UNIVERSITY OF CALIFORNIA, SAN DIEGO

Airborne Oxygen Measurements over the Southern Ocean as an Integrated Constraint of Seasonal Biogeochemical Processes

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Oceanography

by

Jonathan D. Bent

Committee in charge:

Ralph Keeling, Chair
Kathryn Barbeau
Andrew Dickson
Kim Griest
Britton Stephens
Ray Weiss

2014

Copyright

Jonathan D. Bent, 2014

All rights reserved.

The Dissertation of Jonathan Douglas Bent is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2014

Dedication

For Sarah and Wyn Henry

In Memoriam Patricia Gage Holmes

(1915-2014)

Epigraph

It is very windy here [in the Southern Ocean]. It has reaffirmed my faith that the world is round, because if all this air wasn't just doing a lap of Antarctica and coming back round again, then wherever it was coming from would surely have run out by now.

-John A Stephenson, 2004, from All-Geo.org

Table of Contents

List of Abbreviations

ACTM:	The Atmospheric Chemistry Transport Model
APO:	Atmospheric Potential Oxygen (APO ≈ O_2 + 1.1 CO_2)
ATM:	Atmospheric Transport Model
BEC:	Biogeochemical Elemental Cycling model, used in both CCSM3 and CESM ocean model runs
CA:	Curtain Average
CBA:	Cold Bay, Alaska
CCSM3:	Community Climate System Model v.3 (NCAR)
CESM:	Community Earth System Model (NCAR)
CGO:	Cape Grim Observatory, Tasmania
CORE:	Coordinated Ocean-ice Reference Experiments forcing product
DRAKKAR:	(not an acronym) Wind forcing product based on ERA-40 Reanalysis, used in NEMO-CNTRL and WSTIR runs
ECMWF:	European Center for Medium-range Weather Forecasts
ERA-40:	ECMWF Re-Analysis-40 wind modeling product
GFDL:	NOAA Geophysical Fluid Dynamics Laboratory
GK01:	Garcia and Keeling [2001] dissolved O_2 climatology
GV:	Gulfstream Five (airborne laboratory), a.k.a. HIAPER
HIAPER:	High-performance Instrumented Airborne Platform for Environmental Research, a.k.a. GV
HIMIL:	HIAPER Modular InLet
HIPPO:	HIAPER Pole-to-Pole Observations (airborne research campaign)
IH:	Inter-hemispheric
JAMSTEC:	Japanese Agency for Marine-earth Science and TEChnology
MCA:	Meridional Curtain Average
MOM4:	Modular Ocean Model v.4, GFDL, Princeton University
MS:	Mass Spectrometer

NCAR:	National Center for Atmospheric Research
NCEP:	National Centers for Environmental Prediction
NEMO-CNTRL:	Nucleus for European Model of the Ocean, Control Run
NEMO-WSTIR:	Nucleus for European Model of the Ocean, with Wind Stirring parameterization
NIWA:	National Institute for Water and Atmospheric Research, NZ
NOAA:	National Oceanic and Atmospheric Administration
OBM:	Ocean Biogeochemistry Model
ORCA:	(not an acronym) Tripolar gridding scheme for ocean circulation used by NEMO and other models
PCC:	Puyehue-Cordón Caulle volcano complex, Chile
PISCES:	Pelagic Iteration Scheme for Carbon and Ecosystem Studies
POP:	Parallel Ocean Program model (used in CESM)
PSA:	Palmer Station, Antarctica
SOCA:	Southern Ocean Curtain Average (a.k.a. SOMCA)
SOMCA:	Southern Ocean Meridional Curtain Average (a.k.a. SOMA)
SPO:	South Pole Observatory (Amundsen-Scott Base)
TOPAZ:	Tracers Of Phytoplankton with Allometric Zooplankton biogeochemistry model, Princeton University
TM3:	The global atmospheric Tracer Model v.3
VEI:	Volcanic Explosivity Index
VICI:	Valco Instruments Company Incorporated
VUV:	Vacuum Ultraviolet

List of Figures

List of Tables

List of Equations

Acknowledgements

I would like to thank first, and foremost, Professor Ralph Keeling for his patience and sound advice throughout my PhD. I thank him for encouraging me to continue despite several major life changes that threatened to derail my academic progress. His foresight helped me envision the structure of this dissertation, and his insight helped me think through countless scientific problems.

I would also like to thank the huge assistance of my co-advisor Dr. Britton Stephens, who has provided an astounding level of support and editing in the development of this document. His availability to teach me about sampling and measurement systems is responsible for my understanding of most of the material in Chapter 2.

I would like to thank Dr. Sara Mikaloff-Fletcher for an immeasurable amount of help in teaching me about ocean and atmospheric transport models. Sara took time to respond to all of my questions and requests for additional model output, turning around simulation requests, no matter how seemingly unnecessary, in a matter of days or weeks despite a busy schedule. I would also like to thank her for meeting with me in person on two occasions to help me see through some of the more complex features of models, and to answer questions, however misguided.

I would like to thank Dr. Prabir Patra for his great enthusiasm and willingness to help with model output. Prabir happily produced increasingly esoteric modeling products in response to my requests, and always did so with astonishing speed.

I would like to thank Dr. Pavel Romashkin, who oversaw the tremendously complex HIPPO Global sampling campaign as project manager. Over the five month-long HIPPO missions, Pavel was a constant source of support, and always willing to help with instrument problems whenever they threatened to compromise carefully-planned data. I also thank Pavel for guiding me through the logistics of complicated team projects, and for his honest advice when I needed it.

I would like to thank the Harvard QCLS/OMS instrument team: Steve Wofsy, Bruce Daube, Eric Kort, Jasna Pitman, Rodrigo Jimenez, and Greg Santoni, among others, for sharing the N_2O and CO data I used to evaluate samples for stratospheric influence.

I would like to thank my committee for their excellent comments during my qualifier, and for their quick responses to scheduling requests for both the qualifier and defense. I would like to sincerely thank Dr. Mark Thiemens for sitting on my qualifier committee, even though he was unable to be on the defense committee, and I would like to sincerely thank Dr. Kim Griest for stepping in at the last minute to act as a committee member when scheduling conflicts arose.

My tiny contribution to the world of modeling stands on the shoulders of many giants of the ocean modeling world. Their tireless efforts to perfect their and others' models has made this dissertation possible. I can only apologize for taking their incredibly complex and subtle output and reducing it to a mere five points in time. Drs. Keith Rodgers and Olivier Aumont kindly provided output from the NEMO-WSTIR and NEMO-CNTRL ocean model simulations featured in Chapter 3 of this dissertation, even while the output was still in discussion phase. They have provided extremely valuable

feedback on this dissertation, and much greater insight into model results than a literature search could provide. Drs. Scott Doney, Ivan Lima and Matt Long, were responsible for the CCSM3 and CESM model output in this document, and have answered many important questions over the last few years. Dr. Corinne LeQuéré provided the NEMO-PISCES-T model output, and Dr. Erik Buitenhuis provided additional insight into the origins of the model products. Dr. John Dunne kindly provided the MOM4 model output.

I would like to thank the National Center for Atmospheric Research visiting student program for providing me with 5 months of visiting student support that allowed me to spend invaluable time working alongside Dr. Britton Stephens as I wrote portions of this dissertation, and performed follow-up experiments. I would also like to thank the Scripps graduate office for providing me with two quarters of funding when my funding ran out.

I would also like to thank the NCAR support staff, who made my many stays in Colorado pleasant and productive. I'd especially like to thank the technical support personnel who went out of their way to extend work hours on the GV aircraft to accommodate emergent instrument needs in the weeks leading up to the five HIPPO missions. I would also like to thank the pilots, mechanics, and other science crew for their assistance and friendship, especially in high-stress experimental settings, in which I could not have finished my work alone.

Various members of the Scripps O_2 Program and Scripps Carbon Dioxide Group have provided incredible assistance over the entirety of my PhD, guiding me through the process of learning to operate sensitive instrumentation, and helping me assemble and design apparatus of my own. I owe a particular debt of gratitude to Bill Paplawsky for his

support, particularly in the first few years of my PhD. Bill provided me with truly excellent advice and support as I explored various career options, and took long periods of time to help me with both main and side projects. I'd also like to thank Alane Bollenbacher, Kim Spalding, Adam Cox, Sara Afshar, Tim Leuker, and Cindy Uribe for their willingness to run the majority of the HIPPO flask and isotope samples. Alane and Tim also spent large stretches of time patiently teaching me how to run the Optima isotope mass spectrometer in the CO_2 lab.

I would also like to thank my friends for their great support at various stages in this PhD. My physicist roommates Alex Winbow and Yaniv Rosen gave me excellent insight into the physical basis of my work, but also kept the mood light when I needed it to be. Adam Radich, Jared Kleusner, Sandra Kirtland-Turner, James Traer, Mike Pritchard, Peter Barry, Alex Hangsterfer, Cassie Gaston, and many others made learning about oceanography much more enjoyable.

I owe a great debt of gratitude to my family for supporting me unflinchingly throughout a PhD that grew and grew in length and complexity. Advice and support from my parents, siblings and in-laws kept me on track, even when assembling a document of this size seemed like an utterly unachievable goal.

But I owe my greatest debt of gratitude to my wife, Sarah Lebo, whose unquestioning support of this venture kept my spirits high even when I felt like giving up. Sarah's willingness to move to Colorado and back to accommodate my work, and her patience as I slowly developed and wrote this dissertation has been nothing short of extraordinary. On November 18th of 2013, as dissertation writing approached a fever pitch, she gave birth to our son Wyn Henry Bent. I cannot thank her enough for this.

Some of the material in Chapter 2 will be included in a paper in preparation for publication on methods and intercomparison of airborne atmospheric O_2 measurements. Co-authors include: Stephens, Britton, B; Bent, Jonathan D; and Keeling, Ralph F.

Material from Chapter 3 is currently being prepared for submission for publication. Co-authors include: Bent, Jonathan D; Stephens, Britton, B; Keeling, Ralph F; Mikaloff-Fletcher, Sara M; Patra, Prabir K. The dissertation author was the primary investigator and author of this material.

NCAR is sponsored by the National Science Foundation (NSF). HIPPO was supported by NSF grants ATM-0628575, ATM-0628519, ATM-0628388 ATM-0628452, and ATM-1036399, by NOAA, and by NCAR. Interpretive work for this thesis was also supported by NSF grant OCE-1130976.

Vita

2004	Bachelor of Arts, Columbia College, Columbia University
2005	Teaching Assistant, Department of Environmental Engineering Columbia University
2004-2006	Research Assistant, Lamont Doherty Earth Observatory
2006-2014	Research Assistant, Scripps Institution of Oceanography, University of California, San Diego
2011	Visiting Student Scholar, National Center for Atmospheric Research
2014	Doctor of Philosophy, Scripps Institution of Oceanography University of California, San Diego

Publications

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., McManus, J. B., Peischl, J., Ryerson, T. B., Holloway, J. S., Andrews, A. E., Sweeney, C., Hall, B., Hintsa, E. J., Moore, F. L., Elkins, J. W., Hurst, D. F., Stephens, B. B., Bent, J., and Wofsy, S. C.: Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite – CO2, CH4, N2O, and CO – during the CalNex and HIPPO campaigns, Atmos. Meas. Tech., 7, 1509-1526, doi:10.5194/amt-7-1509-2014, 2014.

Graven, H.D., Keeling, R.F., Piper, S.C., Patra, P.K., Stephens, B.B., Wofsy, S.C., Welp, L.R., Sweeney, C., Tans, P.P., Kelley, J.J., Daube, B.C., Kort, E.A., Santoni, G.W. & and Bent, J. (2013). Enhanced seasonal exchange of CO_2 by northern ecosystems since 1960. Science, 341, 1085-1089, doi: 10.1126/science.1239207.

Wofsy, S.C. and the HIPPO science team. (2011). HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369 (1943), 2073 – 2086. doi: 10.1098/rsta.2010.0313

Fields of Study

Marine and atmospheric biogeochemistry

Air-sea gas exchange

Carbon cycle biogeochemistry

Atmospheric CO_2 and O_2 measurements

Ar/N_2 measurements as a tracer of ocean heat and atmospheric circulation

ABSTRACT OF THE DISSERTATION

Airborne Oxygen Measurements over the Southern Ocean as an Integrated Constraint of Seasonal Biogeochemical Processes

by

Jonathan D. Bent

Doctor of Philosophy in Oceanography

University of California, San Diego, 2014

Professor Ralph F. Keeling, Chair

I present airborne in situ and flask measurements of the atmospheric oxygen-nitrogen ratio—$\delta(O_2/N_2)$—and CO_2 over the Southern Ocean made at five points in the seasonal cycle between 2009 and 2011. The flights, which were part of the HIPPO global project, were conducted with repeated profiling along north-south tracks from near the ocean's surface to the lower stratosphere. I use these measurements to define seasonal

variations in meridionally and vertically averaged atmospheric potential oxygen (APO ≈ O_2+1.1*CO_2) over the Southern Ocean—the "meridional curtain average"—and relate these to the underlying oceanic O_2 and CO_2 fluxes. I find that a 1- harmonic fit to the integrated seasonal cycle of APO over the latitudes 65-45°S is better constrained than a 2-harmonic fit, and has an amplitude of 36.4 ±5.6 per meg with a seasonal peak/trough centered at year day 58/260 ±10.8 days. Comparison with nearby station APO records at Cape Grim Observatory and Palmer Station suggest this constitutes roughly 60% of the surface seasonal cycle. I use TM3 and ACTM atmospheric transport model output to compare observations and surface fluxes, and to estimate uncertainties in my approach. The lower boundary condition for these runs is set by flux estimates derived variously from O_2 and CO_2 measurements ("dissolved climatologies") and physical ocean models with biogeochemistry modules. ACTM and TM3 dissolved climatology runs, and TM3 runs with CCSM3, CESM, MOM4, NEMO-PISCES-T, NEMO-CNTRL, and NEMO-WSTIR ocean model simulations overestimate (+) or underestimate (-) the observed amplitude by +49%, +44%, +17%, +45%, +13%, +1%, +4%, -7% respectively, with 1-harmonic phasing that leads (+) or lags (-) observations by +13, +12, -32, -1, -10, -51, -13, and -28 days, respectively. No single model falls within HIPPO error bars by all metrics: amplitude, peak and trough timing, and 2-harmonic trough-peak asymmetry, but MOM4 and NEMO-CNTRL are the most consistently good performers. Analysis of heat fluxes suggests both physical and biological forces are behind differences between observed and modeled seasonal phasing and amplitude. I find that Garcia and Keeling [2001] O_2 flux fields anticipate observations by about two weeks, and need to be re-evaluated based on new gas exchange scaling parameters.

I also present analyses of gradients and seasonal cycles of atmospheric Ar/N_2. I find a measurable vertical Ar/N_2 gradient in the lower stratosphere that corresponds with N_2O data and height. I also find resolvable, phase-opposite seasonal cycles at surface elevations in both hemispheres, and an interhemispheric gradient between the latitudes of Cold Bay Alaska (55°N) and Cape Grim Observatory (41°S) of -5.6 ±4.6 per meg (CBA-CGO) which matches mean annual differences between the two station records well.

1 Introduction

1.1 Southern Ocean biogeochemistry

A chief priority of modern-day carbon cycle biogeochemistry is the resolution of carbon fluxes between earth's four principal carbon reservoirs—the atmosphere, biosphere, hydrosphere and geosphere. A comprehensive understanding of the timescales and magnitudes of these fluxes allows scientists to resolve anthropogenic perturbations to the system, and to quantify the dispersal of recent emissions into earth's reservoirs [*IPCC AR5 WGI*, 2013]. This provides crucial information about humankind's influence on the earth system, and allows models to predict future changes to the earth's climate.

Because most of the earth's carbon sinks are large, heterogeneous, and challenging to sample, scientists often choose to interpret atmospheric measurements of carbon dioxide to infer fluxes between reservoirs. The atmosphere is sufficiently well mixed that measurements taken in remote locations provide information about fluxes averaged over large areas of the earth's surface. However, changes in the atmospheric CO_2 concentration alone provide only a measure of the net effect of the various flux terms. An additional, related variable is required to constrain individual flux components. Measurements of atmospheric oxygen can provide this constraint because O_2 is physically and biologically tied to CO_2 in many distinct ways [*R. F. Keeling et al.*, 1996; *R. F. Keeling and Shertz*, 1992; *Severinghaus*, 1995].

Observations of seasonal cycles in atmospheric CO_2 and O_2 concentrations are valuable for a number of reasons. O_2 has been estimated to exchange with the land

biosphere at an approximate ratio of 1.1:1±0.1 moles O_2 produced per moles of CO_2 consumed [*Severinghaus*, 1995]. Because seasonal air-sea CO_2 exchange is small, knowing the O_2:CO_2 ratio of exchange with the land biosphere allows scientists to resolve the oceanic component of the observed atmospheric seasonal cycle of O_2 [*R. F. Keeling et al.*, 1993]. In combination with understandings of physical forcings, O_2 and CO_2 observations can also provide an estimate of ocean net ecosystem production. In a modeling context, atmospheric data can be compared against simulation output to see whether models are effectively reproducing natural processes, or they can be used as input in model inversions to estimate how global source/sink terms may be changing due to anthropogenic perturbations [*Le Quéré et al.*, 2009].

Efforts to identify the sinks of anthropogenic CO_2 suggest that the Southern Ocean plays a major role in mitigating atmospheric rise [*Law et al.*, 2008; *Mikaloff Fletcher et al.*, 2006; *Sabine et al.*, 2004; *Zickfeld et al.*, 2008]. Oceanic regions south of 44°S are estimated to have taken in 40% of the anthropogenic carbon currently in the ocean [*Khatiwala et al.*, 2009; *Sallee et al.*, 2012]. Strong Westerlies around the Antarctic continent drive northward Ekman transport, exposing deep water rich in dissolved inorganic carbon (DIC), from which CO_2 is released to the atmosphere. As this water moves northward, it absorbs anthropogenic CO_2 before sinking beneath warmer subtropical waters as intermediate and mode water. Because it provides this anthropogenic carbon sink, and because deep water upwelling constitutes a significant "communication" of surface and deep ocean, the Southern Ocean provides an important climate feedback: one whose physical and biogeochemical processes must be accurately described and modeled.

As a natural laboratory, the Southern Ocean (from Antarctica to roughly 45°S) offers an additional experimental incentive—it is far from the world's large Northern Hemisphere anthropogenic sources, and its atmospheric seasonal CO_2 cycle reflects minimal influence from the land biosphere due to the fact that the Southern Hemisphere is roughly 81% ocean. Recent studies [*Stephens et al.*, 2013] suggest that land processes comprise only about 0.7 ppm of the observed CO_2 cycle at Baring Head, New Zealand. Considered in terms of per meg, the units of measurement for O_2/N_2, and in the context of a 1.1:1±0.1 exchange ratio, this is roughly equal to a seasonal cycle of 3.5 per meg in O_2/N_2, about 4% of the roughly 80 per meg O_2/N_2 seasonal signal observed at Cape Grim Observatory. Such a contribution is minute in comparison to Northern Hemisphere stations. Furthermore, the Southern Ocean's water and air masses circulate around Antarctica, so that the patterns of variability in the atmosphere, and to some extent the ocean, are largely consistent about the latitude circles.

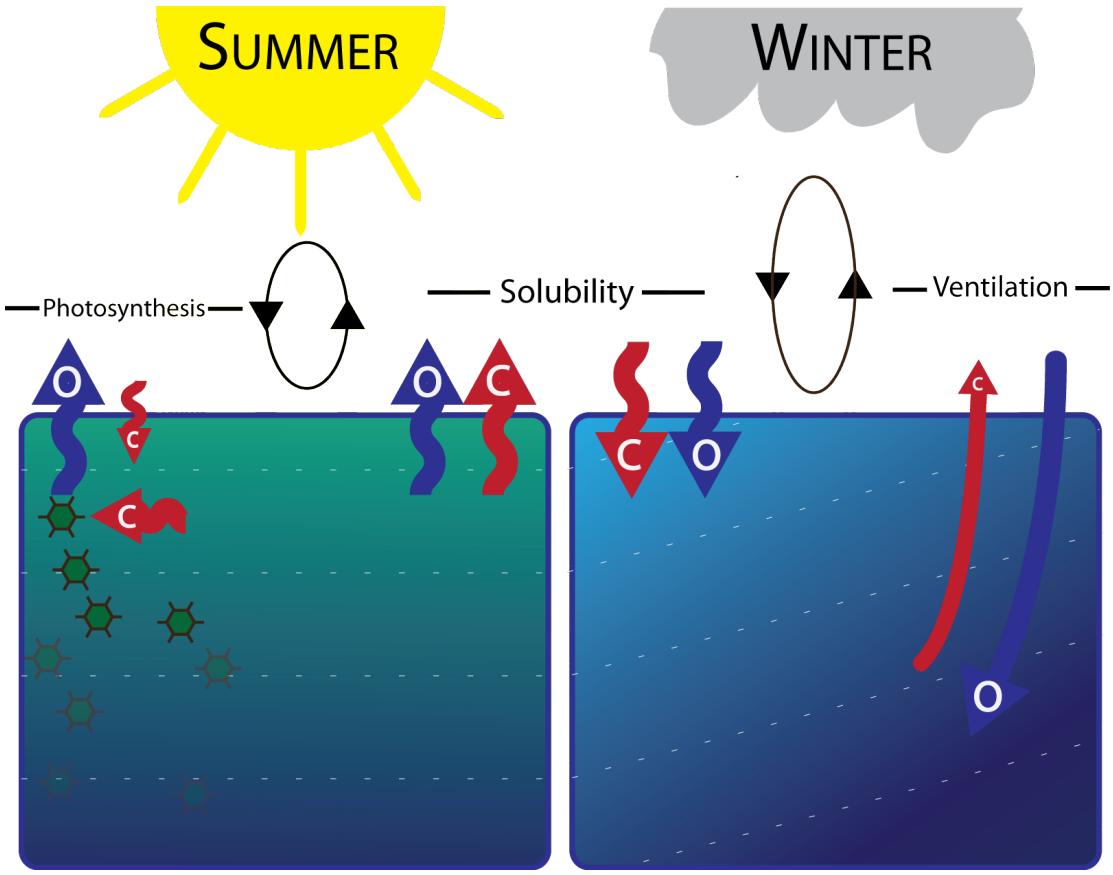

Figure 1.1. Simplified illustration of Southern Ocean open-water seasonal biogeochemistry (anthropogenic and interannual components not shown). Net seasonal fluxes of CO_2 and O_2 are denoted by arrows (red and blue, respectively) at the ocean's surface. The consumption of dissolved inorganic carbon by phytoplankton is indicated as a red arrow pointing to the left in the "summer" scenario. Changes to isopycnal surfaces (shoaling in winter) are noted in light gray. In the atmosphere, vertical mixing is greater in winter than in summer because the ocean is relatively warm compared to the cold air above it.

The Southern Ocean's circulation and biogeochemistry are seasonably dependent. Broadly, during the southern hemisphere summer (left portion of **Figure 1.1**), Southern Ocean surface waters are warmed by sunlight, reducing the solubility of dissolved gases, and driving the efflux of CO_2 and O_2 from the ocean. The warming of the surface establishes a strong thermocline (temperature gradient)—providing relative static stability—which results in stratification of water masses, though the surface mixed layer continues to be significantly deeper than at other latitudes due to continuous wind stirring [*Rodgers et al.*, 2014]. Micronutrients brought to the surface waters in winter along

isopycnals from other regions, and through vertical mixing with the deep ocean, now fuel photosynthesis. The photosynthesis draws down the local levels of dissolved inorganic carbon (DIC), decreasing the partial pressure of CO_2 in the water, and drawing CO_2 into the ocean from the atmosphere, counteracting the solubility effect. At the same time, O_2 produced by phytoplankton supersaturates the surface waters, causing a release of O_2 to the atmosphere [*Najjar and Keeling*, 2000], which reinforces the solubility effect. Some of the carbon sequestered in organic form by phytoplankton eventually rains out of the surface layer as the organisms die, sink or are consumed and excreted, and is subsequently remineralized (returned to DIC), a key part of the biological pump mechanism.

During the Southern Ocean winter (right portion of **Figure 1.1**), insolation is minimal, so the ocean's surface cools, reducing the static stability of the upper water column. As a result, and because of strong westerly winds, isopycnal surfaces shoal, allowing nutrients from deep water masses to reach the ocean's surface. In this fashion, the deep ocean effectively mixes with the surface, allowing oxygen-starved, DIC-rich water to interact with the atmosphere. The cooling winter waters absorb CO_2, drawing down atmospheric concentrations, while the outgassing of CO_2 from the carbon-rich deep waters has the opposite effect. For O_2, however, increased winter solubility coincides with the "ventilation" of the same oxygen-starved deep waters, reinforcing atmospheric O_2 drawdown.

Because atmospheric O_2 variations constitute only a tiny fraction of a much larger atmospheric reservoir, changes are typically measured with high-precision techniques that detect minute changes in the ratio of O_2 to N_2 molecules [*Bender et al.*, 1994]. In

absolute terms, seasonal cycles in atmospheric O_2 are large relative to cycles of CO_2 at high latitudes [*R. F. Keeling and Shertz*, 1992] for the reasons listed above. Additionally, because the solubility-driven efflux and influx of N_2 is largely in phase with O_2, the combined O_2/N_2 signal is in fact slightly smaller than the O_2 signal alone [*R. F. Keeling et al.*, 1993]. Over the Southern Ocean, the strongly reinforcing biological/physical forcings on O_2/N_2 in both summer and winter produce an atmospheric concentration signal that is large relative to other regions and close to sinusoidal, and which tracks sea-surface temperature closely. At the same time, strongly interfering biological/physical forcings in winter and summer produce a seasonal atmospheric CO_2 cycle whose signal is substantially damped and lagged relative to O_2 by a month or more because of the buffering of carbonate chemistry reactions and relatively high solubility. For these reasons, at such latitudes, seasonal cycles of atmospheric O_2/N_2 provide a considerably clearer picture of surface ocean biogeochemistry than CO_2.

However, because atmospheric O_2/N_2 measurements reflect the combination of terrestrial, oceanic, and anthropogenic fluxes, seasonal O_2/N_2 measurements in the atmosphere are not exclusively representative of oceanic processes. One approach to isolating the oceanic component of the O_2 signal is to use the expected terrestrial photosynthetic exchange rate of 1.1:1±0.1 moles O_2 produced per mole CO_2 consumed [*Severinghaus*, 1995] to effectively remove all carbon dioxide from a sample. What remains is the background portion of the O_2 concentration, a value known variously as "oceanic oxygen" or "atmospheric potential oxygen" (APO) [*R. F. Keeling and Shertz*, 1992; *Stephens et al.*, 1998]. I will use the term atmospheric potential oxygen exclusively in this dissertation, as this is the more common term in recent literature. Changes to APO

on a seasonal basis reflect primarily ocean O_2 exchange, while long-term changes reflect the removal of O_2 from the environment by fossil-fuel burning at an exchange ratio greater than 1.1, the long-term uptake of anthropogenic CO_2 by the oceans, and to a smaller degree decreases in ocean solubility due to heating and circulation changes.

1.2 HIPPO: HIAPER Pole to Pole Observations

This dissertation showcases and interprets seasonal APO measurements over the Southern Ocean, which I calculate from CO_2 and O_2 data collected during the HIAPER Pole to Pole Observations global airborne campaign. Using a new metric, the "meridional curtain average," named after the curtain plot (see **Figure 3.6**) whose atmospheric slice it averages over, I compute the mean concentration of APO over a meridional slice of the Southern Ocean at five points in the seasonal cycle. After small adjustments to account for contributing sources of bias, I fit a two-harmonic function to these values in order to quantify the climatological mean amplitude and phase of the seasonal APO cycle, and to compare this cycle against predictions from combined ocean biogeochemistry, atmospheric transport model runs to see how well such models reproduce my observations.

HIPPO was an NSF- and NOAA-funded airborne sampling campaign that involved researchers from Harvard University, the National Center for Atmospheric Research (NCAR), the Scripps Institution of Oceanography (SIO), NOAA, Princeton University, the Rosenstiel School for Marine and Atmospheric Science at the University of Miami (RSMAS), and other collaborating institutions. The HIPPO global campaign

ran between January 2009 and September 2011 aboard the National Science Foundation's Gulfstream Five research jet, "HIAPER" or "GV" (pictured below) and mapped the vertical and meridional distribution of atmospheric carbon cycle gases and other anthropogenic tracers [*Wofsy*, 2011].

Figure 1.2. NSF's HIAPER Gulfstream Five aircraft, a.k.a. "GV" on the tarmac at Midway Island during HIPPO4.

HIPPO consisted of five similarly-structured missions representing five distinct points in the seasonal cycle. Each mission flew from roughly 85°N to 67°S and back along the middle of the Pacific Ocean in a series of a ten or more six-to-nine hour research flights between waypoints of strategic value (see **Figure 1.3**). Research flights typically consisted of two high legs flown at 12-15 km, and five to eight "saw-tooth" maneuvers in between, in which the aircraft "porpoised" between an altitude of approximately 8-9 km, and a near-surface altitude of approximately 150-300 m. South-

and northbound transects in each campaign took place over approximately ten days each, characterizing the meridional and vertical structure of the atmosphere on synoptic time scales.

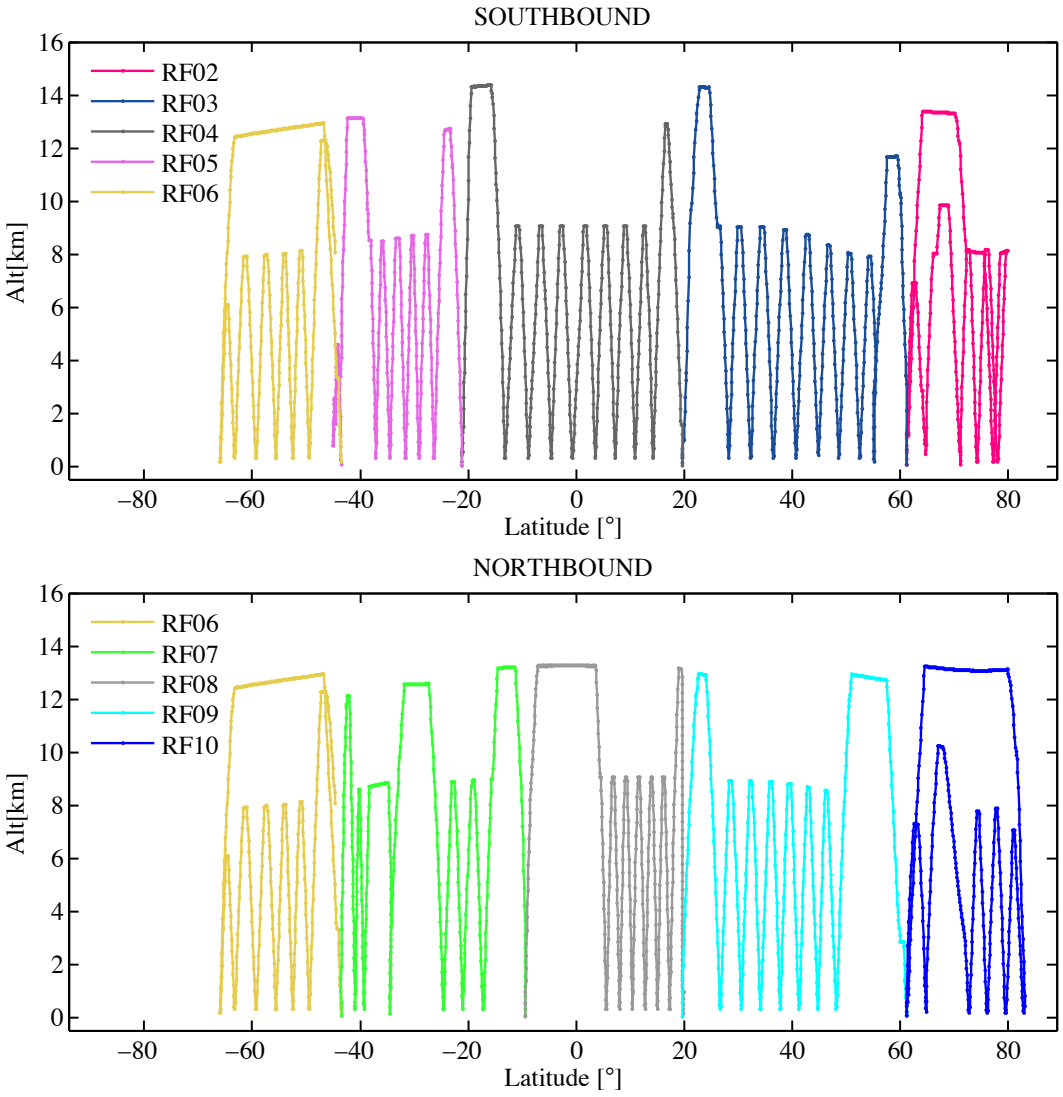

Figure 1.3. South- and northbound flight tracks of the GV over the Pacific Ocean shown by latitude and altitude. Here, HIPPO2 serves as a fair representation of a typical HIPPO mission. Each research flight is denoted by color. RF01 and RF12 are excluded for clarity, as they overlap later flights latitudinally, and are of minimal scientific importance to this study. RF06 (mustard) is shown in both panels, as it is considered part of both the south- and northbound legs. Full flight tracks for all 5 HIPPO missions can be found in Appendix 1 of this dissertation.

