-annual variability were also strong outgassing zones. The region with the strongest inter-annual variability was in the Scotia Sea (between 55 to 65°S and 20 to 60°W) next to the Antarctic Peninsula, probably due to variability in ice melt from year to year. When assessing the variability of all data (Figure 4.10a), the Scotia Sea remains a region of strong variability and was matched in magnitude by the variability seen in the coastal Southern Ocean along Antarctica. From the variability within the seasonal (annual) cycle (Figure 4.10a), it can be seen that the intra-annual variability (variability of the annual cycle) was the largest contributor to the overall variability. The Scotia Sea also had strong variability within the seasonal cycle. The strongest variability within the seasonal cycle was along the Antarctic continent. The variability in the non-seasonal cycle (Figure 4.10b) in the Southern Ocean, contributed equally, if not more, to the overall variability than the seasonal cycle. These results were used to calculate the signal-to-noise ratios (SNR) values in Figures 4.11a and 4.11b.

The SNR has been computed for the 5 day resolution and the daily simulated CO_2 fluxes in the Southern Ocean in Figure 4.11a and Figure 4.11b. Over most of the Southern Ocean the SNR was less than 1. This confirms that non-seasonal variability was contributing significantly to the variability in the region. In these regions it would be difficult to obtain sufficient observations to constrain the seasonal cycle and inter-annual variability. The area around Kerguelen is an exception that returned a very high SNR for both the 5 day resolution model data and the daily resolution model data. The SNR for the 5 day resolution model data was generally greater since the temporal resolution is coarser. Focussing on the daily resolution data, it can be seen that the low variability in the mean annual cycle (seasonal cycle) in certain areas was what contributed to the low SNR. A low variability in the mean annual cycle (seasonal cycle) was generally consistent with a low SNR value rather than the presence of large non-seasonal variability.

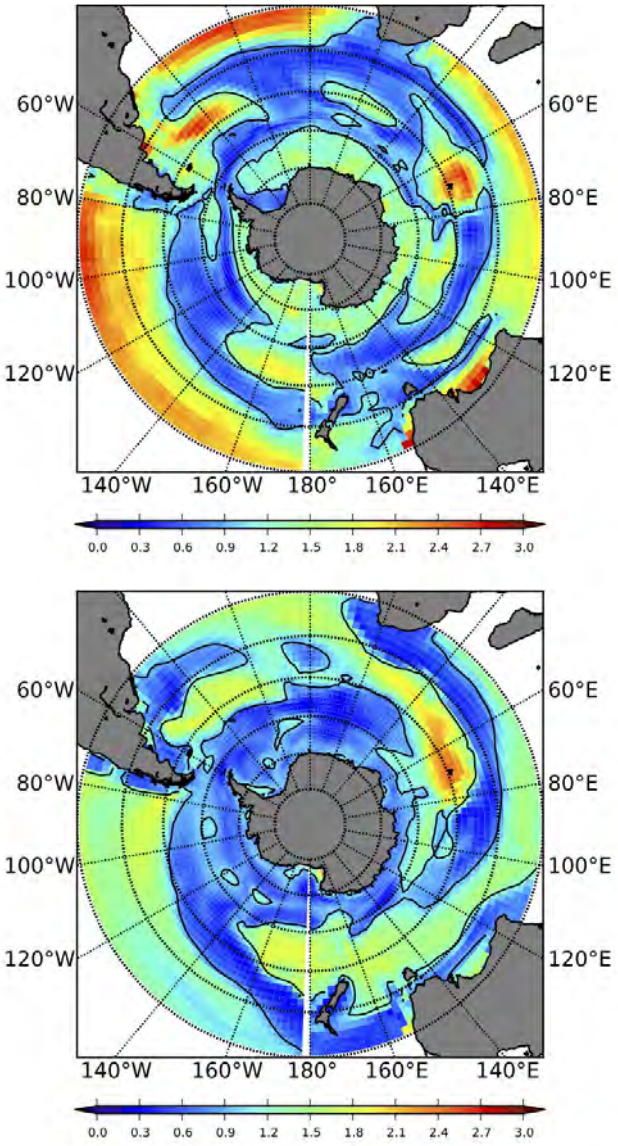

Figure 4.11: Signal-to-noise ratios in the Southern Ocean for a) 5 day CO_2 flux, b) daily CO_2 flux,

Calculating the seasonality (section 3.7) of the CO_2 flux was done according to the method used in Thomalla et al. (2011) and is shown for the 5-day data in Figure 4.12a and the daily resolution data in Figure 4.12b. Defining regions of weak or strong seasonal cycle reproducibility was based on a comparison of the SNR plots and the correlation coefficient plots. Regions with a r^2 value of 0.6 or above were also regions where the SNR value

was above one. From the daily resolution data (Figure 4.12b) it was evident that seasonality (high SNR) was strongest between the Subtropical Front (STF) and the Southern ACC Front (SACCf) with the exception of the Subtropical zone north of the STF where there was low seasonal variability and therefor the the SNR is less than 1.

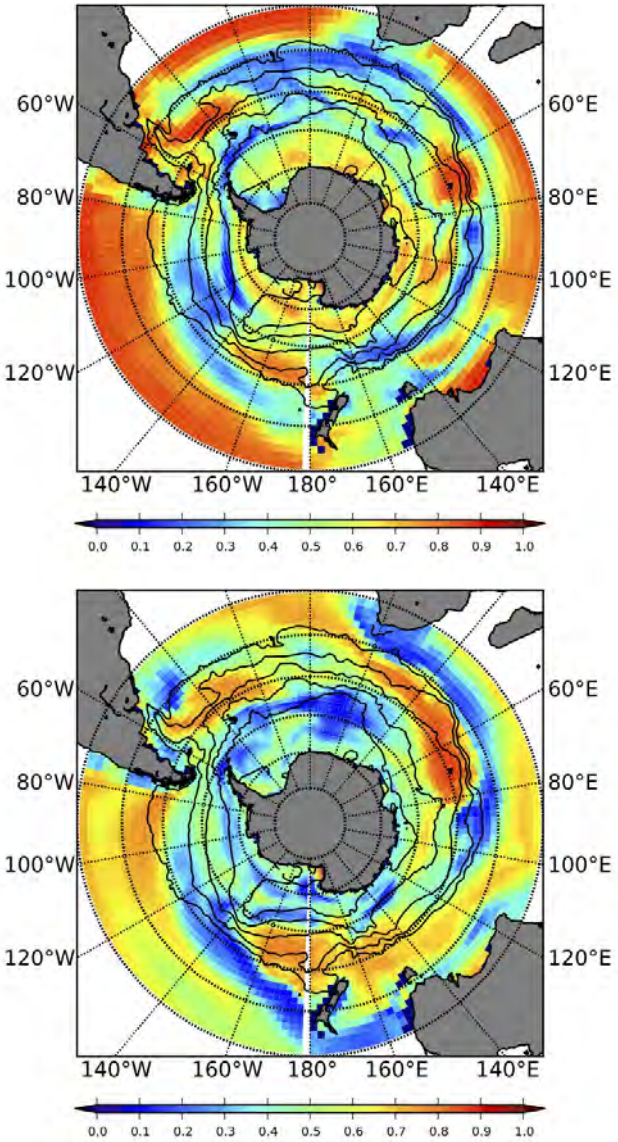

Figure 4.12: Seasonal cycle reproducibility (correlation coefficient) in the Southern Ocean (south of 30°) for a) 5-day mean model output , b) daily seasonality

Plots of time against latitude and time against longitude for the daily resolution model output enable the seasonality and inter-annual characteristics of the data to be visualised. Sections of constant latitude and constant longitude are plotted in Figure 4.13a and 4.13b. The seasonal shift of outgassing and uptake regions from summer to winter can be seen when averaging both over latitude and longitude. In Figure 4.13a we can see that there were 3 latitudinal bands in the Southern Ocean. A very strong, year round band of outgassing can be seen in Figure 4.13a centering on the 60°S line of latitude. North of 55°S the mean sea-air CO_2 flux changed from positive (outgassing) in the summer to negative (uptake) in the winter. South of 65°S the sea-air flux changed from negative (uptake) in the summer to zero flux in the winter due to formation of sea-ice preventing sea-air gas exchange.

The mean sea-air CO_2 flux changed from positive (outgassing) in the summer to negative (uptake) in the winter over most of the lines of longitude.

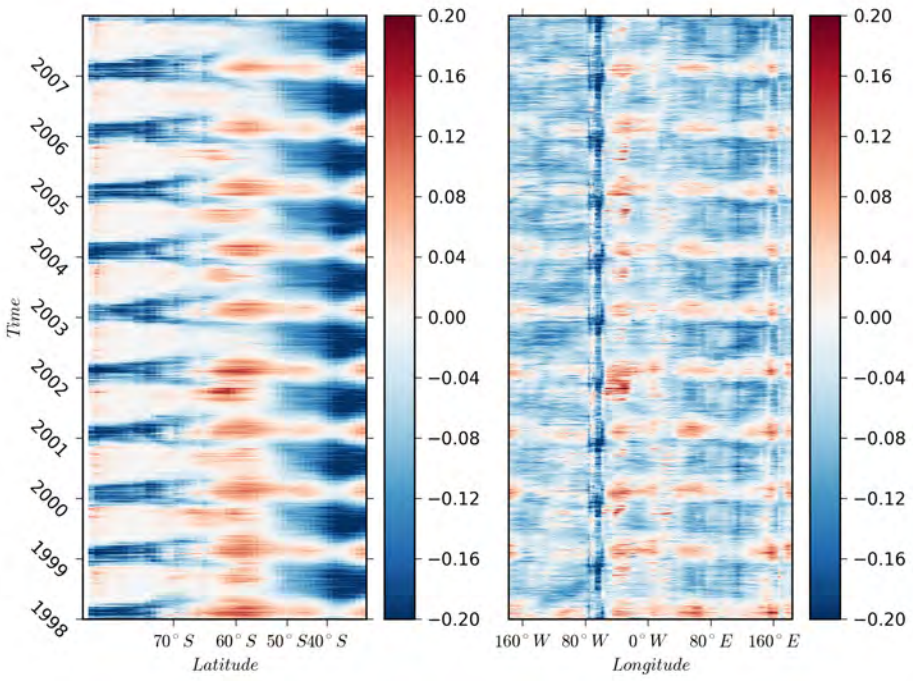

Figure 4.13: Hovmuller plots of time against latitude with longitude averaged (left) for all the data and a plot of time against longitude with latitude averaged (right) for daily resolution CO_2 flux in molC.m^{-2}.month^{-1}

These figures were analysed using 2D Fourier transforms resulting in Figure 4.14a and 4.14b. These figures characterise the spectrum of the simulated flux variability.

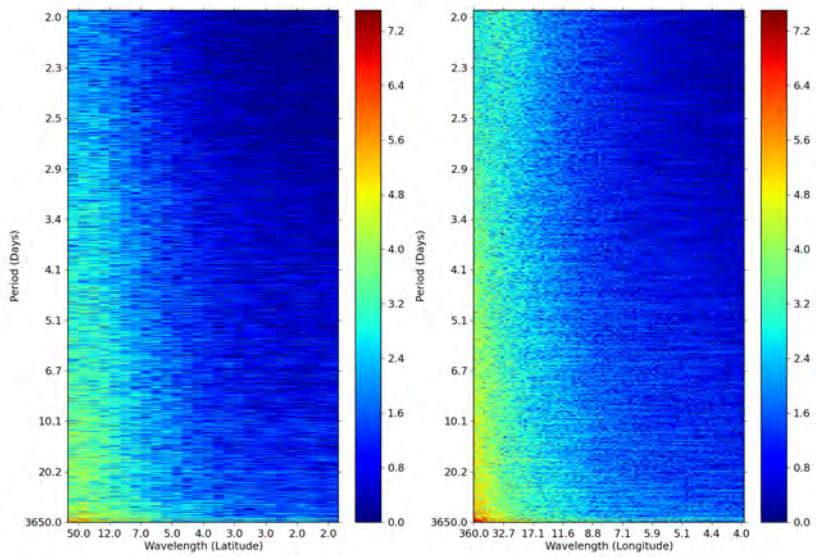

Figure 4.14: Two-dimensional fourier transform of a) longitudinally and b) latitudinally averaged simulated Southern Ocean CO_2 flux for daily resolution data

Figures 4.14 a and 4.14b show that most of the variance was explained by longer periods and wavelengths. The largest values lie consistently across the 10th pixel/row in the v direction. This row represents generalised sinusoids in the y direction with a frequency of 10 per decade or 1 per year. Therefore the strongest signal lies in the seasonal (annual) cycle for all latitudinal and longitudinal wavelengths. The Fourier transform has secondary peaks at the harmonics of the annual cycle. Sampling at a resolution of finer than 20° in latitude will capture most of the variability.

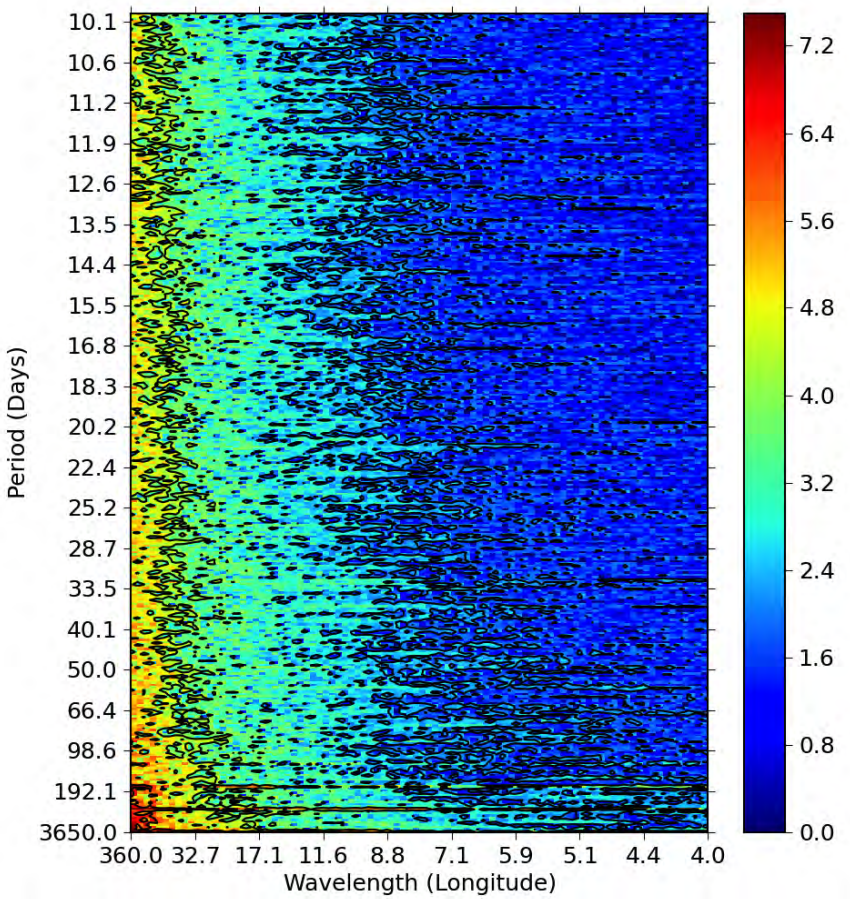

Figure 4.15: Two-dimensional fourier transform of a) longitudinally and b) latitudinally averaged simulated Southern Ocean CO_2 flux for daily resolution data

Figure 4.15 depicts the latitudinally averaged 2D Fourier transform in more detail. Periods larger than 180 days display the most variability but for periods less than 180 days, wavelengths larger than 10° still show a uniform spread of variability over all periods.

4.3 Sensitivity of CO_2 uptake to Sampling strategies using a regular grid

4.3.1 Introduction

To further the understanding of the role of the Southern Ocean in the changing global carbon cycle, the inter-annual change in CO_2 flux needs to be resolved. The uncertainty of current estimates are too high mostly due to the scarcity of observations. In this section, the goal was to reproduce the findings of Lenton et al. (2006) on model simulated data from a ocean biogeochemical model. These findings could then be used as a basis to explore non-structured sampling strategies. To assess the different sampling errors obtained from sampling at different sampling frequencies, an exhaustive search was used to calculate the sampling error obtained when sampling the mean annual cycle using different sampling frequencies. This exhaustive search helped to identify a range of possible sampling frequencies that could be further applied to sampling the entire simulated model data set.

Uncertainty due to inter-annual variability, sampling error and total sampling uncertainty were calculated according to the definitions in Lenton et al. (2006).

Definition of uncertainty due to inter-annual variability:

$$2\sigma_{inter-annual}* = 2\sigma_{inter-annual}/\sqrt{10} \tag{4.1}$$

where θ is the variance across the annual mean uptake.

Definition of sampling error:

$$2\sigma_{sampling}* = 2\sigma_{sampling}/\sqrt{10} \tag{4.2}$$

where θ is the variance across mean values for each permeation.

Definition total sampling uncertainty:

$$2\sigma_{total} = \sqrt{(2\sigma)^2_{inter-annual} * + (2\sigma)^2_{sampling}*}. \tag{4.3}$$

.

4.3.2 Sampling at the model resolution

The simulated annual mean uptake was calculated as 0.348 ± 0.07 PgC/yr at the model resolution. The uncertainty was calculated as

$$2\sigma_{inter-annual}* = 2\sigma_{inter-annual}/\sqrt{10} \tag{4.4}$$

where σ was the variance about the annual mean uptake.

The annual mean uptake from the different sampling strategies that were be investigated were the same as that obtained by sampling at the model resolution when averaged across all realisations of the sampling strategies grid. The variability across each realisation (sampling error) was subsequently combined with the inter-annual variability to estimate the total sampling uncertainty.

4.3.3 Exhaustive search

By using the exhaustive search method (Section 3.9) it was possible to directly investigate the sampling error introduced by sampling at different sampling resolutions. Sampling resolution in space was expressed in degrees longitude and temporal sampling resolution was expressed in days. The mean annual cycle was sampled at each sampling frequency and the sampling error calculated from each sampling frequency is plotted in Figure 4.16.

The results from this method show how to combine sampling in time and in space in order to return the lowest sampling error. Figure 4.16 shows how sampling error decreases with an input of sampling effort, and the suggested sampling strategy can be chosen according to the acceptable sampling error. The figure confirmed the output of the Fourier transform of sea-air CO_2 flux averaged over latitude (figure 4.15), as it is evident that sampling the data less than twice a year and every 120° longitude will result in a high variance or uncertainty. To investigate in more detail, sampling strategies employing 3, 4 and 5 times a year at every 20°, 30° and 40° in longitude were applied to the entire simulated model data set.

4.3.4 inter-annual variability versus Sampling error

According to the model analysis done by Lenton et al. (2006), the inter-annual variability at the proposed sampling strategy contributed more to the error introduced by the coarser resolution of the sampling strategy. This implies that the sampling strategy suggested by Lenton et al. (2006) was sufficient to resolve the seasonal cycle and reduce the sampling uncertainty to an acceptable level.

Combining the sampling error and the uncertainty due to inter-annual variability allowed an estimate for the total uncertainty. The total uncer-

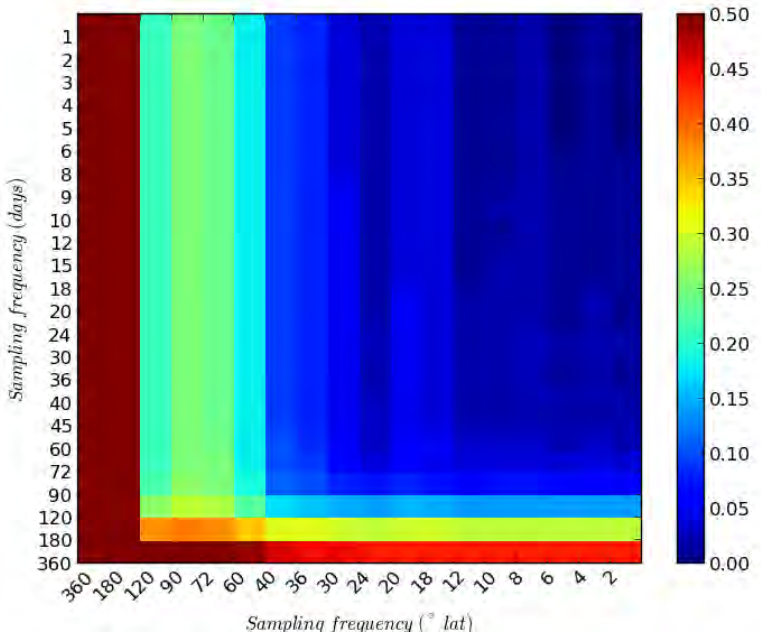

Figure 4.16: Plot of the sampling error (in PgC/yr) obtained when sampling the mean annual cycle at various sampling frequencies. The finest spatio-temporal sampling resolution are in the top right, while coarsest sampling resolution are in the bottom left. Red indicates large sampling errors and blue indicates small sampling errors.

tainty was calculated as

$$2\sigma_{total} = \sqrt{(2\sigma)^2_{inter-annual} + (2\sigma)^2_{sampling}}. \quad (4.5)$$

Additionally, a sampling strategy that results in a sampling error which, in turn does not increase the total sampling error ,would reduce the uncertainty level to one completely dominated by inter-annual variability.

$$2\sigma_{inter-annual} \approx 2\sigma_{total} \quad (4.6)$$

The goal, therefore, is to minimise the $2\sigma_{sampling}$. As with Lenton et al. (2006), a sampling strategy that samples 4 times a year at every 30° results in the sampling error contributing less to the total sampling error than the inter-annual variability. In order for the total sampling error to be unaffected by the sampling error, a sampling strategy that samples 5 times a year at every 24 degrees in longitude is required.

Which sampling strategy can meet the same criteria as suggested by Lenton et al. (2006) when using the PISCES / ORCA model? A table can be constructed for each of the possible sampling strategies in order to assess whether or not the sampling strategy meets these criteria.

4.3.5 Sampling strategies using a regular grid

An area of interest, between sampling twice a year, every 60° and sampling 9 times a year, every 18°, was selected from figure 4.16 in order to investigate different sampling frequencies in more detail. Sampling errors obtained from sampling from 2 to 9 times per year and every 18° to 60° in longitude are summarised in figure 4.17. The tables for each sampling frequency are summarised into Tables A1-A49 (see appendices).

As the sampling density increases above 0.1%, the rate of decrease in sampling error slows. Sampling frequencies with a sampling density of between 0.05% and 0.15% are further investigated and the results are summarised in table 4.3.

Sampling 3 times a year, every 40°, resulted in 2160 realizations or grids per year. This sampling strategy used 1215 sampling locations or a sampling density of 0.058%. This is the equivalent of 27 north-south sampling sections

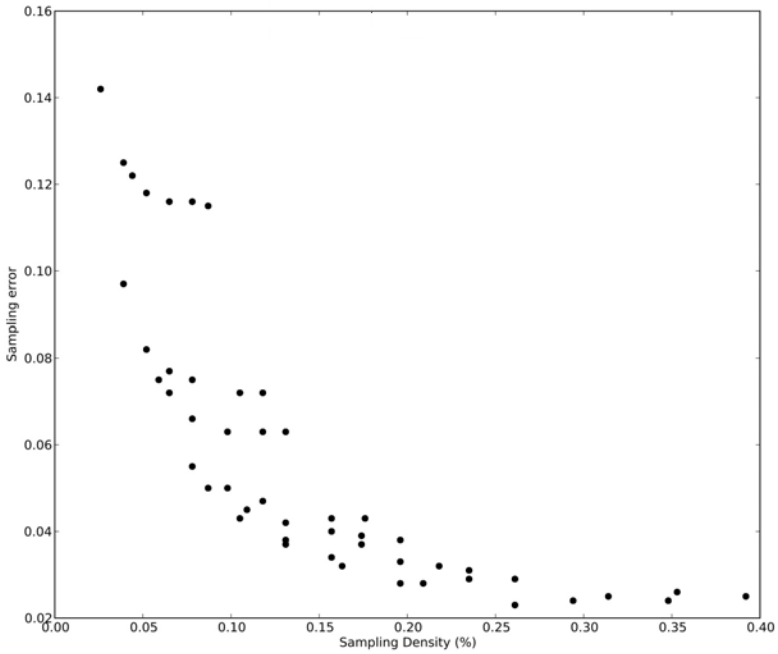

Figure 4.17: Plot of the sampling error obtained when sampling the mean annual cycle at various sampling densities (no. of sampling locations / total no. of locations)

Table 4.3: Summary of different sampling errors and total sampling uncertainty for various sampling frequencies.

Sampling Frequency (per year,longitude)	Annual Mean (PgC/yr)	Sample Error	Total Sampling Uncertainty
3,40	0.347±0.07	0.075	0.103
3,30	0.348±0.07	0.066	0.096
3,20	0.347±0.07	0.063	0.094
4,40	0.347±0.07	0.055	0.089
4,30	0.348±0.07	0.043	0.082
4,20	0.347±0.07	0.040	0.081
5,40	0.347±0.07	0.050	0.086
5,30	0.348±0.07	0.037	0.079
5,20	0.347±0.07	0.033	0.077

per year. The results from applying this sampling strategy are summarized in the table A.41 in the appendices. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution is ± 0.237 PgC/yr for Southern Ocean uptake. The estimated sampling error was ± 0.075 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.347 PgC.yr^{-1} with 0.103 PgC.yr^{-1} being the total sampling error.

Sampling 3 times a year, every 30° (see table A.39), resulted in 1800 realizations or grids per year. This sampling strategy used 1620 sampling locations or a sampling density of 0.078%. This is the equivalent of 36 north-south sampling sections per year. The sampling uncertainty (double the standard deviation (2σ)) from sampling at this resolution was ± 0.209 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.066 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.096 PgC.$^{-1}$yr with 0.096 PgC.yr^{-1} being the total sampling error.

Sampling 3 times a year, every 20° (see table A.37), resulted in 1200 realizations or grids per year. This sampling strategy used 2430 sampling locations or a sampling density of 0.118%. This is the equivalent of 54 north-south sampling sections per year. The sampling uncertainty from sampling

at this resolution was ± 0.200 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.063 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.347 ± 0.094 PgC.yr^{-1}.

Sampling 4 times a year, every 40° (see table A.34), resulted in 2160 realizations or grids per year. This sampling strategy used 1620 sampling locations or a sampling density of 0.078%. This is the equivalent of 36 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.175 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.055 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.089 PgC.yr^{-1}.

Sampling 4 times a year, every 30° (see table A.32), resulted in 1350 realizations or grids per year. This sampling strategy used 2160 sampling locations or a sampling density of 0.105%. This is the equivalent of 48 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.136 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.043 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.082 PgC.yr^{-1}.

Sampling 4 times a year, every 20° (see table A.30), resulted in 900 realizations or grids per year. This sampling strategy used 3240 sampling locations or a sampling density of 0.157%. This is the equivalent of 72 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.126 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.040PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.081 PgC.yr^{-1}.

Sampling 5 times a year, every 40° (see table A.27) resulted in 2160 realizations or grids per year. This sampling strategy used 2025 sampling locations or a sampling density of 0.098%. This is the equivalent of 45 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.157

PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.050 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.086 PgC.yr^{-1} .

Sampling 5 times a year, every 30° (see table A.25), resulted in 1080 realizations or grids per year. This sampling strategy used 2700 sampling locations or a sampling density of 0.131%. This is the equivalent of 60 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.105 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.033 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.077PgC.yr^{-1}.

Sampling 5 times a year, every 20° (see table A.23), resulted in 750 realizations or grids per year. This sampling strategy used 4050 sampling locations or a sampling density of 0.196%. This is the equivalent of 90 north-south sampling sections per year. The sampling uncertainty, double the standard deviation (2σ), from sampling at this resolution was ± 0.105 PgC.yr^{-1} for Southern Ocean uptake. The estimated sampling error was ± 0.033 PgC.yr^{-1}. Combining the sampling error and the uncertainty due to inter-annual variability resulted in an averaged uptake for the model period of 0.348 ± 0.077 PgC.yr^{-1}.

4.4 Sampling Strategies not constrained by a regular grid

4.4.1 Introduction

In the event of pCO_2 sampling by autonomous sampling platforms, the sampling will not be constrained by a regular grid. Therefore it is worth investigating lower sampling densities that are not based on north-south sampling sections. In this scenario the sampling platforms would return a mean with the greatest certainty if the sampling platforms sampled the regions with the highest variability in time and in space. To test the effectiveness of a non-gridded sampling strategy, a genetic algorithm approach (Section 3.10) was used to return a sampling strategy that would best represent the simulated data.

The genetic algorithm was set up to find optimized solutions for a number of locations. An optimized solution was defined as a set of locations where

the fitness function

$$f = (\bar{x}_{population} - \bar{x}_{sample}) + (\sigma^2_{population} - \sigma^2_{sample}) \tag{4.7}$$

is minimised. A number of optimized solutions were selected for each of the sampling density. The sampling error and total uncertainty were calculated in the same way as the sensitivity studies of CO_2 uptake calculated in the previous section.

The genetic algorithm sampled spatial locations at a specific time. In the following experiments, the instantaneous values selected by the genetic algorithm were assembled to represent the decadal mean and the climatological seasonal cycle.

In the first experiment, the decadal dataset was averaged over the time axis to result in a single decadal mean value for each location in space (latitude and longitude). The search space that the GA sampled and selected locations from, was the decadal mean. The genetic algorithm selected locations from the decadal mean for 50, 100, 500 and 1000 locations. These locations were then applied to the annual means for each year of the 5-day averaged simulated model data set in order to quantify the uncertainty.

In the second experiment, the search space was the climatological seasonal cycle. The GA selected locations from the mean climatological cycle. These locations were then used to sample the data for each year in the simulated model dataset.

The mean annual CO_2 flux from the 5-day model data was 1.81±0.07 mmolC.m^{-2}.day^{-1}. The uncertainty was calculated as

$$2\sigma_{inter-annual*} = 2\sigma_{inter-annual}/\sqrt{10} \tag{4.8}$$

where σ was the variance about the annual mean CO_2 flux.

4.4.2 Experiment 1: Sampling the decadal mean to estimate the annual means

The optimized sampling strategies selected by the GA for 50, 100, 500, and 1000 locations were tested on the annual means for each of the years in the model data set. The results are summarised in Table 4.4. Each optimized solution has a fitness function value based on the comparison of the statistics of the sample data and the statistics of the entire data set as explained in the methods section. The mean value from the sample generated data was 1.81±0.07 mmolC.m^{-2}.day^{-1} for the duration of the model data.

Table 4.4: Summary of the sensitivity of the CO_2 mean flux on sampling the annual means of the 5-day mean model simulated data in mmolC.m^{-2}.day^{-1}

Number of Locations	Annual Mean	Sample Error	Total Sampling Uncertainty
50	1.82±0.07	0.13	0.15
100	1.81±0.07	0.07	0.10
500	1.82±0.07	0.04	0.08
1000	1.81±0.07	0.02	0.07

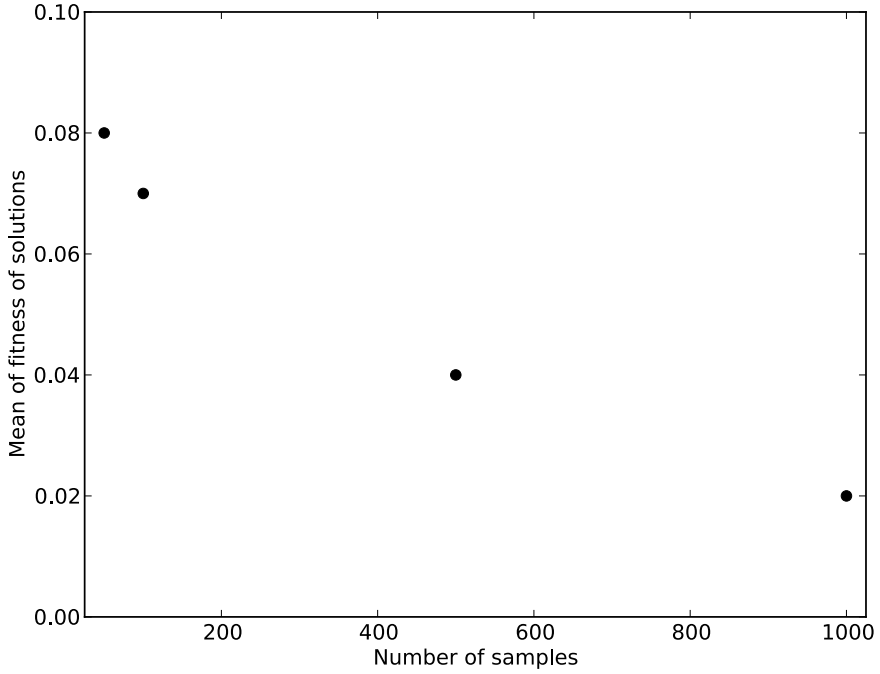

Figure 4.18: Summary of the mean fitness values obtained when sampling the annual means at different sampling frequencies

Sampling the decadal mean using 50 locations resulted in fitness values ranging from 0.01 to 0.3 with a mean fitness of 0.08. This mean value and inter-annual variability was the same as the actual mean of 1.81±0.07. The total sampling uncertainty generated by sampling using this strategy was 0.15 mmolC.m^{-2}.day (see table B.1).