Collaborators measured over 90 species of interest, including CO_2, O_2/N_2, CH_4, CO, N_2O, SF_6, CFCs, HCFCs, O_3, H_2O, black carbon, and hydrocarbons (see **Table 1.1**). Several crucial species, such as CO_2, CO and O_2/N_2 were measured by multiple systems in order to resolve potential offsets between labs and instrument platforms, and to ensure maximal geographic coverage.

Table 1.1. Research teams, their instruments, and species measured.

Team	Instrument(s)	Species Measured
NCAR	AO2, CO, UHSAS, MTP, (MEDUSA)	O_2/N_2, CO_2, CO, aerosols, T profiles, pressure, various environmental and operational variables
SIO	MEDUSA (AO2)	O_2/N_2, CO_2, Ar/N_2, $^{13}CO_2$, $^{14}CO_2$, $C^{18}O^{16}O$
NOAA	PANTHER, NWAS, UCATS, SP2	O_3, black carbon, CFCs, PAN, CH_4, N_2O, H_2O and many others
Harvard	QCLS, OMS	CO_2, CH_4, N_2O, CO
Princeton	VCSEL	H_2O
U. Miami: RSMAS	AWAS	CFCs, HCFCs, methyl halides, solvents, organic nitrates, many more

Researchers from the NCAR/SIO group (Britton Stephens, Jonathan Bent, Andrew Watt and Stephen Shertz) ran the MEDUSA whole air flask sampler and AO2 vacuum ultraviolet oxygen analyzer during the HIPPO campaign. My role in the project included operating the MEDUSA sampler and AO2 instrument on about 60% of the research flights, re-building the expanded MEDUSA sampler, analyzing roughly 25% of the flask samples, quality controlling all flasks, designing the flask analysis code, and doing all subsequent analysis on MEDUSA flasks, and comparisons with AO2 data.

1.3 Thesis goals and structure

The overarching themes of this dissertation are the investigation of ocean biogeochemistry through aircraft observations which reflect large-scale processes, the importance of these observations in complementing ground station measurements, and their ultimate value as a gauge of the amplitude and timing of ocean model air-sea fluxes. Because of the geographic and temporal scope of the HIPPO project, certain measurements and calculations are possible for the first time. Principally, this includes the first meridionally complete measurements of vertical gradients in CO_2 and O_2/N_2, and the first attempts to describe their seasonality over the Southern Ocean using an integrative metric, the *meridional curtain average*. However, I also include some related measurements of the Ar/N_2 ratio in the atmosphere. Though these measurements are only tangentially relevant to the discussion of carbon cycle biogeochemistry in Chapter 3, they are nonetheless scientifically "low-hanging fruit," since they comprise some of the first measurements of their kind sampled and analyzed on a single apparatus at multiple points in both space and time. I show measurements of the $\delta(Ar/N_2)$ interhemispheric difference and the seasonality of this metric, and the tropospheric and lower stratospheric vertical $\delta(Ar/N_2)$ gradients.

I have structured the remainder of this dissertation into three chapters. Chapter 2 is a largely technical paper. It begins by describing the MEDUSA air sampler and the changes made to the system between 2008 and 2011. The chapter continues with a description of the operation of the sampler during the HIPPO campaign. The middle portions of the chapter address the analysis of the resulting flasks by the members of the Scripps O_2 Program at the Scripps Institution of Oceanography. The chapter continues

with a description of the laboratory data workup, the hierarchy of gas standards used to calibrate and operate the system, and a description of the code I developed to assemble and process the HIPPO data. The final sections of the chapter are devoted to explaining inconsistencies in the dataset, and efforts to correct these.

The chief focus of the dissertation is found in Chapter 3, which describes a technique to assess seasonal air-sea exchange of atmospheric oxygen over the Southern Ocean. I interpolate atmospheric potential oxygen (APO ≈ O_2 + 1.1*CO_2) concentrations between saw tooth flight tracks over a slice of the atmosphere sampled by the GV aircraft and calculate a detrended seasonal average at five points in the seasonal cycle. Using atmospheric transport model simulations with data-based climatological flux field boundary conditions, and forced with reanalysis meteorology outputted through the period of the HIPPO flights, I assess possible bias and error from the interpolation of flight tracks in the characterization of the atmospheric slice, and I make adjustments and uncertainty allowances for synoptic and interannual variations in order to produce a climatological meridional curtain average. I then compare this curtain average with atmospheric transport model simulations forced by a collection of ocean model fluxes to determine which ocean model fluxes best fit the measured seasonal APO cycle. I then discuss possible physical and biological reasons for the agreement or disagreement of these models with observations. The final few pages of the chapter focus briefly on transport model differences, comparing the amplitudes of curtain averages and station records for models and observations.

In chapter 4, I discuss the atmospheric Ar/N_2 ratio, and describe what is, to date, the most comprehensive set of measurements of the ratio. Having already discussed new

observations of the correlation of Ar/N_2 with N_2O in the lower stratosphere as part of an O_2/N_2 correction in Chapter 2, I turn to a discussion of the vertical gradient I observe in the troposphere, and compare observations with a simple one-dimensional diffusion model. I then discuss large-scale observations of hemisphere-scale seasonal cycles in the Ar/N_2 measurements. Finally, I show observations that suggest that the interhemispheric difference between the latitudes of distant sampling stations changes noticeably with season, favoring higher Ar/N_2 in the Southern Hemisphere for most of the seasonal cycle.

Readers most interested in the particular technical aspects of this study will likely want to start with Chapter 2, while readers most interested in the scientific results may want to skip to Chapter 3, and consult cross-referenced sections in Chapter 2 as needed to furnish a full understanding of the work.

2 Methods for sampling and analysis of CO_2 and O_2/N_2 data from the MEDUSA air sampler and AO2 vacuum ultraviolet instrument

2.1 Introduction

MEDUSA, or the Multiple Enclosure Device for Unfractionated Sampling of Air, is an automated sampler for aircraft use that collects 32 pressure- and flow-controlled, cryogenically dried air samples in borosilicate glass flasks. MEDUSA is specifically designed to reduce sample fractionation effects to allow for representative samples of the O_2/N_2 ratio, along with the Ar/N_2 ratio, CO_2, and the ^{13}C, ^{18}O, and ^{14}C isotopologues of CO_2. MEDUSA data are particularly relevant to studies of the carbon budget, and to the characterization of the seasonal and interannual exchange of CO_2 and O_2 between the ocean and the atmosphere. MEDUSA collected approximately 1600 samples during the HIAPER Pole to Pole Observations (HIPPO) campaign [*Wofsy*, 2011] between 2009 and 2011 at five points in the seasonal cycle, furnishing one of the most comprehensive datasets of the global distribution of atmospheric O_2/N_2. In this chapter, I describe the design, function and operation of the MEDUSA sampler, as well as the various stages of data quality control, with an emphasis on the correction of O_2/N_2 data using MEDUSA flask measurements of Ar/N_2. I conclude by describing how we use MEDUSA data to anchor concurrent, but higher-rate O_2 and CO_2 data from the AO2 vacuum ultraviolet instrument system.

In Section 2.2, I describe the MEDUSA sampler, explaining its history, structure, and operation. In Section 2.3, I describe the sampling strategy during the HIPPO

campaign. In Section 2.4, I discuss the stages of analysis the MEDUSA flasks undergo. In Section 2.5 I describe the AO2 instrument as an introduction to Section 2.6, which discusses the correction of MEDUSA O_2/N_2 data using MEDUSA Ar/N_2 data, and the use of MEDUSA data in anchoring AO2 results.

2.2 Instrumentation

2.2.1 MEDUSA general description

MEDUSA, pictured below in Figure 2.1-Figure 2.3, and whose flow diagram is shown in Figure 2.4, consists of two identical flask boxes, each of which holds 16 flasks, a dewar which holds cold traps which dry sample air, a pump box which houses most of the pressure control system and the system computer, and a valve box that holds two multi-position gas handling valves. Sample air arrives in the MEDUSA inlet at pressures between 120 and 760 torr, depending on the aircraft altitude; the upstream vacuum pump and pressure controller work in a feedback loop to maintain a constant low pump inlet pressure, whose value is below the lowest expected atmospheric pressure, in order to allow constant inflow of sample air. After passing through the inlet pump, sample air is immediately cooled and dried in 2 electropolished stainless steel cryotraps immersed in a slurry of dry ice and Fluorinert™ at -78°C. The dried air is then directed to the inlet of one of the 32 flasks by a series of valves. A second pump and pressure controller downstream of the flask outlets control sample pressure to approximately one atmosphere.

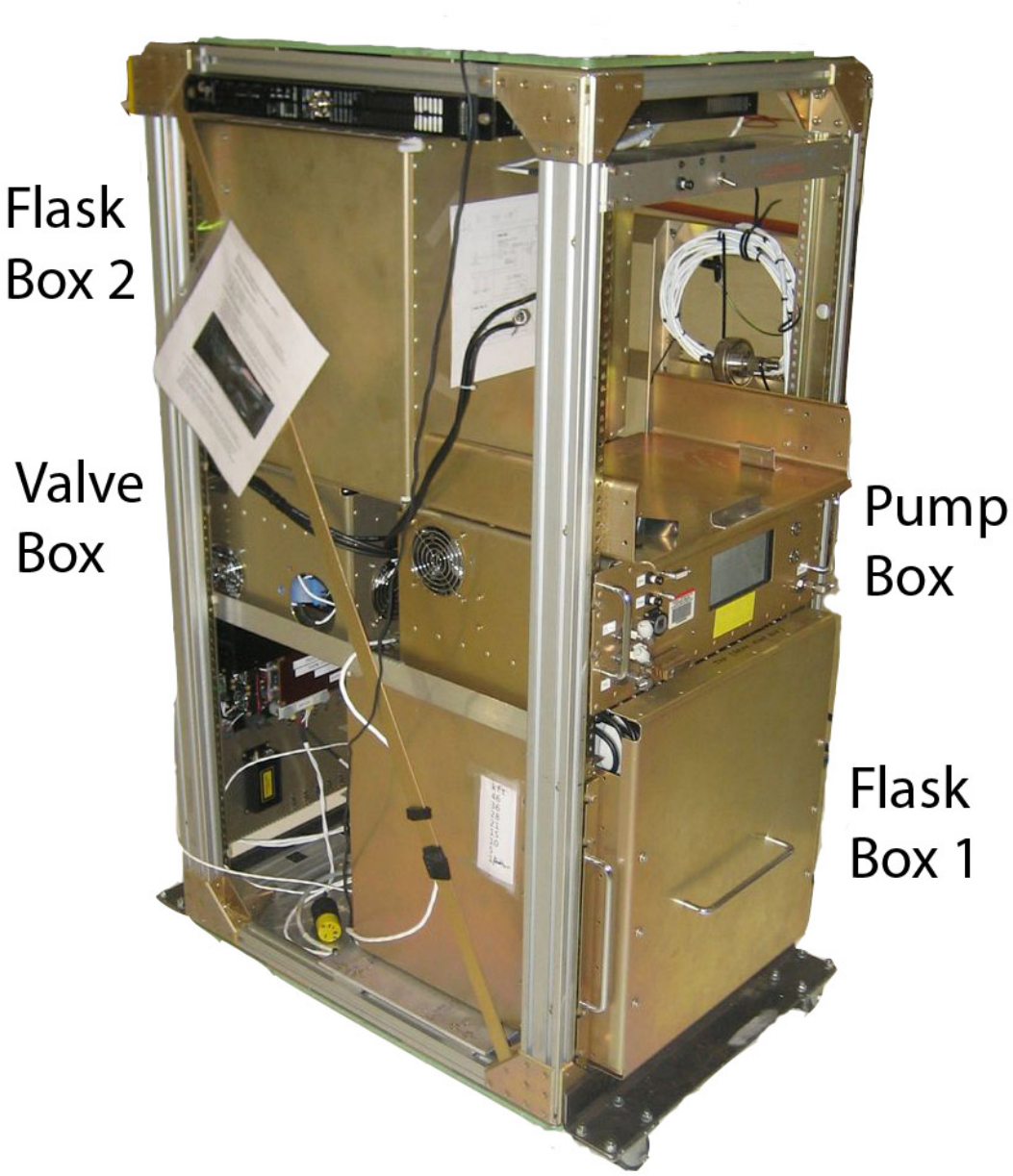

Figure 2.1. The MEDUSA rack, as stored between HIPPO missions, showing the locations of each of the four boxes. The dewar is located on the rear of the rack (at the left of the image), and can be seen in Figure 2.3.

Figure 2.2. The front of the MEDUSA sampler, as seen by the operator, showing a laptop workstation (not part of MEDUSA), the pump box beneath it (gold color) and Flask Box 1 near the floor, with splinter shield removed to show the arrangement of the flasks.

Figure 2.3. MEDUSA sampler from the rear, showing Flask Box 2 at the top (splinter shield removed to show the flasks), the valve box below it (gold color), and the dewar at the bottom left with dual cryotraps immersed in a dry ice slurry.

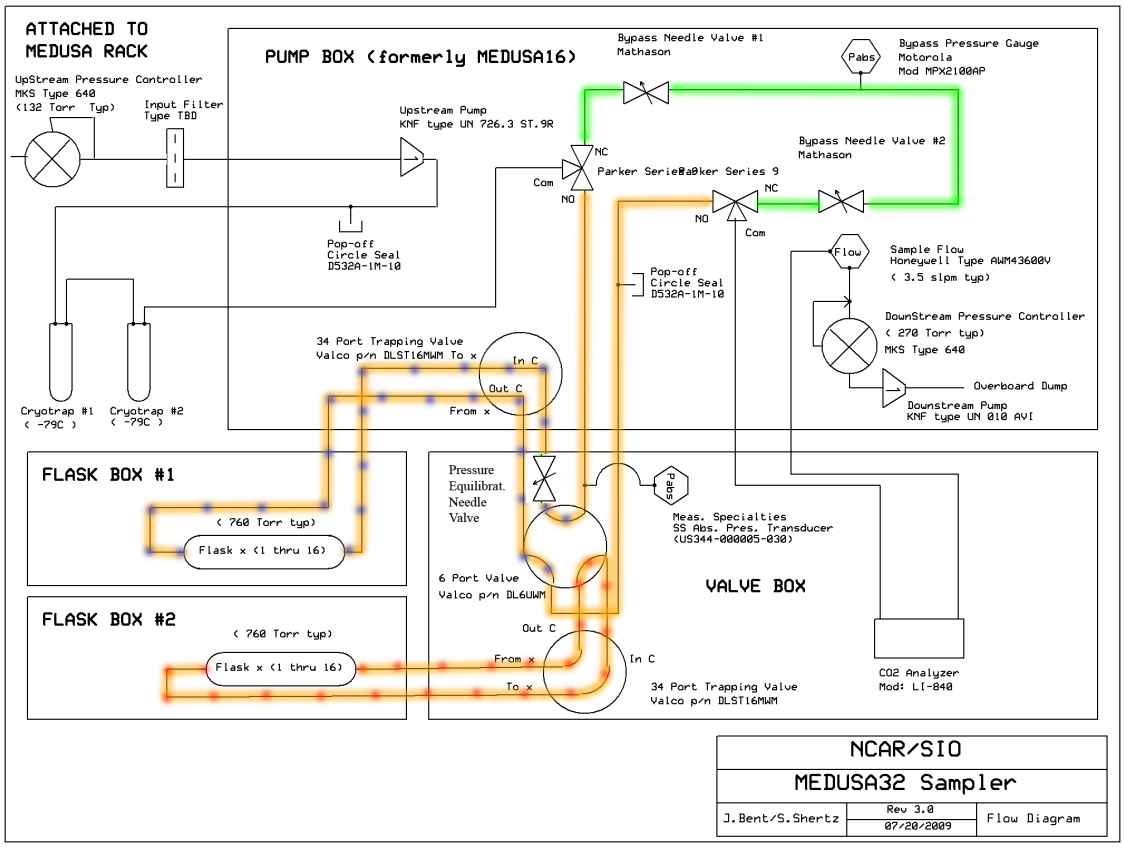

Figure 2.4. MEDUSA flow diagram. Green: bypass loop. Orange: sample loop. Blue dots: lines to/from to Box 1. Red dots: lines to/from Box 2. Schematic: S. Shertz and J. Bent.

Each MEDUSA flask (pictured below in **Figure 2.5**) contains 1.5 liters of gas at standard pressure. Air enters from the inlet port (lower in photo, shown with a blue cap on it) when the stopcock is opened to allow flow, and proceeds into the flask body. Flow exits by the diptube, which extends to the end of the flask, and out the outlet port, which is closed manually with an identical stopcock after sampling. Accordingly, air enters at the near end of the flask, and is removed at the far end to ensure that air that entering via the inlet isn't immediately removed at the outlet. Testing in a laboratory setting in 2009 showed that flasks purged with a near "perfect mixing" regime.

The flask body and valve stems are made from borosilicate, while the O-rings are made from Viton™ . Both of these materials have shown low chemical reactivity with

air, and Viton has been shown to minimize permeation effects on O_2/N_2 [*R. F. Keeling et al.*, 1998b].

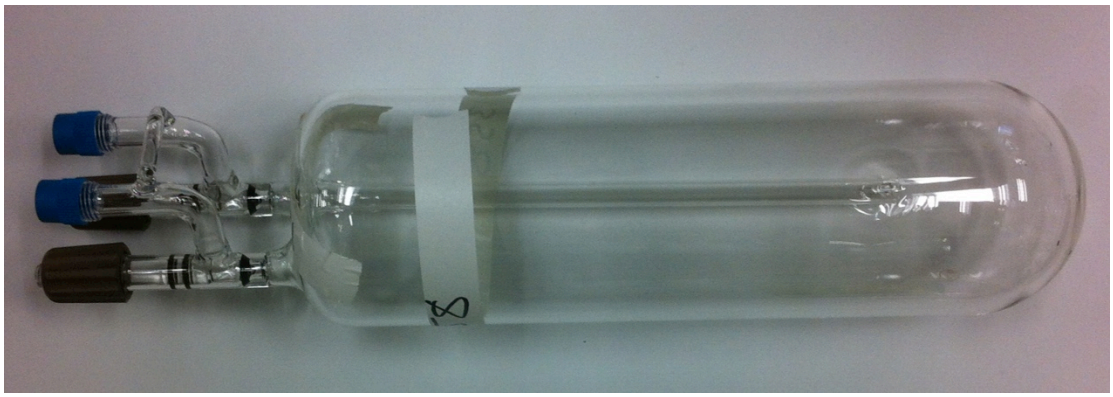

Figure 2.5. MEDUSA flask, showing stopcocks (brown handled, to the left) and diptube, which extends almost all the way to the end of the flask.

2.2.2 MEDUSA operation and electronics

The MEDUSA system is designed to minimize flask pressure/temperature fluctuations and dead volume, in order to minimize sampling artifacts and achieve the highest possible measurement precision and comparability. Enrichment or depletion of sample species may occur when transient temperature, pressure, and moisture fluctuations cause adsorption or desorption of gases such as O_2 and CO_2 from surfaces within the sample line. Similarly, fractionation, the preferential sampling of one gas species over another can occur if thermal or pressure gradients are not stable, or exist at key sampling orifices, such as system inlets and flask outlets [*R. F. Keeling et al.*, 2004]. We attempt to counteract adsorption/desorption effects by minimizing the surface area that undergoes pressure fluctuations. We achieve this by placing the first "upstream" pressure controller as close to the system inlet as possible, here within the "lasagna pan," roughly 30 cm from the HIAPER Modular InLet, or HIMIL (see **Figure 2.8**, and for an

image of the inlet itself, see Appendix 1, Figures A1-06 and A1-07). All inlet lines to and from this pressure controller are furnished in electropolished stainless steel. Furthermore, all sample-ready flasks are filled with cryogenically dried air at sea-level pressure before use, and are "pre-purged" shortly after takeoff during each research flight with dried ambient sample air, also at sea-level pressure. Finally, we place all flasks within an insulating foam to somewhat stabilize sampling temperature conditions.

MEDUSA is equipped with a bypass flow loop—whose volume is less than 1 cm^3—for periods of time when the system must maintain flow and pressure conditions, but not flush flasks (**Figure 2.4**). The bypass loop also allows the operator to briefly, and automatically, isolate the flask loops as one of two 16-position sampling valves switches from position to position, which prevents the overpressurizing of flasks, tubing and valves. Downstream of the cryotraps, dry sample gas is routed to a three-way "upstream bypass" solenoid valve. During "sampling mode" (in **Figure 2.4**, orange highlighter, either to Box 1 (blue dots), or Box 2 (red dots)), this valve sits in its normal position, and allows air to flow through the normally open port to the 6-port valve (which selects one or the other flask box), to a 16-position (34-port) valve, through a flask, back through the 16-position valve, back through the 6-port valve, and back to a three-way "downstream bypass" solenoid valve. From here, the air flows to a LiCor 840 single-cell CO_2/H_2O sensor, through to the downstream flow meter and back-pressure control loop, through to the downstream pump, and out to the communal air waste.

In bypass mode (highlighted in green in **Figure 2.4**) the upstream and downstream bypass solenoid valves are energized, diverting sample air away from the various sample valves and flasks, and redirecting it through a small volume to the downstream bypass

solenoid, and on to the LiCor and downstream hardware. Between samples, this switching isolates the sample loop for roughly 5 seconds, enough time to move one or more of the sample valves to direct air to the next flask. In order to avoid large changes in the pressure of the system, the bypass loop is equipped with a pressure sensor and two needle valves which are adjusted at the beginning of a campaign to ensure that the flow impedance of the bypass loop matches the average impedance across the flask sample loop.

The electronics of the MEDUSA system are designed to control the majority of system operations automatically, allowing for one-touch operator sampling intervention. The operator communicates with the system computer through the front panel of the pump box, seen in **Figure 2.6**, using the two silver buttons at the right of the display to toggle between operation menus, and to alter system parameters such as pressure control set points, sample mode and clock time. The majority of interfacing with the instrument happens during flask swap-out and system leak testing on maintenance days and before research flights. Once a research flight has started, the operator need only start a pre-purge cycle to flush all flasks with sample air, and then push the sample button at the end of each sample flush throughout the flight.

Certain important controls are also located on the front panel to simplify diagnostics, and to allow for intervention if necessary. These include power breakers, and the control switches for the pressure controllers, seen in the upper left of **Figure 2.6**. A large D-type connector at the lower left allows a cable to be connected to the box, enabling direct communication between the onboard computer and a laptop to allow the operator to download system metadata at the end of a research flight, while a second

smaller connector inside the box allows the operator to alter the software that runs the sampler. MEDUSA metadata are collected variously by the onboard computer, by the aircraft data system as a collection of analog signals, and by the operator, and are merged at a later date.

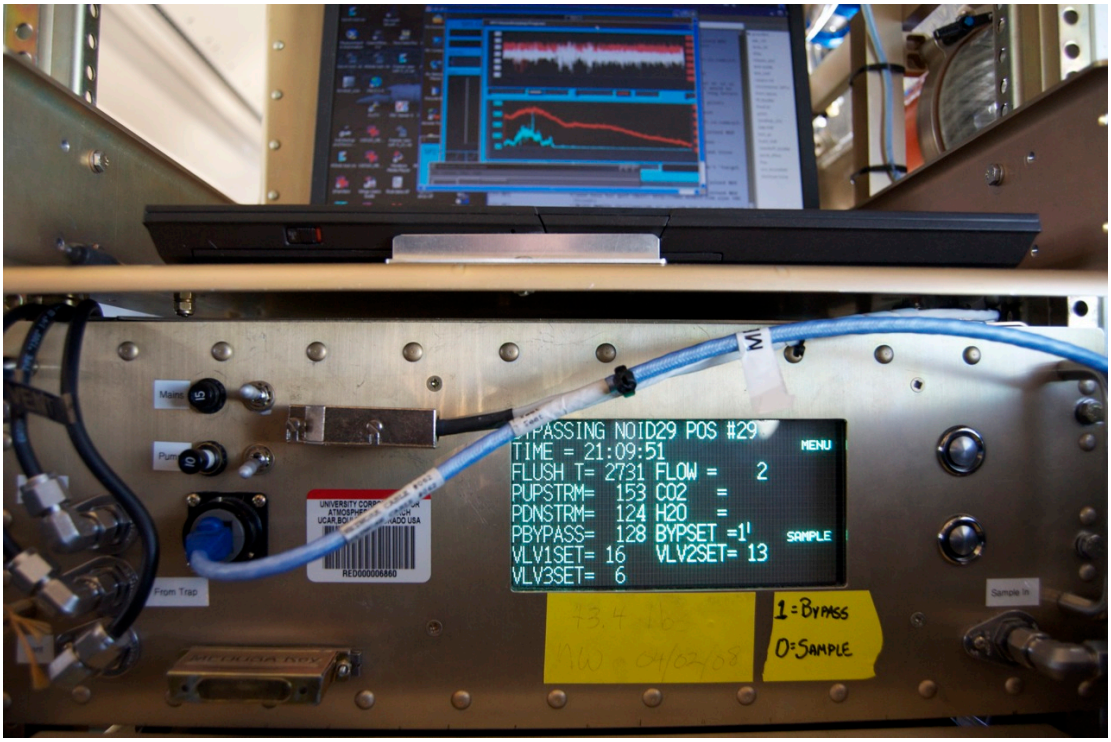

Figure 2.6. MEDUSA Pump Box as seen by operator. Select signals are displayed on the pump box screen, while others are visible only through the AEROS (onboard data system software) interface on connected laptops.

2.2.3 MEDUSA history and alterations

MEDUSA first flew on the University of North Dakota Citation II in the COBRA-2000 and COBRA-2003 campaigns [*Kort et al.*, 2008]. It was repackaged for START-08, and then expanded to hold 32 flasks, rather than 16, for the HIPPO1 mission. The redesign of MEDUSA after START-08 involved the installation of a second 16-position valve to fill the second box of flasks, and a 6-position valve to allow for

switching of airflow from Box 1 to Box 2. The Valve Box depicted in the MEDUSA flow diagram (**Figure 2.4**) was created to house both of these valves, and a LiCor 840 CO_2/H_2O sensor used primarily to detect leaks during pre-flight and operation. The Valve Box is located behind the main Pump Box, and can be seen in **Figure 2.2**.

HIPPO1 gave the MEDUSA sampler its first exposure to polar and tropical conditions, and illustrated how difficult sampling these diverse environments could be. We noted that the first of the two cryotraps was routinely plugging up at a thin section of tubing at the inlet to the first trap during tropical flights due to higher moisture content in the ambient air. So, shortly before HIPPO2, we replaced all of the original thin, inset cryotrap O-rings with thicker O-rings to help thermally insulate the upper portion of the trap from the lower temperatures of the slurry beneath. We also installed a small heater on the first cryotrap to maintain a controlled 4.5˚ C temperature at the inlet diptube.

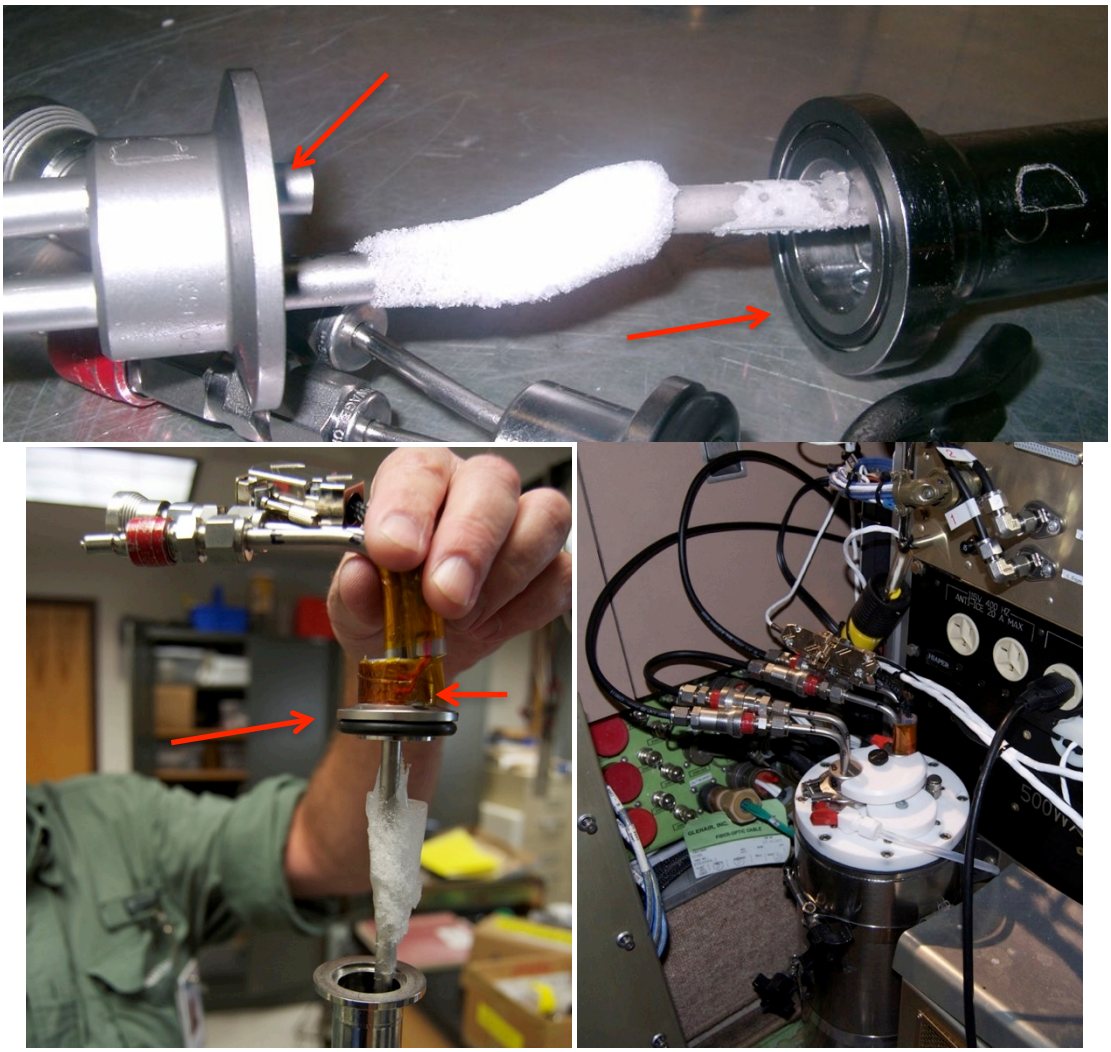

Figure 2.7. TOP: First generation MEDUSA trap design, showing small, custom inset O-ring groove and O-ring (right hand side) and unheated stainless steel inlet line (top left). BOTTOM LEFT: Second generation MEDUSA trap showing thicker O-ring. Shown is the upstream trap, which is equipped with a small surface heater (pointed to), seen encased in orange insulating tape near the top of the image. This maintains a non-freezing temperature at the short inlet diptube (obscured behind the longer out diptube) to prevent this area from freezing and stopping system flow. Both changes were instituted between HIPPO1 and HIPPO2. BOTTOM RIGHT: Second generation cryotraps installed in MEDUSA dewar. Sample air enters the upstream trap at the right from the pump box (not shown), circulates through the first trap and continues via a jumper to the downstream cryotrap at the left, and finally returns to the pump box.