Sampling the decadal mean using 100 locations resulted in fitness values ranging from 0.02 to 0.1 with a mean fitness of 0.07. The mean value from the sample generated data was 1.81±0.07 for the duration the model data. This mean value and inter-annual variability was the same as the actual mean of 1.81±0.07 with a mean fitness of 0.08. The total sampling uncertainty generated by sampling using this strategy was 0.10 mmolC.m^{-2}.day (see table B.2).

Sampling the decadal mean using 500 locations resulted in fitness values ranging from 0.02 to 0.07 with a mean fitness of 0.04. The mean value from the sample generated data was 1.81±0.07 for the duration the model data. This mean value and inter-annual variability was the same as the actual mean of 1.81±0.07. The total sampling uncertainty generated by sampling using this strategy was 0.08 mmolC.m^{-2}.day (see table B.3).

Sampling the decadal mean using 1000 locations resulted in fitness values ranging from 0.00 to 0.03 with a mean fitness of 0.02. The mean value from the sample generated data was 1.81±0.07 for the duration the model data. This mean value and inter-annual variability was the same as the actual mean of 1.81±0.07. The total sampling uncertainty generated by sampling using this strategy is 0.07 mmolC.m^{-2}.day (see table B.4).

4.4.3 Experiment 2: Sampling the Annual Cycle to estimate the data

The optimized sampling strategies selected by the GA for 1000, 2000, 5000, and 7000 locations were tested on the entire data for each of the years in the model data set. The results are summarised in table 4.5.

Table 4.5: Summary of the sensitivity of the CO_2 mean flux on sampling the annual means of the 5-day mean model simulated data in mmolC.m^{-2}.day^{-1}

Number of Locations	Annual Mean	Sample Error	Total Sampling Uncertainty
1000	1.81±0.06	0.10	0.073
2000	1.81±0.06	0.08	0.071
5000	1.81±0.07	0.05	0.070
7000	1.81±0.07	0.04	0.070

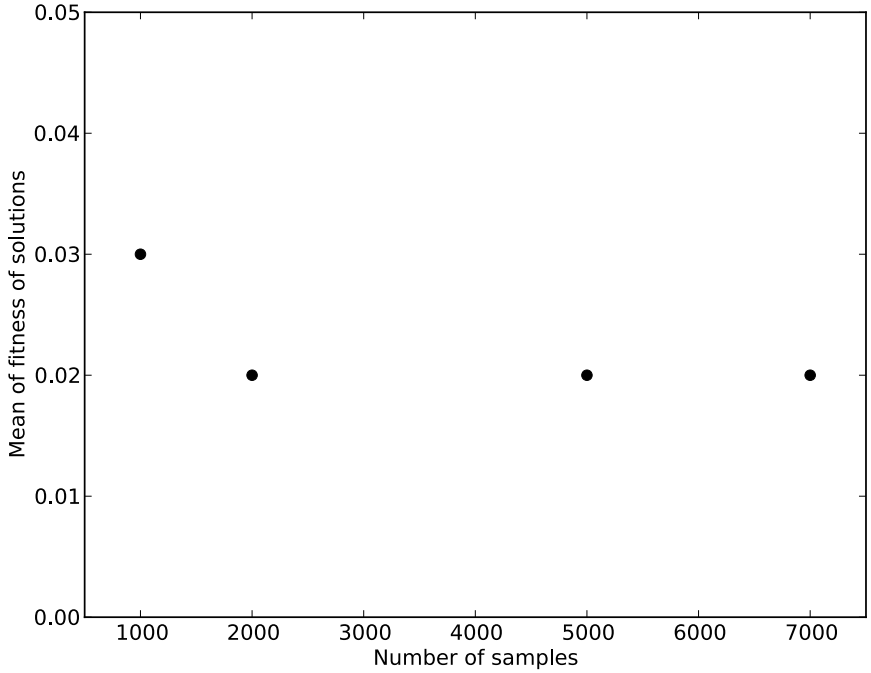

Figure 4.19: Summary of the mean fitness values obtained when sampling at different sampling frequencies

Sampling the annual mean using 1000 locations resulted in fitness values ranging from 0.01 to 0.07 with a mean fitness of 0.03. Sampling using 1000 locations, the mean value for the CO_2 flux was 1.81±0.06 mmolC/m^2/day similar to the mean value of 1.81±0.07 mmolC/m^2/day from the entire data set. Sampling using 1000 optimized locations resulted in a sampling uncertainty of 0.10. The total sampling uncertainty when using this sampling strategy was 0.07, which was the same uncertainty as sampling at the model resolution (see table B.5).

Sampling the annual mean using 2000 locations resulted in fitness values ranging from 0.01 to 0.04 with a mean fitness of 0.02. Sampling using 2000 locations, the mean value for the CO_2 flux was 1.81±0.06 mmolC/m^2/day similar to the mean value of 1.81±0.07 mmolC/m^2/day from the entire data set. Sampling using 2000 optimized locations resulted in a sampling uncertainty of 0.08. The total sampling uncertainty when using this sampling strategy was 0.07, which was the same uncertainty as sampling at the model resolution (see table B.6).

Sampling the annual mean using 5000 locations resulted in fitness values ranging from 0.01 to 0.04 with a mean fitness of 0.02. Sampling using 5000 the mean value for the CO_2 flux was 1.81±0.06 mmolC/m^2/day similar to the mean value of 1.81±0.07 mmolC/m^2/day from the entire data set. Sampling using 5000 optimized locations resulted in a sampling uncertainty of 0.05. The total sampling uncertainty when using this sampling strategy was 0.07, which was the same uncertainty as sampling at the model resolution (see table B.7).

Sampling the annual mean using 7000 locations resulted in fitness values ranging from 0.00 to 0.04 with a mean fitness of 0.02. Sampling using 7000 the mean value for the CO_2 flux was 1.81±0.06 mmolC/m^2/day similar to the mean value of 1.81±0.07 mmolC/m^2/day from the entire data set. Sampling using 7000 optimized locations resulted in a sampling uncertainty of 0.04. The total sampling uncertainty when using this sampling strategy was 0.07, which was the same uncertainty as sampling at the model resolution (see table B.8).

4.4.4 Using Radial Basis Interpolation to determine an optimized sampling strategy

The radial basis function interpolation method was used to show which locations are most important to sample in order to sample the model data set with the lowest Root mean Squared Error (RMSE) between interpolation and the model data set. Figure 4.20 shows how the RMSE is reduced as the number of locations used to sample is increased.

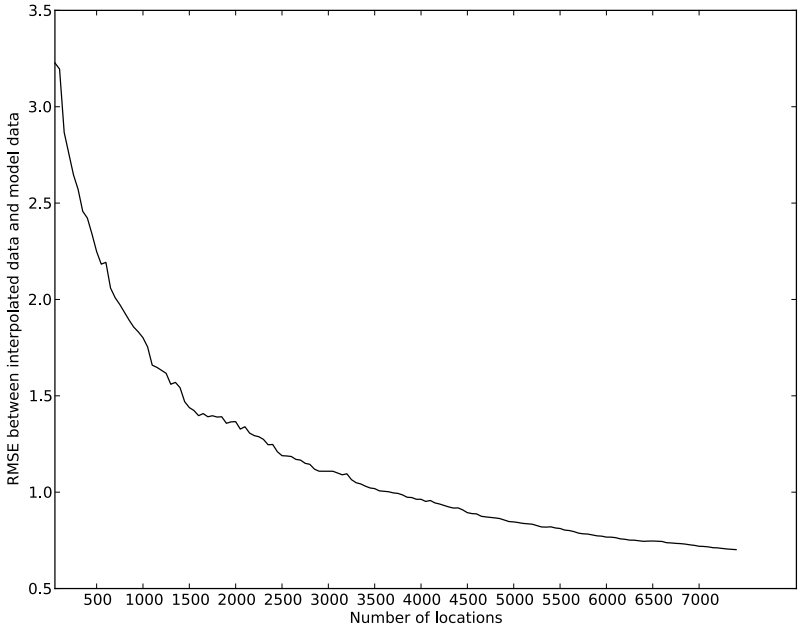

Figure 4.20: The RMSE calculated from the RBF interpolations for each number of sampling locations used

The investigation of optimized sampling strategies that do not rely on regular gridded sampling suggested that sampling with as few as 1000 model locations was sufficient to estimate the annual sea-air CO_2 flux with a low uncertainty. The Root Mean Squared Error (RMSE) between the values predicted by the interpolation and the values observed in the model was 1.8 mmolC.m^{-2}.day^{-1}.

In order to test whether the algorithm favoured any time, latitude or longitude band, histograms were plotted to see the frequency of samples made for each time, latitude or longitude band.

The longitudinal bands (Figure 4.21) which have the greatest frequency of samples are regions of high variability but they are also mainly regions around Antarctica allowing for more possible sampling locations in those areas.

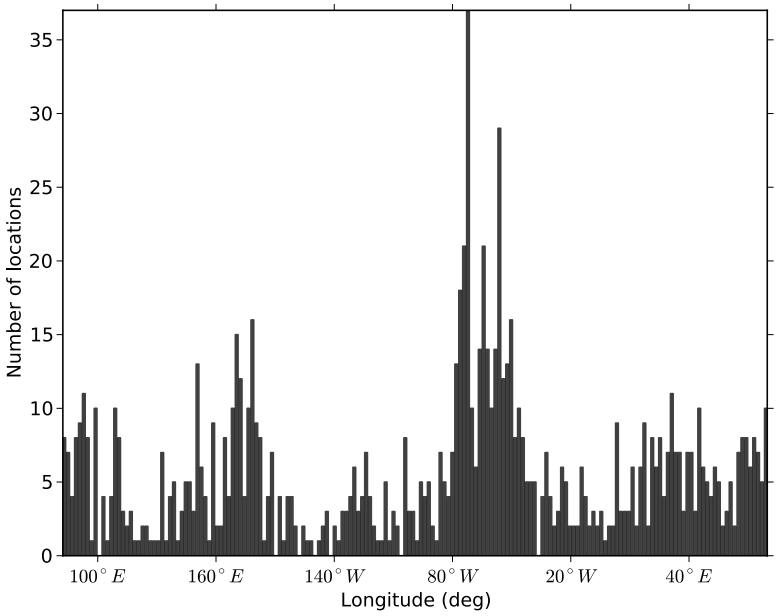

Figure 4.21: Number of sampling locations selected per 2° in longitude

The same can be said about the histogram of sampling frequency in latitudinal bands (Figure 4.22) where locations further north are favoured. Due to Antarctica taking up much of the area of the Southern part of the Southern Ocean it is possible that the algorithm has less probability of choosing locations further south. There is a spike at ±65° which corresponds to the zone of winter maximum outgassing and sea-ice. There are also lesser peaks at 40°, corresponding to the Sub-antarctic front, and 50°, corresponding to the Polar Frontal Zone.

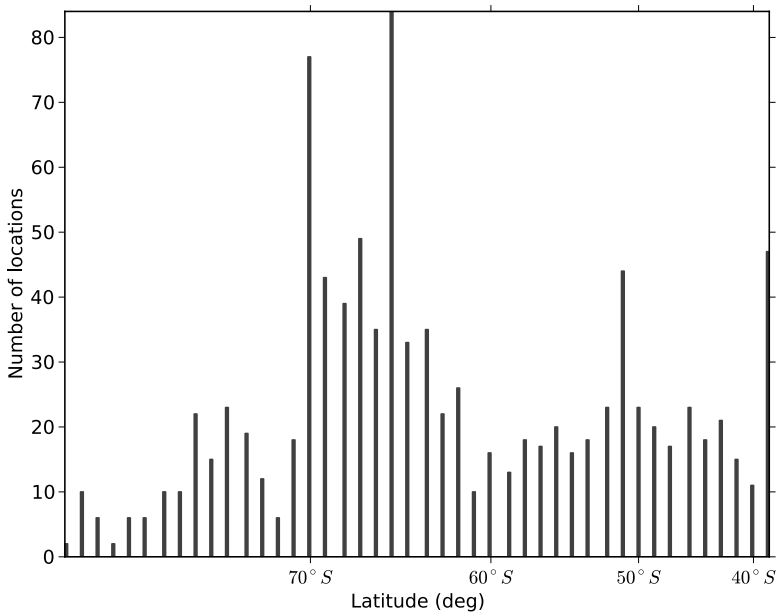

Figure 4.22: Number of sampling locations selected per 2° in latitude

From figure 4.23, the frequency of locations sampled in the summer was consistently less than the locations sampled in the winter. Due to ice forming in the winter, there was both less variability and fewer sampling locations available. However, during this time of the year the outgassing is at a maximum along the marginal ice zone, and in order for the algorithm to capture the range of the data set it would have to sample in this period.

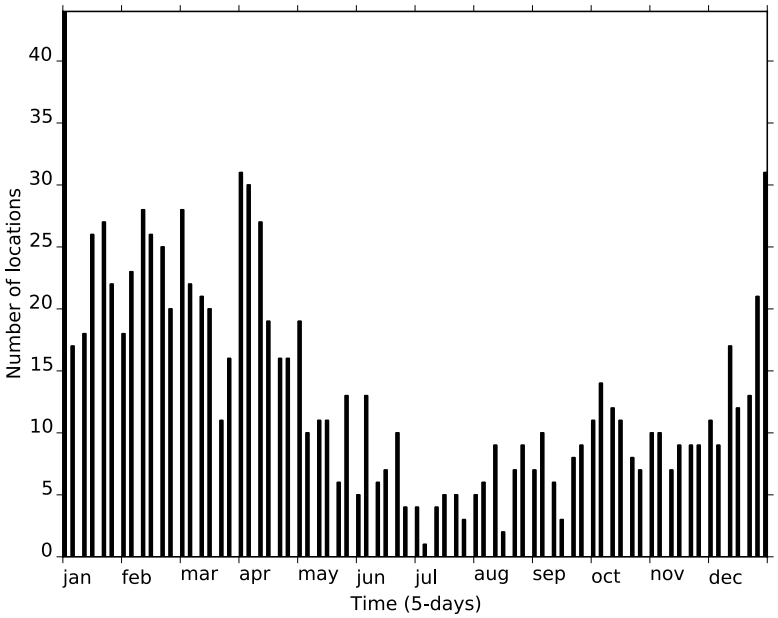

Figure 4.23: Number of sampling locations selected per 5 day period

To assess the accuracy of sampling locations chosen by the RBF method, 1000 locations chosen by the RBF method are plotted against a map of standard deviation distribution of the CO_2 data. Figure 4.24 shows that the sampling locations favoured by the RBF method were mostly found in areas of high variability in time. Plotting these same locations on latitudinally and longitudinally averaged plots (figure 4.25a and 4.25b) gives the same results i.e. zones that have high variability also have the highest concentration of sampling locations..