Figure 2.8. “Lasagna pan” at the bottom of the MEDUSA HIMIL, as seen by the sampler’s operator from the cabin. Air passes from the HIMIL inlet along electropolished ¼” stainless steel tubing to pass-through bulkhead fittings on the right, and to the upstream pressure controller (PC). All sample air downstream of the PC is kept at a constant pressure, which is determined by one of 2 flow regimes (high flow/low alt; low flow/high alt).

HIPPO1 Ar/N_2 values also exhibited more scatter than expected—based on flasks collected routinely at ground stations in the Scripps O_2 Program—particularly in the first of the two 16-flask boxes (seen at bottom in **Figure 2.2**). Previous studies suggest that the mechanism behind these issues might cause similar aliasing of O_2/N_2 data [*R. F. Keeling et al.,* 2004]. In the absence of a demonstrated culprit, we chose to improve an area of the system that had the potential to cause the observed effects—the 15 feet of tubing between the aircraft bulkhead and the upstream pressure controller, where pressure swings might cause adsorption and desorption of gas species. During HIPPO1, the upstream pressure controller was located in the pump box, so before HIPPO2, we moved it to the aircraft bulkhead (light gray box in **Figure 2.8**, above), reducing the pressure-unregulated volume

by roughly 90%. At this time, we also replumbed the inlet with electropolished stainless steel tubing to minimize surface effects. Finally, we switched from plumbing alternate flasks in by the dip tube and out by the diptube, and instead plumbed all flasks out by the diptube. This last measure theoretically serves to minimize potential for fractionation at the flask outlet: the diptube opening is deep in the insulated flask box where it is less likely to experience temperature gradients than the valve-end of the flask, which is exposed to cabin air. These changes did largely result in more consistent behavior at each flask port in flask box 1. This difference can be seen between HIPPO1 and HIPPO2 Ar/N_2 distributions by flask position in Appendix 2 (A2-05 and A2-06). Additional fractionation problems in all HIPPO missions are discussed later in this chapter.

2.3 MEDUSA preflight and in-flight operation, sampling strategies

For HIPPO, each research flight cycle consisted of a maintenance day, during which flasks were switched out and tested for leak-tightness, a 2-hour preflight period for system preparation, and six to nine hours of research flight.

During the maintenance day, flasks that were closed after sampling from the previous flight were carefully removed and replaced with a new set. The cryotraps, which were removed at the end of the previous flight, were replaced with new, dry traps. To test for the leak-tightness of both the seal to the Synflex^{TM} (minimally reactive tubing material) lines, and of the flasks' internal O-rings, all flasks were left closed, and the entire system was evacuated by closing the inlet, running the pumps for several minutes, and rotating the 16-position valves to evacuate the upstream and downstream portions of

the sample tubing. Pressures were recorded in real time by the onboard data-logging software. The pumps were then turned off, the inlet and outlet were closed, and the system was allowed to sit for one or more hours in bypass mode. After an hour or more, the system was restored to sample mode, and the valves were stepped every ten seconds from position to position. Leaks in individual flasks or individual flask loops manifested as excursions from a system background pressure. The operator could subsequently investigate whether the leak was occurring at the flask stem (inlet/outlet) or internally at a stopcock O-ring.

The next day, during "pre-flight" instrument preparation, dry ice was loaded into the cryotrap dewar. MEDUSA flasks were opened, and the system was turned on in bypass mode to determine whether system pressures were as they should be, and whether sample air was flowing properly. The operator then set up the operating program to run the flask "pre-purge sequence" at the beginning of the flight.

Immediately after takeoff, the operator initiated the flask pre-purge sequence. The program purged two volumes (i.e. 3 liters) of dry sample air at the expected sample pressure and temperature through each of the 32 flasks in sequence. At the end of the pre-purge sequence, which took roughly half an hour, the system returned the air flow to the first flask and switched to sample mode. From this point on, samples could be taken after roughly 1-2 minutes of purging. The flasks were isolated by the movement of the 16-position valves. During the first two HIPPO campaigns, flask stopcocks were closed at the end of the flight, but during HIPPO3-5, flasks were typically closed within a minute of the isolation of the sample to minimize the effect of any leaks that might be causing

fractionation of the sample. An investigation of fractionation effects can be found in Section 2.6.1.

It is important to note that the timing, pressure set points, and flows of the system were not strictly fixed. Small changes on order 1-10 torr in the calibration of the upstream and downstream pressure controllers were typical over the course of a mission that ran from poles to tropics, and these could not be adjusted and recalibrated daily because of the time required to do so. Instead, the operator noted the system flow and pressure at the beginning of every flight to determine whether it was necessary to change the upstream and downstream set points to return pressure and flow variables to an acceptable range. These set-points were typically determined at the beginning of a given mission and adjusted up or down once or twice, as needed, later on in the mission.

2.3.1 Sampling strategy

Most system metadata was recorded by either the MEDUSA onboard computer (sample times, purge durations, date) or by the aircraft data system (real-time flow, position, sample pressure, CO_2, H_2O, elevation, etc., all at 1-Hz). The sampler operator nonetheless recorded a back-up version of most of these numbers at the time of flask collection in case of digital data loss. The operator was also responsible for determining the optimal distribution of flasks throughout the flight so as to capture the entire meridional profile, and features of particular scientific interest. To this end, a few flasks were typically sampled as soon as possible after the initial pre-purge sequence while the aircraft was descending from an initial “deep profile” (maximum elevation). Subsequent

flasks were sampled on roughly one of every three profiles (half saw-teeth) during the remainder of the flight (see **Figure 1.3**, or **Figure 3.3**), and on the final "deep" profile. For consistency's sake, the operator typically sampled at the minimum altitude of roughly 300 m, and at roughly 1.5 km, 3 km, 4.5 km, 6.5 km, and 8 km. During "deeper" profiles, the operator typically sampled at a mid-height of roughly 10 km, and the maximum altitude of 12-15 km. When targeting a particular altitude, the operator typically waited for 30-45 seconds (i.e. ~1 full purge of the flask at typical flow rates) after the plane passed the altitude of interest so as to allow air from that altitude to reach and fill the flask.

The operator typically chose from one of two appropriate flow regimes, depending on the operating altitude of the aircraft: low inlet pressure, low flow for high altitude; higher inlet pressure, higher flow for low altitude. The choice of two regimes reflected the desire to sample most flasks at the highest possible flow rate, but the necessity that the flow rate had to be dropped at high elevations, where pulling sample air into the aircraft became much more difficult for the inlet pump. The high flow regime was usual, while the low flow regime was reserved for deep profiles into the upper troposphere and lower stratosphere. Minimum fill time for flasks changed accordingly, and was controlled by the MEDUSA computer. During high flow, system flow rate was nominally 3.3 SLPM, and samples could be isolated after roughly 150 seconds, or 5.5 volume flushes. During low flow, system flow rate was nominally 1.8 SLPM, and samples could be isolated after roughly 300 seconds, or 6 volume flushes. These were minimal flush times and many flasks were flushed for considerably longer if there was an intentional pause between samples. The operator waited 10 minutes after switching

between high and low flow modes to allow the system to "re-equilibrate" before sampling in case changes to system pressure had caused ingassing or outgassing of sample gases to/from sample cell surfaces.

2.4 Analysis and data reduction

Sealed flasks were removed from the MEDUSA rack the day after a given research flight, and packed into shipping boxes. These were sent to the Scripps Institution of Oceanography nominally within a week, where they were stored in a temperature-controlled space. Members of the Scripps O_2 Program typically analyzed flasks within a month of collection, though some were run up to four months after collection due to competing demands on the analysis rack.

Analysis proceeded in four stages. Though only the first stage pertains directly to the methods associated with this dissertation, the remaining steps have not yet been characterized in literature, so I mention them here very briefly.

First, up to 15 MEDUSA flasks at a time were attached to the Scripps O_2 Program's analysis rack. The stems of these flasks were closed off, evacuated, and then the stopcocks were opened to backfill the stems with flask air. Flasks were then allowed to sit for half an hour to come to an equilibrium temperature. Each flask was then purged from the end of the flask forwards (in through the dip tube) with a known "push" gas at a rate of 30 sccm for 5 minutes. Due to the slow flow and the design of the flasks, this procedure operated in a largely plug-flow regime, minimizing sample dilution. The extracted air passed first through a LiCor 6252 non-dispersive infrared CO_2 analyzer,

then to a pick-off T, where a small portion of the sample was directed through an ISOPRIME sector-magnet mass spectrometer to measure O_2/N_2, Ar/N_2 and other isotope ratios using a differential measurement technique first described by Bender [1994], and adapted for Scripps by R. F. Keeling [2004]. The remaining portion of the 30 sccm flow was wasted to a vacuum pump. The subsequent flask analysis software looked for consistent peaks, and flagged any flasks whose CO_2 concentration started to drop during analysis. CO_2 concentrations were reported in parts per million by dry air mole fraction, while all ratiometric measurements were reported in "per meg", the relative deviation in parts per million of a sample from a known reference standard [*R. F. Keeling*, 1988], by the equation:

$$\delta(O_2/N_2) = \left(\frac{(O_2/N_2)_{sample}}{(O_2/N_2)_{reference}} - 1\right) * 10^6 \qquad \textbf{Eq. 2.1}$$

Second, flasks were allowed to sit overnight to allow the "push" gas to mix with the remaining sample in the flask. For subsequent analyses of the remaining flask air it was necessary to determine the amount of dilution by the push gas. To do this, the flask was sampled again on the same rack with only the outlet stopcock open, and without a "push" gas, to determine the CO_2 concentration after dilution. Subsequent isotopic measurements could then be adjusted by the equations (^{13}C shown as an example):

$$F_{push} = \frac{(CO2_{sniff} - CO2_{sample})}{(CO2_{push} - CO2_{sample})} \qquad \textbf{Eq. 2.2}$$

$$\Delta\delta^{13}C_{sample} = \frac{\left(\delta^{13}C_{sniff} - F_{push} * \delta^{13}C_{push}\right)}{1 - F_{push}} \quad \textbf{Eq. 2.3}$$

where "CO2" and "δ^{13}C" refer to the CO_2 concentration in ppm and δ^{13}C value in per mil of *sample* (undiluted), *sniff* (sample diluted by push gas) or *push* (push gas). *F* refers to the fraction of the total sniffed sample which came from the push tank. The push tank isotopic and total CO_2 concentrations ($CO2_{push}$ and $\delta^{13}C_{push}$) were well-characterized, and the CO_2 concentration of the push gas (~350 ppm) was well below ambient values (~390 ppm at the time) to allow precise determination of the dilution factor. This technique was first used during the COBRA campaign [*Kort et al.*, 2008] for subsequent analysis by NOAA and University of Colorado INSTAAR laboratories.

Third, flask isotope concentrations for equation Eq. 2.3 were determined in the Scripps Carbon Dioxide Monitoring group's laboratories: the CO_2 from each flask was cryogenically extracted into a glass flame-off tube using liquid nitrogen, and was subsequently analyzed on an OPTIMA mass spectrometer which measured $^{13}C/^{12}C$ and $^{18}O/^{16}O$ ratios of the sample [*Guenther et al.*, 2001]. Finally, all remaining sample was frozen back into a second flame-off tube, and stored for future ^{14}C analysis.

2.4.1 Analytical technique and jog integration of samples

All sample measurements performed by the Scripps O_2 Program are characterized relative to a hierarchy of gas references which relate the span gases used to analyze the sample back through secondary references, and ultimately to long-term primary references [*R. F. Keeling et al.*, 1998b]. LiCor 6252 CO_2 measurements (short-term

reproducibility of 0.2 ppm [*LiCor*, 1996]) are first characterized by the delta (measured signal difference) between the sample and a working gas, which flows before and after each sample to provide a "baseline" value (green and red points, respectively, in **Figure 2.9**). The raw voltage of this baseline working tank changes slightly with time in response to LiCor drift, so frequent comparison cancels this drift. The average of the "delta" between the sample and the front and back baselines is then related to the measured difference between the same working tank and two span tanks, one high and one low, as characterized just before the flask samples are run. This second "delta" is preserved, and the ultimate measured value of the sample is calculated based on longer-term trends in the value of these span tanks relative to well-established primary and secondary tanks.

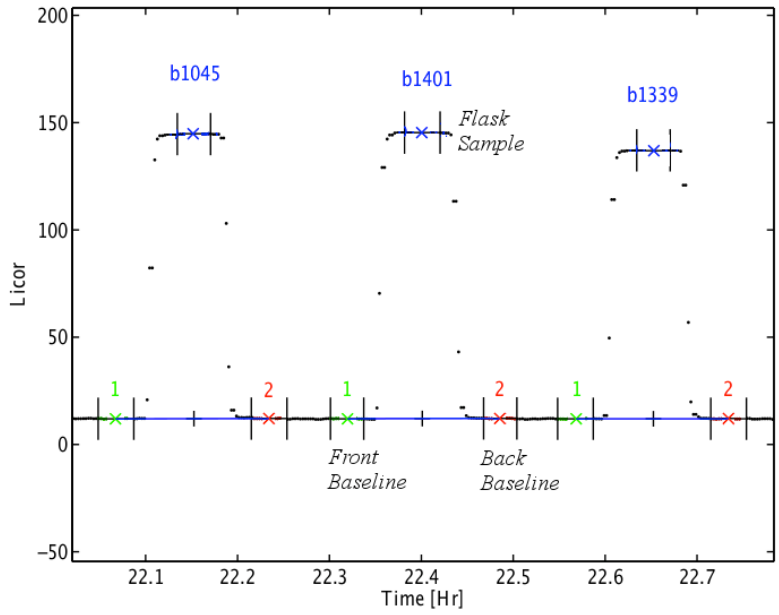

Figure 2.9. LiCor raw signals (mV) for a portion of a typical HIPPO MEDUSA flask analysis run (here from HIPPO2, RF03, showing three flasks), as shown by the Scripps O_2 Program jog integrator analysis software. Dots (which resemble lines at this scale) show LiCor CO_2 measured values in voltage units against time. Blue letters and numbers indicate the ID of a flask being analyzed. Green and red "1" and "2" markers show the "front" and "back" baseline measurements of the working tank, respectively. In each case, an "x" indicates the mean value of the accepted points. The vertical black bars on either side of each segment of the run bracket the data whose values have stabilized sufficiently to calculate an average. Points that fall within these ranges acquire the color associated with the portion of the run, though this is difficult to distinguish at this resolution, and is more apparent in Figure 2.10. A single "measurement" is constituted by the front baseline (green), sample (blue) and back baseline (red), and is adjusted by the blue line, which shows the short-term movement (over a few minutes or hours) of the working tank signal.

Ratiometric O_2/N_2 and Ar/N_2 measurements from the ISOPRIME mass spectrometer are characterized using the same hierarchy of standards technique, with one crucial difference. The mass spectrometer is equipped with a changeover valve just upstream of the mass spec pick-off T, which switches every 5 seconds between sample and working tank to overcome short-term instrument drift [*R. F. Keeling et al.*, 2004], so the "measured" value (i.e. the dots in **Figure 2.10**) represents the amplitude of the resulting square wave and is actually twice the delta between these two gases. The recent precision

of the O_2/N_2 measurement on the mass spectrometer is on order 2 per meg [*R. F. Keeling*, pers. comm.].

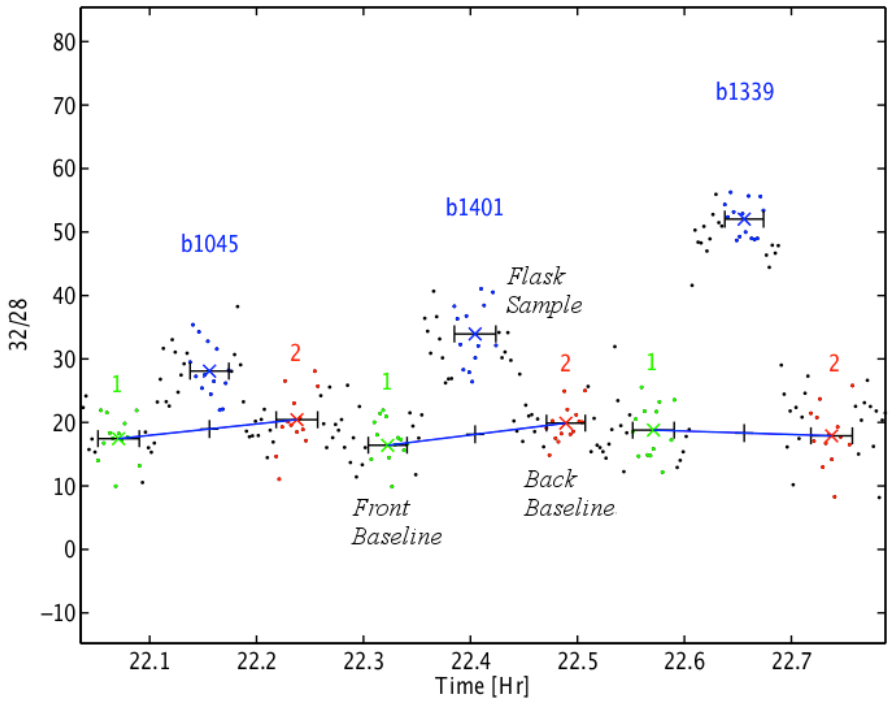

Figure 2.10. Mass spectrometer output for isotope ratio 32/28 (O_2/N_2) in unadjusted per meg units, showing the same portion of the MEDUSA flask run shown in Figure 2.9. In this particular instance the drift in the working tank baseline is much more visible than in the CO_2 signal, as the measurement is coming from the mass spectrometer (the O_2/N_2 signal to noise ratio is lower than the LiCor's CO_2 signal to noise ratio). This demonstrates more clearly the importance of the front and back baseline reference technique.

2.4.2 Merging data, kernel files

Code I wrote in Matlab for the HIPPO project imported the flask data, situational variables from the GV data system and other metadata, and performed a series of corrections and calculations specific to the operation of the MEDUSA sampler. Importantly, flasks were sampled over a time window, so they are best characterized by a weighting over time. To address this, I created a weighting kernel, a function that ascribes a weight to each second of the sampling time. Multiplying such a function by the

1-Hz metadata allowed me to compute the "representative" sample latitude, altitude, etc. The weighting function could also be applied to the output from other 1-Hz instruments over the sample time to compare flask data with higher-resolution in situ data.

To derive a weighting function, I began by calculating the average flow rate during the sample from our 1-Hz metadata. Flask start and end times were known from metadata, because valve position is recorded at 1-Hz with other variables. I masked out the first ten seconds of flow data for each flask as these represent the changeover, and a slight 3-5 second delay in air moving between the inlet and flasks, and I mask out any non-real values (i.e. NaNs in the aircraft data stream) so as not to affect the average. I then computed the mean of the remaining values. Dividing this mean by the size of the flasks (~1500 cm^3) gave the flushing time of the individual sample. Laboratory tests showed that the flasks flushed with nearly perfect mixing, so I calculated the weighting kernel by:

$$w(t) = \left(\frac{1}{\tau}\right) e^{-\left(\frac{t_f - t}{\tau}\right)} \Big/ \left[1 / \sum_{t_0}^{t_f} w(t)\right] \qquad \textbf{Eq. 2.4}$$

where *w(t)* is the weighting of any 1-second time increment *t* between the switch to the sample, and the switch to the next sample, τ is the flushing time in seconds, and t_f is the finishing (closing) time of the flask sample purge. The weight at any increment t is then adjusted up by the total weight of the function from t_0 to t_f to account for possible missing values (NANs) in the variable to which the kernel is being applied. For any given variable *C*, the weighting function was applied to calculate a "representative" value:

$$C_{rep} = \frac{\sum_{t_0}^{t_f} C(t) * w(t)}{\sum_{t_0}^{t_f} w(t)} \quad \textbf{Eq. 2.5}$$

For all C(t) values that are not NANs, where t_0 is the time the flask started purging, and where the division by the denominator reflects the need to adjust by the actual sum of weights, which may be <1 when a given variable *C* is not characterized for all time steps. The analysis scripts produced two final files—the averaging kernel and a final table of flask values. I included metadata about the time of analysis, corrections, measurement units, formatting and about the instruments and individuals involved in a NASA Ames formatted header. Flask and kernel files were later uploaded to the NCAR EOL website (https://www.eol.ucar.edu/field_projects/hippo) in a final NASA Ames format for public access, and to unite them with measurements from other HIPPO instruments. A flat text file with all MEDUSA data and with all in situ values merged in with the MEDUSA sampling kernels was made publicly available at the CDIAC website (http://hippo.ornl.gov), with the naming format:

HIPPO_medusa_flasks_merge_insitu_YYYYMMDD.tbl

MEDUSA data in this dissertation come from the most recent HIPPO Archive file on the Bluemoon server in the Scripps O_2 Program lab, as this data is updated more often than the CDIAC database. These can be found on:

/data/instrument/hippo_archive/2014-10-02-08:43/output_files/flask_data

/data/instrument/hippo_archive/2014-10-02-08:43/output_files/kernel

AO2 data in this dissertation come from the May 21, 2014 release. All data files are marked with version number “v140521”.

2.5 AO2 general description

The NCAR Airborne Oxygen Instrument (AO2) measures O_2 concentration using a vacuum ultraviolet absorption technique. AO2 is based on earlier shipboard [*Stephens et al.*, 2003] and laboratory instruments using the same technique, but has been designed specifically for airborne use to minimize motion and thermal sensitivity, and with a pressure- and flow-controlled inlet system. AO2 flew on the Wyoming King-Air during the ACME-2007 campaign and on the NCAR GV during START-08 before flying on the HIPPO campaigns. To achieve the high levels of precision needed, AO2 switches between sample gas and air from a high-pressure reference cylinder every 2.5 seconds. AO2 has a precision of ±2 per meg on a 5 second measurement. At typical HIPPO operational speeds of 150 m/s or climb/descent rates of 500 meters/minute, 5 seconds correspond to a horizontal resolution of 750 meters and a vertical resolution of 40 meters. The AO2 system, seen in **Figure 2.11**, consists of a pump module, a cylinder module, an instrument module, and a dewar. A more detailed description of the AO2 instrument and its operation is in preparation for publication by Dr. Britton Stephens.

Figure 2.11. The operator side of the AO2 instrument seen from the aisle of the GV. The cylinder box sits at the bottom of the rack, with the pump box and instrument boxes above it. The system is run by an internal computer, accessed remotely from the data display laptop shown retracted in the laptop tray, though some interaction is also possible through the front panel of the instrument box.

2.6 Correcting O_2/N_2 data: using Ar/N_2 and N_2O data to prepare a single, consolidated high-resolution O_2/N_2 product

An important application of the MEDUSA results is to help identify and correct for systematic drift in the continuous AO2 oxygen measurement. Testing has shown that the AO2 measurements have time-dependent biases that appear related to slow recovery from pre- and during-flight plumbing humidity changes and to small contributions of fractionated cabin air at or immediately downstream of the inlet. Comparisons of station, MEDUSA and AO2 O_2/N_2 values show that MEDUSA values are, indeed, closer to station values than AO2 values. The following section details the multi-stage technique I apply to correct the AO2 data to be consistent with the MEDUSA flask data. This process is complex, because I must first correct MEDUSA data for known sampling artifacts. The flow chart below (**Figure 2.12**) serves as a visual guide to the following sections, which describe the correction process.

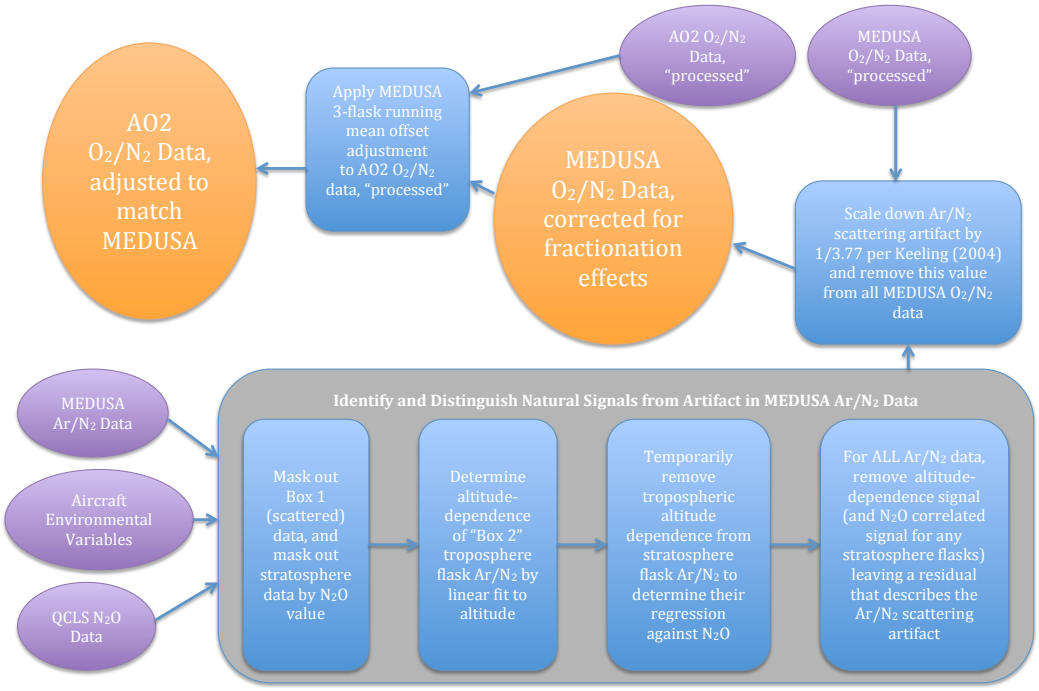

Figure 2.12. Flow chart showing the corrections required to MEDUSA and AO2 data to produce finalized "corrected" and "adjusted" data. Processes are denoted by blue squares. Data products are denoted by ovals (purple for intermediary, orange for final).

2.6.1 O_2/N_2 fractionation

Perhaps the most persistent impediment in establishing accurate O_2/N_2 measurements is the occurrence of diffusive fractionation—in response to pressure or thermal gradients [*Blaine et al.*, 2006; *R. F. Keeling et al.*, 2004; *R. F. Keeling et al.*, 1998b; *Steinbach*, 2010]. Because the magnitude of the seasonal and interannual atmospheric oxygen signal corresponds to changes at the level of parts per million in the O_2 mole fraction, against the 21% O_2 background, even very small relative changes due to sampling protocol can significantly impact findings. As a result, O_2/N_2 measurements

are subject to various sampling artifacts that are typically negligible for gases whose natural signals (seasonal, interannual etc.) comprise a greater portion of their abundance. Even larger relative effects of the same sort are typically present in measurements of the Ar/N_2 ratios.

Ar/N_2 is known to vary in several systematic ways in the atmosphere. Station measurements show the Ar/N_2 ratio varies seasonally on order 5-20 per meg because of thermally-driven exchanges of Ar and N_2 with the ocean [*Battle et al.*, 2006; *R. F. Keeling et al.*, 2004]: because Ar is more soluble than N_2 [*Weiss*, 1970], the ratio in the atmosphere increases when seawater warms, and decreases when it cools. Keeling et al. [2004] estimate an additional synoptic variability in the troposphere on order 2.5-5 per meg. Stratospheric samples from recent studies [*Ishidoya et al.*, 2008; *Ishidoya et al.*, 2013] suggest a vertical gradient in the Ar/N_2 ratio may exist in the stratosphere. All characterized natural tropospheric signals are relatively small, suggesting that seasonally-independent signals in excess of 10-25 per meg in tropospheric measurements are almost certainly artifact.

Figure 2.13 shows uncorrected MEDUSA Ar/N_2 data from all five HIPPO missions. Because I am aware that stratospheric samples may exhibit a different Ar/N_2 signature due to gravitational settling in the upper atmosphere [*Ishidoya et al.*, 2008], N_2O values from the Harvard QCLS instrument are shown by color, to give a rough sense for whether a sample is stratospheric. All 2009-2011 N_2O values are detrended relative to 2009 for consistency; accordingly, samples with values below 318 ppb, a rough cutoff for the stratosphere in 2009 based on observations of tropopause height, are shown with dark outlines. The data show considerable scatter in Ar/N_2 ratio, with a 1-sigma of ±21 per

meg, with stratospheric samples (indicated by low N_2O) showing low Ar/N_2. Locally, the stratospheric Ar/N_2 and N_2O variations are altitude dependent, but while the relationship to N_2O is preserved over the full data set, spatial tropopause height variations mask the strong altitude dependence in **Figure 2.13**. Additionally, a small altitude-dependent gradient appears present for the troposphere, with Ar/N_2 values highest at the earth's surface, and decreasing with height. Against this trend, stratospheric samples stand out clearly, as they are depleted in Ar with respect to tropospheric samples at similar elevations, giving the overall profile a striking "Florida panhandle" look. The Ar/N_2 depletion at high elevations anti-correlates well with the N_2O value of the sample, which is expected to decrease with age in the stratosphere due to photolysis and reaction with $O(^1D)$ [*Andrews et al.*, 2001]. The scatter in the tropospheric data is larger than the 10-25 per meg signals from seasonal and synoptic variability [*Battle et al.*, 2006; *R. F. Keeling et al.*, 2004], suggesting that much of the tropospheric variability around a given altitude is due to sampling or analysis artifacts.