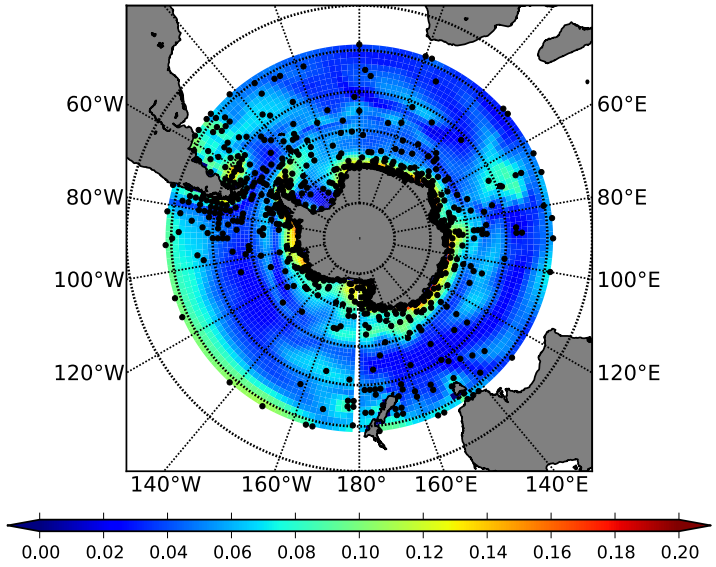

Figure 4.24: The location of 1000 sampling locations selected plotted on the standard deviation of CO_2 flux data for the Southern Ocean south of 40°S. The plot shows that sampling locations have been selected in regions of high variability.

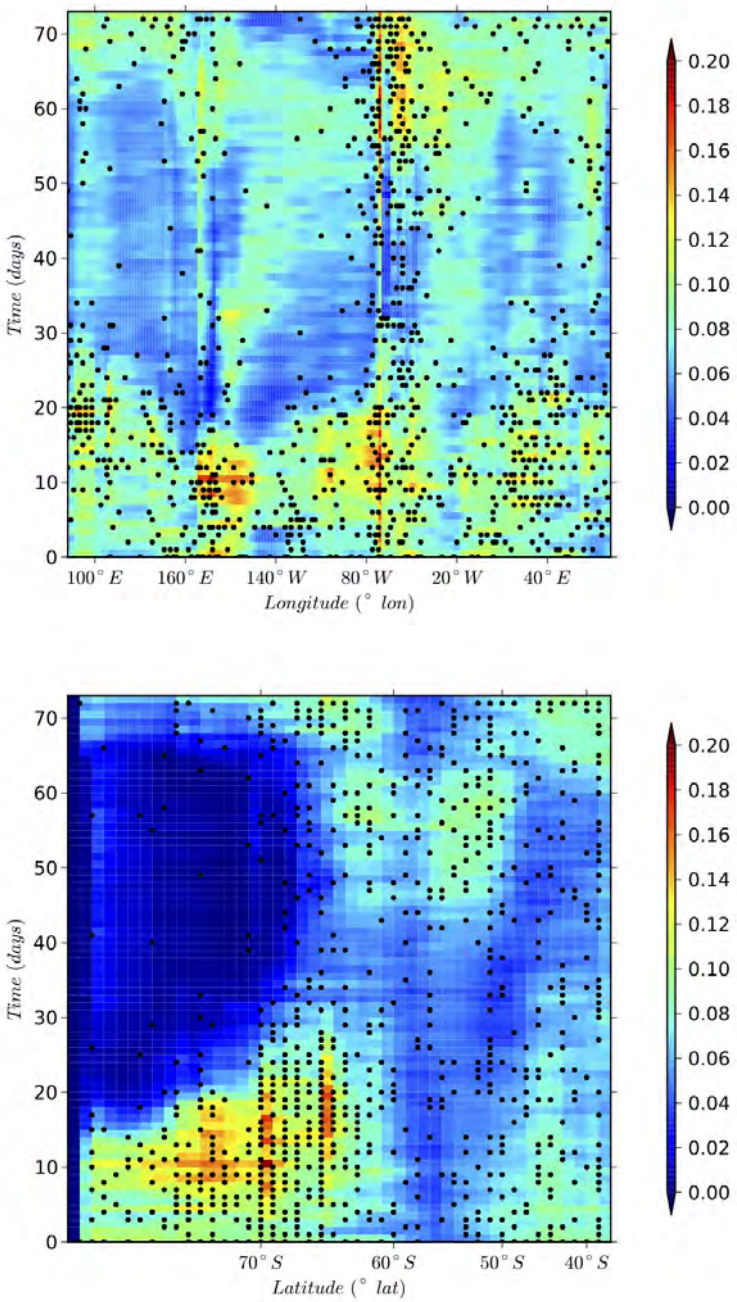

Figure 4.25: The location of 1000 sampling locations selected plotted on the standard deviation for a) latitudinally averaged, and b) longitudinally averaged CO_2 flux data for the Southern Ocean south of 40°S

Plotting the sampling locations on the mean CO_2 flux time-averaged map

(figure 4.26) shows that the sampling locations were adequately distributed to capture the spatial distribution of the data. Repeating this for latitudinally and longitudinally averaged plots (figure 4.27a and 4.27b) shows that the locations capture the distribution along all axes.

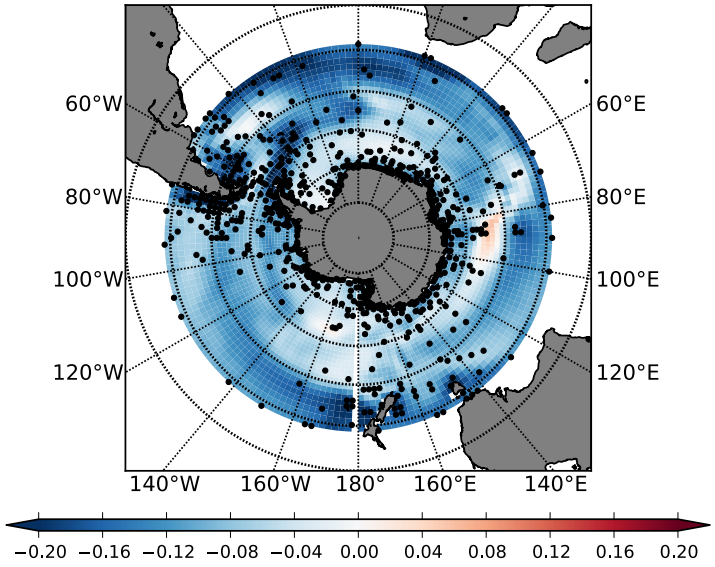

Figure 4.26: The location of 1000 sampling locations selected plotted on the mean of interpolated CO_2 flux data for the Southern Ocean south of 40°S. Sampling locations that have been selected are shown to be in regions where extreme fluxes occur.

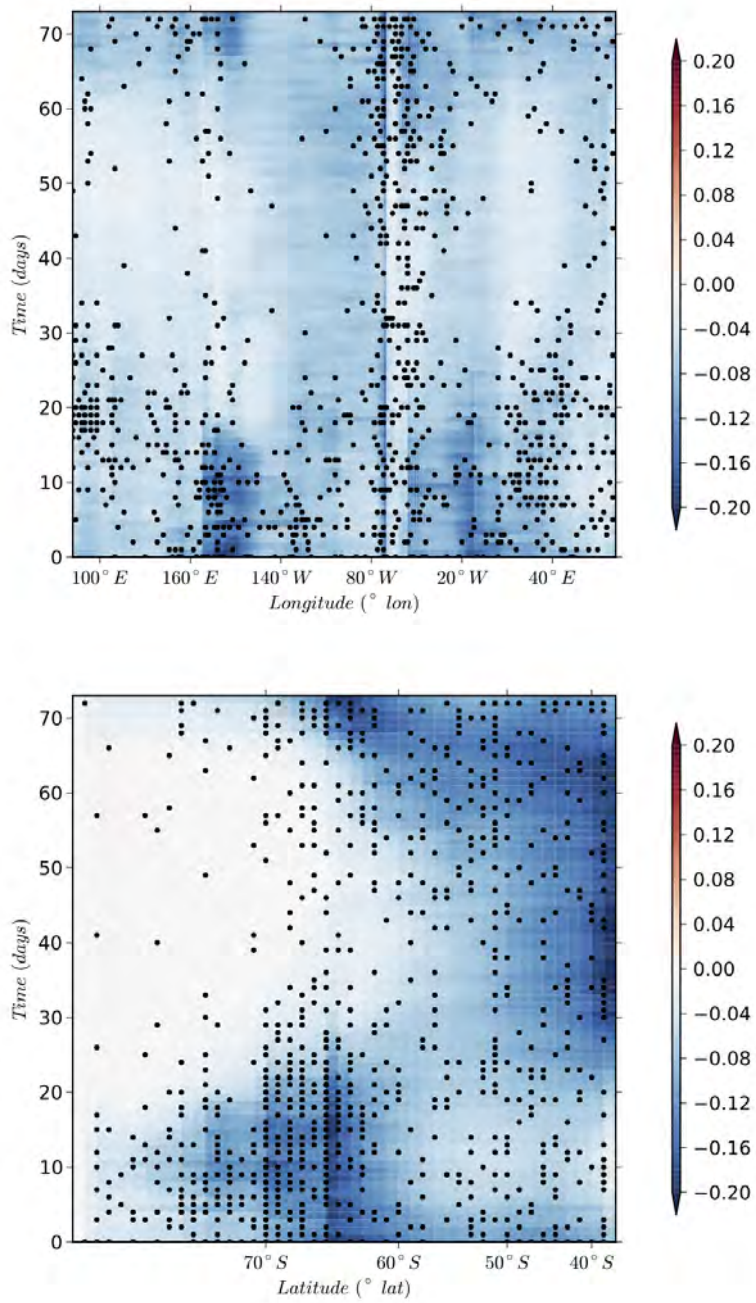

Figure 4.27: The location of 1000 sampling locations selected plotted on the mean for a) latitudinally averaged, and b) longitudinally averaged CO_2 flux data for the Southern Ocean south of 40°S

The locations selected by the radial basis function interpolation ensured

that the difference between the simulated model data and the RBF interpolated data was kept to a minimum. In order to achieve this, the locations with extreme values were preferred.

5 Discussion

5.1 Introduction

The Southern Ocean lacks sufficient coverage of pCO_2 observation to resolve the inter-annual change of the CO_2 flux with low uncertainty. Present sampling is inadequate (Wanninkhof et al., 2012), and needs to increase. The most effective way to determine how to increase sampling efforts needs to be determined. The main aim of this dissertation is to investigate sampling strategies for measuring Southern Ocean CO_2 flux that can resolve the seasonal sea-air CO_2 flux within the natural variability of the system. The goal is to find a sampling strategy that optimally balances sampling effort and sampling uncertainty. Methods that can determine the efficacy of sampling strategies to achieve this goal are explored, and this allows the different sampling strategies to be compared. Simulated model data from an ocean biogeochemical model (ORCA2/PISCES) are used to create an idealised world and results from different sampling strategies can be applied to this world and compared. 5-day mean model output and daily resolution model output are both used in this study. The daily resolution model output is used to replicate a study done by Lenton et al. (2006) The genetic algorithm and radial basis function experiments use .he 5-day mean model output.

5.2 Model Data Analysis - Characteristics of Variability

In order to assess whether or not the simulated model data captures the large scale variability present in the observations, the two model data sets are analysed and compared to observation based on a a reference year 2000 from Takahashi et al (2009). The analysis shows that the simulated data does capture the general variability present in the observations. The main difference between the three datasets are that the 5-day mean model output results in little outgassing, and therefore overestimates the role of the Southern Ocean as a atmospheric CO_2 sink. The 5-day model output data shows the western Weddell gyre and Scotia sea as a strong drawdown region

whereas the observational data (Takahashi et al., 2009) and the daily resolution model output show these regions as strong sources for atmospheric CO_2.

The 5-day mean data is smoother than the daily resolution data but has a larger amplitude in the seasonal cycle. The difference is due to an error in the model run for the 5 day data. The daily resolution data gives an amplitude in the mean annual cycle which is more similar to the observational data. It should be noted that using the 5-day mean model data does not have a negative effect on the experiments that were used to determine the effectiveness of the sampling strategies. Repeating the experiments on the 5 day mean data obtained from the daily model run was not possible due to time constraints. Similarly, attempting to do the experiments on the daily resolution data would not be possible due to computational cost.

The daily resolution model data (Figure 4.3) reproduces results from studies conducted around the Weddel Gyre, the Ross Sea and coastal Antarctica which found the northeastern Weddell Gyre to be a strong source of atmospheric carbon and the coastal Antarctic is a strong sink (Arrigo et al., 2008; Bakker et al., 1997; Bakker et al., 2008; Robertson and Watson, 1995).

Analysis of the variability of the daily resolution data (Figure 4.9) showed a small contribution by inter-annual variability. This contribution is smaller than the contribution from inter-annual variability seen in Lenton et al. (2006). This has implications with regard to determining the effectiveness of the sampling strategies if using the same definitions as Lenton et al (2006). The method used by Lenton et al (2006) and in this study compares uncertainty due to inter-annual variability with uncertainty from sampling and uses the comparison to quantify the effectiveness of sampling strategies (Section 4.2.4).

The contribution variability by the annual cycle and the non-seasonal variability (Figure 4.10) are of the same magnitude. The climatological seasonal cycle is responsible for similar CO_2 flux variance as the spatial variability. Areas of low and high seasonal and non-seasonal variability generally correspond (Figure 4.10). This has important implications for the effectiveness of different sampling strategies. A sampling strategy that mostly selects locations in areas where non-seasonal and seasonal variability are large will perform better than a sampling strategy that is regularly spaced spatially and temporally.

The regions in which the model shows the greatest variability (Figure 4.9) are the eastern Weddell Gyre (Scotia Sea) and the coastal Southern Ocean. According to observational studies by Bakker et al. (1997, 2008), the eastern Weddell Gyre is an outgassing region that has large variability due to Circumpolar Deep Water bringing CO_2 rich water to the winter mixed layer. Significant amounts of CO_2 are released when the ice cover recedes. Variability is also affected by the biological drawdown of CO_2 in the region. The shelf region of the Southern ocean is a strong sink for atmospheric CO_2 due to summer biological productivity, intense winds and efficient ventilation (Arrigo, 2008). The time scales at which these occur result in a high variability in CO_2 fluxes. The model investigation undertaken into the daily resolution model output shows similar results to observational and model studies. This means that sampling locations resulting from this study are not based solely on the the data from a single model.

Having defined the annual cycle as the signal and the non-seasonal variability as the noise, the signal-to-noise (SNR) ratio of the Southern Ocean CO_2 flux is determined as in Lenton et al. (2006). The SNR is generally higher in the 5-day mean resolution. This is thought to be due to averaging over 5 days which would smoothe the data and removes part of the synoptic non-seasonal variability. The higher amplitude in the mean annual cycle for the 5-day resolution model output also contributes to the higher SNR due to larger values for the annual cycle variability.

Irrespective of this potential shortcoming, the SNR is less than 1 in most of the Southern Ocean for both the daily resolution model output and the 5-day mean model output datasets. In regions where there is a low signal-to-noise ratio (SNR$<$1), it would be difficult to separate the seasonal cycle variability from the non-seasonal variability. However, the regions that have large variability in the mean annual cycle are the same as the regions with large non-seasonal variability (Figure 4.10). This is important in determining which type of sampling strategy to use.

Sampling regularly (4 times a year) will not be the best sampling strategy due to the SNR value being less than 1 in most of the Southern Ocean. This suggests the occurrence of shorter timescale events contributing to the variability. The objective sampling strategy selects locations in areas where both the non-seasonal and seasonal variability is large.

The seasonal cycle reproducibility of each pixel is also calculated as in

Thomalla et al (2011). Areas with low seasonal cycle reproducibility are areas where the seasonal cycle is not key in attempting to explain the overall variance, i.e. the annual seasonal cycle is different to the climatological seasonal cycle. This corresponds strongly to the regions with a SNR of less than 1.

An ideal sampling strategy would be to sample in areas with high variability and low SNR. A non-structured, objective sampling strategy would be more suitable in achieving this ideal sampling strategy than a regularly spaced sampling strategy. The ORCA2/PISCES biogeochemical model is used to reproduce the findings of Lenton et al (2006) and explore regularly spaced sampling strategies. The results where then used as a basis for the exploration of non-structured sampling strategies. The outcomes from these two approaches (structured and unstructured) to sampling strategies will now be discussed as well as their implications.