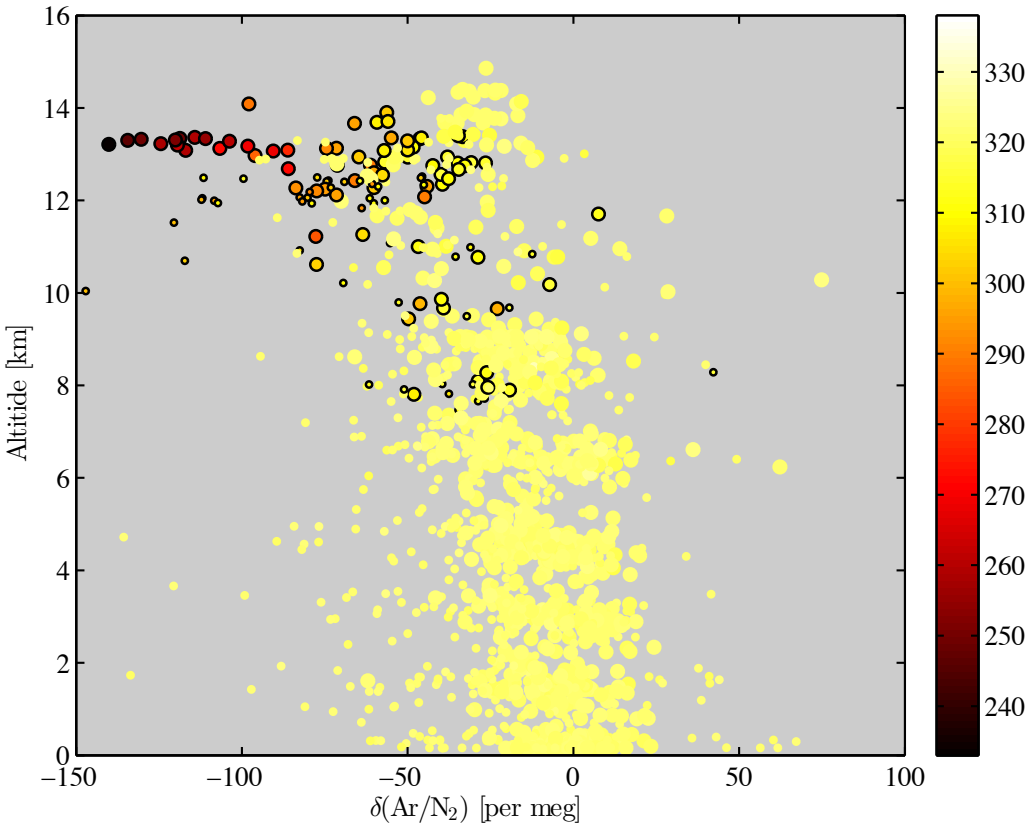

Figure 2.13. MEDUSA flask $\delta(Ar/N_2)$ by altitude. Color denotes detrended N_2O value (relative to 2009) in parts per billion, a quantity known to decrease with residence time in the stratosphere, as N_2O is consumed in the stratosphere. Small circles come from Flask Box 1 (positions 1-16, noisier data), and large circles from Flask Box 2 (positions 17-32). All flasks with normalized N_2O values of, or below 318 ppb are shown with a dark outline to indicate stratospheric influence. Strongly stratospheric samples "veer off" to the left at the top of the plot, exhibiting depleted $\delta(Ar/N_2)$ values, while non-stratospheric, high altitude samples continue the fairly linear drop-off with height seen in lower elevation flasks.

Previous studies have suggested several mechanisms that may contribute to artifacts in Ar/N_2 and associated O_2/N_2 data. Fractionation during analysis is one of these possible mechanisms, and could systematically affect flasks analyzed on certain laboratory rack ports. Analysis of the mass spectrometer rack used to analyze the flask samples suggests this is not a likely cause of the observed variation. **Figure 2.14** and **Figure 2.15** show violin plots (vertically aligned frequency distributions mirrored along their main axis) of uncorrected Ar/N_2 and O_2/N_2 data against analysis port on the mass

spectrometer rack. Small differences in mean (red cross) and median (green square) exist on order of ±5 per meg for both species, but means for the two species show no systematic correlated offsets at individual positions.

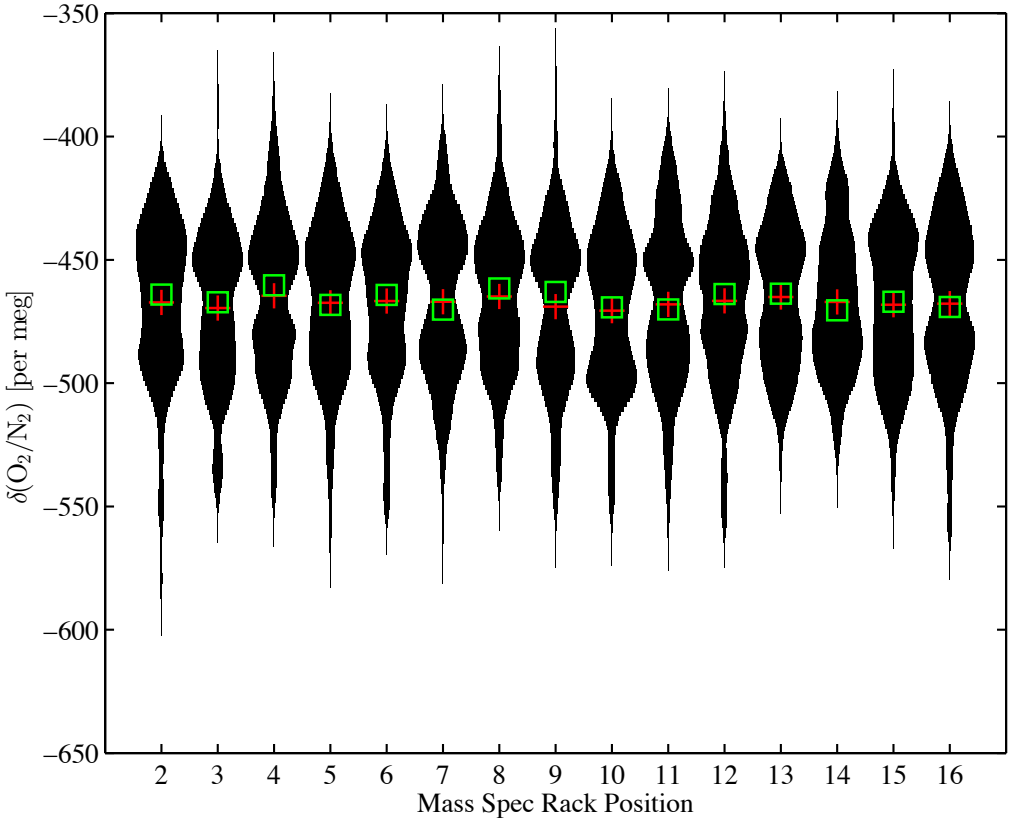

Figure 2.14. Violin plot of MEDUSA flask $\delta(O_2/N_2)$ by mass spectrometer rack position. Red crosses show mean value, while green squares show median. Each "violin" shows a scaled frequency distribution for all flasks analyzed on a given position. Position 1 is reserved for gas handling.

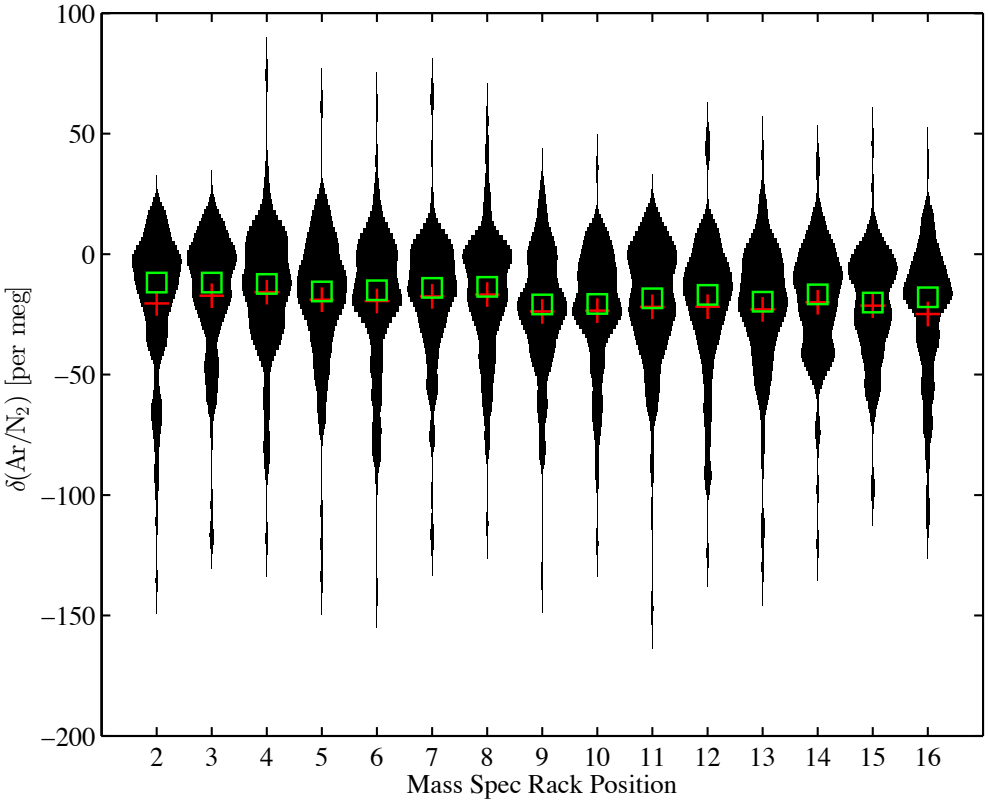

Figure 2.15. Violin plot of MEDUSA flask Ar/N_2 ratios by mass spectrometer rack position. Red crosses show mean value, while green squares show median.

A second effect [*Blaine et al.*, 2006; *Manning*, 2001] involves thermal fractionation at the sample intake, which may occur if the intake is radiatively heated or cooled relative to the ambient air. This effect is strongest with low sampling flows, and is not expected to contribute to MEDUSA measurements due to the higher sample flow of the airborne system. Additionally, flow across the sample inlet, which is receded in a HIMIL (HIAPER Modular InLet), is shielded from solar heating. Finally, the HIMIL system directs a large amount of air around the same inlet, mimicking the effects of the aspirated inlet described by Blaine [*Blaine et al.*, 2006; *Manning*, 2001], which was shown to minimize the thermal fractionation effect.

Another fractionation mechanism [*Steinbach*, 2010] is related to the pressure nozzle effect, and is caused by pressure gradients normal to the sample streamlines upstream of the inlet. The pressure gradients are induced by ram pressure, and may cause sampled Ar/N_2 to decrease with elevation for rear-facing inlets, and increase with elevation for forward-facing inlets. This mechanism may explain the gradient I see with elevation if fractionation appears to scale with ram pressure or airspeed, however, HIMIL inlets are designed to slow flow internally which would reduce the occurrence of such fractionation effects at the internal inlet, and ram pressure during HIPPO did not vary systematically with altitude. Though the pressure nozzle effect could conceivably be responsible for vertical gradients, it doesn't explain the noise around a mean seen in the Ar/N_2 data, or the Ar/N_2 depletion in the stratosphere relative to tropospheric samples at the same elevation, and in Chapter 4, I show that neither airspeed nor ram pressure correlates with Ar/N_2 depletion in a statistically significant way.

A fourth mechanism, plumbing leaks, could fractionate sample gas to cause either enrichment or depletion of sample Ar/N_2 [*R. F. Keeling et al.*, 2004; *R. F. Keeling et al.*, 1998b]. In HIPPO, the tubing upstream of the inlet pump was sub-ambient, while the plumbing between the inlet pump and flasks was super-ambient. Pressure-induced fractionation upstream of the inlet pump would manifest equally in both boxes as a depletion of Ar relative to N_2, as an inward leak would preferentially introduce lighter molecules to the sample gas. Such a leak would also show an enrichment of CO_2 due to human respiration and dry ice sublimation in the aircraft cabin (measured at 600-1000 ppm CO_2) and depletion of O_2 due to human respiration and fractionation.

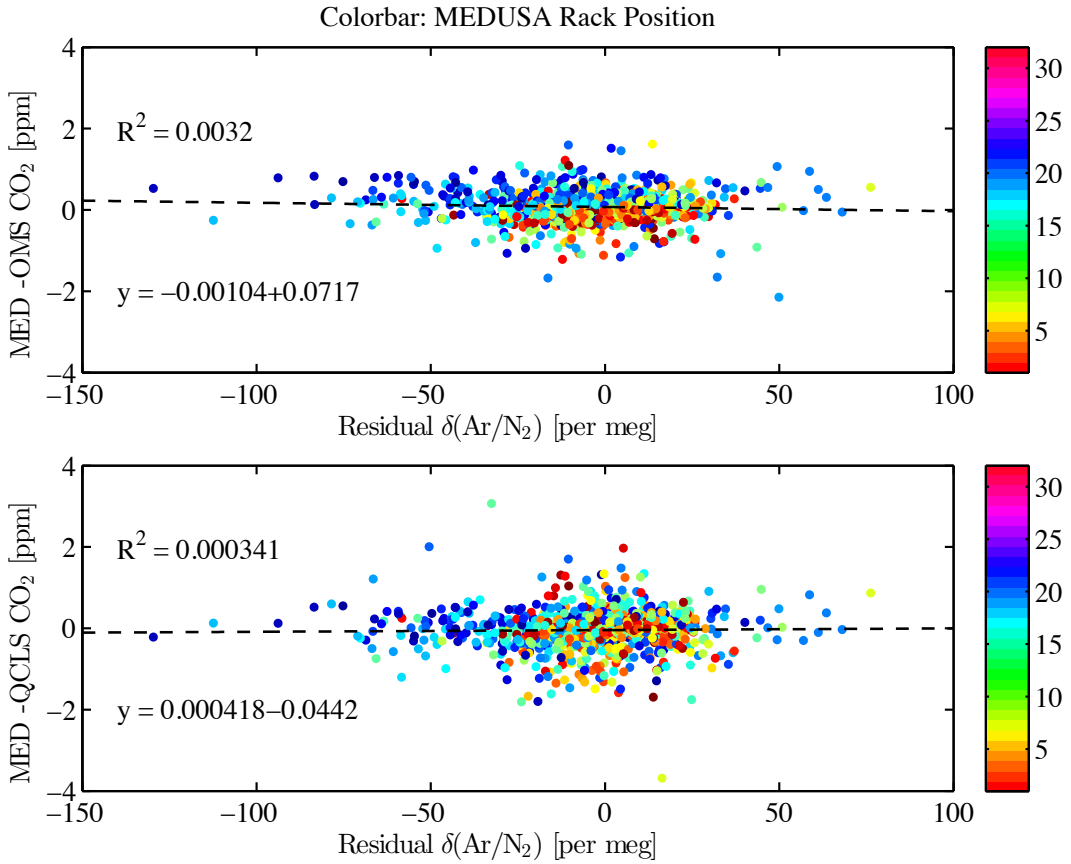

Figure 2.16. Difference between MEDUSA CO_2 and OMS (top) and QCLS (bottom) CO_2 measurements plotted against tropospheric $\delta(Ar/N_2)$ with vertical gradient removed. A strong negative correlation would indicate an inwards leak of high-CO_2 air to MEDUSA flasks correlated with a low $\delta(Ar/N_2)$ ratio. However, correlations are statistically insignificant for both comparisons.

However, comparisons of tropospheric Ar/N_2 residuals (i.e. with the observed vertical gradient removed) against the difference between MEDUSA and other (Harvard QCLS and OMS instrument) CO_2 measurements suggests no statistically significant correlation, suggesting that depleted Ar/N_2 values are not the result of inward leaks, though analysis by Keeling et al. [1998b] suggests such an effect might be hard to observe in CO_2. Furthermore, because cabin pressure decreases with altitude, the magnitude of the leak would be expected to decrease with height if it related to the section of tubing between the inlet pressure controller and upstream pump, not increase.

An outward leak between the six-port valve and the flasks on Box 1 would favor the loss of N_2 to the cabin, which would be worse at lower cabin pressure, which is the opposite of the observed effect at greater altitude and in the stratosphere, and extensive testing turned up no consistent leaks in this section of the system. Two tests of the inlet lines between the HIMIL and the upstream pressure controller also showed no evidence of leaks.

A final mechanism may involve thermal fractionation at the flask outlet [*R. F. Keeling et al.*, 2004] related to temperature differences between the tip of the diptube outlet and the air contained in the flask. The thermal environment in MEDUSA during the HIPPO campaign made it difficult to control the temperature gradients in the flasks during sampling. Cabin temperatures often dropped by 10° C in the first few hours of flight and an air conditioning vent at the base of the MEDUSA rack blew cold air at the bottom box throughout research flights. Differences between individual flask thermal gradients may lead to scatter, and possibly bias if the thermal gradient is consistently one-directional. Under such conditions, both Ar/N_2 and O_2/N_2 would be affected. Studies suggest that the magnitude of the O_2/N_2 fractionation signal will scale in a predictable way with the magnitude of the Ar/N_2 fractionation. Here, I use a scaling factor of 3.77, based on the measurements of Keeling et al. [2004]. I believe that this effect led to the difference between diptube-in and diptube-out flasks in HIPPO1 (see Appendix 2, A2-05). I also believe that this is a background effect that increases the overall noise of flasks from Box 1 (see **Figure 2.17**), and which causes a low bias for early Box 1 positions, which are sampled when flask temperatures are changing the most and which are located on the lower outer side of the MEDUSA rack, closest to the air conditioning vents.

Unfortunately, though a cabin temperature measurement exists for the HIPPO flights, this captures only the air temperature at one point, which is kept relatively constant by the aircraft AC system and does not define temperature gradients. The instrumentation and flasks take considerably longer to warm up or cool down due to thermal mass, an effect for which I do not have a direct measure on each research flight. However, testing during the second test flight of HIPPO4 showed a difference of approximately 6°C between the bottom (8.6°C) and top (14.3°C) of Box 1, and a difference of almost 11°C between the bottom of Box 1 and the top of the instrument rack (19.7°C).

Examination of MEDUSA Ar/N_2 data suggest a combination of effects is responsible for the variations in the Ar/N_2 signal. A systematic vertical gradient, which could be due to system inlet fractionation is apparent in all data, regardless of box. A larger gradient in the stratosphere is related to air age. A box-specific effect is also apparent when the Ar/N_2 values are plotted by box. **Figure 2.17** shows the vertical distribution of Ar/N_2 for Flask Box 1 (positions 1-16) and Flask Box 2 (positions 17-32) separately. **Figure 2.18** shows these same values by frequency. Both figures suggest that Box 1 data have greater scatter than those from Box 2, and that they exhibit a lower mean.

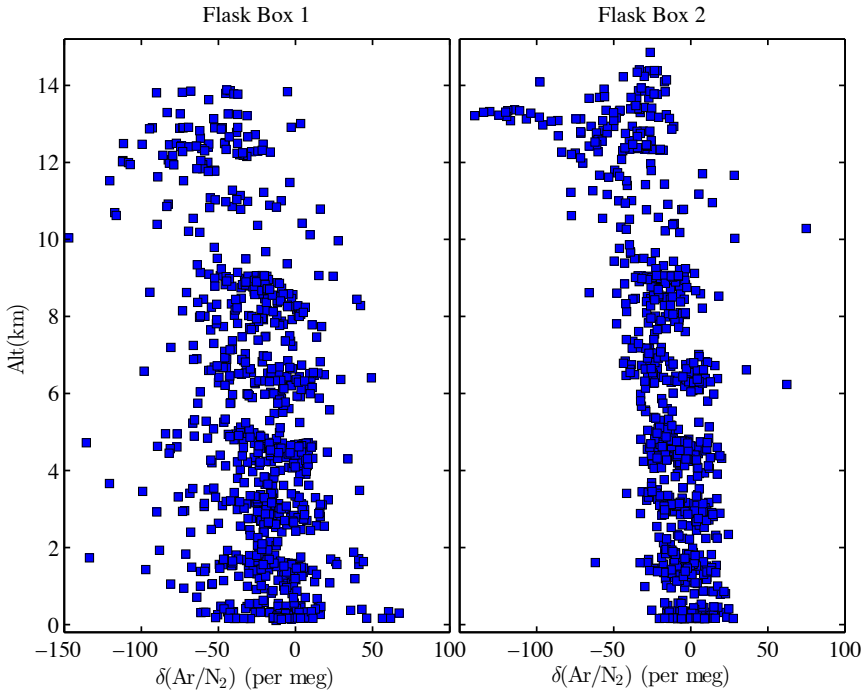

Figure 2.17. Flask Box 1 vs. Flask Box 2 $\delta(Ar/N_2)$ vs. Altitude for all HIPPO missions. Both boxes show a vertical gradient. Box 1 $\delta(Ar/N_2)$ values have a much greater spread, and a lower mean. The lowest stratospheric samples occur exclusively on Box 2 because the sampling procedures resulted in sampling the highest profile at the end of each flight, when the aircraft was lightest.

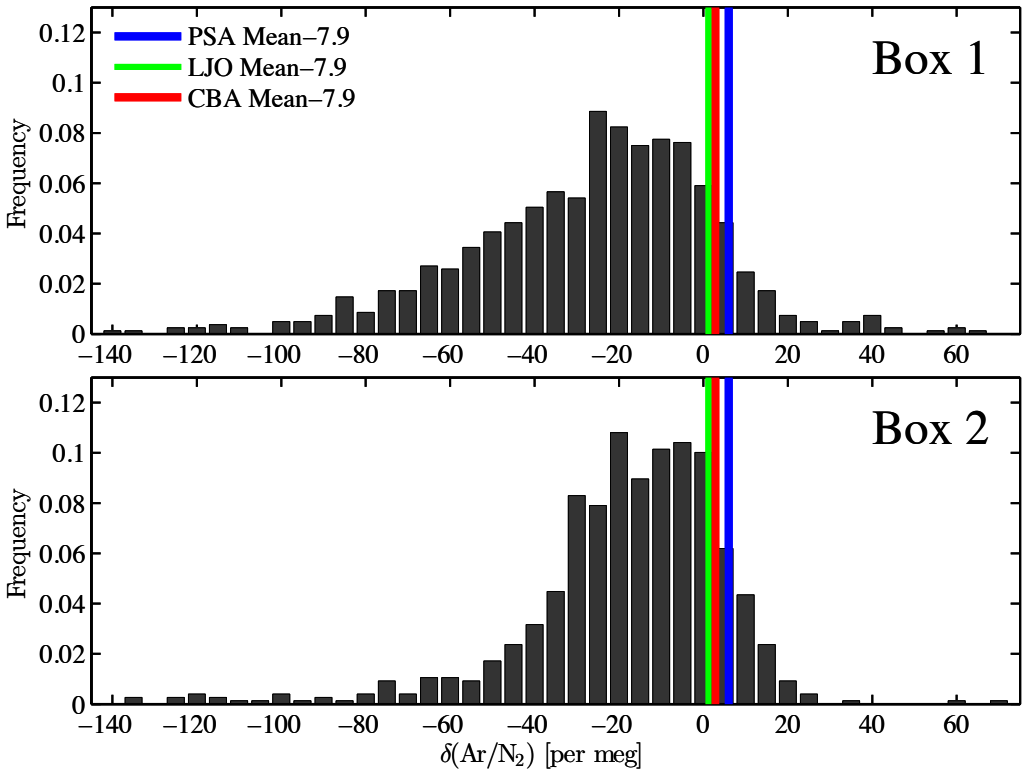

Figure 2.18. Distribution of all $\delta(Ar/N_2)$ values for Box 1 and 2 against (normalized) frequency. Box 1 values tend to center around a lower mean, with more scatter. Palmer Station, La Jolla, and Cold Bay yearly mean values are shown for reference. Some system-to-system offset (which I calculate in Chapter 4 to be 7.9 per meg (Station Round Bottom-Station B flask) exists due to the different handling of working and calibration tanks between the two flask racks in the Scripps O_2 Program lab, so I have adjusted all station values down by this offset to better compare with the HIPPO data.

Unfortunately, because the errors and biases associated with fractionation effects are masked by vertical, stratospheric and seasonal Ar/N_2 signals, I was only able to diagnose all the contributing factors after the entire HIPPO campaign was over. However, knowledge of these competing effects allows me to correct for them reasonably well.

Figure 2.19 below shows a violin plot of the distribution of tropospheric $\delta(Ar/N_2)$ at each of the 32 MEDUSA positions after the altitude dependence has been removed (as detailed in 2.6.2). I exclude stratospheric samples as this signal can bias the mean and is

less easily removed to allow background scatter to be visible. Mean (red crosses) and median (green squares) values for each position are shown on top of the distribution. Removing the altitude dependence from the signal shows the relatively low scatter in Box 2 flasks, and shows that Box 1 flasks tend to show a time-in-flight dependence that I believe is due to temperature equilibration as the MEDUSA rack and flasks cool down to a stable temperature over the first few hours of the research flight. Flask positions 1-4 are also along the outer edge of the lower flask box, and thus most influenced by the air conditioning vent at the outer floor of the GV cabin. The degree of this cooling, which varies from flight to flight, contributes to the larger scatter in the Ar/N_2 signal in Box 1.

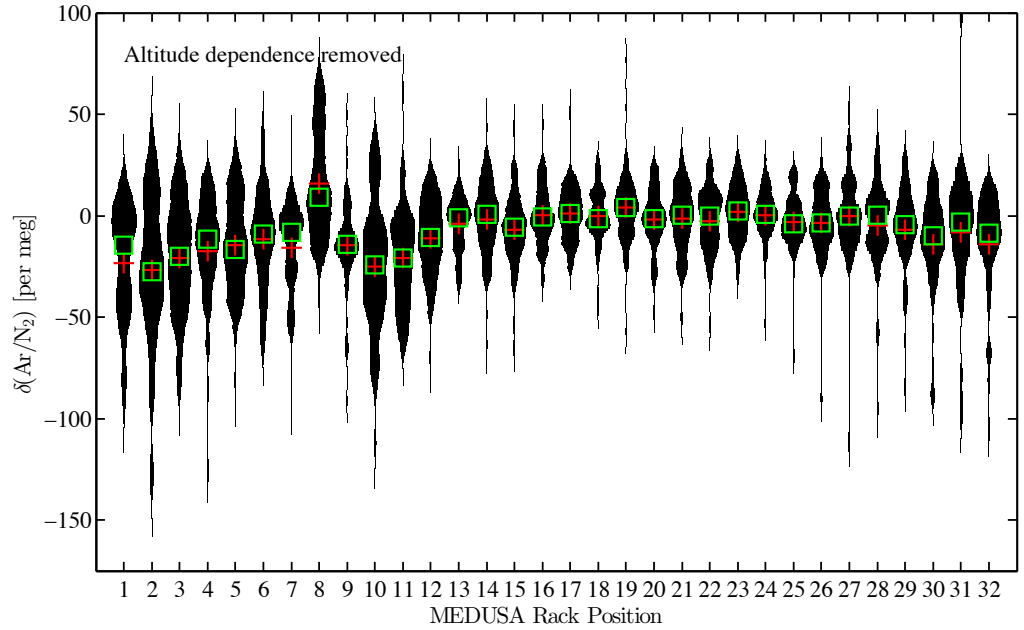

Figure 2.19. Violin plots of tropospheric $\delta(Ar/N_2)$ data seen with the altitude-dependence (discussed in coming sections) removed to show the residual scatter. Box 1 values (positions 1-16) show a lower mean and greater scatter relative to Box 2 (positions 17-32).

One way to determine whether the Box 1 low-biased flasks are low because of thermal diffusive fractionation effects is to compare the O_2/N_2 measurements between MEDUSA and AO2 to see whether the offset in O_2 within a given flight relates to the magnitude of the Ar/N_2 depletion signal by the scaling factor mentioned above from Keeling et al. [2004]. The figure below shows this relationship by plotting the difference between AO2 in situ O_2 measurements and MEDUSA flasks against the Ar/N_2 signal. The expected slope indicating fractionation is shown in cyan. Individual flask points are plotted as numbers and letters to distinguish them by position. Positions 1-9 are marked as numbers, while subsequent positions are marked with the letters A-W. If low Ar/N_2 in early flasks relates to low O_2/N_2, the two will show a related fractionation scaling of roughly 1/3.77 (0.26). This figure and similar figures for other flights show that early

tropospheric flasks (shown in green, not gray) tend to fall along the expected relationship line, suggesting that the flask depletions are consistent with fractionation, and may be corrected for accordingly.

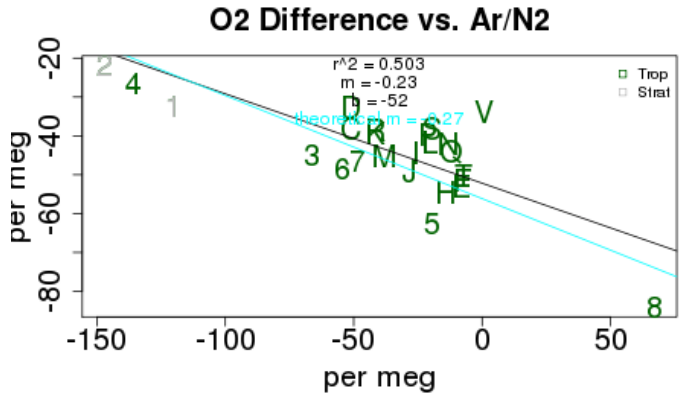

Figure 2.20. Regression of HIPPO3 RF03 $\Delta[\delta(O_2/N_2)_{AO2} - \delta(O_2/N_2)_{MEDUSA}]$ vs. MEDUSA $\delta(Ar/N_2)$.The average relationship for all 32 flasks is shown as a black line, while the expected slope of 1/3.77 (0.26) is shown in cyan. Flasks positions 1-9 are marked as numbers, while subsequent positions 10-32 are marked with the letters A-W. Early flasks, which show the most low bias and scatter tend to fall along the expected relationship for thermal diffusive fractionation (figure from B. Stephens).

Finally, as mentioned briefly above, a detailed analysis of the differences in Box 1 flasks by mission also reveals an additional distinct mission-related effect in HIPPO1. The residual scatter for each of the missions can be seen in Appendix 2 (Figures A2-05:A2-09). HIPPO1 flights show the most position-to-position offsets, with consistently high even position means and low odd position means. This feature is not seen in subsequent missions, and is due to the original choice to plumb odd positions diptube-out, and even positions diptube-in for Box 1, and odd positions diptube-in and even positions diptube-out in Box 2. Doing so exposed Box 1 even position outlets and Box 2 odd position outlets to the more variable, cooler temperatures at the flask outlets, while the remaining flasks' outlets sat deeper in the box where temperature was more

consistent. This marked even/odd difference led to the decision to plumb all flasks diptube-out for the remainder of the missions.

Various possible fractionation mechanisms could influence the scatter and bias seen in Box 1 data, and the vertical gradient seen in all data. For now, because the vertical gradient isn't systematic to one box or another, I have temporarily disregarded this effect, which I will discuss in the coming section and in Chapter 4. Some of the scatter in HIPPO1 flasks was due to our original plumbing convention: flasks showed exaggerated mean differences between odd and even positions because of the unstable thermal regime the flask valves experience in flight. I have also demonstrated that Box 1 and 2 have very different scatter and means. I believe this scatter is largely related to artifact. I also believe that the greater scatter in Box 1 is due to thermal equilibration effects in early flask positions. Analysis of the difference between MEDUSA and AO2 O_2 measurements vs. Ar/N_2 depletion suggests the effect comes from thermal diffusive fractionation. Fortunately, because Box 2 flasks appear to show considerably less scatter, and fewer flasks with low bias, I am able to use these flasks in the remainder of the dissertation to assess natural Ar/N_2 signals, and to determine to what extent MEDUSA O_2/N_2 data should be corrected.

2.6.2 Correcting MEDUSA O_2/N_2 data

Several recent papers [*Battle et al.*, 2003; *R. F. Keeling et al.*, 2004] have examined Ar/N_2 data in detail; by establishing a relationship in the tendency of Ar and O_2 to fractionate relative to the lighter species N_2, researchers have been able to infer otherwise undetectable measurement artifacts in O_2/N_2 signals. Deviations in the Ar/N_2

ratio are likely to be larger than those in the O_2/N_2 ratio because of the comparative molecular mass difference between the measured species:

$$\frac{m_{Ar} - m_{N_2}}{m_{O_2} - m_{N_2}} = \frac{40 - 28}{32 - 28} = 3 \qquad \textbf{Eq. 2.6}$$

The aforementioned studies have observed and quantified the fractionation of Ar/N_2 relative to O_2/N_2 under conditions in which pressure and thermal gradients encourage different rates of diffusion for the two gas ratios, and all find a fractionation scaling factor f_s ($f_s=f_{Ar/N2}:f_{O2/N2}$) between 2 and 4 [*Battle et al.*, 2003; *Battle et al.*, 2006; *R. F. Keeling et al.*, 2004]. Keeling [2004] measured a divided gas stream under a thermal gradient at one atmosphere and found the value of f_s to be 3.77±0.04. Battle [2006] approached the process of quantifying this number differently, endeavoring not to understand whether pressure or thermal effects were dominating the fractionation process, but only to quantify the fractionation scaling factor empirically based on the actual conditions in their sampling apparatus. Battle [2006] found this value to be 3.3±0.3, consistent with both thermal and mass-dependent fractionation within errors. Because the general scatter in MEDUSA's Box 1 is considerably worse than in Box 2, suggesting the point of fractionation is within the flasks, and because the fractionation appears to correlate with thermal gradients in the MEDUSA rack, I choose a scaling factor based on Keeling [2004] of 3.77. In the subsequent correction process, all scatter around a mean and all fractionation effects that are not due to avoidable natural signals in Ar/N_2 are corrected out from the O_2/N_2 data by removing a scaled down version (1/3.77) of the Ar/N_2 scatter around the expected mean value at a given altitude (discussed below). This correction does not adjust MEDUSA measurements for any system mean

biases that are the result of slightly different analytical conditions between MEDUSA B-flasks and Scripps O_2 Program round-bottom station flasks (estimated to be on order 3 per meg in O_2/N_2). I explore this offset in Chapter 4 for Ar/N_2, but leave a more detailed assessment of the influence on O_2/N_2 for another study since this study does not concern itself with absolute O_2/N_2 measurements—rather their seasonal cycles.

Since other studies have observed that some natural signals in the Ar/N_2 ratio do exist, I must first characterize each of these separately as "components" of the Ar/N_2 signal, so that natural atmospheric variations in Ar/N_2 are not treated as artifact and introduced to the O_2/N_2 ratio erroneously during the correction process. I find a fast diminution in the Ar/N_2 ratio in the stratosphere of, on average -11 per meg/km, a relationship which accelerates with elevation. This relationship is nominally consistent with Ishidoya [2013], who described a ~24 per meg/km decrease in the ratio at higher elevations, between 15 and 30 km, over Japan (slower at the bottom, and accelerating toward the top).

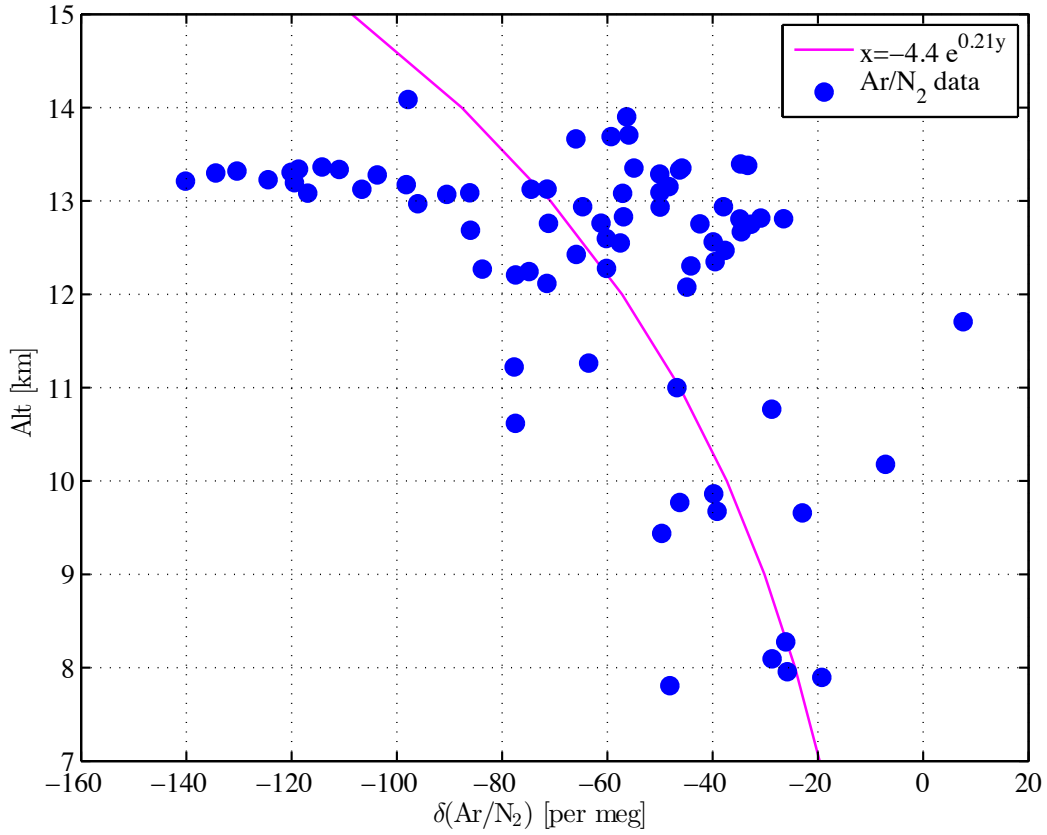

Figure 2.21. Stratospheric (determined by detrended N_2O value) Ar/N_2 flask samples against altitude. The magenta line is an exponential consideration of the height-dependence, showing an inverse correlation of altitude and Ar/N_2.

Though Ar/N_2 values can be regressed against height, as seen above in **Figure 2.21**, the decrease depends strongly upon latitude, with higher latitude flasks much more depleted at lower elevations than low latitude flasks. However, observations of N_2O provide a latitude-independent estimate of whether the flask is stratospheric, and how long the sample has been in the stratosphere. Accordingly I factor this relationship into my correction model. I describe the model as follows, removing natural Ar/N_2 phenomena when possible so that Ar/N_2 signals aren't introduced erroneously into the O_2/N_2 measurements. I start by describing the Ar/N_2 observations as the combination of three terms:

$$\delta(Ar/N_2)_{obs} = (Ar/N_2)_{art} + (Ar/N_2)_{vert} + (Ar/N_2)_{strat} \quad \textbf{Eq. 2.7}$$

where *obs* refers to the observed (uncorrected) value, *art* to the artifact that creates the scatter (and bias in Box 1) in the signal, *vert* to the full mission / tropospheric altitude-dependent component, and *strat* to the N_2O-correlated depletion effect noted in the stratosphere. I model the vertical component according to:

$$\delta(Ar/N_2)_{vert} = (\alpha x - \beta) \quad \textbf{Eq. 2.8}$$

where x is the altitude, and α and β are linear fit coefficients. For stratospheric flasks, I describe a stratospheric component by:

$$for\ N_2O \leq 318, \quad \delta(Ar/N_2)_{strat} = \gamma * (N_2O - 318) \quad \textbf{Eq. 2.9}$$

$$for\ N_2O > 318, \quad \delta(Ar/N_2)_{strat} = 0$$

where γ is a regression coefficient for stratospheric samples (with the full-profile vertical Ar/N_2 gradient removed) only, as determined by correlation with N_2O detrended relative to 2009 values.

Because flasks from Box 2 show considerably less scatter, and because they appear not to exhibit low bias due to temperature and moisture effects, I choose to quantify the stratospheric and altitude-dependent signals using only Box 2 data. Using QCLS instrument N_2O data which has been linearly detrended relative to 2009, I determine which flasks demonstrate stratospheric influence. Flasks whose detrended N_2O is less than or equal to 318 ppb are considered stratospheric and masked out temporarily.

I fit a line to the remaining flasks against altitude, as seen in **Figure 2.22**. The slope and intercept of this line are used to define the coefficients α and β.

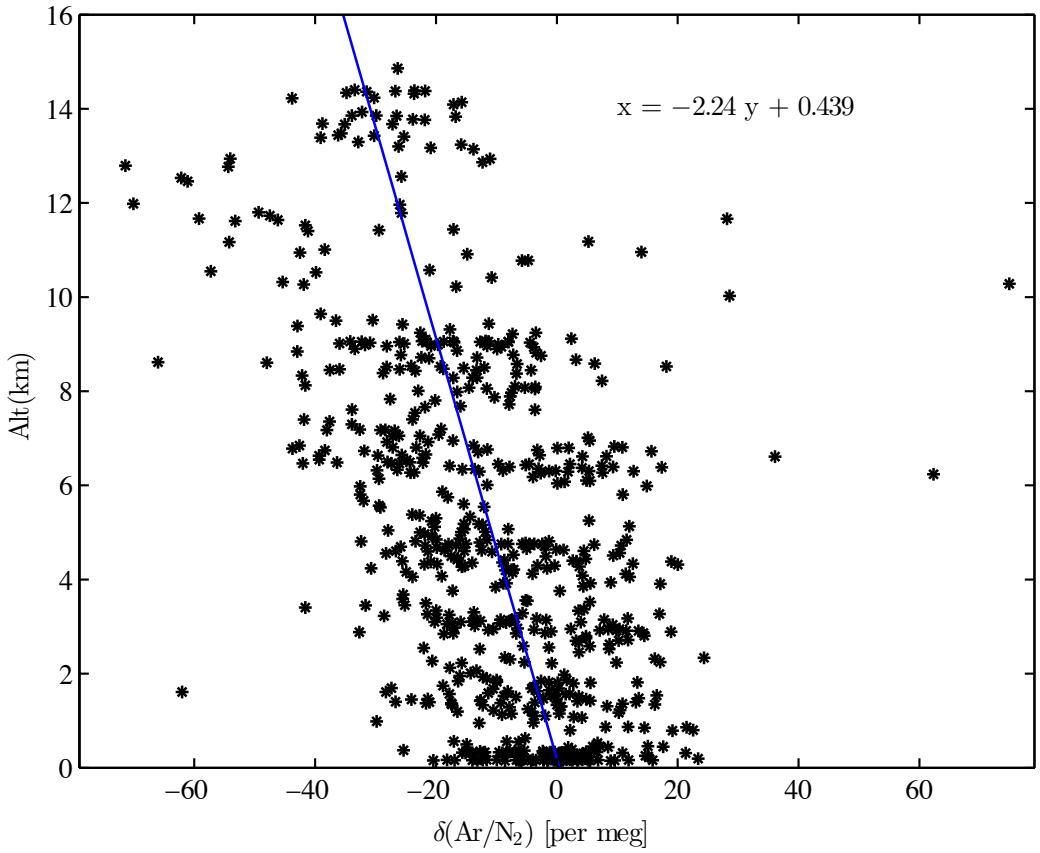

Figure 2.22. Correlation of flask altitude with $\delta(Ar/N_2)$ value for non-stratospheric samples. Black asterisks show MEDUSA Flask Box 2 $\delta(Ar/N_2)$ values, while the blue line shows the linear fit. Here, for consistency with previous plots that show altitude on the y axis, x is the dependent variable.

Next, to determine the stratospheric component independently, I remove this altitude-dependence from all stratospheric flasks and fit the residuals against concurrent N_2O data from the QCLS instrument which has been detrended relative to 2009. The slope of this line, seen in **Figure 2.23**, is used to define the value of γ.

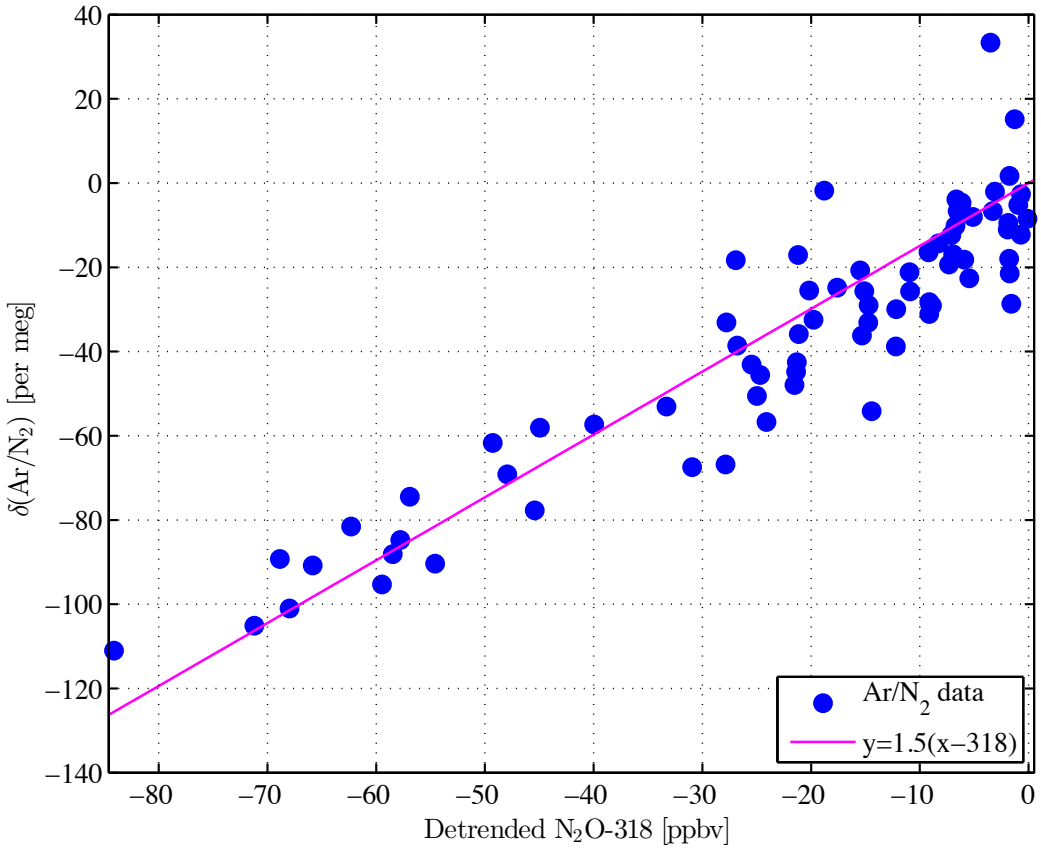

Figure 2.23. The linear regression of Ar/N_2 with detrended N_2O-318 for stratospheric samples (blue dots). Stratospheric Ar/N_2 data were corrected for hypothesized altitude-dependence (see previous figure).

Next, I determine how these features relate to the correction of O_2/N_2. The relationship of the artifactual scatter in O_2/N_2 to Ar/N_2 is given by:

$$\delta(O_2/N_2)_{art} = \frac{\delta(Ar/N_2)_{art}}{f_s} \qquad \textbf{Eq. 2.10}$$

where f_s is the scaling factor 3.77 from Keeling [2004]. I then solve Eq. 2.7 for $\delta(Ar/N_2)_{art}$, and substitute this on the right hand side of Eq. 2.10, giving:

$$\delta(O_2/N_2)_{art} = \frac{\delta(Ar/N_2)_{obs} - (Ar/N_2)_{vert} - (Ar/N_2)_{strat}}{f_s} \qquad \textbf{Eq. 2.11}$$

Corrected $\delta(O_2/N_2)$ can be considered as the difference between the observed and artifact values, or:

$$\delta(O_2/N_2)_{corr} = \delta(O_2/N_2)_{obs} - \delta(O_2/N_2)_{art} \quad \textbf{Eq. 2.12}$$

Substituting in the right hand side of Eq. 2.11 for $\delta(O_2/N_2)_{art}$, and substituting in the values of $\delta(Ar/N_2)_{vert}$ and $\delta(Ar/N_2)_{strat}$ from Eq. 2.8 and Eq. 2.9 gives the full correction is as follows:

$$for\ N_2O \leq 318, \quad \delta(O_2/N_2)_{corr} = \delta(O_2/N_2)_{obs} - \frac{[\delta(Ar/N_2)_{obs} - (\alpha x - \beta) - (\gamma * (N_2O - 318))]}{f_s} \quad \textbf{Eq. 2.13}$$

$$for\ N_2O > 318, \quad \delta(O_2/N_2)_{corr} = \delta(O_2/N_2)_{obs} - \frac{[\delta(Ar/N_2)_{obs} - (\alpha x - \beta)]}{f_s}$$

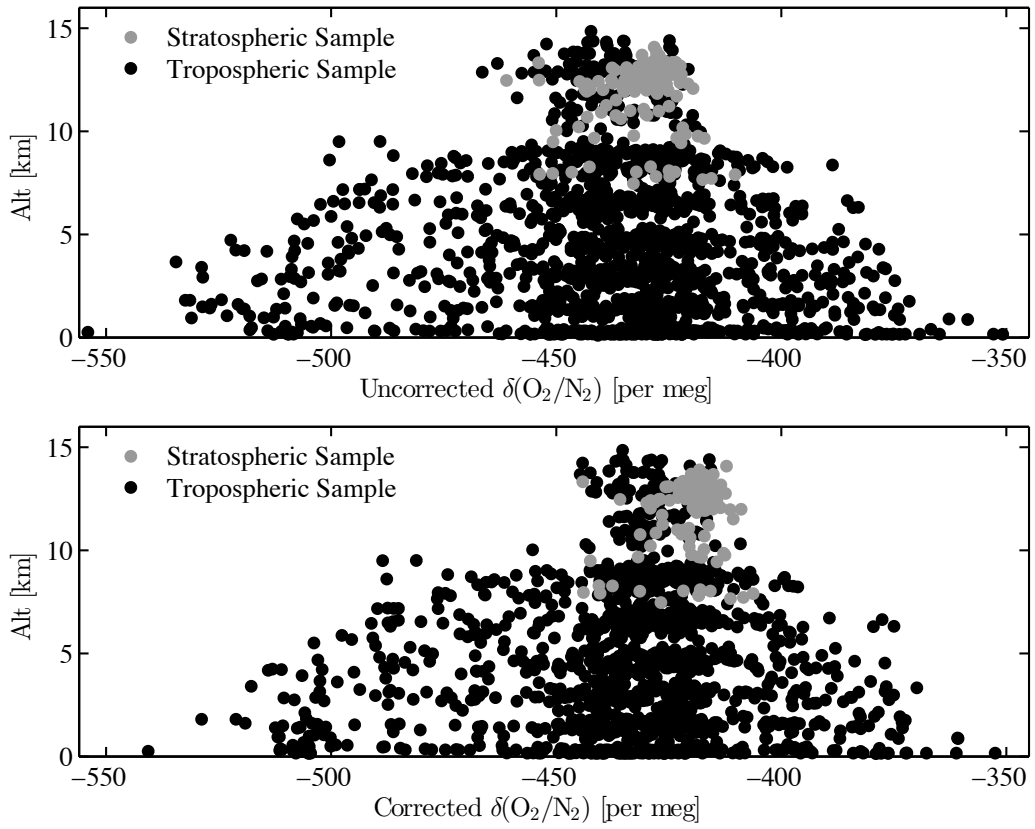

Figure 2.24. Uncorrected (top panel) and corrected (bottom panel) $\delta(O_2/N_2)$ values. Stratospheric samples are shown in gray, while tropospheric samples are black.

The correction of O_2/N_2 data is shown in **Figure 2.24** for all O_2/N_2 flasks, regardless of MEDUSA box number. The general trend is that values are shifted up; the mean of the absolute value of the adjustment is 5.98 per meg, with a standard deviation of ±5.25 per meg. The higher O_2/N_2 values associated with the corrected stratospheric samples reflect previous observations that stratospheric samples, which are largely older than their tropospheric counterparts, typically show higher O_2/N_2 due to the long-term negative fossil-fuel trend in atmospheric O_2.

2.6.3 Summary of fractionation and correction processing

The model in Eq. 2.13 quantifies the two observed gradients in the Ar/N_2 signal, removing them before quantifying the residual scatter of the signal around a vertically-dependent mean, and its concomitant representation of scaled-up O_2/N_2 scatter. Because the stratospheric gradient is nominally in agreement with the observations of Ishidoya et al. [2013], and because it scales so cleanly with N_2O, I believe this is a natural effect. I choose to leave a discussion of the tropospheric gradient until Chapter 4, where doing so won't slow the narrative of this technical paper. However, I believe it may also be a real effect, one that I show doesn't scale with airspeed or ram pressure, and which differs subtly by latitude and season. In any case, the handling of the tropospheric gradient has negligible importance for the APO results in Chapter 3 because an altitude correction would increase the value of all five curtain average points equally, offsetting them from a mean that I already normalize to zero.

One final note about Ar/N_2 corrections is that this protocol treats the natural seasonal cycles in atmospheric Ar/N_2 as artifact, because the cycle is not explicitly accounted for and removed from the Ar/N_2 data in Eq. 2.7. Station records suggest this cycle is as great as ±10 per meg (**Figure 4.15**) depending on latitude/location, which theoretically introduces a bias to the O_2/N_2 results of up to ±2.7 per meg. I choose not to explicitly account for this signal in this equation because doing so would require a climatological understanding of time- and space-varying (at least 3-D) Ar/N_2 cycles, something that is presently beyond our capability. Fortunately, though, on a case-by-case basis this bias can usually be determined very effectively by accounting for the seasonal cycle of Ar/N_2 at sampling stations near to the O_2/N_2 observations of interest, assuming

minimal seasonality in Ar/N_2 vertical gradients, by scaling down the Ar/N_2 cycle by 3.77, and by scaling it down further by the ratio of the full-column signal to the observed surface signal at stations (~0.6/1). Adding this value back to the O_2/N_2 seasonal cycle of interest effectively amplifies the O_2 cycle in much the way it was diminished in the correction process. This approach is feasible for the curtain average calculations of Chapter 3, so I prefer to follow this protocol. I believe it is preferable to applying no correction to the O_2/N_2 data at all, because Ar/N_2 data (±22.6 per meg 1σ around a mean by altitude) suggests the various forms of bias in the signal could introduce an equivalent, scaled-down bias in the O_2/N_2 measurements with a 1σ of ±6 per meg.

2.6.4 Anchoring AO2 to MEDUSA

AO2 oxygen measurements show biases relative to MEDUSA that drift during the course of a research flight and change from mission to mission. These may be related to fractionating leaks with greater impact at the low AO2 flow rate, or to residual humidity effects, as portions of tubing dry out after being exposed to ambient humidity during pre-flight trap change procedures, and other portions of tubing experience variable humidities between calibration and sample cycles. Laboratory research into these effects and potential empirical corrections is ongoing and subsequent releases of the AO2 data set will be revised accordingly. MEDUSA experiences these effects to a much lesser degree, because the flow rates are much higher, and because the sampler does not require calibration gases. Comparisons to temporally interpolated station data in the Scripps O_2 Program network show that MEDUSA O_2/N_2 data have a mean offset value of -1.6 ±3.5

per meg from stations, while uncorrected AO2 values differ by 10s of per meg, depending on flight, as shown in Figure 2.25.

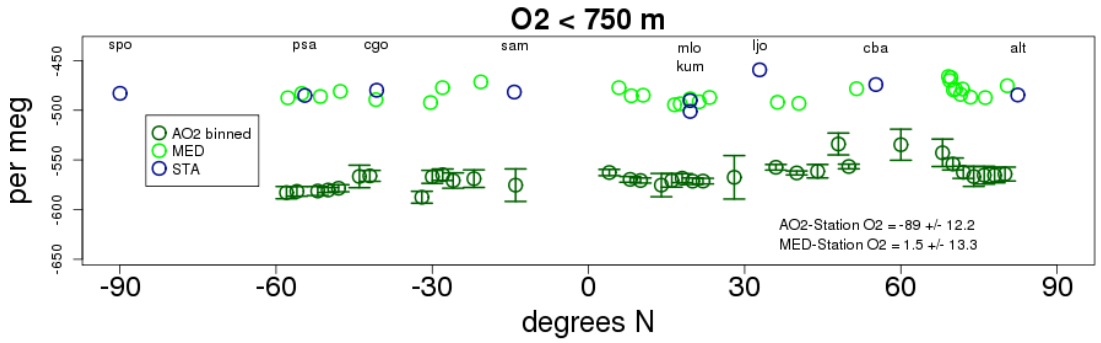

Figure 2.25. Comparison of station (blue) MEDUSA (light green) and AO2 (dark green with 1-sigma variability) O_2/N_2 values during the northbound leg of HIPPO4. Station values are interpolated in time to the date of the HIPPO flight from the nearest station flask triplicate mean values. MEDUSA values are below 750 m, and AO2 data are the 5-degree binned means of AO2 values below 750 m. For the difference calculations, the MEDUSA and binned AO2 values were spatially interpolated to the latitude of the station. Figure from B. Stephens.

The AO2 offset (here -89 ±12.2 per meg) from station values is substantially larger than the MEDUSA offset (1.5 ±13.3 per meg), suggesting that most AO2-MEDUSA differences are primarily linked to AO2 biases. Anchoring AO2 measurements to MEDUSA flasks, which are analyzed by the Scripps O_2 Program, also brings any calibration offsets in line with the Scripps O_2 scale.

Raw data from the AO2 instrument is converted into calibrated data with the use of measurements of calibration gases, which are run roughly every 40-60 minutes during the flight. Once the data are calibrated, some offset is still observed relative to MEDUSA flasks, and the magnitude of this offset can be seen to change with time in flight, and from flight to flight and mission to mission. It is this time dependent bias that MEDUSA flasks may help with. The difference between MEDUSA flasks and AO2 data is ascertained by applying the MEDUSA kernel to AO2 1-Hz data, and calculating the

offset between the two measurements. In order to apply a smoothed offset, rather than a step-wise offset that changes greatly at the instant of each flask sample, a three-flask running mean offset is calculated, and a program linearly interpolates between this mean offset value in time, so that a smoothed offset value can be applied to the AO2 data with time. The resulting AO2 data differs slightly at the moment of flask sampling due to the 3-flask running mean offset, but the overall fit to MEDUSA data is fairly smooth and continuous.

The left panel of **Figure 2.26** shows the difference between AO2 and MEDUSA O_2 values by flight within the HIPPO2 campaign (shown variously as up-pointing or down-pointing green triangles or squares depending on whether the aircraft is ascending, descending or level, respectively), while the right panel shows the corrected O_2 data. The wandering in the mean offset—as much as ±30 per meg in the left panel—is typical for AO2, given the challenging and rapidly-changing environmental conditions that the instrument faces. I leave a more detailed discussion of this anchoring process to a methods paper in preparation by Dr. Britton Stephens.

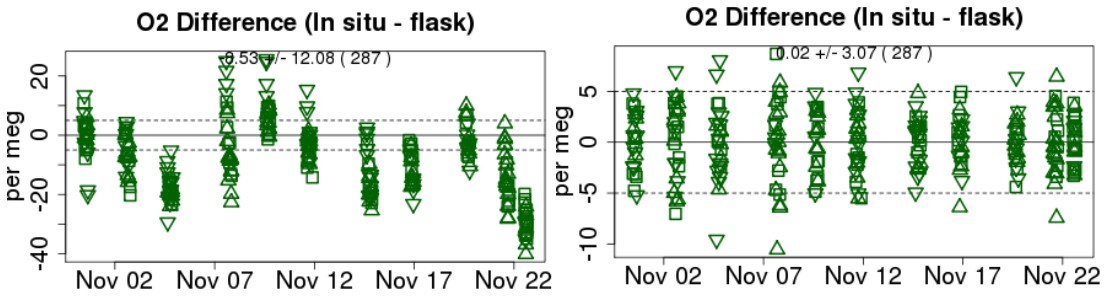

Figure 2.26. O_2/N_2 differences between AO2 and MEDUSA for each flight (columns of green symbols) in the HIPPO2 mission. Left: Differences using uncorrected AO2 data. Right panel: Differences after AO2 O_2/N_2 is anchored to a running mean of MEDUSA O_2/N_2. Figure from B. Stephens.

2.7 Conclusion

In this chapter, I have provided an extensive description of the MEDUSA sampler and its pre- and in-flight operating procedures. I have described the analysis and data reduction processes that generate analyzed and corrected MEDUSA O_2/N_2, CO_2 and Ar/N_2 values, and which produce NASA Ames format flask data and averaging kernel output files.

I have shown that systematic low bias and scatter present in MEDUSA Box 1 Ar/N_2 results, which are likely due to thermal fractionation effects stemming from a slow temperature equilibration time of the MEDUSA system, make the data a less useful diagnostic of the difference between natural signals and measurement artifact. Fortunately, data from Box 2 provide a sufficient picture of Ar/N_2 vertical distributions by themselves to allow me to distinguish signals from artifact. And, because the low-biased and scattered Box 1 flasks are likely scattered because of thermal diffusive fractionation, I am able to correct out the effect in the concomitant O_2/N_2 data by adjusting the O_2/N_2 data by a scaled down version of the Ar/N_2 bias [*R. F. Keeling et al.*, 2004]. Accordingly, I have described the process through which Box 1 data is quality controlled, and the model that I have developed to correct all MEDUSA O_2/N_2 data for noise based on its simultaneously-measured Ar/N_2 data.

I find that two potentially natural signals must be removed from the Ar/N_2 data before it is representative of a scaled-up version of only the noise in the O_2/N_2 data. These signals include a strong depletion of the Ar/N_2 ratio in the lower stratosphere due to gravitational settling of heavier molecules in the stratosphere, whose degree can be ascertained by the concomitant reduction in stratospheric N_2O measurements, and a

smaller vertical gradient in the troposphere that is well-described by a simple linear fit with elevation. The nature of the latter signal is discussed in proper depth in Chapter 4. I also recognize a limitation in the correction: that because our Ar/N_2 data does not have high enough resolution to clearly resolve seasonal cycles as a function of latitude, time and altitude, the correction does not explicitly account for this known phenomenon, and that as such, the noise correction removes a scaled-down version of this small signal from the O_2/N_2 data. Accordingly, in any future application of the O_2/N_2 results, some accounting must be made to re-introduce this signal. I address any associated concerns for our application in Chapter 3, where I discuss the process of reintroducing this signal to the curtain average values based on a scaled-down version of the observed seasonal cycle of Ar/N_2 at sampling stations in and near the Southern Ocean.

Finally, I have described the use of the corrected MEDUSA O_2/N_2 data to anchor AO2 O_2/N_2 data according to a running mean of flask values.

Restricting a description of these complex methods to this chapter helps facilitate the conveyance of a scientific narrative for the remaining results presented in Chapters 3 and 4. I encourage readers to consult Chapter 4 for a description of the scientific significance of our Ar/N_2 observations, and Chapter 3 for a discussion of O_2/N_2 data over the Southern Ocean. In Chapter 3, I discuss the development of a new metric for quantifying the seasonal flux of oxygen into and out of the Southern Ocean, as a means toward constraining the various biogeochemical processes that contribute to these observations.