5.3 Design of observational strategy using regular grids (structured approach)

The structured approach uses regular grids (regular sampling in time and space) to sample the Southern Ocean CO_2 flux. The daily resolution model output is used for this approach.

Uncertainty due to inter-annual variability, sampling error and total sampling uncertainty are calculated according to the definitions in Lenton et al. (2006). The results are summarised in table 4.3 in the results section. Uncertainty is expressed as variance. The three terms Lenton et al (2006) use are:

1. Uncertainty due to inter-annual variability:

$$2\sigma_{inter-annual}* = 2\sigma_{inter-annual}/\sqrt{10} \tag{5.1}$$

2. Sampling error:

$$2\sigma_{sampling}* = 2\sigma_{sampling}/\sqrt{10} \tag{5.2}$$

3. Total sampling uncertainty:

$$2\sigma_{total} = \sqrt{(2\sigma)^2_{inter-annual} * + (2\sigma)^2_{sampling}*} \tag{5.3}$$

Using the current sampling efforts, from Lenton et al. (2006), of sampling regularly every 3 month, every 2° in latitude and every 60° in longitude, results in a mean CO_2 uptake of 0.352±0.11 PgC.yr^{-1} compared to sampling at the model resolution in which the uptake is 0.348±0.07 PgC.yr^{-1}. This sampling strategy is an overestimate of the actual current sampling efforts.

In this study, all sampling strategies have a meridional resolution of 2°. This is the model resolution and thus a model constraint. According to Lenton et al. (2006), sampling at every degree instead of every 3° in latitude does not change the estimate of sampling uncertainty. Also, according to Lenton et al. (2006), sampling with or without sub-grid scale variability does not change the estimates in uptake or sampling error. This is in agreement with Mahadevan (2004) and Resplandy et al. (2009) who suggest that large scale CO_2 flux variability is not affected by mesoscale and sub-mesocale processes. This implies that variability is not affected by mesoscale variability. Sampling strategies are differentiated according to the daily and meridional sampling resolution.

A number of criteria are used in this study to determine the suitability of a sampling strategy in calculating the net Southern Ocean CO_2 uptake. These criteria use the definitions in Equations (5.1) - (5.3).

The first criterion that is used to assess whether or not a sampling strategy can produce suitable flux estimates is whether the sampling error of the strategy is less than the uncertainty introduced by the inter-annual variability, i.e.

$$2\sigma_{sampling}* < 2\sigma_{inter-annual}* \tag{5.4}$$

This is not considered a good criterion to assess sampling strategy efficiency but it is useful to identify at which sampling resolution this occurs.

The results show that sampling regularly every 4 months (3 times per year), at every 2° in latitude and every 30° in longitude is sufficient to make the total sampling uncertainty to approximately equal to inter-annual variability. The sampling uncertainty at this resolution is 0.66 PgC.yr^{-1} (see Figure 5.1).

The second criterion used to determine the effectiveness of a sampling strategy when calculating suitable flux estimates is to test if the sampling strategy can calculate the net CO_2 exchange with an uncertainty approximately equal (equal to within one significant figure) to the uncertainty when calculating net flux at the model resolution (i.e. uncertainty due to inter-

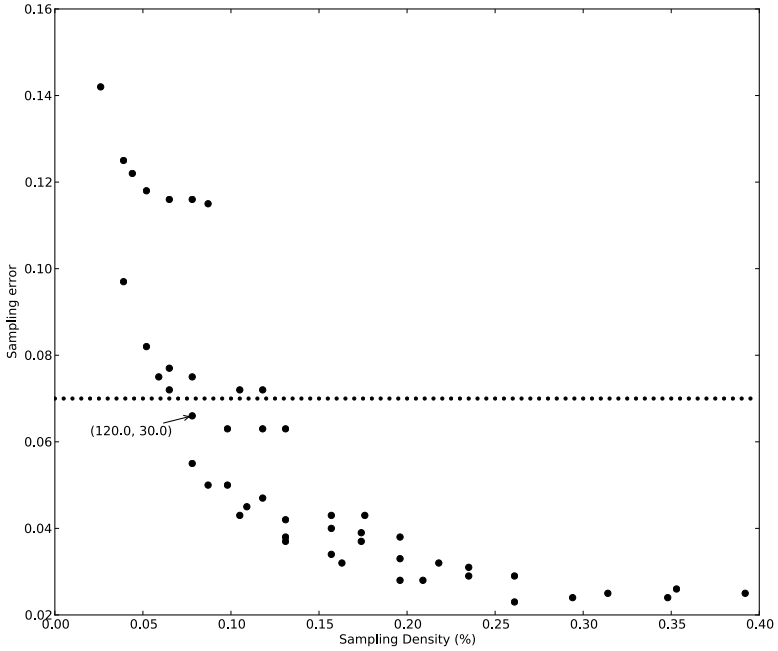

Figure 5.1: Plot of the sampling error obtained when sampling the mean annual cycle at various sampling frequencies highlighting the sampling strategy sampling 3 times a year, every 30°. The dotted line represents the uncertainty due to inter-annual variability.

annual variability). The second criteria is expressed as:

$$2\sigma_{sampling} * << 2\sigma_{inter-annual} *$$
$$2\sigma_{total} \approx 2\sigma_{inter-annual} * \tag{5.5}$$

Figure 5.2 shows that sampling regularly every 2 months, at 2° in latitude and every 24° in longitude is sufficient determine the net CO_2 uptake with a total uncertainty approximately equal to sampling at the model resolution.

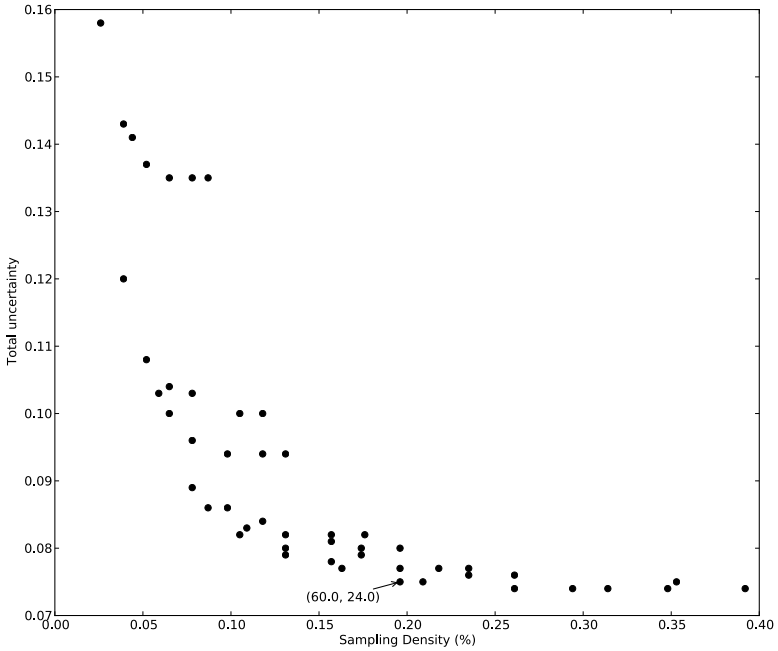

Figure 5.2: Plot of the total sampling uncertainty obtained when sampling the mean annual cycle at various sampling frequencies highlighting the total sampling uncertainty obtained when sampling 6 times a year, every 24°.

Reducing the total uncertainty to below 20% of the regional CO_2 flux estimate could not be achieved because the uncertainty due to inter-annual variability is already 20% of the regional estimate. Although not ideal, it was decided to use the global mean ocean uptake. According to Takahashi et al. (2009) the global mean ocean uptake is 2.0±0.1 PgC.yr^{-1}.

In order to keep the sampling uncertainty below 15% of the global mean ocean uptake, sampling regularly every 3 months, at every 1° in latitude and

every 40° in longitude is sufficient (see Figure 5.3). None of the sampling strategies can constrain the total uncertainty to less than 10% of the global mean ocean uptake, however, sampling every 3 months, at 1° in latitude and 40° in longitude is also sufficient to constrain the total uncertainty to less than 15% of the global mean ocean uptake.

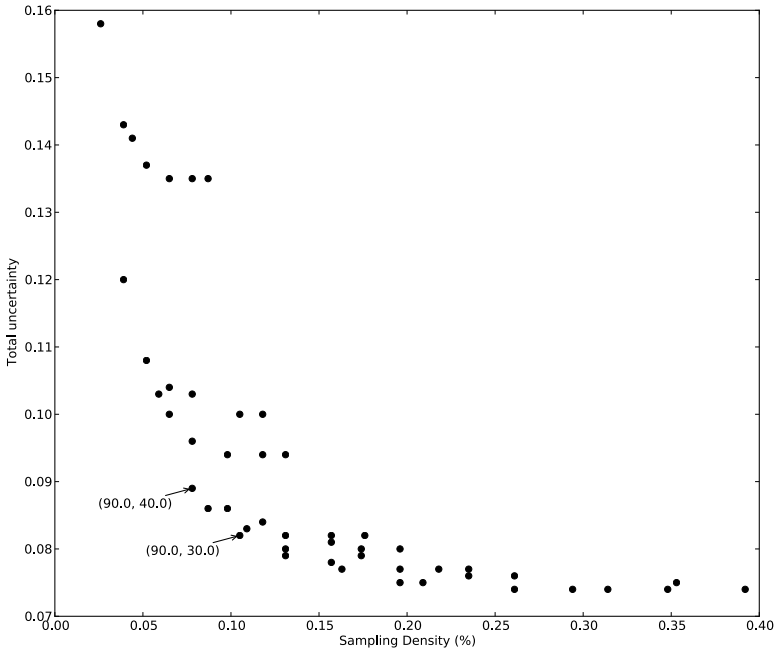

Figure 5.3: Plot of the total sampling uncertainty obtained when sampling the mean annual cycle at various sampling frequencies highlighting the total sampling uncertainty obtained when sampling 4 times a year, every 30°, and 4 times a year, every 40°

Figure 5.3 shows that sampling regularly every 3 months, at every 2° in latitude and every 30° in longitude is a acceptable compromise between these different criteria. This is because the total sampling uncertainty is at 0.082 PgC.yr^{-1} which is 0.012 PgC.yr^{-1} greater than sampling at the model resolution.

To design an observational strategy for the Southern Ocean uptake of CO_2, Lenton et al. (2006) applied two-dimensional Fourier transforms and signal-to-noise ratios to simulated model air-sea CO_2. In the present study

an exhaustive search (see Section 3) on the seasonal cycle of simulated model data is used to compare the sampling error obtained at each different sampling resolutions (see Figure 4.16). The results shown in Figure 4.16 of the exhaustive search show that sampling more than 3 times a year and more than 6 times in longitude (every 60°) reduces the uncertainty due to sampling error.

To compare the results with that of Lenton et al. (2006), the relevant sampling strategies are repeated using the simulated model data that has been remapped to a 360° meridional resolution. The results do not differ more than ± 0.001 PgC.yr^{-1}.

Lenton et al. (2006) tested the present sampling strategy on an ocean biogeochemical model driven with NCEP-R1. They calculated a mean annual uptake for the Southern Ocean of 0.6 ± 0.2PgCyr^{-1}.

Lenton et al. (2006) designed and proposed a sampling strategy that sampled the Southern Ocean at a resolution of 3 months, 3° in latitude and 30° in longitude. They tested their proposed sampling strategy on the same simulated model data and calculated a mean annual uptake for the Southern Ocean of 0.6 ± 0.1PgCyr^{-1}. The error introduced by sampling at their proposed sampling strategy was half of that of the current sampling strategy.

The mean annual uptake calculated in this study is less than Lenton et al., 2006, but still falls within the range of estimates from model and observational studies (0.2 to 0.8 PgC.yr^{-1} for the 1990s).

Inter-annual variability is substantially greater in Lenton et al. (2006). The primary difference between the calculations of uncertainty in Lenton et al., 2006 lie in the difference between the inter-annual variability in the model used by Lenton et al., 2006 and the model used in this study.

The difference in models is as follows:

The total sampling uncertainty of the present (current sampling effort) Southern Ocean sampling strategy calculated in this study is less than that calculated by Lenton et al. (2006). However the sampling error is significantly larger than the uncertainty due to inter-annual variability. The results in this study confirm that sampling at the present sampling strategy cannot result in flux calculations that are as accurate or precise as sampling at the model resolution.

Using the proposed sampling strategy of Lenton et al. (2006), sampling 4

times a year, at every 30°, the total sampling uncertainty is larger than the uncertainty obtained by sampling at the model resolution, unlike the results obtained by Lenton et al. (2006). This is due to the greater inter-annual variability in Lenton et al. (2006) contributing more to the total sampling uncertainty than in this study. Sampling using the proposed Lenton et al. (2006) sampling strategy on the daily resolution model output does not result in a situation where the criteria for Equation 5.5 are met.

In order to meet the same criteria (Equation 5.5) as Lenton et al. (2006), i.e. sampling every 2 months, every 24°, would result in 90 north-south sampling sections per year. This is almost twice the number of the 48 sampling sections that would be used by implementing the proposed Lenton et al., 2006 sampling strategy.

Sampling every 3 months, at every 1° in latitude and every 40° in longitude (see Table 5.1) is the equivalent of 24 north-south sampling sections per year and results in an uncertainty of within 15% of the global mean CO_2 uptake. Although the sampling strategy increases the uncertainty by more than 0.019 compared to sampling at the model resolution, the number of sampling sections made per year is half that of Lenton et al., 2006. If the results from Lenton et al., 2006 are calculated to the same accuracy as this study (i.e. using uncertainty due to inter-annual variability as 0.12, sampling error as 0.068, and rounding off to two significant figures), then the uncertainty is closer to 0.139 which increases the uncertainty of 0.1 $PgC.yr^{-1}$ by 0.019, the same increase that sampling every 3 months at every 40° in longitude does in this study. Therefore it is proposed to sample every 3 months, at every 2° in latitude and every 40° in longitude as a basis to optimize effort and reduce uncertainty.

Table 5.1: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.055 and the total sampling uncertainty is 0.089

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.189
1999	0.199	0.199	0.182
2000	0.26	0.26	0.157
2001	0.224	0.224	0.2
2002	0.412	0.412	0.162
2003	0.419	0.419	0.172
2004	0.363	0.363	0.193
2005	0.426	0.426	0.172
2006	0.443	0.443	0.165
2007	0.522	0.522	0.154
1998-2007	0.348±0.07	0.347±0.07	0.175

The advantages of investigating sampling strategies with a regular grid is that it is easier to quantify uncertainty as the mean. Each of the permutations of each grid resolution can be used to quantify uncertainty. Regularly spaced north to south sampling lines are effective for ship based sampling strategies but according to the model, large areas in the Southern Ocean have high mean annual cycle and non-seasonal variability. Using regular spaced sampling strategies are likely to result in areas of low variability being oversampled and areas of high variability being under sampled. This could be in a waste of sampling effort and invokes the need for a scale sensitive sampling strategy.

5.4 Design of observational strategy using Genetic Algorithm

It is contended that optimized sampling strategies that do not adhere to a regularly spaced grid should be able to more efficiently sample the model data set. An optimized strategy would sample preferentially in areas of high variability and less in areas of low variability.

With these optimized strategies, the sampling frequency is dependent on the number of sampling locations allowed by the GA to sample the data set. Each sampling frequency requires a number of permutations in order to compare the sampling frequency in the same way as done by the regularly spaced data. The result should be a group of solutions that represent the data optimally.

A genetic algorithm is used to characterise sampling strategies that are not based on regular sampling in time or space. Genetic algorithms are relatively straightforward to implement and are designed to find successively better solutions than a random solution. Therefore they meet the needs of this study. The limitations of Genetic Algorithms (Section 2.8.6) are thought to have a low impact since the fitness function can be very simple and because local optima are sufficiently good solutions

In order to allow the genetic algorithm to perform efficiently, some compromises are made (Section 3.10.1) with regards to sampling the data. This allows for many solutions to be obtained. Firstly, the data is regridded onto a 640 day, by 64° latitude by 128° longitude grid. Then the last value is removed to obtain a data set with a size of 2^n-1. As explained in the methods section, this reduces the number of possible locations by half. Secondly, sea-air fluxes over land (which are 0) are included in possible sampling locations. Thirdly, it is possible that the genetic algorithm can select the same location twice. Better solutions would avoid sampling the same location twice as the additional sampling point can be utilised better at another location .