3 The curtain average: large-scale Southern Ocean biogeochemical constraints

3.1 Introduction

The Southern Ocean plays an important role in controlling climate by exposing CO_2-rich deep ocean water to the atmosphere and subducting anthropogenic CO_2 [*Sarmiento et al.*, 1998]. Alterations to the heat balance, wind patterns and chemistry of the Southern Ocean by anthropogenic climate change could alter the magnitude of the sink or source of CO_2 caused by this overturning, with implications for future feedbacks to climate change [*Le Quéré et al.*, 2009; *Le Quéré et al.*, 2007; *Sabine and Tanhua*, 2010]. Assessing the rate at which the Southern Ocean may be changing requires not only good estimates of interannual trends in air-sea gas and heat exchange, but a good understanding of, and models that can reproduce, seasonal ocean biogeochemical processes such as primary productivity, ventilation of deep waters, and solubility changes and their impacts on O_2 and CO_2. Such seasonal-scale exchanges have an additional utility: that they provide an excellent laboratory for investigating and characterizing the very processes that lead to long-term climate feedbacks. Measurement of the air-sea fluxes that reflect the impact of these driving processes is poor, however. Southern Ocean sampling is restricted by rough weather, and high winds push the limits of our interpretation of air-sea gas exchange parameterizations [*R. Wanninkhof*, 1992; *Rik Wanninkhof and McGillis*, 1999]. At the same time, remote sensing measurements, such as those of ice cover, sea surface chlorophyll and temperature—many of which are crucial tools for modelers—are limited by heavy cloud cover.

Traditional methods of estimating ocean biogeochemical processes have often relied on measurements of atmospheric CO_2, which represent a well-mixed picture of surface ocean processes beneath. Measurements of Southern Ocean atmospheric carbon dioxide by many research groups are well-established [*C D Keeling*, 1960; *Law et al.*, 2010; *Stephens et al.*, 2013] at sites such as Palmer Station, Syowa and South Pole, Antarctica; Cape Grim and Macquarie Island, Australia; and Baring Head, New Zealand. Records at South Pole extend back as far as 1957 [*C D Keeling*, 1960], while other stations' records span 20 or more years. Such records show a consistent seasonal cycle of atmospheric CO_2 of approximately 1 ppm at high southern latitudes. The cycle is small because of a lack of strong terrestrial signals, and due to the interference of ocean signals: summer warming drives CO_2 out of the surface ocean, while productivity draws it back in; winter cooling draws in CO_2, while upwelling releases it to the atmosphere [*Stephens et al.*, 1998]. This presents a challenge to estimating biogeochemical processes and their resulting contribution to the total air-sea flux, because small errors in component fluxes can lead to large errors in the residual net flux. The carbonate chemistry of the surface oceans also diminishes the magnitude of the seasonal signal, and adds complexity to the equation: photosynthesis removes dissolved inorganic carbon from the surface ocean, but the corresponding uptake of CO_2 is approximately $1/10^{th}$ as big due to the Revelle Factor, damping and delaying the observed phase of the carbon cycle in the atmosphere. Adding more complexity again, the atmospheric CO_2 signal over the Southern Ocean reflects the influence of anthropogenic and terrestrial exchange [*Stephens et al.*, 2013; *Stephens et al.*, 1998], signals that must be removed to quantify ocean signals alone.

Because the consumption and production of CO_2 and O_2 are linked through photosynthesis and respiration in the surface ocean, and because the two experience solubility changes based on surface temperature, scientists have turned to observations of the atmospheric O_2/N_2 ratio [*Garcia and Keeling*, 2001; *Najjar and Keeling*, 2000] to estimate the magnitude of ocean biogeochemical processes (biological+physical). The atmospheric seasonal cycle of oxygen over the Southern Ocean is larger than that of CO_2, as biological and thermal flux components reinforce each other, and changes in surface waters are rapidly reflected in atmospheric observations because of oxygen's relatively low solubility and lack of buffering chemistry in water. In order to remove the influence of land photosynthesis, oxygen and carbon dioxide measurements can be combined into a conservative tracer called atmospheric potential oxygen [*R. F. Keeling et al.*, 1993; *Stephens et al.*, 1998]. APO ($APO \approx O_2 + 1.1CO_2$) can be thought of as the expected concentration of oxygen in the atmosphere if all CO_2 were removed by land photosynthesis. Because it removes the terrestrial signal, APO is a conservative tracer of ocean biogeochemistry, and can be compared directly against ocean models' estimates of seasonal exchange, as long as small anthropogenic signals are well accounted for.

Researchers who first attempted to estimate seasonal fluxes of APO distributed CO_2 and O_2 fluxes from ocean models through atmospheric transport models to test whether output agreed with data at surface stations on the seasonal cycle and spatial gradients of APO [*Stephens et al.*, 1998]. Subsequent studies [*Battle et al.*, 2006; *Naegler et al.*, 2007] have continued the approach of Stephens et al. [1998], using newer models and different model combinations. The same, and additional [*R. F. Keeling et al.*, 1998a] studies have also used dissolved climatologies of CO_2 and O_2—empirical estimates of

mean seasonal cycles based on parameterizations of measurements of oceanic pCO_2 [*Takahashi et al.*, 2002; *Takahashi et al.*, 2009] and surface ocean O_2 anomaly [*Garcia and Keeling*, 2001]. However, studies have so far been unable to assess how much data-model disagreement stems from air-sea flux estimates and atmospheric transport models (ATMs), individually, because disagreement may reflect incorrect fluxes to begin with, or the improper mixing of these signals by ATMs [*Naegler et al.*, 2007].

A large portion of the uncertainty in ATMs stems from difficulty in determining rates of vertical mixing. Accurate APO fluxes that are too vigorously lofted into the upper atmosphere (center left box in **Figure 3.1**) will nonetheless appear to underestimate the amplitude of the seasonal cycle at surface stations, while inaccurately small fluxes could appear to reproduce seasonal cycles well if APO is trapped too much in the planetary boundary layer (bottom right box). Since most station data is taken at the earth's surface, assessment of vertical mixing uncertainty has been unsuccessful, leading some studies [*Naegler et al.*, 2007] to conclude that the interpretive power of APO model-data comparisons is limited.

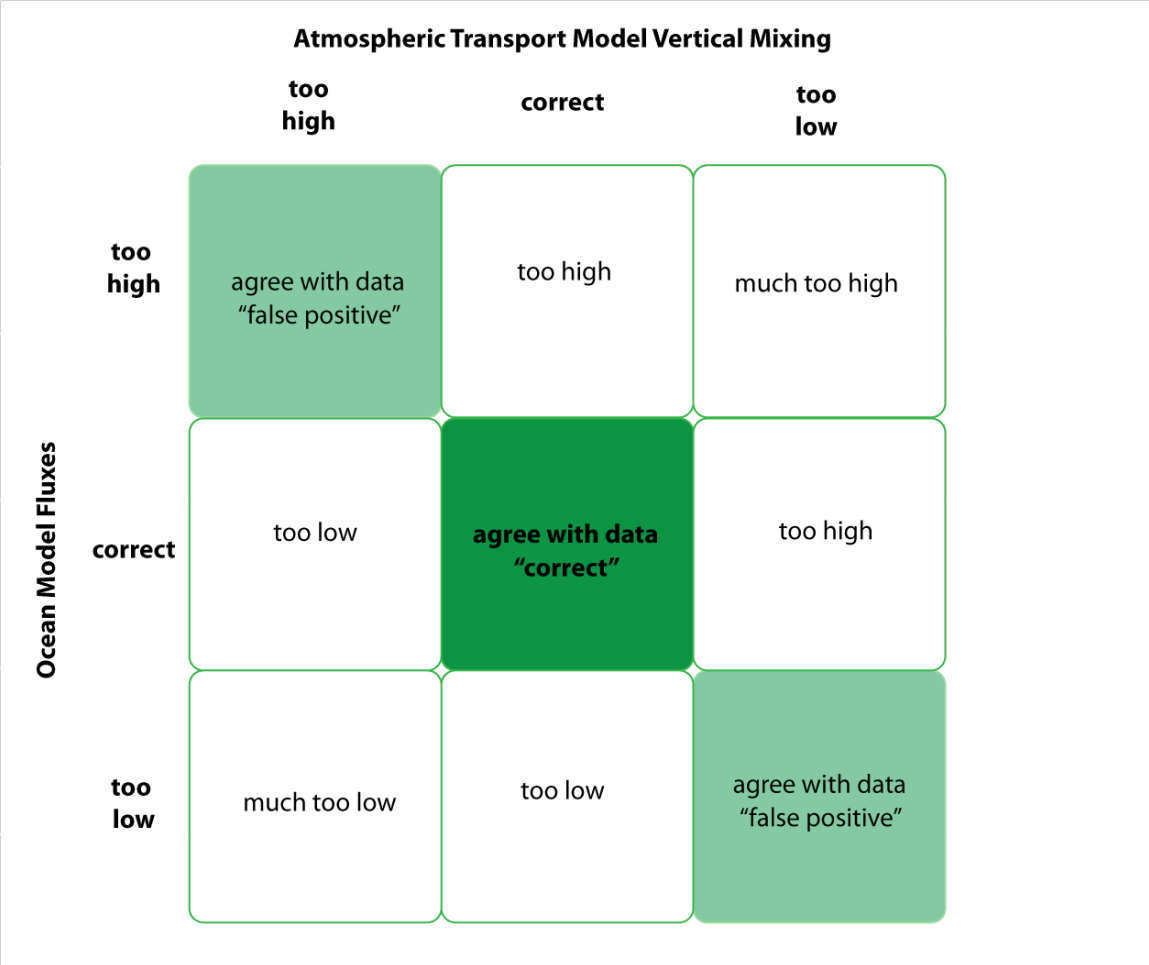

Figure 3.1. Comparison of possible model-data agreement and disagreement scenarios for APO seasonal cycles at surface sampling stations. Text in boxes indicates whether modeled surface station amplitudes are theoretically correct or not. The dark green central panel is the desired result, but without an estimate of whether vertical mixing in transport models and fluxes are correct, it is impossible to tell if perceived agreement in the amplitude of models' and station observations' seasonal cycles is due to correct transport and fluxes, or a "false positive" (light green) combination of too-large fluxes with too-weak vertical mixing or too-small fluxes with too-vigorous vertical mixing.

Recent airborne measurements from the HIPPO global sampling campaign offer a solution to this modeling impasse by providing vertical profiles of the atmosphere between the surface and the lower stratosphere at five points in the seasonal cycle. Changes in the vertical column averages are less impacted by vertical mixing than surface observations, affording an improved constraint on surface fluxes. Toward this end, I compute a metric that I call the Southern Ocean meridional curtain average

(SOMCA, or more broadly just CA), a pressure- and latitude-weighted mean of atmospheric potential oxygen over latitudes from 45˚ to 65˚S near the dateline. The rapid zonal mixing in the atmosphere suggests that this curtain average generally constrains the three-dimensional average over the entire zone (see **Figure 3.10**). If the meridional atmospheric mixing across 45˚ and 65˚S is known, this would also constrain the zonally averaged ocean fluxes. However, for my purposes, I am largely interested only in whether models can reproduce this CA transect.

In this chapter, I discuss the comparison of HIPPO APO data, expressed as a curtain average, to combined ocean-atmosphere model output, and implications for ocean and atmosphere modeling. Because this study involves both evaluation of ocean models, and data corrections using combined ocean-atmosphere model output, I have chosen to structure this chapter in a slightly non-traditional way: some data results are given before modeling methods are discussed because some understanding of sampling biases, measurement uncertainty and the curtain average method are necessary to understand the motivation for the modeling methods. Section 3.2 gives a brief overview of the Southern Ocean HIPPO flights, instruments, corrections to data, and methodology for computing the curtain average. Section 3.3 shows HIPPO data that reflect instrumental corrections, but not any modeling adjustments, and provides preliminary impressions of the curtain averages and their relationship to observations at surface stations. In Section 3.4, I detail the calculation of the curtain average, and the detrending of data based on observations at surface stations in the Scripps sampling network. In Section 3.5, I introduce combined ocean-atmosphere model runs and explain how I use various model simulations to estimate spatial and temporal sampling biases and uncertainty. In Section 3.6, I discuss

assessments of interannual variability and possible volcanic influence by consulting Scripps sampling network data. In Section 3.7, I compare data and model-output curtain averages and discuss possible reasons for model disagreement. In Section 3.8, I discuss assessments of the Garcia and Keeling [2001] O_2 climatology fluxes, and how recent understandings of their phasing and amplitude may bring them into closer agreement with observations. In Section 3.9, I very briefly introduce station data and compare these with model station output as a first look at how transport models mix flux signals vertically. In Section 3.10, I revisit a discussion of NEMO ocean model output in light of the transport model discussion in the previous section. Finally, in Section 3.11, I conclude with a discussion of future work needed to expand upon the present study. Readers interested in a more complete explanation of the quality control and correction of O_2 data may consult Chapter 2 of this dissertation.

3.2 Methods

3.2.1 Instrumental methods

The southernmost flights of the HIPPO campaign departed from Christchurch, New Zealand (43.5˚S), returning either to Christchurch or to Hobart, Tasmania. The resulting dataset provides a seasonal picture of the structure of the atmosphere over the Southern Ocean. Data from the NCAR AO2 vacuum ultraviolet oxygen instrument, and the NCAR/Scripps MEDUSA whole-air flask sampler, which ran on all of these flights, provide crucial, and new information about the Southern Ocean carbon cycle. These measurements comprise the majority of the data in this chapter, though I have gleaned some additional information from N_2O data (Harvard University's QCLS instrument) to

help in the Ar/N_2-based correction of our O_2 measurements (see Chapter 2). The five HIPPO missions were spaced approximately two and a half months apart in the seasonal cycle with an expectation, based on simulations of randomly-chosen 5-point fits to time series data, that detrended data could characterize up to two harmonics of the climatological mean seasonal cycles for atmospheric tracers. Any remaining interannual variability might be accounted for by comparison against model output or station data. **Table 3.1** provides a summary of the timing of the five HIPPO missions with an emphasis on the Southern Ocean flights covered in this chapter. HIPPO1 and 3 bracketed the seasonal O_2 maximum, as seen in Southern Ocean station records, while HIPPO5 and 2 bracketed the minimum. HIPPO4 characterized a transitional point between the maximum and minimum in early austral winter.

Table 3.1. HIPPO mission metadata, emphasizing the timing and seasonality of Southern Ocean flights.

Mission	Dates	SO Research Flight #	SO Flight Year Day	Austral Season
HIPPO1	Jan 2009	7	20	Mid Summer
HIPPO2	Oct-Nov 2009	6	315	Mid Spring
HIPPO3	Mar-Apr 2010	6	95	Early Fall
HIPPO4	Jun-Jul 2011	6	179	Early Winter
HIPPO5	Aug-Sep 2011	9	241	Late Winter

Figure 3.2 shows the flight paths of the five Southern Ocean flights (full flight paths are included in Appendix 1). HIPPOs 1-3 and 5 departed from and returned to Christchurch, NZ. Paths differed between missions because of meteorological, safety and air-traffic considerations and in the case of HIPPO4, to allow for a different destination point in Tasmania, and to avoid ash from the June 3, 2011 eruption of the Puyehue-Cordón Caulle volcanoes in Chile.

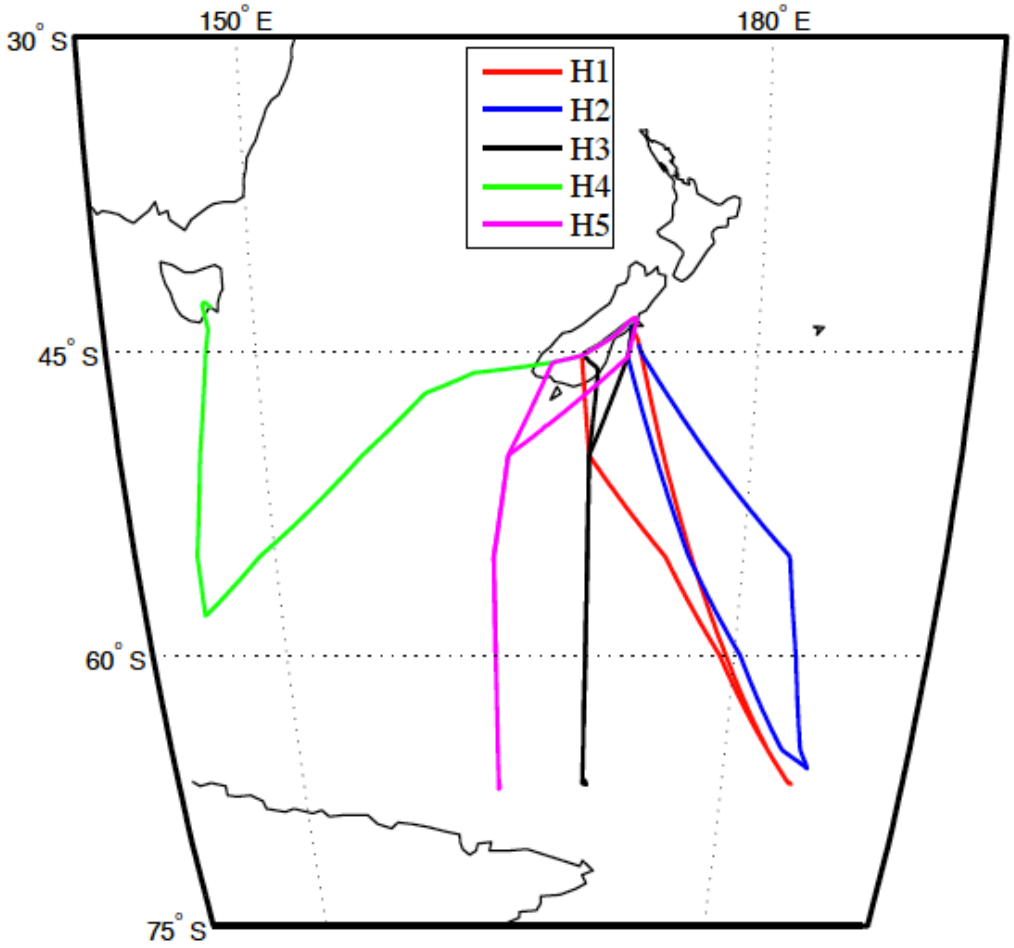

Figure 3.2. HIPPO 1-5 flight tracks over the Southern Ocean.

Typical flights, such as RF09 of HIPPO5 seen in **Figure 3.3**, traveled south from Christchurch, ascending initially to 12+ km, and then "porpoised" between 150-300 m above the ocean's surface and an altitude of 8-9 km. At the turnaround point of ~67˚S, the GV typically ascended to 12-14 km and remained at elevation until the final descent back into Christchurch or Tasmania. Such a flight path allowed for high-resolution characterization of tropospheric gradients, with the ability to at least partially constrain the stratospheric and high-tropospheric distribution of gases of interest.

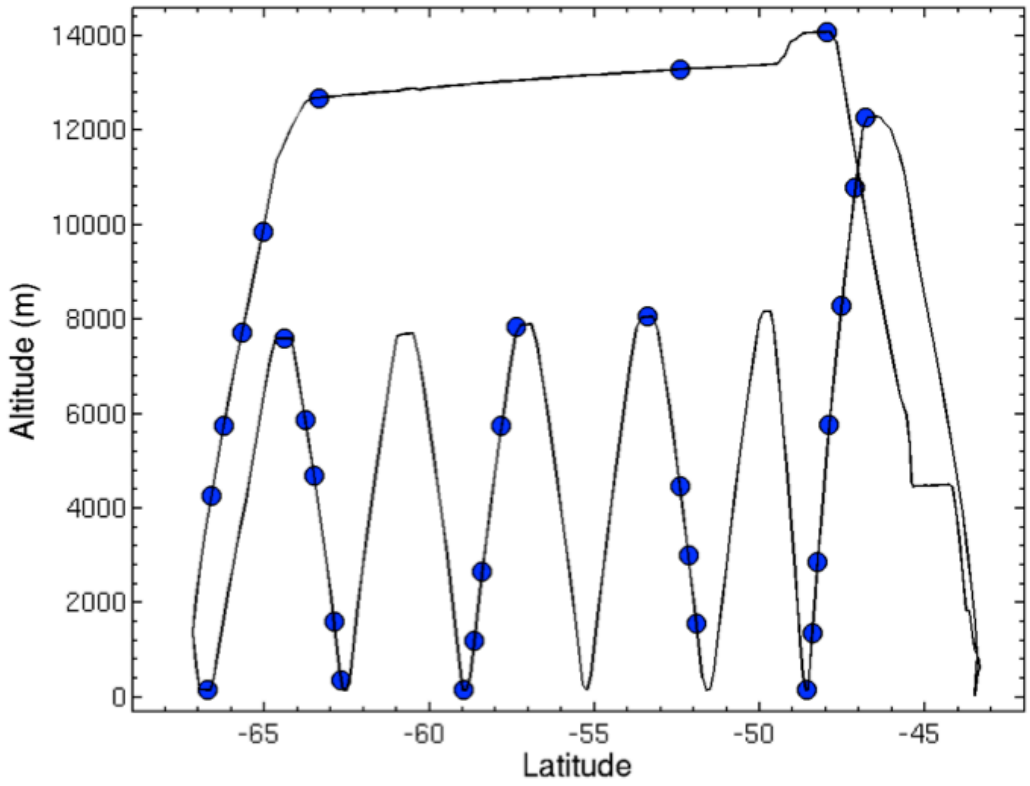

Figure 3.3. A typical Southern Ocean "return" flight track, HIPPO5 RF09, with locations of MEDUSA flask samples noted in blue. The GV aircraft started from Christchurch, NZ, flying saw-tooth maneuvers south, and returned to Christchurch at altitude.

3.2.2 Initial corrections to MEDUSA O_2/N_2 data

Flask O_2/N_2 measurements are known to be sensitive to thermal fractionation effects during sampling [*Blaine et al.*, 2006; *R. F. Keeling et al.*, 2004]. The magnitude of this effect can be estimated based on the deviation from an expected value of concurrently-sampled Ar/N_2 data. In chapters 2 and 4 I explore the relationship of scatter in the Ar/N_2 and O_2/N_2 signals, and explain the investigation I undertook to preserve natural signals in the Ar/N_2 data. I recognize two effects that I believe should be preserved in the Ar/N_2 data, so as not to bias the O_2/N_2 data correction. The first is a

depletion signal in all data that correlates with altitude. The second is a depletion signal in the stratosphere, which stems from differences in molecular diffusivity between atmospheric gas constituents in a low-convection environment [*Ishidoya et al.*, 2013]. This stratospheric effect depends on tropopause height, and is most directly accounted for by its relationship with concurrent measurements of N_2O content. I fit a single linear trend to the vertical gradient in the troposphere, and a trend with N_2O for stratospheric flasks (after removing the altitude-dependent signal derived from tropospheric data). I then correct the O_2/N_2 values based on the residual scatter around these fits using a scaling factor based on empirical observations of thermal fractionation effects in Ar/N_2 and O_2/N_2 signals [*R. F. Keeling et al.*, 2004]. Though such a correction will add a small amount of analytical error from the Ar/N_2 measurement to the O_2/N_2 measurement—approximately ±2.7 per meg 1σ—it removes a greater imprecision (i.e. scatter) of ±6 per meg 1σ. I show, in Section 3.6.3, that the uncertainty of this value and associated flask-specific errors can actually be substantially reduced to a value of 1.2 per meg for each curtain average value, because the error associated with each flask is minimized by the averaging of multiple flasks. I include this number in **Table 3.5**, where I sum the uncertainties associated with the curtain average technique.

Of note, though I do observe a natural seasonal cycle in atmospheric Ar/N_2, based on small differences in mean Ar/N_2 over the five HIPPO missions, I choose not to directly factor this into the equations I have produced to determine the vertical gradients of Ar/N_2 in the atmosphere, as I would need separate latitude-appropriate estimates of the seasonal gradient and stratospheric gradient for each hemisphere at each of the five points

in the seasonal cycle. I explain the method of assessing and correcting for this remaining bias in Section 3.5.4.

3.2.3 Adjustments to AO2

The AO2 analyzer furnishes the majority of the dataset discussed in this chapter because it offers much higher resolution than MEDUSA flasks, and thus resolves small vertical and meridional features MEDUSA would otherwise miss. However, the AO2 instrument shows slowly varying time-dependent biases relative to MEDUSA. These are likely related to small inlet leaks or residual humidity effects on tubing walls and optical components, but research to fully explain the effects is still ongoing at NCAR. With a goal of obtaining a single high-resolution dataset whose value is tied to the Scripps O_2 scale, I anchor AO2 O_2 values (as described in Section 2.6.4) to a 3-point running mean of offsets to MEDUSA flasks, applying the kernel weighting to AO2 when calculating offsets. The offset between the two measurements is subtracted or added to the AO2 measurement, as needed, and linearly interpolated between 3-point running means of MEDUSA sample comparisons to allow for a continuous correction.

3.2.4 Calculation of APO

The atmospheric $\delta(O_2/N_2)$ measurement is calculated as the deviation in the ratio of O_2 to N_2 molecules in a sample relative to an established reference gas by the equation:

$$\delta(O_2/N_2) = \left(\frac{(O_2/N_2)_{\text{sample}}}{(O_2/N_2)_{\text{reference}}} - 1\right) \text{x } 10^6 \qquad \textbf{Eq. 3.1}$$

This is reported in units of "per meg", i.e. deviations of one part in 10^6 from a standard. APO—the O_2/N_2 ratio of an air sample if all its CO_2 were drawn down by land photosynthesis—is calculated by the equation [*Stephens et al.*, 1998]:

$$\mathrm{APO} = \delta(\mathrm{O_2/N_2}) + \frac{1.1}{\mathrm{x_{O2}}}([\mathrm{CO_2}] - 350) \qquad \textbf{Eq. 3.2}$$

where [CO_2] is the CO_2 mole fraction in μmol/mol of dry air, and where $\delta(O_2/N_2)$ is the measured atmospheric oxygen to nitrogen ratio deviation in per meg. 1.1 is the approximate terrestrial exchange ratio in moles of O_2 per moles of CO_2 either produced or consumed [*Severinghaus*, 1995], and X_{O2} is 0.2094, the dry air mole fraction of O_2 [*Tohjima*, 2000]. The factor X_{O2} converts ppm CO_2 to its equivalent in per meg because per meg is reported relative to the concentration of O_2 in the atmosphere. The subtraction of 350 ppm before the calculation of APO is standard in literature, and generates a number that is easier to handle. I use 1.1 here for the terrestrial exchange ratio in keeping with past studies [*Battle et al.*, 2006; *Nevison et al.*, 2008; *Stephens et al.*, 1998]. The uncertainty in this ratio is estimated to be around ±0.1 mol/mol. Over the Southern Ocean, however, the seasonal cycles in APO are dominated by O_2/N_2, with only a very small contribution from CO_2. The uncertainty in the 1.1 ratio is therefore negligible for this study. I do not include this uncertainty in the error budget table (**Table 3.5**) for the HIPPO curtain average values: an exchange value of 1.0 or 1.2 mol/mol only changes the amplitude of the curtain average seasonal cycle by ±0.1%, on order 0.04 per meg.

I calculate APO for ocean model runs, also, since this is the quantity I'm interested in comparing with observations. In this context, APO is calculated slightly

differently, as contributions to the tracers are individualized, with oceanic O_2 and N_2 fields reported separately, and with CO_2 divided into ocean and fossil fuel components. Furthermore, atmospheric transport models are generally intended to simulate the transport of trace gases and not set up to account for changes in the total number of moles in an air parcel in response to a surface flux. Following Stephens et al. [1998], I use the ATMs to simulate the tiny absolute deviations in O_2 and N_2 concentrations as if they were trace gases against an arbitrary background, and then calculate APO as:

$$\mathrm{APO} = \left[\frac{\Delta \mathrm{O2_{Oc}}}{\mathrm{X_{O2}}}\right] - \left[\frac{\Delta \mathrm{N2_{Oc}}}{\mathrm{X_{N2}}}\right] + \left[\frac{1.1}{\mathrm{X_{O2}}}\Delta \mathrm{CO2_{Oc}}\right] + \left[\frac{1.1 - 1.4}{\mathrm{X_{O2}}}\Delta \mathrm{CO2_{FF}}\right] \qquad \textbf{Eq. 3.3}$$

where $\Delta O2_{Oc}$ and $\Delta N2_{Oc}$ and $\Delta CO2_{Oc}$ are the simulated oceanic components of the atmospheric O_2, N_2 and CO_2 signals in ppm equivalent units, respectively, and where X_{O2} and X_{N2} are the mole fractions of O_2 and N_2 in the atmosphere (0.2094 and 0.7809 respectively), and $\Delta CO2_{FF}$ is the abundance of fossil fuel CO_2 in the run. This last quantity is adjusted in two ways—it is scaled up by 1.1 as per the APO calculation above, but it is also equivalently scaled down by -1.4 to reflect the number of moles of O_2 that are removed from the atmosphere each time a mole of CO_2 is produced through combustion for the average global fuel mixture [*R. F. Keeling*, 1988; *Manning*, 2001]. I choose to add in the fossil fuel signals of CO_2 and O_2 because it is not possible to correct these out of the observed APO value. However, I do not include land biosphere tracers, because terrestrial exchange has theoretically been removed from the data through the calculation of APO. In dissolved climatology runs—i.e. runs using measurement-based estimates of air-sea fluxes—mean O_2 [*Gruber et al.*, 2001] and N_2 [*Gloor et al.*, 2001] ocean inversion terms (i.e. large-scale net source/sink terms that don't change from year

to year) are added to the O_2 and N_2 components in the numerator of the first and second terms, respectively.

The reader should be aware that plots of seasonal cycles can be visually confusing because of the need to emphasize certain aspects of the cycle. I show individual atmospheric tracers in any of three configurations: as [ppm], [per meg] or [per meg contribution] to the APO signal. The first cycle appears smaller than the second because the units haven't been scaled up to per meg in APO (by 1/0.2094 for O_2, -1/0.7809 for N_2, and 1.1/0.2094 for CO_2, see Eq. 3.3). The third value shows the (numerically larger) per meg cycle, but in the case of N_2 is sign-flipped about the X-axis to reflect the fact that N_2 diminishes the amplitude of the seasonal cycle of O_2/N_2 (or APO) relative to what the cycle would be for just atmospheric O_2 concentration (or APO calculated using just O_2 concentration). I show tracers this final way only when they are plotted alongside APO so that the individual tracers add up visually to produce the APO signal. For quick visual reference, I adopt the convention of concluding all figure and table captions with a parenthetical reference to the type (or types) of units being used. Accordingly, I write, "Units: [ppm]," "Units: [per meg]," or "Units: [per meg cont.]". Additionally, in all plots whose X-axis is month of year, I adopt the convention of placing the month letter at the beginning of the month. Finally, for plots that show mean seasonal cycles, I often show 18 months' values. The second January-June is an exact replication of the first six months, but allows the reader's eye to notice crucial features across the December-January divide that may not be obvious in a standard 12-month plot.

Figure 3.4 gives a sense for the contributions in per meg of each of the major tracers to the seasonal cycle of APO for the Southern Ocean meridional curtain

average—in this case from the TM3 atmospheric transport model [*Gurney et al.*, 2000] run by Dr. Sara Mikaloff Fletcher, using a series of dissolved gas climatology fluxes, which I discuss in Section 3.5. I choose to leave fossil fuel and mean N_2 and O_2 signals out of this figure as these contributions are imperceptible when presented in this context. Notably, the major contributor to APO variations in the Southern Ocean is O_2 fluxes. The N_2 cycle is the second largest contributor, reducing the APO cycle by roughly 15%, followed by the oceanic CO_2 signal, which contributes ~2-10%, depending on which model or climatology is being used to estimate CO_2 fluxes. Accordingly, the greatest driver in the differences between various simulations (and data) will almost always be air-sea O_2 flux.

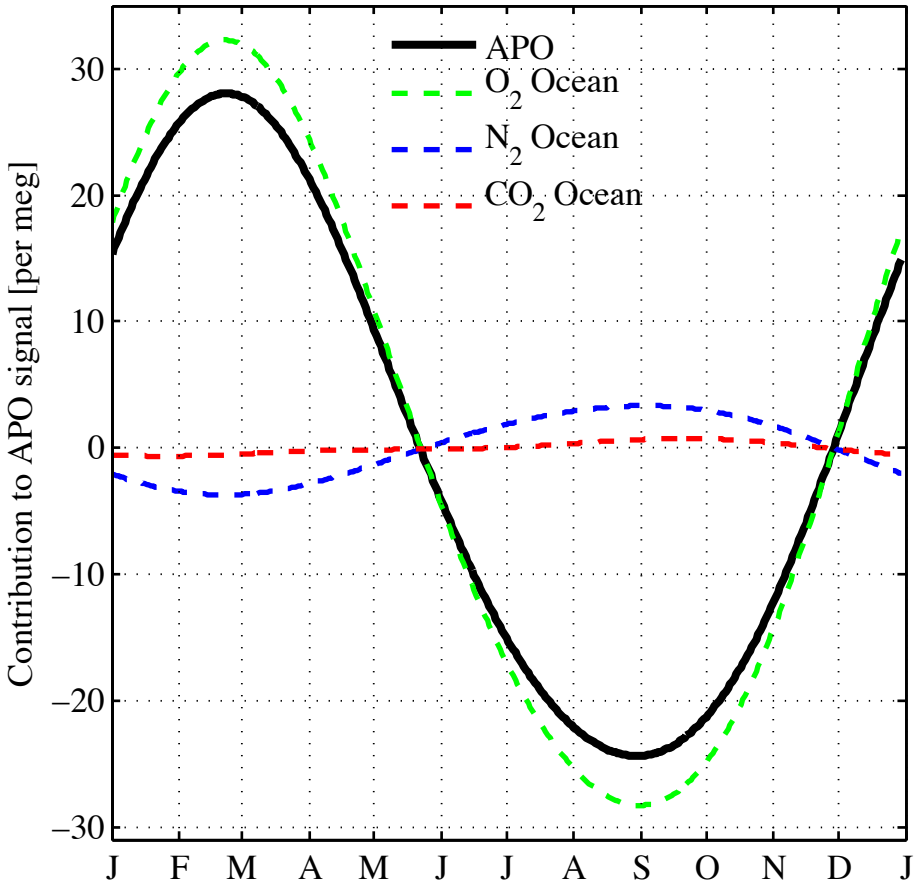

Figure 3.4. Contributions in per meg to the 2-harmonic Southern Ocean curtain average APO seasonal signal (TM3-GK01 weekly-resolution output) by oceanic O_2, N_2, and CO_2. Since APO by definition removes the terrestrial signal, I ignore terrestrial CO_2 and O_2 signals in this plot. And because fossil fuel and mean O_2 and N_2 components are negligibly small in this context, I also do not plot these, as they would appear as flat, overlapping lines in the middle of the plot. All fits are 2-harmonic to the average of each component over the Southern Ocean slice (180˚W at 65˚S-45˚S to 300 mb). Units: [per meg cont.]

3.2.5 Interpolation of flight tracks

To compute the curtain average, we start by interpolating flight track data to a regular grid using the bivariate triangulation interpolation scheme of Akima [1978], so that a weighted average of the individual grid points can be computed. The Akima interpolation scheme uses a fifth-degree polynomial to establish smooth functions between triangular cells in the x-y plane. Though the function is smooth (i.e. it and its

first order partial derivative are continuous at triangle boundaries), it is not smoothed, so the original AO2 values at the points of sampling are preserved, while points in between saw teeth are filled in based on the interpolation scheme and some outside the saw teeth by limited extrapolation. Analysis in Section 3.5.5 of this chapter suggests that the choice of interpolation and extrapolation scheme has relatively little effect on the expected curtain average value for each of the HIPPO missions, with differences on order 1 per meg or less over three different interpolation schemes (see **Figure 3.15**).

Several additional steps were taken to prepare the data for the Akima interpolation scheme. First, terrestrial boundary layer data, and any other data during takeoffs or landings that show signs of localized anthropogenic sources are masked out. For the flights considered here, this is only during take off and landing in Christchurch or Hobart. Next, a very small vertical extrapolation is applied to match altitudes of all lower and upper saw-tooth boundaries (which are already more or less the same altitude) to avoid the interpolation scheme extending information from one profile above or below an adjacent profile. All values are then bin-averaged as evenly-spaced gridded points with a resolution of 20 mb x 0.25° latitude. Next, the Akima interpolation is applied to interpolate between flight tracks within the flight envelope, and simple horizontal extrapolation is used outside of the flight envelope. After this, any areas that are more than 4°, or 25 mb distant from actual observations are masked out (see **Figure 3.6**).

3.2.6 Fitting the seasonal cycle of APO

I choose to fit all seasonal signals in data and model output to a combination of sine and cosine functions because seasonal cycles of individual gases in the atmosphere tend to approximate sinusoids with a period of a year. I variously choose to fit one or two harmonics, with a fundamental frequency of 1 year and a second frequency of 6 months. The two harmonic version can be expressed as:

$$C_{fit} = a_1 + a_2 t + a_3 \cos(2\pi t) + a_4 \sin(2\pi t) + a_5 \cos(4\pi t) + a_6 \sin(4\pi t) \quad \textbf{Eq. 3.4}$$

where t is time in years, a_1 is a constant offset, a_2 is the constant linear trend component over a small period of time (3-5 years), and where a_3:a_6 are amplitudes of the various harmonic components. (The one-harmonic equivalent of the equation lacks the last two terms.) For HIPPO curtain averages, values a_1 and a_3:a_6 are solved for directly, while a_2 is provided as the deseasonalized linear interannual component of the Scripps South Pole Observatory station record for data, or the same component from model output at the South Pole.

I choose to fit up to two harmonics (fundamental harmonic+2nd harmonic) because I am limited to five data points to determine five variables (a_1, a_3:a_6) when fitting the HIPPO curtain averages. For models, though they are output with higher temporal resolution, I still fit only 1 or 2 harmonics for consistency with HIPPO fits. I choose to evaluate the second harmonic of data and models because I have observed that a second harmonic to this fundamental frequency exists in the data, based on Fourier analysis of APO time series data from Southern Ocean stations.

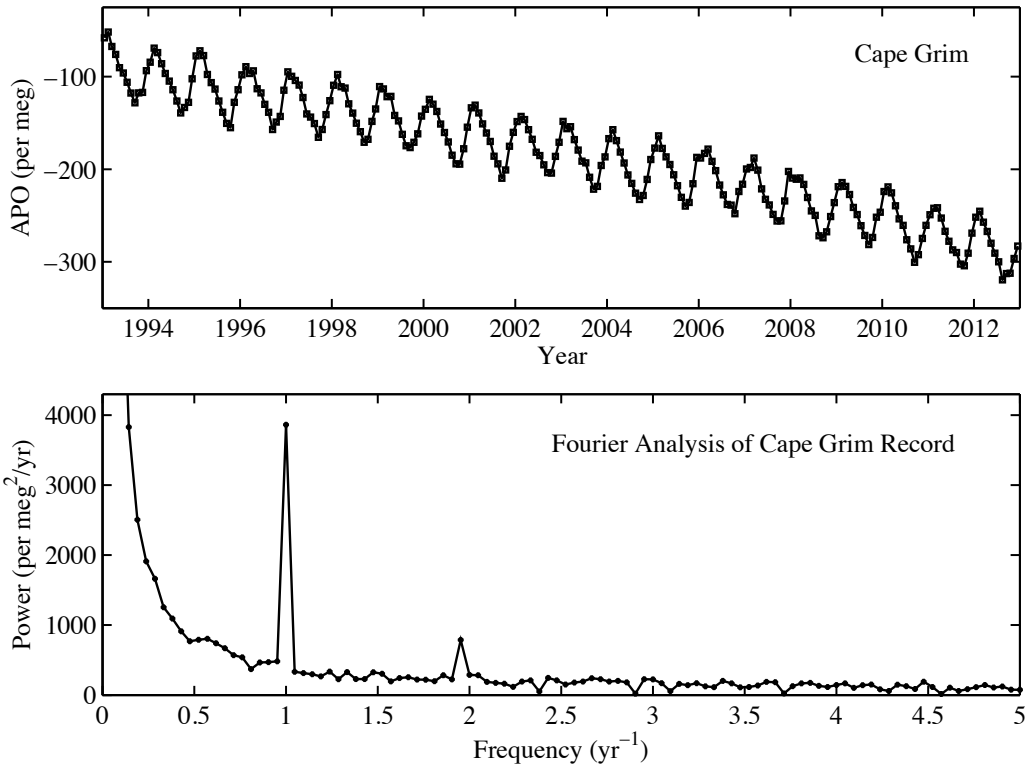

Figure 3.5. Top Panel: Cape Grim Observatory record of monthly mean atmospheric potential oxygen data between 1993 and 2013, from the Scripps Oxygen program. Bottom Panel: Fourier analysis of the record above, showing annual and biannual frequency peaks.

Figure 3.5 shows the Scripps O_2 Program APO record from Cape Grim over a twenty year time span. Fourier analysis shows large but seasonally insignificant peaks below 0.5 yr^{-1} that are related to the trend in the data. Frequencies around 1 and 2 correspond to annual and biannual components. The exact nature of the biannual frequency is beyond the scope of this dissertation, so I note it here simply as justification of my methodology; however, contributing factors like phase offsets between the northern and southern hemisphere cycles, and between individual tracers that constitute the APO signal are likely sources. I address a related question—whether five points are capable of accurately capturing the phase and amplitude of the seasonal cycle—in Section 3.5.3, after I have discussed my modeling methods.

It bears mentioning that I choose to show both 1- and 2-harmonic fits in this chapter because at the outset it is unclear whether the HIPPO data, adjusted for biases and with error bars, are sufficient to constrain the second harmonic of the seasonal cycle. The majority of bias and error estimation in this chapter is done using 2-harmonic fits, as these allow me to best capture seasonal temporal asymmetry, and fit most closely the shape of known station records. I show comparisons of 2-harmonic data and model output fits in Section 3.7, however, given the fewer degrees of freedom in the 1-harmonic fits, I also show these results, as these provide an arguably less ambiguous picture of the timing of the seasonal cycle.

3.2.7 Interannual detrending

HIPPO data are taken at five points in the seasonal cycle over three calendar years. The burning of fossil fuels (and to a lesser extent changes in ocean sink terms and land use) leads to a secular positive trend in atmospheric CO_2, coupled with a secular negative trend in atmospheric O_2. Because the concentration of these two gases changes with time, I must remove these trends from the data to quantify only the climatological signals associated with seasonal changes. Detrending data and output, however, is complicated by several factors. First, no single computed atmospheric "mean" value exists for the entire earth's atmosphere, and no single secular APO trend exists, as all sample stations report subtly different rates. Second, model output produces an interannual trend that is imperfectly representative of the observed trend. Third, detrending data or output at individual sites, or range of latitudes independently produces

seasonal cycles symmetrical about zero, but removes any annual mean offsets in the cycles that might be due to annual mean sinks or sources.

I avoid most of these complications by choosing to detrend all data and model output by a long-term deseasonalized linear fit to the corresponding observed or modeled South Pole record, similar to previous studies [*R. F. Keeling et al.*, 1998a; *Stephens et al.*, 1998]. When considering interannual and synoptic variability in model output and data (pertinent to assessing bias in Section 3.5.5), I remove the appropriate linear deseasonalized South Pole trend from the data (3-5 year trend, depending on the instance), preserving some long-term and interannually varying spatial gradients and all seasonal signals. When computing a seasonal mean cycle, however, I remove the linear SPO trend *and* fit a one-, or two-harmonic fit to what remains: the seasonal + interannual residuals. For model output, such a fit, applied over multiple years, averages out interannual differences, producing a mean seasonal cycle which is described by the fit coefficients. For the 5 HIPPO observations, any interannual variability is treated as seasonal variability by the fit procedure, and is accounted for in the error estimate for the five HIPPO curtain average values in **Table 3.5**.

3.3 HIPPO Southern Ocean results

Figure 3.6 and **Figure 3.7** show "curtain" plots—or contours along a slice of the atmosphere, here a meridian—of measured and detrended APO, respectively, over the Southern Ocean near the dateline to a little over 200 mb for the five HIPPO missions (here arranged by year day as HIPPO1, 3, 4, 5 and 2 for consistency with subsequent

plots). Contours show interpolated AO2 data, while filled colored circles show MEDUSA flasks. Because AO2 values are anchored to a 3-flask running mean of MEDUSA values, agreement between the two measurements at flask sample locations is close, but not exact. To give a sense for the scale of the secular trend in APO, I first show the data as sampled—variously in 2009, 2010, and 2011 (**Figure 3.6**)—and then with the interannual trend removed (**Figure 3.7**). Note that, as expected from an interannual trend of -11 per meg/year over almost three years [*R. F. Keeling*, 2013], the range in the APO data is approximately 30 per meg smaller in the detrended plot.

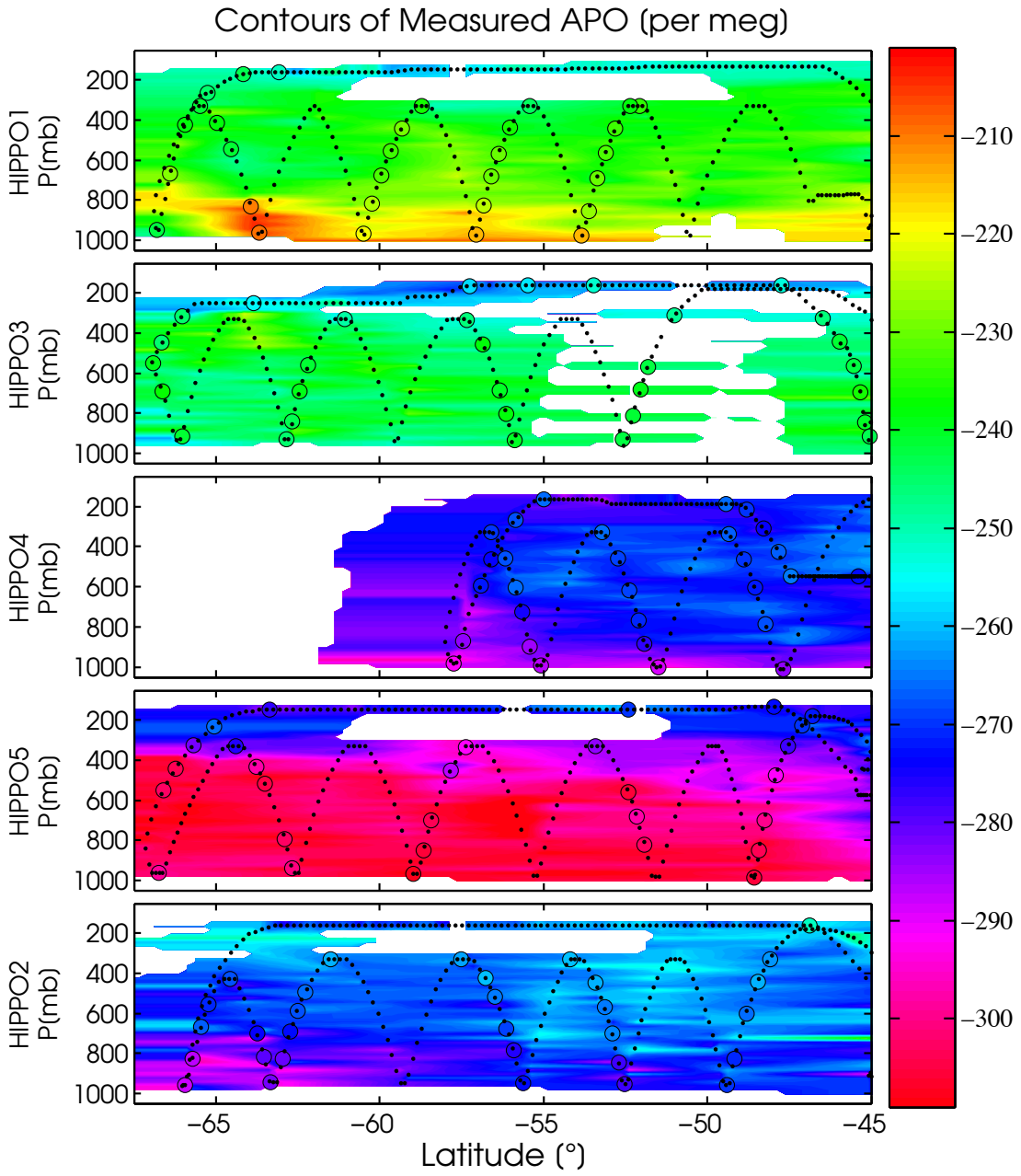

Figure 3.6. Curtain plots of MEDUSA flask and AO2 APO from the Southern Ocean flights as measured. Interpolated AO2 values are shown as color contours, while MEDUSA flasks are shown as colored circles. Flight tracks are shown as small black dots. For consistency, I plot in order of year day, here showing missions in the following order: HIPPO1 (YD20), HIPPO3 (YD95), HIPPO4 (YD179), HIPPO5 (YD241), and HIPPO2 (YD315). The gap in HIPPO3 is due to a pressure-control issue, which led to a very noisy signal; I have excluded it for the present, and Dr. Britton Stephens has interpolated these values using MEDUSA flasks. Units: [per meg].

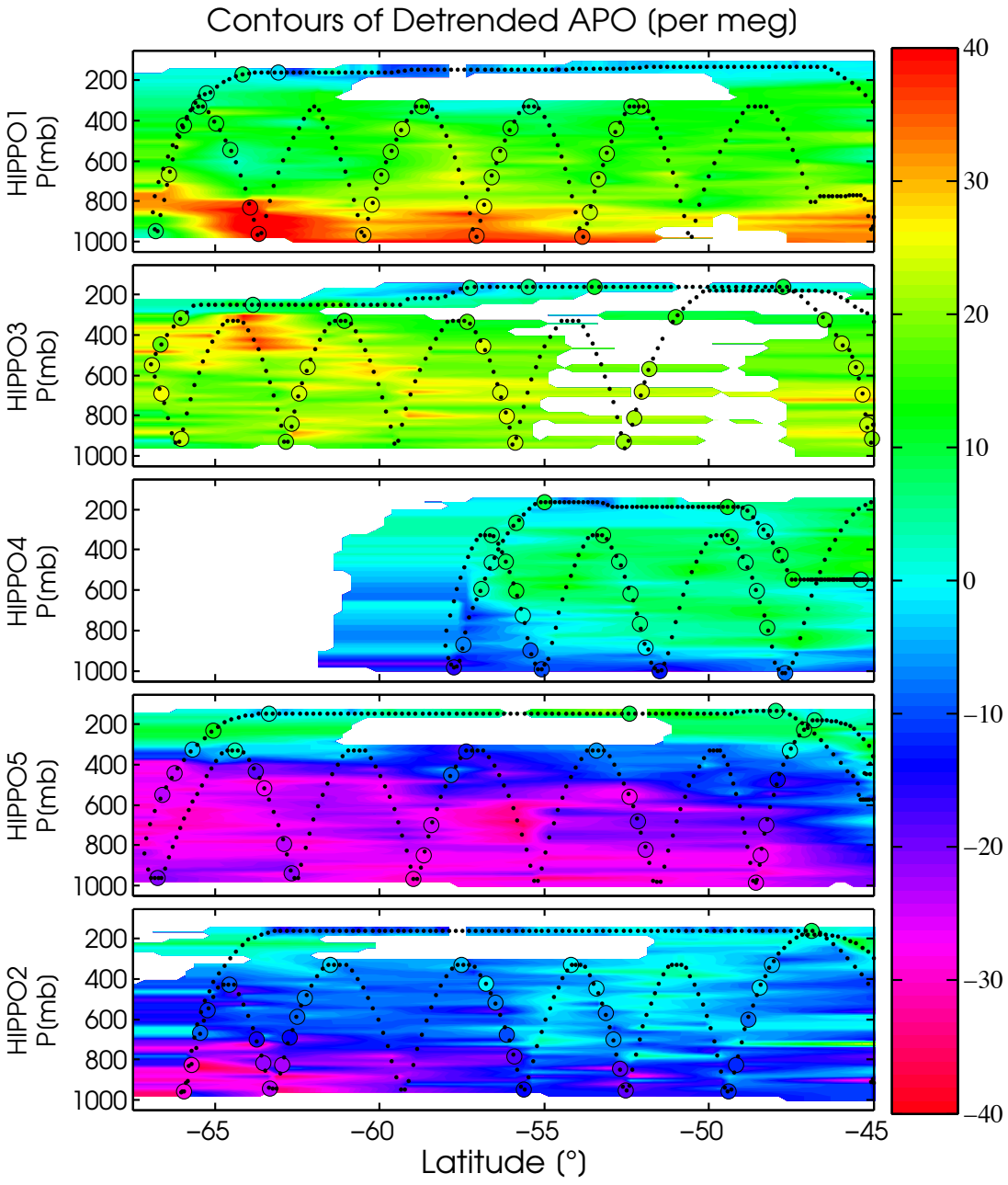

Figure 3.7. The same plot as above, but for interannually detrended MEDUSA flask and AO2 APO. All data have been detrended by the deseasonalized South Pole APO record. Units: [per meg].

Because I am interested principally in climatological differences between the seasons, the following paragraphs refer to the detrended data shown in **Figure 3.7**.

The HIPPO1 curtain plot corresponds to mid austral summer, shortly before the seasonal APO maximum at the ocean's surface as shown by Palmer and Cape Grim time series records [*Hamme and Keeling*, 2008]. At this point in the seasonal cycle, phytoplankton are producing oxygen strongly, while consuming dissolved inorganic carbon, encouraging the strong outgassing of O_2 and (smaller) ingassing of CO_2 from the atmosphere. Relatively high surface ocean temperatures also encourage the outgassing of both species to the atmosphere. The curtain plot shows a strong vertical gradient in APO, of roughly 25-30 per meg (here I define a positive gradient as higher at the surface than at elevation, and refer to the difference between surface and 300 mb) which is greatest near the southern extent of the plot. The southernmost profile also shows a brief, but markedly APO-depleted air mass which is reflected both in a MEDUSA flask and unadjusted AO2 data.

HIPPO3 took place in early austral fall when surface productivity has dropped, and the ventilation of shoaling subsurface waters may have begun in earnest. The curtain plot shows the first signs of an inverted profile near its southernmost extent on order 5-10 per meg, with high APO at mid altitudes, and lower APO below. The AO2 instrument suffered from an intermittent pressure-control related noise problem on portions of this flight. It may be possible to recover more of this data with more detailed filtering of the raw data, but for now I have excluded it, patching in MEDUSA flask observations instead. Data on either side of the masked-out region, plus the MEDUSA data within this region, indicate little or no horizontal gradient in APO, suggesting that an average of the

entire plot is likely to characterize the mean conditions well, and is unlikely to miss a significantly different air mass (this is confirmed in **Figure 3.15**).

HIPPO4 occurred in early austral winter, and shows a generally inverted profile with vertical gradients as large as -25 per meg. The gradient is relatively uniform over the sampled area, though it increases slightly at its southernmost point. Unfortunately, the GV was unable to perform a full transect to 67˚S due to the ash cloud from the Puyehue-Cordón Caulle volcanoes, and the pilots were forced to turn back due to restricted airspace at 57˚ South.

HIPPO5 occurred in late austral winter, just before the seasonal minimum as seen by surface stations. The average atmospheric concentration of APO here had dropped roughly 25 per meg since HIPPO4, only two months earlier, over a period of time where ventilation of subsurface water is expected to pull O_2 out of the atmosphere, while releasing CO_2 at a slower rate, and at a time when cooling of surface water is expected to pull both out of the atmosphere. The curtain plot shows that the vertical gradient has diminished from HIPPO4 with tropospheric gradients of roughly -10 to -15 per meg near the north edge of the curtain plot, and roughly 0 to -5 per meg near its southern extent.

HIPPO2 occurred in austral spring when the Southern Ocean is transitioning from an O_2 sink into an O_2 source. Though a small inverted profile still exists, the mean APO is higher, suggesting the beginning of primary productivity, and the net outgassing of oxygen from the Southern Ocean. Notably, the mean APO at low latitude mid elevations is higher than at the surface at high latitude, suggesting that production and warming in lower latitudes has commenced before production near Antarctica, probably due to the availability of sunlight, and to a lesser extent, the absence of surface ice. The higher APO

at the southern extent of the profile at mid elevations is probably the result of the mixing in of signals from the north, or from another Southern Ocean region with greater productivity and warming than the region influencing the local boundary layer.

Overall, I observe a vertical gradient in APO that has a greater absolute magnitude in austral summer than in winter, which is in keeping with observations that the rate of change in air-sea flux is greatest during the spring and summer due to the rapid onset of primary production. Gradients in austral fall and winter are less pronounced, consistent with both reduced rates of change in fluxes, and the deepening of the planetary boundary layer due to relatively warm surface waters and cool atmospheric conditions [*C D Keeling et al.*, 1989]. Such conditions indicate a seasonal cycle at the surface with a rapid spring increase and a slow autumn/winter decline, suggesting that the seasonal cycle is asymmetrical with respect to time. Due to the mixing of this signal through the tropospheric column, however, this asymmetry may, or may not be reflected as clearly in the full curtain average.

3.4 Calculating the curtain average

I compute a data-based weighted mean atmospheric potential oxygen for each of the five Southern Ocean AO2 transects. As I show in **Figure 3.7**, this output is interpolated between flight track "saw teeth" using Akima's [1978] bivariate triangulation in an effort to better characterize the slice. I choose an interpolated resolution of 20 mbar x 0.25˚ latitude grid, mask out any regions more than 4 degrees of latitude or 25 mb from an actual observation, and calculate the latitude-weighted mean value of the remaining APO

based on these grid points between 45˚ to 65˚S, from the surface to 300 mb. For a given grid point *A(i,j)*—with indices of latitude and pressure altitude, respectively—weight, i.e. the relative contribution to the total meridional slice, is computed following the equation:

$$w(i,j) = \frac{p(i,j)_{top} - p(i,j)_{bottom}}{p(i)_{surface}} * \frac{|\cos(lat(j)_S) - \cos(lat(j)_N)|}{|\cos(lat_{sliceS}) - \cos(lat_{sliceN})|} \quad \textbf{Eq. 3.5}$$

Here, *i* and *j* refer to the meridional and vertical components of the southern ocean slice. $p(i,j)_{top}$ and $p(i,j)_{bottom}$ refer to the pressure altitude at the midpoint between *A(i,j)* and the grid points above and below it, while $p(i)_{surface}$ refers the surface pressure at the bottom of the given column. Because the upper and lower boundaries of each box are 20 mb apart, each box receives equal pressure weighting. Variables $lat(j)_S$ and $lat(j)_N$ refer to the latitude at the midpoint between *A(i,j)* and the grid points to the south and north of it, while lat_{sliceS} and lat_{sliceN} refer to the southern and northern latitudes of the entire meridional slice. I choose to evaluate the contributing weight of each column to the total based on the delta of the cosine of the latitude to reflect the diminishing volume of the atmosphere over a given latitude band as one approaches the poles, and thus make the average more reflective of spatially integrated surface fluxes. I choose not to extrapolate above 300 mb, because the seasonal cycle in the stratosphere is largely decoupled from the tropospheric cycle, and data indicate the seasonal cycle in the stratosphere is minimal (<5 per meg). I also cut off data and models at the same elevation for consistency—300 mb is within the GV's reach.

For the particular meridional slice I have chosen, the mean concentration of a gas is thus calculated as:

$$\mathrm{MI(C)} = \frac{\sum_{i=-65^{\circ}}^{-45^{\circ}} \sum_{j=p_{surface}}^{300mb} C(i,j) * w(i,j)}{\sum_{i=-65^{\circ}}^{-45^{\circ}} \sum_{j=p_{surface}}^{300mb} w(i,j)} \quad \textbf{Eq. 3.6}$$

where *w* is the weight of each box as computed above, and *C* is the concentration of a given species. This approach entails only limited extrapolation outside of the measured pressure/latitude envelope. As a result, some "boxes" (e.g. the southern portion of HIPPO4, as seen in **Figure 3.7**) do not include data due to differences in flight path or instrument issues. Similarly, I've chosen not to interpolate between points that are more than 4 degrees, or 25 mb apart (e.g. the middle latitudes of HIPPO3, as seen in **Figure 3.7**). Such boxes are masked out, and these and any other boxes with missing data, receive a weight of zero. I divide the sum in the numerator by the sum of the boxes' weights (which is typically <1) to account for such gaps in coverage. I evaluate the error associated with spatial undersampling and with this method of extrapolating and masking the measurements in Section 3.5.5.

The focus on the 45˚-65˚S latitude band is based on several considerations. First, TransCom atmospheric transport simulations (multiple ATMs with a single set of fluxes) discussed below suggest that this zone has a large seasonal O_2/N_2 signal, and the various TransCom atmospheric models agree well on the magnitude and phase of this signal. Also, surface station data exist close to the latitudes of these boundaries for the purposes of intercomparison (Palmer Station, 64.8˚S; Cape Grim 40.7˚S). I choose 65˚S specifically because this is the southernmost well characterized latitude of the Southern Ocean flights, and I choose 45˚S because HIPPO data just north of this point tend to show land and industrial signals from New Zealand. I also choose these latitudes because the air over the Southern Ocean exhibits very little seasonality that can be attributed to

terrestrial and industrial seasonal exchanges, minimizing the error these may introduce to the APO calculation.

Because the five HIPPO Southern Ocean flights are snapshots of conditions at the time of sampling, and not average conditions, I apply a number of adjustments to the data to best translate the five data values into estimates of climatological means for the season of each HIPPO flight. These include the detrending mentioned in the previous section, and model-based estimates of the bias that results from local synoptic-scale atmospheric transport effects. I also assess temporal and spatial representivity in the measurements using a combination of modeling output and time series data. The original raw, the detrended and the fully adjusted values for the five HIPPO Southern Ocean meridional curtain averages are shown after a discussion of the various contributing factors in **Table 3.4**. The various uncertainty contributions to the error bar are shown in **Table 3.5**.

Plotting curtain average values for each of the HIPPO missions with station data gives a sense for how well the height/latitude averages compare with data from established surface time series sites. Note, I do not expect these to agree, as one is a height/latitude average and the other representative of only surface observations. In the top panel of **Figure 3.8**, I compare the five AO2 curtain averages to monthly mean station data and 2-harmonic fits from Cape Grim, Tasmania (CGO) and Palmer Station, Antarctica (PSA). The lower panel of the same figure shows the same data with the deseasonalized linear South Pole trend removed, and the mean cycles for station data, based on the detrending described in Section 3.2.7. The 2-harmonic fit to the unadjusted data has a peak-to-peak amplitude of 40.1 per meg, which constitutes about 64% and 58% of the amplitudes at CGO and PSA, which are 63 and 69 per meg, respectively.