The most important aspect of the genetic algorithm is the fitness function f (Equation 5.6). With the population defined as the data being sampled, and the sample being defined as the location selected by the genetic algorithm, the fitness function selected for the study calculates the sum of the absolute difference between the data mean and the sample mean, and the difference between the data standard deviation and the sample standard deviation i.e.

$$f = (\bar{x}_{population} - \bar{x}_{sample}) + (\sigma^2_{population} - \sigma^2_{sample}) \tag{5.6}$$

The genetic algorithm aims to select a solution, or set of locations, that minimise this fitness function.

As shown in experiment 1 (Section4.4.2), the genetic algorithm is set up to estimate the decadal mean. The decadal mean is sampled with varying sampling densities. The genetic algorithm selects a number of locations that

minimise Equation 5.6. These locations are suggested as good locations for moorings. Using different sampling densities one can determine how many moorings one would need to capture the annual means with a reasonable degree of certainty, as well as the ideal locations. These sampling locations are then tested on the annual mean for each year in the data to determine whether the sampling strategy selected by the genetic algorithm can accurately capture the inter-annual variability over the years and the spatial variability within the annual means. The results are summarised in table 4.4.

The mean fitness values (from Equation. 5.6) of the sampling solutions are <5% of the annual mean. This indicates that these solutions are a good representation of the overall data.

In order to quantify the sampling uncertainty, and therefore the total uncertainty generated by sampling with this many solutions, ten solutions are selected by the genetic algorithm. The tables generated by these experiments are available in Appendix B.

The sampling error of the annual means is calculated. This sampling error can be compared with the sampling errors calculated by using sampling strategies based on regular grids.

Table 4.4 shows a summary of the sampling error and total sampling uncertainty for Experiment 1. Sampling the annual mean with 50 locations results in a mean CO_2 flux of 1.82±0.15. Sampling the annual mean with 100 locations results in a mean CO_2 flux of 1.81±0.10. Sampling the annual mean with 500 locations results in a mean CO_2 flux of 1.82±0.08 (0.077). Sampling with this many (optimized) locations returns a mean CO_2 flux with a total sampling uncertainty approximately the same as that of sampling at the model resolution. Sampling the annual mean with 1000 locations results in a mean CO_2 flux of 1.81±0.07. In spite of sampling with double the number of locations, the total sampling strategy is approximately the same as sampling with 500 locations. Once the sampling error is small, the total uncertainty cannot be reduced without a reduction in the inter-annual variability.

The genetic algorithm is then set up to sample the mean annual cycle (Experiment 2: Section 4.4.2). Locations are selected by the genetic algorithm. These locations are tested on the data for each year to determine whether or not the sampling strategy can capture the inter-annual variabil-

ity as well the spatial and seasonal variability within each year (the standard deviation of the data). The results are summarised in table 5.5.

In order to quantify the uncertainty, solutions are selected by the genetic algorithm. The sampling error of the annual means is calculated. This sampling error can be compared to the sampling errors calculated by using sampling strategies based on regular grids.

The sampling errors show a significant improvement over sampling with regular grids. The sampling error obtained from sampling with 1000 locations is significantly smaller (0.089 vs 0.073) than sampling regularly 5 times a year, every 2° in latitude and every 20° in longitude. Sampling regularly at this time and space frequency would require 4050 locations.

Sampling the annual mean with 100, 2000, 5000, and 7000 locations results in a mean flux of 1.81±0.0725, 1.81±0.0712, 1.82±0.0703, 1.81±0.0701 respectively. The total sampling uncertainty is insensitive to the number of locations because the sampling error is already low in comparison to inter-annual variability.

Using fewer sampling locations, the optimized sampling strategies still return levels of uncertainty similar to sampling strategies that use regular grids. Even using 1000 locations, the sampling uncertainty is approximately the same as sampling at the model resolution. The sampling error is low enough using 1000 locations to meet the conditions in Equation 5.5, i.e. the sampling error is sufficiently smaller than inter-annual variability such that total sampling error is approximately equal to inter-annual variability. Using more sampling locations does not result in a significant reduction in the total sampling error. This shows that the use of optimized sampling strategies chosen by a genetic algorithm is more effective at reducing the uncertainty of Southern Ocean CO_2 flux sampling than using regular grids.

This is expected from the analysis of the model data which showed that certain regions experienced greater variability than others (Figure 4.9, 4.10, 4.11 and 4.24). Sampling strategies that sample more in these areas and less in areas that do not have a lot of intraseasonal variability are more efficient at estimating the Southern Ocean CO_2 flux with a low uncertainty.

5.5 Design of observational strategy using iterative maximum error reduction

As it has been shown, is i spossible to sample with fewer sampling locations using optimized, non-grid sampling strategies. As a basis to further optimize unstructured sampling of high variability areas, RBFs are now used to return a highly optimized sampling strategy that will enable an investigation addressing which areas are the most important to sample.

A radial basis function interpolation is used to interpolate the sample data and determine where the greatest error is. As stated in Section 3.11, starting with four locations, the algorithm finds the location with the greatest error and then does the next interpolation with that location included until the number of locations used for the interpolation reaches the number required to fulfil the required sampling density. For the Experiment 2 in the Genetic Algorithm Section, it is decided that 1000 optimized locations is sufficient to sample the Southern Ocean CO_2 flux with low uncertainty. The Root Mean Squared Error (RMSE) between the values predicted by the interpolation and the values observed in the model is 1.8 mmolC.m$^{?2}$ .day$^{?1}$.

Plotting the chosen locations against standard deviation along all axes (Figure 4.24 and 4.25) shows that this method selects locations where the variability is highest. Most of the locations are along the Antarctic Continent and the South American Continent. These are strong uptake regions in the summer months (Figure 4.7a and 4.8a). More locations are selected in the summer months as many locations are covered by ice in winter. There are also many locations in the Scotia Sea, which is a strong outgassing region (Figure 4.3). The goal of this radial basis function method is to perform an interpolation using the information from locations that will minimise the difference (RMSE) between the sample data and the population data. To do this it must capture the variability in the data. In order to minimise the error between the interpolated data set and the original data set, areas with extreme flux values will be selected. The reasons for high variability in these areas are due to variability in sea-ice interaction, productivity and upwelling (Arrigo et al., 2008; Bakker et al., 1997; Bakker et al., 2008; Robertson and Watson, 1995). Sea ice retreat leads to a strong outgassing in the eastern Weddel Gyre as the CO_2 rich Circumpolar Deep Water comes into contact with the atmosphere (Bakker et al. 1997, 2008). This is counteracted by biological drawdown which makes this region highly variable as

well as extreme. In the coastal Southern Ocean, biological drawdown due to increased productivity results in this region being a strong CO_2 sink in summer whereas in winter there is no flux due to sea-ice-coverage (Robertson et al, 2005;Arrigo et al. 2008).

This method (using RBF interpolation to successively reduce RMSE) is good because it finds specific locations to get the most efficient sample. The disadvantage is that it is dependent on the model data, which may not represent the real world well.

5.6 Further work and issues

The methods employed in this study would benefit from being implemented on a finer resolution data-set in both time and space. Investigation into the appropriate fitness functions to determine the optimized solutions from the Genetic Algorithm would be advantageous. A platform specific approach, similar to Heaney et al. (2007), would benefit observational oceanography. Different combinations of sampling platforms could be tested on fine resolution model data in order to determine which combination is the best.

6 Conclusion

Using an ocean biogeochemical model and methods developed by both Lenton et al. (2006) and this study, regularly spaced sampling strategies are investigated in order to determine a sampling strategy that returns the same level of uncertainty as sampling at model resolution.

This level of uncertainty is determined to be attained when the uncertainty due to sampling is significantly less than the uncertainty due to interannual variability, such that, the total sampling uncertainty is approximately equal to the uncertainty due to interannual variability.

It is determined that sampling regularly every 3 months, at every 2° in latitude and every 40° in longitude is sufficient to estimate the annual mean CO_2 uptake with a low uncertainty.

This sampling strategy is infeasible due to the limitations of ship availability and high intraseasonal variability. However, using a regular grid approach ensures that a sampling strategy is likely to perform robustly should Southern Ocean behaviour differ from the model.

In order to determine if sampling effort can be reduced by using a targeted sampling strategy instead of a regular grid approach, a genetic algorithm is utilized to determine targeted or optimal sampling solutions.

Unlike the regular grid approach, the sampling strategy elected by the genetic algorithm resemble a sampling strategy that could be used by a multi-platform observation network including platforms such as gliders and floats.

For various sampling densities, a number of targeted sampling strategies are selected by the genetic algorithm. The sampling uncertainties of these sampling strategies are calculated and return lower uncertainties than sampling strategies that rely on a regular grid. Sampling the data using 1000 optimized sampling locations is enough to estimate the annual mean CO_2 flux with a low uncertainty. Beyond this, the extra effort does not return a significant gain.

Although the method allows more freedom in selecting sampling locations as well as requiring less sampling effort than using a regular grid approach, real-life application of this method would require a thorough understanding of the variables being measured and a good confidence in the model used to generate the sampling strategy.

The genetic algorithm approach indicates a sufficient sampling density to estimate the annual mean CO_2 uptake. In order to suggest the best possible locations to sample at, a method is developed that calculates the maximum error between the data and a set of sample locations. The method uses a radial basis function (RBF) interpolation to successively find optimized sampling locations. The sampling locations are investigated to determine the locations that are most important to sample when attempting to sample with low sampling densities.

The sampling locations returned from this method closely follow areas that exhibit a high variability in the seasonal cycle and non-seasonal variability. Using RBF interpolation to successively find optimized sampling locations can therefor be considered a good tool to develop a sampling strategy for estimating mean annual CO_2 flux in the Southern Ocean.

The methods described to design a sampling strategy have their advantages and disadvantages. This study shows that, for a given model dataset, it is possible to reduce both sampling effort and sampling uncertainty by targeting sampling locations.

References

Arrigo, Kevin R, Gert van Dijken, and Matthew Long (2008). "Coastal Southern Ocean: A strong anthropogenic CO2 sink". In: *Geophysical Research Letters* 35.21, p. L21602.

Bakker, D.C.E., H.J.W. De Baar, and U.V. Bathmann (1997). "Changes of carbon dioxide in surface waters during spring in the Southern Ocean". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 44.12, pp. 91 –127.

Bakker, DCE et al. (2008). "A rapid transition from ice covered CO 2–rich waters to a biologically mediated CO 2 sink in the eastern Weddell Gyre". In: *Biogeosciences* 5.5, pp. 1373–1386.

Bracewell, Ronald Newbold and RN Bracewell (1986). *The Fourier transform and its applications*. Vol. 31999. McGraw-Hill New York.

Broecker, Wallace S and T-H Peng (1974). "Gas exchange rates between air and sea1". In: *Tellus* 26.1-2, pp. 21–35.

Budin, Leo, Marin Golub, and Andrea Budin (2010). "Traditional techniques of genetic algorithms applied to floating-point chromosome representations". In: *sign* 1, pp. 11–52.

Chierici, Melissa et al. (Apr. 2009). "Algorithms to estimate the carbon dioxide uptake in the northern North Atlantic using shipboard observations, satellite and ocean analysis data". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8-10, pp. 630–639.

Danckwerts, PV (1951). "Significance of liquid-film coefficients in gas absorption". In: *Industrial & Engineering Chemistry* 43.6, pp. 1460–1467.

Falkowski, Paul et al. (2000). "The global carbon cycle: a test of our knowledge of earth as a system". In: *science* 290.5490, pp. 291–296.

Floater, Michael S and Armin Iske (1996). "Multistep scattered data interpolation using compactly supported radial basis functions". In: *Journal of Computational and Applied Mathematics* 73.1, pp. 65–78.

Friedlingstein, By P et al. (2003). "How positive is the feedback between climate change and the carbon cycle ?" In: pp. 692–700.

Friedrich, T. and a. Oschlies (Oct. 2009). "Basin-scale pCO 2 maps estimated from ARGO float data: A model study". In: *Journal of Geophysical Research* 114.C10, p. C10012.

Garçon, Véronique C et al. (1992). "Gaining insight into the seasonal variability of CO2 at ocean station P using an upper ocean model". In: *Deep Sea Research Part A. Oceanographic Research Papers* 39.6, pp. 921–938.

Gershenfeld, Neil A (1999). *The nature of mathematical modeling*. Cambridge university press.

Gruber, Nicolas and Charles D Keeling (2001). "An improved estimate of the isotopic air-sea disequilibrium of CO2: Implications for the oceanic uptake of anthropogenic CO2". In: *Geophysical Research Letters* 28.3, pp. 555–558.

Gurney, Kevin Robert et al. (Feb. 2002). "Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models." In: *Nature* 415.6872, pp. 626–30.

Ho, David T et al. (2011). "Toward a universal relationship between wind speed and gas exchange: Gas transfer velocities measured with 3He/SF6 during the Southern Ocean Gas Exchange Experiment". In: *Journal of Geophysical Research: Oceans (1978–2012)* 116.C4.

Hoppema, Mario, Michel H.C Stoll, and Hein J.W de Baar (2000). "CO2 in the Weddell Gyre and Antarctic Circumpolar Current: austral autumn and early winter". In: *Marine Chemistry* 72.24, pp. 203 –220.

Indermühle, A et al. (1999). "Holocene carbon-cycle dynamics based on CO2 trapped in ice at Taylor Dome, Antarctica". In: *Nature*.

Keeling, CD et al. (1995). "Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980". In: *Nature* 375.6533, pp. 666–670.

Keeling, Charles D. et al. (Feb. 2011). "Evolution of natural and anthropogenic fluxes of atmospheric CO2 from 1957 to 2003". In: *Tellus B* 63.1, pp. 1–22.

Keeling, Ralph F. and Stephen R. Shertz (Aug. 1992). "Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle". In: *Nature* 358.6389, pp. 723–727.

Le Quéré, Corinne et al. (2007). "Saturation of the Southern Ocean CO2 sink due to recent climate change". In: *science* 316.5832, pp. 1735–1738.

Lenton, a (Apr. 2009). "Strategies for high-latitude northern hemisphere CO2 sampling now and in the future". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8-10, pp. 523–532.

Lenton, Andrew, Richard J. Matear, and Bronte Tilbrook (Nov. 2006). "Design of an observational strategy for quantifying the Southern Ocean uptake of CO 2". In: *Global Biogeochemical Cycles* 20.4, pp. 1–11.

Lenton, Andrew, Laurent Bopp, and Richard J Matear (2009). "Strategies for high-latitude northern hemisphere CO2 sampling now and in the future". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8, pp. 523–532.

Lenton, Andrew et al. (2013). "Sea-air CO2 fluxes in the Southern Ocean for the period 1990–2009". In: *Biogeosciences Discussions* 10, pp. 285–333.

Liss, Peter S and Liliane Merlivat (1986). "Air-sea gas exchange rates: Introduction and synthesis". In: *The role of air-sea exchange in geochemical cycling*. Springer, pp. 113–127.

Mahadevan, A. (2004). "Mesoscale variability of sea surface pCO2 : What does it respond to?" In: *Global Biogeochemical Cycles* 18.1.

Matear, R. J. and a. Lenton (Nov. 2008). "Impact of Historical Climate Change on the Southern Ocean Carbon Cycle". In: *Journal of Climate* 21.22, pp. 5820–5834.

Matsumoto, K. (2004). "Evaluation of ocean carbon cycle models with data-based metrics". In: *Geophysical Research Letters* 31.7, p. L07303.

McCready, MJ and TJ Hanratty (1985). "Effect of air shear on gas absorption by a liquid film". In: *AIChE journal* 31.12, pp. 2066–2074.

McNeil, Ben I et al. (2003). "Anthropogenic CO2 uptake by the ocean based on the global chlorofluorocarbon data set". In: *Science* 299.5604, pp. 235–239.

McNeil, Ben I. et al. (Aug. 2007). "An empirical estimate of the Southern Ocean air-sea CO 2 flux". In: *Global Biogeochemical Cycles* 21.3, pp. 1–16.

Mémery, Laurent et al. (Jan. 2002). "The relevant time scales in estimating the airsea CO2 exchange in a mid-latitude region". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 49.11, pp. 2067–2092.

Metzl, N., B. Tilbrook, and A. Poisson (1999). "The annual fCO2 cycle and the airsea CO2 flux in the sub-Antarctic Ocean". In: *Tellus B* 51.4, pp. 849–861.

Metzl, N. et al. (2006). "Summer and winter airsea CO2 fluxes in the Southern Ocean". In: *Deep Sea Research Part I: Oceanographic Research Papers* 53.9, pp. 1548 –1563.

Metzl, Nicolas (2009). "Decadal increase of oceanic carbon dioxide in Southern Indian Ocean surface waters (1991–2007)". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8, pp. 607–619.

Monteiro, P et al. (2010). "A global sea surface carbon observing system: Assessment of changing sea surface CO 2 and air-sea CO 2 fluxes". In: *Proceedings of the OceanObs* 9.

Nightingale, PD (2009). "Air-sea gas exchange". In: *Geophysical Monograph Series* 187, pp. 69–97.

Orsi, Alejandro H, Thomas Whitworth, and Worth D Nowlin (1995). "On the meridional extent and fronts of the Antarctic Circumpolar Current". In: *Deep Sea Research Part I: Oceanographic Research Papers* 42.5, pp. 641–673.

Paillard, D and F Parrenin (2004). "The Antarctic ice sheet and the triggering of deglaciations". In: *Earth and Planetary Science Letters*.

Petit, JR et al. (1999). "Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica". In: *Nature*.

Pickard, George L and William J Emery (1990). *Descriptive physical oceanography: an introduction*. Elsevier.

Rangama, Y. (2005). "Variability of the net airsea CO 2 flux inferred from shipboard and satellite measurements in the Southern Ocean south of Tasmania and New Zealand". In: *Journal of Geophysical Research* 110.C9, p. C09005.

Raven, John et al. (2005). *Ocean acidification due to increasing atmospheric carbon dioxide*. The Royal Society.

Resplandy, L. et al. (Mar. 2009). "Impact of submesoscale variability in estimating the air-sea CO 2 exchange: Results from a model study of the POMME experiment". In: *Global Biogeochemical Cycles* 23.1, GB1017.

Revelle, Roger and Hans E Suess (1957). "Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO2 during the past decades". In: *Tellus* 9.1, pp. 18–27.

Rio, MH, S Guinehut, and G Larnicol (2011). "New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry, and in situ measurements". In: *Journal of Geophysical Research: Oceans (1978–2012)* 116.C7.

Robertson, Jane E. and Andrew J. Watson (1995). "A summer-time sink for atmospheric carbon dioxide in the Southern Ocean between 88W and

80E". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 42.45, pp. 1081 –1091.

Sabine, CL et al. (2004). "The oceanic sink for anthropogenic CO2". In: *Science*.

Sarmiento, JL and M Bender (1994). "Carbon biogeochemistry and climate change". In: *Photosynthesis Research*.

Sarmiento, JL and N Gruber (2006). *Ocean biogeochemical dynamics*.

Schiller, a., S. E. Wijffels, and G. a. Meyers (Oct. 2004). "Design Requirements for an Argo Float Array in the Indian Ocean Inferred from Observing System Simulation Experiments". In: *Journal of Atmospheric and Oceanic Technology* 21.10, pp. 1598–1620.

Schimel, D S et al. (Nov. 2001). "Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems." In: *Nature* 414.6860, pp. 169–72.

Sigman, DM and EA Boyle (2000). "Glacial/interglacial variations in atmospheric carbon dioxide". In: *Nature*.

Sweeney, Colm et al. (2002). "Spatial and temporal variability of surface water pCO2 and sampling strategies". In: *A large-scale CO2 observing plan: In situ oceans and atmosphere, edited by: Bender, M., Doney, S., Feely, RA, et al., National Technical Information Service, Springfield, Virginia, USA*, pp. 155–175.

Takahashi, T et al. (Aug. 1997). "Global air-sea flux of CO2: an estimate based on measurements of sea-air pCO2 difference." In: *Proceedings of the National Academy of Sciences of the United States of America* 94.16, pp. 8292–9.

Takahashi, Taro and Colm Sweeney (2002). "Errors in sea-air CO 2 flux due to time-space ocean sampling strategies for sea-air pCO 2 difference". In: *A Large-Scale CO* 2, pp. 177–183.

Takahashi, Taro et al. (2002). "Global seaair CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 49.910, pp. 1601–1622.

Takahashi, Taro et al. (Apr. 2009). "Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8-10, pp. 554–577.

Thomalla, SJ et al. (2011). “Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean”. In: *Biogeosciences* 8.10, pp. 2849–2866.

Volk, Tyler and Martin I Hoffert (1985). “Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes”. In: *The Carbon Cycle and Atmospheric CO: Natural Variations Archean to Present*, pp. 99–110.

Wanninkhof, R et al. (2012). “Global ocean carbon uptake: magnitude, variability and trends”. In: *Biogeosciences Discussions* 9.8, pp. 10961–11012.

Wetzel, P. (2005). “Sea-to-air CO 2 flux from 1948 to 2003: A model study”. In: *Global Biogeochemical Cycles* 19.2, GB2005.

Whitley, Darrell (2001). “An overview of evolutionary algorithms: practical issues and common pitfalls”. In: *Information and software technology* 43.14, pp. 817–831.

1 Appendix A

Table A1: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.025 and the total sampling uncertainty is 0.074

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.094
1999	0.199	0.199	0.077
2000	0.26	0.259	0.074
2001	0.224	0.224	0.092
2002	0.412	0.412	0.082
2003	0.419	0.419	0.081
2004	0.363	0.363	0.075
2005	0.426	0.426	0.065
2006	0.443	0.443	0.076
2007	0.522	0.522	0.072
1998-2007	0.348±0.07	0.347±0.07	0.079

Table A2: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.026 and the total sampling uncertainty is 0.075

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.097
1999	0.199	0.199	0.079
2000	0.26	0.259	0.077
2001	0.224	0.224	0.106
2002	0.412	0.411	0.084
2003	0.419	0.419	0.085
2004	0.363	0.363	0.081
2005	0.426	0.426	0.066
2006	0.443	0.443	0.082
2007	0.522	0.522	0.076
1998-2007	0.348±0.07	0.347±0.07	0.083

Table A3: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.024 and the total sampling uncertainty is 0.074

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.091
1999	0.199	0.199	0.082
2000	0.26	0.26	0.067
2001	0.224	0.224	0.085
2002	0.412	0.412	0.077
2003	0.419	0.419	0.075
2004	0.363	0.363	0.068
2005	0.426	0.427	0.061
2006	0.443	0.443	0.072
2007	0.522	0.522	0.07
1998-2007	0.348±0.07	0.348±0.07	0.075

Table A4: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.029 and the total sampling uncertainty is 0.076

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.098
1999	0.199	0.199	0.093
2000	0.26	0.26	0.079
2001	0.224	0.224	0.103
2002	0.412	0.412	0.105
2003	0.419	0.419	0.087
2004	0.363	0.364	0.086
2005	0.426	0.427	0.09
2006	0.443	0.444	0.091
2007	0.522	0.522	0.082
1998-2007	0.348±0.07	0.348±0.07	0.092

Table A5: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.038 and the total sampling uncertainty is 0.08

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.137
1999	0.199	0.198	0.147
2000	0.26	0.259	0.133
2001	0.224	0.223	0.143
2002	0.412	0.412	0.109
2003	0.419	0.419	0.106
2004	0.363	0.363	0.105
2005	0.426	0.426	0.106
2006	0.443	0.443	0.097
2007	0.522	0.522	0.107
1998-2007	0.348±0.07	0.347±0.07	0.119

Table A6: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.043 and the total sampling uncertainty is 0.082

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.145
1999	0.199	0.199	0.124
2000	0.26	0.26	0.105
2001	0.224	0.224	0.16
2002	0.412	0.412	0.118
2003	0.419	0.419	0.156
2004	0.363	0.363	0.148
2005	0.426	0.426	0.126
2006	0.443	0.443	0.137
2007	0.522	0.522	0.133
1998-2007	0.348±0.07	0.347±0.07	0.135

Table A7: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 40 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.072 and the total sampling uncertainty is 0.1

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.237
1999	0.199	0.203	0.208
2000	0.26	0.264	0.202
2001	0.224	0.229	0.24
2002	0.412	0.417	0.231
2003	0.419	0.423	0.235
2004	0.363	0.368	0.231
2005	0.426	0.432	0.254
2006	0.443	0.447	0.191
2007	0.522	0.526	0.234
1998-2007	0.348±0.07	0.352±0.07	0.226

Table A8: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.024 and the total sampling uncertainty is 0.074

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.081
1999	0.199	0.199	0.077
2000	0.26	0.259	0.082
2001	0.224	0.224	0.083
2002	0.412	0.412	0.081
2003	0.419	0.419	0.068
2004	0.363	0.363	0.08
2005	0.426	0.426	0.07
2006	0.443	0.443	0.068
2007	0.522	0.522	0.068
1998-2007	0.348±0.07	0.347±0.07	0.076

Table A9: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.025 and the total sampling uncertainty is 0.074

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.085
1999	0.199	0.199	0.079
2000	0.26	0.259	0.085
2001	0.224	0.224	0.098
2002	0.412	0.411	0.083
2003	0.419	0.419	0.073
2004	0.363	0.363	0.086
2005	0.426	0.426	0.069
2006	0.443	0.443	0.074
2007	0.522	0.522	0.073
1998-2007	0.348±0.07	0.347±0.07	0.08

Table A10: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.023 and the total sampling uncertainty is 0.074

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.078
1999	0.199	0.199	0.082
2000	0.26	0.26	0.074
2001	0.224	0.224	0.075
2002	0.412	0.412	0.076
2003	0.419	0.419	0.062
2004	0.363	0.363	0.073
2005	0.426	0.427	0.065
2006	0.443	0.443	0.062
2007	0.522	0.522	0.065
1998-2007	0.348±0.07	0.348±0.07	0.071

Table A11: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.028 and the total sampling uncertainty is 0.075

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.086
1999	0.199	0.199	0.091
2000	0.26	0.26	0.086
2001	0.224	0.224	0.098
2002	0.412	0.412	0.105
2003	0.419	0.419	0.076
2004	0.363	0.364	0.093
2005	0.426	0.427	0.095
2006	0.443	0.444	0.084
2007	0.522	0.522	0.08
1998-2007	0.348±0.07	0.348±0.07	0.09

Table A12: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.037 and the total sampling uncertainty is 0.079

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.131
1999	0.199	0.198	0.148
2000	0.26	0.259	0.136
2001	0.224	0.223	0.14
2002	0.412	0.412	0.109
2003	0.419	0.419	0.098
2004	0.363	0.363	0.11
2005	0.426	0.426	0.113
2006	0.443	0.443	0.091
2007	0.522	0.522	0.103
1998-2007	0.348±0.07	0.347±0.07	0.118

Table A13: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.043 and the total sampling uncertainty is 0.082

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.14
1999	0.199	0.199	0.122
2000	0.26	0.26	0.112
2001	0.224	0.224	0.159
2002	0.412	0.412	0.121
2003	0.419	0.419	0.151
2004	0.363	0.363	0.156
2005	0.426	0.426	0.13
2006	0.443	0.443	0.135
2007	0.522	0.522	0.132
1998-2007	0.348±0.07	0.347±0.07	0.136

Table A14: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 45 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.072 and the total sampling uncertainty is 0.1

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.234
1999	0.199	0.203	0.209
2000	0.26	0.264	0.208
2001	0.224	0.229	0.241
2002	0.412	0.417	0.232
2003	0.419	0.423	0.23
2004	0.363	0.368	0.235
2005	0.426	0.432	0.256
2006	0.443	0.447	0.19
2007	0.522	0.526	0.236
1998-2007	0.348±0.07	0.352±0.07	0.227

Table A15: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.029 and the total sampling uncertainty is 0.076

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.11
1999	0.199	0.199	0.075
2000	0.26	0.259	0.088
2001	0.224	0.224	0.114
2002	0.412	0.412	0.084
2003	0.419	0.419	0.081
2004	0.363	0.363	0.1
2005	0.426	0.426	0.096
2006	0.443	0.443	0.094
2007	0.522	0.522	0.086
1998-2007	0.348±0.07	0.347±0.07	0.093

Table A16: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.031 and the total sampling uncertainty is 0.077

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.112
1999	0.199	0.199	0.079
2000	0.26	0.259	0.091
2001	0.224	0.224	0.126
2002	0.412	0.411	0.086
2003	0.419	0.419	0.085
2004	0.363	0.363	0.105
2005	0.426	0.426	0.095
2006	0.443	0.443	0.098
2007	0.522	0.522	0.09
1998-2007	0.348±0.07	0.347±0.07	0.097

Table A17: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.028 and the total sampling uncertainty is 0.075

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.109
1999	0.199	0.199	0.082
2000	0.26	0.26	0.083
2001	0.224	0.224	0.11
2002	0.412	0.412	0.079
2003	0.419	0.419	0.075
2004	0.363	0.363	0.094
2005	0.426	0.427	0.094
2006	0.443	0.443	0.091
2007	0.522	0.522	0.084
1998-2007	0.348±0.07	0.348±0.07	0.09

Table A18: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.034 and the total sampling uncertainty is 0.078

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.117
1999	0.199	0.199	0.092
2000	0.26	0.26	0.098
2001	0.224	0.224	0.128
2002	0.412	0.412	0.109
2003	0.419	0.419	0.089
2004	0.363	0.364	0.112
2005	0.426	0.427	0.118
2006	0.443	0.444	0.109
2007	0.522	0.522	0.098
1998-2007	0.348±0.07	0.348±0.07	0.107

Table A19: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.042 and the total sampling uncertainty is 0.082

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.153
1999	0.199	0.198	0.151
2000	0.26	0.259	0.143
2001	0.224	0.223	0.163
2002	0.412	0.412	0.116
2003	0.419	0.419	0.105
2004	0.363	0.363	0.131
2005	0.426	0.426	0.132
2006	0.443	0.443	0.115
2007	0.522	0.522	0.117
1998-2007	0.348±0.07	0.347±0.07	0.133

Table A20: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.047 and the total sampling uncertainty is 0.084

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.16
1999	0.199	0.199	0.127
2000	0.26	0.26	0.119
2001	0.224	0.224	0.179
2002	0.412	0.412	0.124
2003	0.419	0.419	0.16
2004	0.363	0.363	0.167
2005	0.426	0.426	0.147
2006	0.443	0.443	0.15
2007	0.522	0.522	0.143
1998-2007	0.348±0.07	0.347±0.07	0.148

Table A21: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 60 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.075 and the total sampling uncertainty is 0.103

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.248
1999	0.199	0.203	0.212
2000	0.26	0.264	0.212
2001	0.224	0.229	0.262
2002	0.412	0.417	0.239
2003	0.419	0.423	0.24
2004	0.363	0.368	0.245
2005	0.426	0.432	0.27
2006	0.443	0.447	0.202
2007	0.522	0.526	0.243
1998-2007	0.348±0.07	0.352±0.07	0.237

Table A22: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.032 and the total sampling uncertainty is 0.077

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.118
1999	0.199	0.199	0.09
2000	0.26	0.259	0.097
2001	0.224	0.224	0.102
2002	0.412	0.412	0.105
2003	0.419	0.419	0.123
2004	0.363	0.363	0.09
2005	0.426	0.426	0.122
2006	0.443	0.443	0.083
2007	0.522	0.522	0.085
1998-2007	0.348±0.07	0.347±0.07	0.102

Table A23: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.033 and the total sampling uncertainty is 0.077

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.12
1999	0.199	0.199	0.092
2000	0.26	0.259	0.099
2001	0.224	0.224	0.116
2002	0.412	0.411	0.108
2003	0.419	0.419	0.125
2004	0.363	0.363	0.094
2005	0.426	0.426	0.122
2006	0.443	0.443	0.089
2007	0.522	0.522	0.089
1998-2007	0.348±0.07	0.347±0.07	0.105

Table A24: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.032 and the total sampling uncertainty is 0.077

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.117
1999	0.199	0.199	0.097
2000	0.26	0.26	0.092
2001	0.224	0.224	0.098
2002	0.412	0.412	0.103
2003	0.419	0.419	0.119
2004	0.363	0.363	0.083
2005	0.426	0.427	0.122
2006	0.443	0.443	0.08
2007	0.522	0.522	0.085
1998-2007	0.348±0.07	0.348±0.07	0.1

Table A25: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.037 and the total sampling uncertainty is 0.079

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.124
1999	0.199	0.199	0.106
2000	0.26	0.26	0.105
2001	0.224	0.224	0.119
2002	0.412	0.412	0.129
2003	0.419	0.419	0.13
2004	0.363	0.364	0.105
2005	0.426	0.427	0.142
2006	0.443	0.444	0.103
2007	0.522	0.522	0.1
1998-2007	0.348±0.07	0.348±0.07	0.116

Table A26: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.045 and the total sampling uncertainty is 0.083

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.161
1999	0.199	0.198	0.159
2000	0.26	0.259	0.152
2001	0.224	0.223	0.157
2002	0.412	0.412	0.136
2003	0.419	0.419	0.147
2004	0.363	0.363	0.126
2005	0.426	0.426	0.155
2006	0.443	0.443	0.109
2007	0.522	0.522	0.12
1998-2007	0.348±0.07	0.347±0.07	0.142

Table A27: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.05 and the total sampling uncertainty is 0.086

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.171
1999	0.199	0.199	0.137
2000	0.26	0.26	0.131
2001	0.224	0.224	0.175
2002	0.412	0.412	0.143
2003	0.419	0.419	0.186
2004	0.363	0.363	0.162
2005	0.426	0.426	0.168
2006	0.443	0.443	0.147
2007	0.522	0.522	0.147
1998-2007	0.348±0.07	0.347±0.07	0.157

Table A28: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 72 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.077 and the total sampling uncertainty is 0.104

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.257
1999	0.199	0.203	0.22
2000	0.26	0.264	0.222
2001	0.224	0.229	0.256
2002	0.412	0.417	0.249
2003	0.419	0.423	0.262
2004	0.363	0.368	0.245
2005	0.426	0.432	0.283
2006	0.443	0.447	0.202
2007	0.522	0.526	0.249
1998-2007	0.348±0.07	0.352±0.07	0.244

Table A29: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.039 and the total sampling uncertainty is 0.08

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.141
1999	0.199	0.199	0.143
2000	0.26	0.259	0.127
2001	0.224	0.224	0.139
2002	0.412	0.412	0.127
2003	0.419	0.419	0.092
2004	0.363	0.363	0.127
2005	0.426	0.426	0.126
2006	0.443	0.443	0.107
2007	0.522	0.522	0.094
1998-2007	0.348±0.07	0.347±0.07	0.122

Table A30: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.04 and the total sampling uncertainty is 0.081

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.144
1999	0.199	0.199	0.145
2000	0.26	0.259	0.13
2001	0.224	0.224	0.149
2002	0.412	0.411	0.128
2003	0.419	0.419	0.097
2004	0.363	0.363	0.132
2005	0.426	0.426	0.126
2006	0.443	0.443	0.111
2007	0.522	0.522	0.098
1998-2007	0.348±0.07	0.347±0.07	0.126

Table A31: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.038 and the total sampling uncertainty is 0.08

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.14
1999	0.199	0.199	0.147
2000	0.26	0.26	0.125
2001	0.224	0.224	0.136
2002	0.412	0.412	0.127
2003	0.419	0.419	0.09
2004	0.363	0.363	0.126
2005	0.426	0.427	0.125
2006	0.443	0.443	0.105
2007	0.522	0.522	0.094
1998-2007	0.348±0.07	0.348±0.07	0.121

Table A32: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.043 and the total sampling uncertainty is 0.082

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.149
1999	0.199	0.199	0.156
2000	0.26	0.26	0.136
2001	0.224	0.224	0.153
2002	0.412	0.412	0.149
2003	0.419	0.419	0.103
2004	0.363	0.364	0.14
2005	0.426	0.427	0.147
2006	0.443	0.444	0.122
2007	0.522	0.522	0.109
1998-2007	0.348±0.07	0.348±0.07	0.136

Table A33: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.05 and the total sampling uncertainty is 0.086

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.18
1999	0.199	0.198	0.196
2000	0.26	0.259	0.174
2001	0.224	0.223	0.186
2002	0.412	0.412	0.154
2003	0.419	0.419	0.124
2004	0.363	0.363	0.156
2005	0.426	0.426	0.159
2006	0.443	0.443	0.13
2007	0.522	0.522	0.129
1998-2007	0.348±0.07	0.347±0.07	0.159

Table A34: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.055 and the total sampling uncertainty is 0.089

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.189
1999	0.199	0.199	0.182
2000	0.26	0.26	0.157
2001	0.224	0.224	0.2
2002	0.412	0.412	0.162
2003	0.419	0.419	0.172
2004	0.363	0.363	0.193
2005	0.426	0.426	0.172
2006	0.443	0.443	0.165
2007	0.522	0.522	0.154
1998-2007	0.348±0.07	0.347±0.07	0.175

Table A35: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 90 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.082 and the total sampling uncertainty is 0.108

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.272
1999	0.199	0.203	0.255
2000	0.26	0.264	0.241
2001	0.224	0.229	0.277
2002	0.412	0.417	0.261
2003	0.419	0.423	0.25
2004	0.363	0.368	0.265
2005	0.426	0.432	0.289
2006	0.443	0.447	0.218
2007	0.522	0.526	0.257
1998-2007	0.348±0.07	0.352±0.07	0.259

Table A36: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.063 and the total sampling uncertainty is 0.094

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.198
1999	0.199	0.199	0.196
2000	0.26	0.259	0.219
2001	0.224	0.224	0.195
2002	0.412	0.412	0.191
2003	0.419	0.419	0.202
2004	0.363	0.363	0.18
2005	0.426	0.426	0.205
2006	0.443	0.443	0.184
2007	0.522	0.522	0.208
1998-2007	0.348±0.07	0.347±0.07	0.198

Table A37: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.063 and the total sampling uncertainty is 0.094

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.2
1999	0.199	0.199	0.198
2000	0.26	0.259	0.22
2001	0.224	0.224	0.203
2002	0.412	0.411	0.192
2003	0.419	0.419	0.205
2004	0.363	0.363	0.183
2005	0.426	0.426	0.206
2006	0.443	0.443	0.186
2007	0.522	0.522	0.21
1998-2007	0.348±0.07	0.347±0.07	0.2

Table A38: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.063 and the total sampling uncertainty is 0.094

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.199
1999	0.199	0.199	0.201
2000	0.26	0.26	0.218
2001	0.224	0.224	0.195
2002	0.412	0.412	0.19
2003	0.419	0.419	0.202
2004	0.363	0.363	0.18
2005	0.426	0.427	0.206
2006	0.443	0.443	0.184
2007	0.522	0.522	0.209
1998-2007	0.348±0.07	0.348±0.07	0.198

Table A39: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.066 and the total sampling uncertainty is 0.096

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.205
1999	0.199	0.199	0.209
2000	0.26	0.26	0.226
2001	0.224	0.224	0.207
2002	0.412	0.412	0.207
2003	0.419	0.419	0.21
2004	0.363	0.364	0.191
2005	0.426	0.427	0.221
2006	0.443	0.444	0.197
2007	0.522	0.522	0.216
1998-2007	0.348±0.07	0.348±0.07	0.209

Table A40: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.072 and the total sampling uncertainty is 0.1

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.23
1999	0.199	0.198	0.241
2000	0.26	0.259	0.256
2001	0.224	0.223	0.234
2002	0.412	0.412	0.214
2003	0.419	0.419	0.22
2004	0.363	0.363	0.205
2005	0.426	0.426	0.231
2006	0.443	0.443	0.203
2007	0.522	0.522	0.23
1998-2007	0.348±0.07	0.347±0.07	0.226

Table A41: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.075 and the total sampling uncertainty is 0.103

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.237
1999	0.199	0.199	0.232
2000	0.26	0.26	0.24
2001	0.224	0.224	0.247
2002	0.412	0.412	0.219
2003	0.419	0.419	0.251
2004	0.363	0.363	0.232
2005	0.426	0.426	0.242
2006	0.443	0.443	0.225
2007	0.522	0.522	0.245
1998-2007	0.348±0.07	0.347±0.07	0.237

Table A42: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 120 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.097 and the total sampling uncertainty is 0.12

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.309
1999	0.199	0.203	0.291
2000	0.26	0.264	0.303
2001	0.224	0.229	0.32
2002	0.412	0.417	0.305
2003	0.419	0.423	0.316
2004	0.363	0.368	0.301
2005	0.426	0.432	0.339
2006	0.443	0.447	0.272
2007	0.522	0.526	0.322
1998-2007	0.348±0.07	0.352±0.07	0.308

Table A43: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 18 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.115 and the total sampling uncertainty is 0.135

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.407
1999	0.199	0.199	0.325
2000	0.26	0.259	0.364
2001	0.224	0.224	0.333
2002	0.412	0.412	0.406
2003	0.419	0.419	0.426
2004	0.363	0.363	0.411
2005	0.426	0.426	0.356
2006	0.443	0.443	0.302
2007	0.522	0.522	0.314
1998-2007	0.348±0.07	0.347±0.07	0.364

Table A44: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 20 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.116 and the total sampling uncertainty is 0.135

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.408
1999	0.199	0.199	0.328
2000	0.26	0.259	0.366
2001	0.224	0.224	0.338
2002	0.412	0.411	0.407
2003	0.419	0.419	0.428
2004	0.363	0.363	0.413
2005	0.426	0.426	0.356
2006	0.443	0.443	0.304
2007	0.522	0.522	0.316
1998-2007	0.348±0.07	0.347±0.07	0.366

Table A45: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 24 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.116 and the total sampling uncertainty is 0.135

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.408
1999	0.199	0.199	0.329
2000	0.26	0.26	0.365
2001	0.224	0.224	0.335
2002	0.412	0.412	0.407
2003	0.419	0.419	0.427
2004	0.363	0.363	0.412
2005	0.426	0.427	0.357
2006	0.443	0.443	0.303
2007	0.522	0.522	0.316
1998-2007	0.348±0.07	0.348±0.07	0.366

Table A46: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 30 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.118 and the total sampling uncertainty is 0.137

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.415
1999	0.199	0.199	0.335
2000	0.26	0.26	0.372
2001	0.224	0.224	0.344
2002	0.412	0.412	0.417
2003	0.419	0.419	0.434
2004	0.363	0.364	0.419
2005	0.426	0.427	0.369
2006	0.443	0.444	0.313
2007	0.522	0.522	0.324
1998-2007	0.348±0.07	0.348±0.07	0.374

Table A47: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 36 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.122 and the total sampling uncertainty is 0.141

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.207	0.429
1999	0.199	0.198	0.361
2000	0.26	0.259	0.39
2001	0.224	0.223	0.363
2002	0.412	0.412	0.421
2003	0.419	0.419	0.441
2004	0.363	0.363	0.43
2005	0.426	0.426	0.377
2006	0.443	0.443	0.318
2007	0.522	0.522	0.333
1998-2007	0.348±0.07	0.347±0.07	0.386

Table A48: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 40 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.125 and the total sampling uncertainty is 0.143

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.206	0.434
1999	0.199	0.199	0.355
2000	0.26	0.26	0.386
2001	0.224	0.224	0.373
2002	0.412	0.412	0.425
2003	0.419	0.419	0.459
2004	0.363	0.363	0.442
2005	0.426	0.426	0.384
2006	0.443	0.443	0.338
2007	0.522	0.522	0.348
1998-2007	0.348±0.07	0.347±0.07	0.394

Table A49: Comparison of the total Simulated Uptake With the Uptake from Our Proposed(Sampling every 180 days and every 60 degrees in longitude) Sampling and the Sampling Error Introduced. The sampling error is 0.142 and the total sampling uncertainty is 0.158

Year	Total Simulated Uptake, PgC/yr	Sample estimate of Uptake, PgC/yr	Sampling uncertainty (2σ), PgC/yr
1998	0.207	0.211	0.483
1999	0.199	0.203	0.408
2000	0.26	0.264	0.433
2001	0.224	0.229	0.437
2002	0.412	0.417	0.482
2003	0.419	0.423	0.503
2004	0.363	0.368	0.487
2005	0.426	0.432	0.457
2006	0.443	0.447	0.372
2007	0.522	0.526	0.412
1998-2007	0.348±0.07	0.352±0.07	0.447

2 Appendix B

Table B1: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the annual mean with 50 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.28
2001	1.63	1.64	0.23
2002	1.83	1.85	0.25
2003	1.80	1.78	0.17
2004	1.83	1.84	0.14
2005	1.87	1.88	0.24
2006	1.91	1.90	0.23
2007	1.98	2.00	0.24
2008	1.82	1.81	0.14
2009	1.81	1.84	0.20
2000-2009	1.81±0.07	1.82±0.07	0.21

Table B2: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the annual mean with 100 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.13
2001	1.63	1.65	0.15
2002	1.83	1.80	0.13
2003	1.80	1.81	0.12
2004	1.83	1.84	0.12
2005	1.87	1.87	0.06
2006	1.91	1.92	0.10
2007	1.98	1.97	0.15
2008	1.82	1.82	0.13
2009	1.81	1.80	0.13
2000-2009	1.81±0.07	1.81±0.07	0.12

Table B3: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the annual mean with 500 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.63	0.05
2001	1.63	1.64	0.05
2002	1.83	1.84	0.06
2003	1.80	1.81	0.05
2004	1.83	1.84	0.05
2005	1.87	1.88	0.07
2006	1.91	1.93	0.05
2007	1.98	2.00	0.06
2008	1.82	1.83	0.05
2009	1.81	1.81	0.06
2000-2009	1.81±0.07	1.82±0.07	0.06

Table B4: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the annual mean with 1000 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.04
2001	1.63	1.63	0.03
2002	1.83	1.83	0.04
2003	1.80	1.81	0.04
2004	1.83	1.83	0.03
2005	1.87	1.87	0.05
2006	1.91	1.91	0.04
2007	1.98	1.99	0.05
2008	1.82	1.83	0.02
2009	1.81	1.81	0.05
2000-2009	1.81±0.07	1.81±0.07	0.04

Table B5: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the simualted data with 1000 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.63	0.10
2001	1.63	1.65	0.05
2002	1.83	1.83	0.09
2003	1.80	1.82	0.09
2004	1.83	1.82	0.11
2005	1.87	1.86	0.12
2006	1.91	1.92	0.10
2007	1.98	1.98	0.13
2008	1.82	1.85	0.12
2009	1.81	1.80	0.09
2000-2009	1.81±0.07	1.81±0.06	0.10

Table B6: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the simualted data with 2000 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.07
2001	1.63	1.62	0.07
2002	1.83	1.82	0.08
2003	1.80	1.80	0.09
2004	1.83	1.84	0.08
2005	1.87	1.87	0.06
2006	1.91	1.90	0.09
2007	1.98	1.97	0.08
2008	1.82	1.82	0.08
2009	1.81	1.82	0.06
2000-2009	1.81±0.07	1.81±0.06	0.08

Table B7: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the simualted data with 5000 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.06
2001	1.63	1.62	0.04
2002	1.83	1.83	0.06
2003	1.80	1.80	0.03
2004	1.83	1.84	0.04
2005	1.87	1.88	0.05
2006	1.91	1.92	0.05
2007	1.98	1.99	0.04
2008	1.82	1.82	0.06
2009	1.81	1.81	0.05
2000-2009	1.81±0.07	1.81±0.07	0.05

Table B8: Comparison of the simulated annual mean CO_2 with the annual mean from sampling the simualted data with 7000 locations.

Year	Model Mean	Sample Mean	Sampling Uncertainty 2σ
2000	1.61	1.61	0.03
2001	1.63	1.63	0.04
2002	1.83	1.83	0.04
2003	1.80	1.80	0.06
2004	1.83	1.82	0.05
2005	1.87	1.87	0.05
2006	1.91	1.91	0.03
2007	1.98	1.97	0.05
2008	1.82	1.82	0.05
2009	1.81	1.80	0.04
2000-2009	1.81±0.07	1.80±0.07	0.04