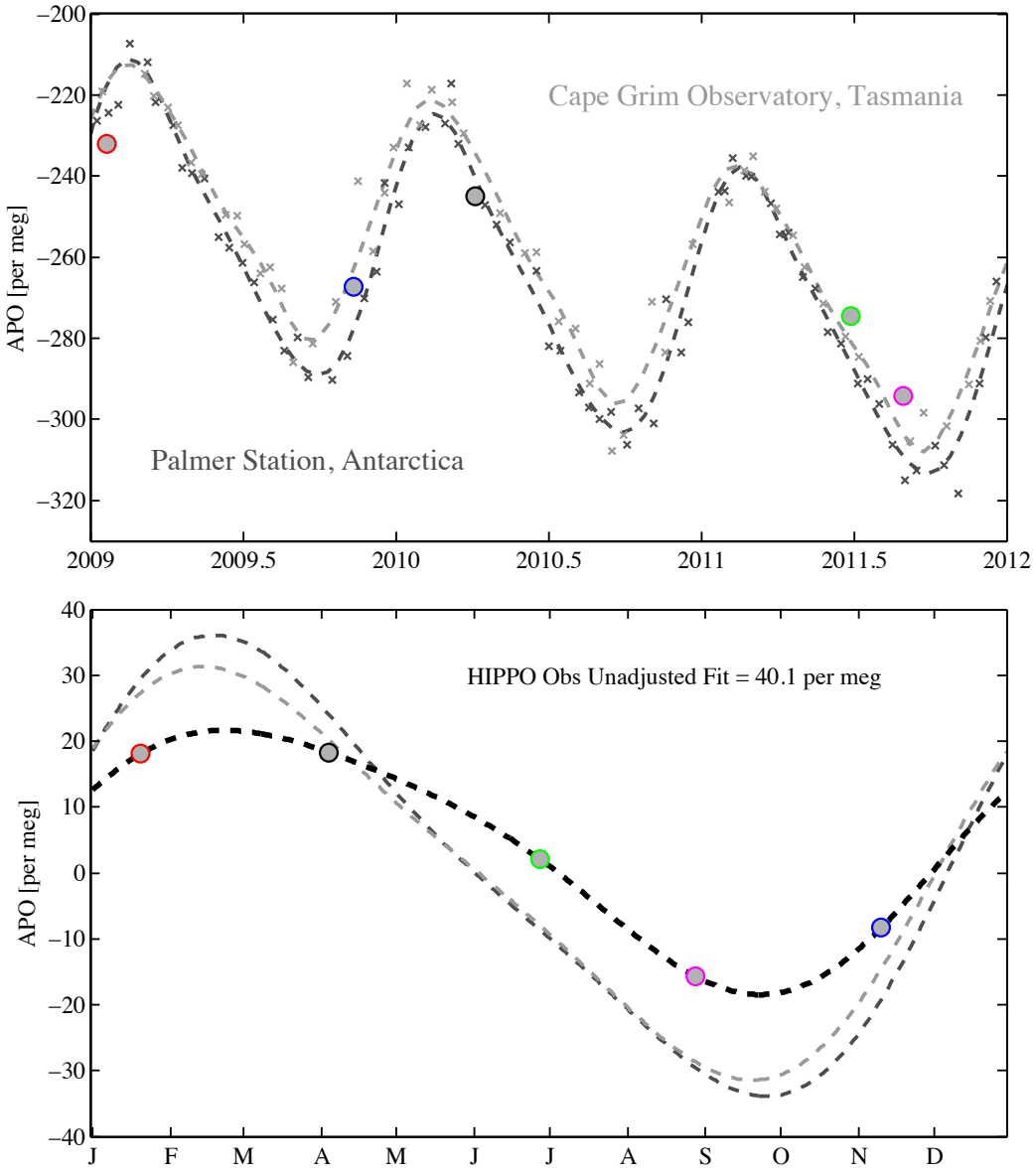

Figure 3.8. TOP: Comparison of 2009-2011 APO from surface station round bottom flask data (exes) and their 2-harmonic fits (lines) with HIPPO curtain averages (filled circles). Palmer Station values are shown in dark gray, and Cape Grim values in light gray. The five HIPPO missions are shown in order, HIPPO1 (red outline), HIPPO2 (blue outline), HIPPO3 (black outline), HIPPO4 (green outline), HIPPO5 (magenta outline). BOTTOM: Collapsed onto a single detrended climatological year: comparison of the two-harmonic fits to the PSA and CGO data and the unadjusted HIPPO AO2 data (anchored to MEDUSA) with 2-harmonic fit (thick black line). The peak-to-peak amplitude of this fit is 40.1 per meg. Units: [per meg]

3.5 Atmospheric transport model simulations: assessment of methods and transport of flux estimates

I compare HIPPO data with atmospheric transport model forward simulations whose air-sea fluxes are provided variously by ocean models and dissolved gas climatologies. These comparisons serve three purposes. I initially use a collection of atmospheric transport model output from the TransCom3 APO project to test the assumption that a curtain average of APO from 45° to 65°S is less sensitive to uncertainty in atmospheric transport than surface data alone, allowing the performance of ocean models and climatologies to be evaluated without the uncertainty introduced by transport model vertical mixing. I then use multiple-year atmospheric transport simulations, run by Dr. Sara Mikaloff-Fletcher at the National Institute of Water and Atmospheric Research (NIWA, New Zealand) and Dr. Prabir Patra at the Japanese Agency for Marine-earth Science and TEChnology (JAMSTEC, Japan) to estimate the error due to the nature of HIPPO sampling and my CA calculation methodology. Finally, I compare observations to atmospheric transport model output to evaluate which ocean flux estimates appear to best predict the phasing and amplitude of the seasonal APO signal over the Southern Ocean.

I test the assumption that a curtain average is less sensitive to transport uncertainty using output from the TransCom3 [*Baker et al.*, 2006; *Gurney et al.*, 2003] O_2/N_2 forward simulations coordinated by Blaine [2005], which the authors called the "TransCom APO Tracer Transport Experiment". I will refer to this experiment hereafter as the "TransCom O_2 Experiment" since this is slightly more accurate, and simpler to express. These simulations use surface O_2 fluxes from Garcia and Keeling [2001] and

Blaine's own N_2 fluxes. I perform the same curtain average on output from six transport models over the Southern Ocean to determine whether an atmospheric weighted average brings TransCom models into better agreement with each other than output at surface stations.

I then compare HIPPO APO results against a suite of ocean model fluxes as run through the TM3 [*Heimann and Körner*, 2003] and ACTM [*Patra et al.*, 2005] transport models. These runs are detailed variously in **Table 3.2** and **Table 3.3**. I rely primarily on TM3 runs to observe the differences in a curtain average between multiple ocean models, and compare TM3 output with ACTM for dissolved climatology runs—whose reanalysis winds are available through the dates of the five HIPPO missions—to assess the impact of atmospheric transport differences. All runs were allowed to spin up for at least four years. Of the spun-up model years, I compare five that are common to all runs, 1999-2003.

TM3 simulations by Sara Mikaloff-Fletcher were run with 6-hourly NCEP2 reanalysis meteorology and at a spatial resolution of 3.75˚ lat x 5˚ lon x 19 levels. These simulations use annual-mean CDIAC fossil fuel emission estimates (using the 1995 distribution), which I scale to the time of the HIPPO flights by the ratio of CDIAC global emission estimates for the two periods. I calculate the associated depletion of O_2 during combustion by removing 1.4 moles of O_2 for every mole of CO_2 added to the air [*R. F. Keeling*, 1988; *Manning*, 2001].

ACTM simulations by Prabir Patra were also run with 6-hourly NCEP2 reanalysis meteorology, but at a spatial resolution of 2.5˚ lat x 2.5˚ lon x 32 levels [*Kanamitsu et al.*, 2002]. These simulations used EDGAR 4.2 fossil fuel emission estimates [*Olivier et al.*,

2012]. Differences in fossil-fuel flux estimates, and the use of a single O_2:CO_2 scaling factor instead of applying ratios by fuel type, account for only exceedingly small differences in the calculated Southern Ocean meridional curtain average values (<1% of seasonal cycle). The same ACTM output are produced along the HIPPO flight tracks at the local noon time of the individual HIPPO missions as a means toward estimating spatial sampling bias, as detailed in 3.5.5.

The oceanic O_2, CO_2 and N_2 fluxes used in the TM3 simulations come from six ocean model runs, which I refer to as: CCSM3, CESM, MOM4, NEMO-PISCES-T, NEMO-CNTRL and NEMO-WSTIR.

The CCSM3 [*Collins et al.*, 2006] and CESM [*Long et al.*, 2013] runs refer to forced runs of the penultimate and present generations of the ocean component of the NCAR Community Earth System Model. Officially, the ocean component of CESM is from CCSM4, however, for consistency with previous presentations of this data, and with a recent paper by the developers of the model, using a version of the model [*Long et al.*, 2013] forced with the CORE V2 reanalysis [*Large and Yeager*, 2009], I choose to adopt the convention of referring to the forced CCSM4 runs as "CESM". The CCSM3 and CESM models use similar, related versions of the Biogeochemical Elemental Cycling module [*Doney et al.*, 1996], which is run within two quite different versions of the POP physical oceanography model—CCSM3 runs v.1.4.3, while CESM runs v.2 [*Smith et al.*, 2010]. Version 2 implements numerous improvements, including a time- and space-varying mesoscale eddy-induced advection coefficient which improves the circulation response to variable wind forcings (particularly important in the Southern Ocean), an improved boundary-layer parameterization which implements latitudinally-varying

background internal wave diffusivity, and improved spatial and temporal resolution [*Long et al.*, 2013].

MOM4 is a recent published version of NOAA Geophysical Fluid Dynamics Laboratory's Modular Ocean Model [*Griffies et al.*, 2000] coupled with the TOPAZ biogeochemistry model [*Dunne et al.*, 2010], and has featured in IPCC assessment reports 4 and 5, and in the Coupled Model Intercomparison Project (CMIP5, [*Taylor et al.*, 2012]). MOM4 uses a tri-polar grid to resolve arctic flows at a nominal 1 degree resolution, and includes various physical parameterizations including fresh water input to surface ocean, neutral and sub-grid-scale diffusion, and an explicit surface mixed layer [*Griffies et al.*, 2005]. MOM4 employs a relatively new third-order upwind-biased tracer scheme that improves upon previous second-order schemes. MOM4 also accounts for solar shortwave radiation differently than in previous iterations of the model. In a coarser-resolution model (50 m surface depth), a single deep surface layer would absorb all of the solar shortwave, but with finer layers (10 m), up to 20% of the flux can penetrate into multiple layers beneath 10 m. Not accounting for this creates a spurious accumulation of heat in the upper cell only. To combat this, the modelers decided to generate a season- and space-varying climatology of short-wave radiation penetration depths to give optimized model results.

NEMO-PISCES-T, NEMO-CNTRL and NEMO-WSTIR all use versions of the NEMO model (Nucleus for European Modelling of the Oceans) with versions of the NEMO-OPA physical oceanography model [*Madec et al.*, 1998], and versions of the PISCES biogeochemistry model [*Aumont et al.*, 2003] with the ORCA2 trip-polar gridding scheme, which produces self-consistent circumpolar flows. The NEMO-

PISCES-T simulation is an older, but high profile version of the NEMO model run by Corinne Le Quéré [*Le Quéré et al.*, 2007] using NCEP-reanalysis-forced dynamics, and with a variant of the original PISCES biogeochemistry model—PISCES-T—that allowed for temperature-dependent remineralization and separate characterizations of micro- and mesozooplankton [*Buitenhuis et al.*, 2006]; NEMO-CNTRL and NEMO-WSTIR simulations were run by Rodgers et al. [2014] using NEMO 3.2 and PISCES [*Aumont and Bopp*, 2006], and using the DRAKKAR upper-ocean forcing set #4.1 [*Brodeau et al.*, 2010] based on ERA-40 wind reanalysis [*Uppala et al.*, 2005]. The WSTIR run includes a new wind stirring parameterization which forces model mixed layer depth to be consistent with ARGO-float observations by latitude, and which allows for the transfer of wind energy to the water column below the mixed layer, to test the sensitivity of Southern Ocean carbon exchange to wind stirring depth. CNTRL reflects the standard "control" configuration against which the parameterization is compared. This study comprises one of the first, albeit brief, examinations of the NEMO-WSTIR output, which has kindly been availed to Sara Mikaloff Fletcher and me at this early stage by Drs. Keith Rodgers and Olivier Aumont.

The long (multi-multi-century) spin-up of NEMO-PISCES (for the CNTRL and WSTIR runs mentioned in this chapter) was conducted using the CNTRL configuration of the physical state model NEMO, with PISCES being tuned to represent the global biogeochemical state following the description in Aumont and Bopp [2006]. This CNTRL configuration was run through the year 2006. The WSTIR physical perturbation run was split from the CNTRL run abruptly in 1958 (the onset of the ERA-40 reanalysis forcing cycle) without re-tuning PISCES, but with a tuning of mixed layer depths to

better match the observed seasonal cycle. For this reason, one should exercise caution in interpreting either of these experiments of being a best estimate or state estimate, since by construction neither of these runs include a consistent tuning of both the biogeochemical state and the physical state. It is not clear a priori which of the two runs will better represent the observed seasonal cycle in air-sea fluxes of O_2, given that each of the runs may be said to have one of its components (physical or biogeochemical) tuned. In fact, it may be expected that the two runs will exhibit heterogeneity in skill assessment against data constraints, with the runs exhibiting different skill in different regions [*K. Rodgers*, pers. comm.].

The sixth TM3 run uses dissolved climatology-based flux fields: the seasonal-only O_2 and N_2 flux fields of Garcia and Keeling [2001] and Blaine [2005], respectively, used in the TransCom O_2 Experiment, and Takahashi et al. [2009] seasonal CO_2 flux fields. Garcia and Keeling [2001] computed monthly oxygen fluxes based on sea surface oxygen anomaly measurements. In order to fill in an otherwise sparse O_2 dataset, Garcia and Keeling (hereafter often referred to as "GK01") computed a weighted linear least squares fit to heat flux anomaly based on observations that the two quantities were correlated in time. The resulting O_2 anomalies were multiplied by the Wanninkhof [1992] gas-exchange velocity to calculate air-sea O_2 fluxes. Blaine [2005] calculated seasonal N_2 fluxes from heat flux estimates [*Gibson et al.*, 1997] using the temperature derivative of solubility relationship laid out by Keeling and Shertz [1992]. The Takahashi et al. [2009] fluxes are based on roughly three million historical measurements of pCO_2 in surface waters (normalized to a mean year of 2000), and are produced using air-sea exchange parameterizations and a spatial interpolation method. The dissolved climatology runs also

include ocean inversion-derived annual mean flux fields from Gloor et al. [2001] and Gruber et al. [2001] to provide global annual flux balances for N_2 and O_2, respectively, as Garcia and Keeling and Blaine fluxes are seasonally mass-balanced within each grid box. For simplicity, I refer to combinations of these fields as the "GK01" runs, or sometimes the "dissolved climatology runs".

When comparing a mean climatological cycle for ACTM output, I show ACTM dissolved climatology runs over the same shared five-year period between 1999 and 2003. However, I also use output from 2009 to 2011 (inclusive) to assess spatial and synoptic bias in the HIPPO measurements because these years reflect the wind conditions (from NCEP2 reanalysis) at the time of sampling.

3.5.1 Assessment of the curtain average metric

I begin with a proof of principle application of the curtain average (CA) method using the TransCom [*Baker et al.*, 2006; *Gurney et al.*, 2003] O_2 Experiment coordinated by Blaine [2005] to determine whether an atmospheric weighted average brings transport models into greater agreement than at fixed points on the globe. Blaine submitted monthly oxygen [*Garcia and Keeling*, 2001] and nitrogen [*Blaine*, 2005] fluxes to participating TransCom modelers. The modelers ran a simplified 3-year forward simulation of the fluxes using a single year of repeated winds (which varied between models) and an initial tracer value of 300 ppm, for which fluxes were turned on during year one, and not during years 2 and 3. Years 1-3 are then summed, with the background value removed, and the resulting sum represents a single climatological year.

The TransCom O_2 experiment neglects the contribution of air-sea CO_2 flux to the seasonal APO variations, but for my purposes this is negligibly important, since my intent is to test whether models agree on the distribution of a tracer over large spatial scales. This tracer could effectively be any species with a seasonal cycle, as long as the cycle was based on air-sea exchange.

I compute the O_2/N_2 ratio using Blaine's monthly O_2 and N_2 anomaly concentrations, taking care to multiply all N_2 output by -1 to address a sign convention in the N_2 flux input files that was carried through to the concentration values. I then sum the values from each of the run's three years to reflect the intended methodology of Blaine, and remove the background value. I then isolate the region of interest, and identify all grid boxes that lie partly or completely within it. Boxes that lie partly within the region are ascribed a contributing weight according to the percentage of inclusion and the cosine of the latitudes of the upper and lower bounds of the box itself. I then compute the contribution of the O_2/N_2 in each box to the regional total, as described in Section 3.4 for APO, and calculate a mean O_2/N_2 ratio curtain average.

I choose to show output from the six models which met the requirements of my study—that the metadata associated with station data explicitly document which vector corresponds to which station to avoid accidentally plotting the wrong station (a problem with one model that would require a rerun since the documentation is missing); that their O_2 and N_2 fields be produced separately to allow correction of the sign convention in the N_2 fluxes mentioned above, and that all 36 months be characterized, so as to not bias results for a given month (one model lacks January data for year 1). These models include GCTM [*Levy*, 1982; *Mahlman and Moxim*, 1978], GISS [*Hansen et al.*, 1997],

JMA [*Iwasaki*, 1998; *Taguchi*, 1996], MATCH [*Law and Rayner*, 1999; *Rasch et al.*, 1997], NIES [*Maksyutov and Inoue*, 2000], and NIRE [*Taguchi*, 1996].

Figure 3.9 shows TransCom output for these six models at Palmer Station and Cape Grim Observatory compared to a weighted average of the atmosphere over the dateline between 65°S and 45°S. The modeled amplitudes at Palmer and Cape Grim vary across models by ± 10.5% and 18.7% (1σ), respectively, while the amplitudes of the curtain averages agree to within a tighter 6.4%. Small differences in phasing between models can be seen, which can be attributed primarily to the particular interpolation scheme the individual modeler chose to interpolate between mid-monthly flux values. The remaining differences in the amplitudes of the curtain averages may be attributed to slight differences in how these models mix horizontally across the 65°S and 45°S boundaries, which relates to the choice of winds used by each model. It is also very possible that the use of only 9 fairly coarse output levels for the TransCom models is driving the remaining difference, as any trapping of signals below the lowest level will not be captured, and I note that the same comparison for ACTM and TM3 with higher resolution vertical output shows almost perfect agreement (see **Figure 3.28**).

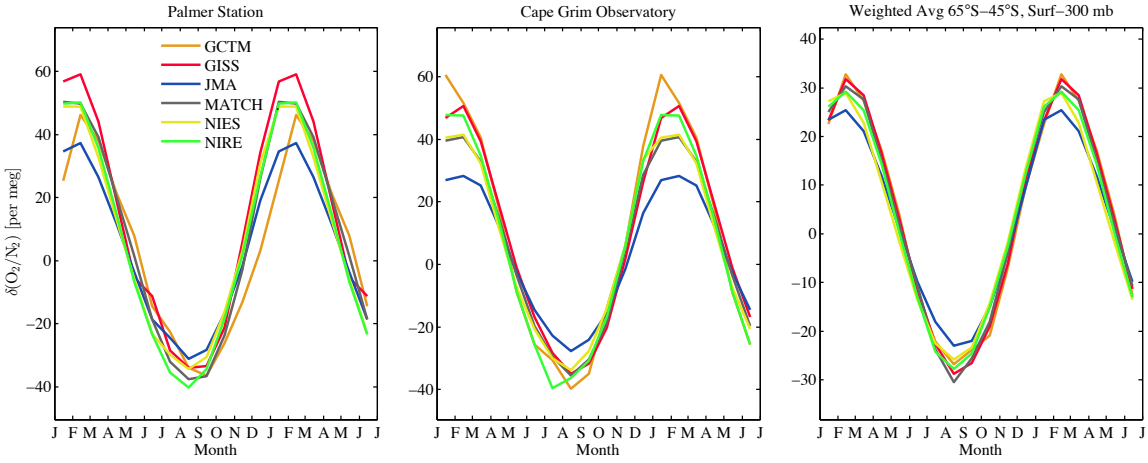

Figure 3.9. TransCom model output: Palmer and Cape Grim mean O_2/N_2 climatology and mean climatology for the "curtain average" of the atmosphere over the Southern Ocean between 65°S and 45°S, from the surface to 300 mb, at the dateline. Model output is from the TransCom O_2 Experiment, and details of the individual models can be found in Table 3.2 here and in Table 2.1 of Blaine [2005]. Models are shown over 18 months to highlight the December-January transition (i.e. months 13-18 are simply repeats of months 1-6). The Y-axis is scaled to show 1.25x the amplitude of the model which shows the largest O_2/N_2 seasonality. This means that for the atmospheric weighted average, shown in the right hand panel, the scale is zoomed in by approximately a factor of 2. Units: [per meg].

Table 3.2. TransCom O_2 Experiment models used in this study [*Blaine*, 2005; *Gurney et al.*, 2000], all of which ran Garcia and Keeling [2001] O_2 fluxes and Blaine N_2 [2005] fluxes. Additional details about the models can be found in Table 2.1 of Blaine [2005]. I have included only models that met the requirements of the present study, as mentioned above. Number of levels in the TransCom models varies. However, the TransCom protocol required that output be interpolated to 9 pressure levels for consistency between models.

Transport Model	Winds	Output Resolution (lat /lon/ layers)	Winds	Submitted By	Citation
GCTM	ZODIAC GCM	2.4° x 2.4° x 9	ZODIAC GCM	D. Baker	[*Levy*, 1982; *Mahlman and Moxim*, 1978]
GISS	GISS GCM II	5° x 4° x 9	GISS GCM II	I. Fung	[*Hansen et al.*, 1997]
JMA-CDTM	JMA	2.5° x 2.5° x 9	JMA	T. Maki	[*Iwasaki*, 1998; *Taguchi*, 1996]
MATCH	MACCM 2	5.6° x 2.8° x 9	NCEP	R. Law	[*Law and Rayner*, 1999; *Rasch et al.*, 1997]
NIES	ECMWF (97)	2.5° x 2.5° x 9	ECMWF (97)	S. Maksyutov	[*Maksyutov and Inoue*, 2000]
NIRE	ECMWF (95)	2.5° x 2.5° x 9	ECMWF (95)	S. Taguchi	[*Taguchi*, 1996]

Table 3.3. Forward simulations with reanalyzed winds, using TM3 and ACTM transport models, analyzed in this study. If fluxes of mean and seasonal components were run separately, both are noted, otherwise mean (M) and seasonal (S) fluxes come from same source.

Model Run	O_2 Fluxes (seas/mean)	CO_2 Fluxes (seas/mean)	N_2 Fluxes (seas/mean)	Fossil Fuel Fluxes	TM Resolut. (lat/lon/ layers)
ACTM+ Dissolved Climat.	**S**=[*Garcia and Keeling*, 2001] **M**=[*Gruber et al.*, 2001]	**S+M** = [*Takahashi et al.*, 2009]	**S**=Scaled to heat [*Blaine*, 2005] **M**=[*Gloor et al.*, 2001]	**CO_2**= EDGAR4.2 **O_2** = -1.4 * EDGAR4.2	2.5ºx2.5ºx 32
TM3+ Dissolved Climat.	**S**=[*Garcia and Keeling*, 2001], **M**=[*Gruber et al.*, 2001]	**S+M** = [*Takahashi et al.*, 2009]	**S**=Scaled to heat [*Blaine*, 2005] **M**=[*Gloor et al.*, 2001]	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ CESM	**S+M**=CESM [*Long et al.*, 2013]	**S+M**=CESM	**S+M**=Scaled to CESM heat fluxes	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ CCSM3	**S+M**=CCSM3 [*Collins et al.*, 2006]	**S+M**=CCSM3	**S+M**=scaled to CCSM3 heat fluxes	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ MOM4	**S+M**=MOM4 [*Dunne et al.*, 2010]	**S+M**=MOM4	**S+M**=scaled to MOM4 heat fluxes	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ NEMO-CNTRL	**S+M**=NEMO-CNTRL [*Rodgers et al.*, 2014]	**S+M**=NEMO-CNTRL	**S+M**=scaled to NEMO-CNTRL heat fluxes	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ NEMO-WSTIR	**S+M**=NEMO-WSTIR [*Rodgers et al.*, 2014]	**S+M**=NEMO-WSTIR	**S+M**=scaled to NEMO-WSTIR heat fluxes	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19
TM3+ NEMO-PISCES-T	**S+M**= NEMO-PISCES-T [*Le Quéré et al.*, 2007]	**S+M**=NEMO-PISCES-T	**S+M**=provided by C. LeQuéré	**CO_2**= CDIAC **O_2** = -1.4 * CDIAC	3.75ºx5ºx 19

Table 3.3. Forward simulations with reanalyzed winds, using TM3 and ACTM transport models, analyzed in this study. If fluxes of mean and seasonal components were run separately, both are noted, otherwise mean (M) and seasonal (S) fluxes come from same source. (continued)

Model Run	Yrs. run	Winds	Run By	Transport Model Citation
ACTM+ Dissolved Climat.	1995-2011	NCEP2 6-hour	Prabir Patra (JAMSTEC)	[*Patra et al.*, 2005]
TM3+ Dissolved Climat.	1995-2004	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ CESM	1991-2006	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ CCSM3	1995-2004	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ MOM4	1991-2005	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ NEMO-CNTRL	1995-2006	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ NEMO-WSTIR	1995-2006	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]
TM3+ NEMO-PISCES-T	1995-2003	NCEP2 6-hour	Sara Mikaloff-Fletcher (NIWA)	[*Heimann and Körner*, 2003]

3.5.2 Assessment of zonal representivity

Because HIPPO samples were taken over a longitudinally limited slice of the Southern Ocean, I was curious to determine how well these samples represent a zonal mean, and whether their interpretive power can be extended to constraining zonally integrated Southern Ocean seasonal net outgassing. I approach this question by investigating how a latitude-weighted curtain average from TransCom model output between 65°S and 45°S at the dateline—a rough approximation of the HIPPO SO flight tracks—compares with the mean curtain average of the entire zone.

In **Figure 3.10**, I plot the difference between the monthly mean zonal and dateline curtain averages. Comparing to the magnitude of the CA shown in **Figure 3.9**, TransCom output suggests that the agreement of the two quantities is very close, with most differences under one per meg on a monthly basis, and with all deltas below 2.5 per meg, a very small component of a modeled mean seasonal CA amplitude of ~46 per meg (**Figure 3.9**).

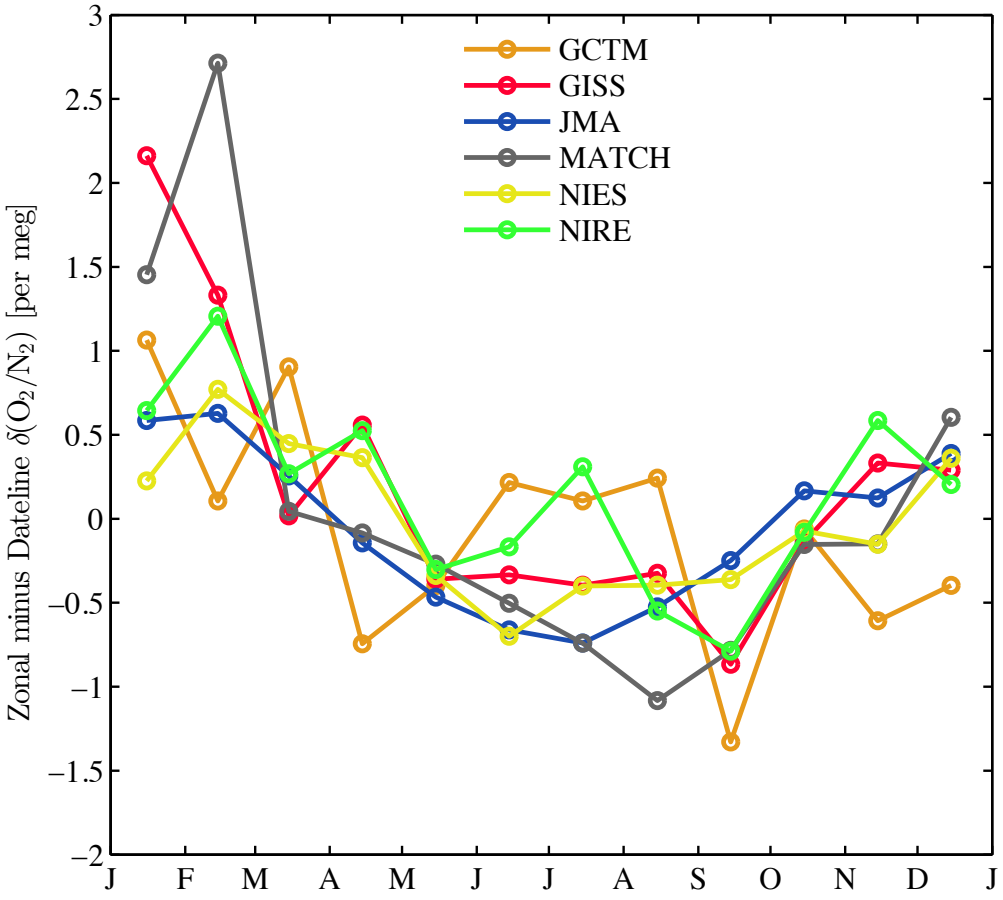

Figure 3.10. Differences between curtain averages between (65-45°S, surface to 300 mb) of TransCom O_2/N_2 output taken at 180˚W and zonally at twelve points in the seasonal cycle for each of the six models that produced output suitable for this study. The axes for this figure are zoomed by about a factor of 20 relative to the third panel in Figure 3.9 to show small differences. Units: [per meg].

3.5.3 Assessing how well five points capture the seasonal signal

I hypothesize that five points is sufficient to capture at least the first, and possibly the second harmonic of a seasonal cycle with a known frequency within reasonable limits. Initial experiments run before the HIPPO campaign to sub-sample seasonal O_2/N_2 data from the Scripps O_2 Network indicated that seasonal phase could be represented to within a week or two of the observed cycle with as few as five points taken from data with natural background synoptic variability. I have taken this analysis further using

daily-resolution curtain average output (180°W, 65°S-45°S, surface-300 mb, as described above) from the ACTM transport model using GK01 fluxes to estimate how well a 1- or 2-harmonic five point fit alone can capture seasonal amplitude and phase of data with background synoptic variability preserved. Importantly, this assessment of *fit* uncertainty differs from an assessment of the uncertainty of each of the five HIPPO curtain average points, detailed in Section 3.5.5, as individual points don't allow assessment of phase or amplitude. I remove the trend from a single year (2009) of output and perform Monte Carlo simulations using a series of 1000 randomly-chosen five-point sets to which I fit individual 1 and 2-harmonic curves. I then compare each of these against the data themselves, and a fit to all 365 daily resolution points in order to see how each of the five-point fits performs.

Because the choice of points from the daily resolution output (i.e. daily resolution output curtain averages) strongly influences the ability to interpret the seasonal signal—bunched points (e.g. YD 106:110) provide virtually no information—I choose to select points which fall within a two-week window of the Southern Ocean flights. This gives the greatest sense for whether the five HIPPO points in time can capture timing and amplitude well. Accordingly, I choose five points (small black crosses in **Figure 3.11**) at random—one each from the window of time around year days 20, 95, 179, 241 and 315, respectively—and fit them with one and two harmonics, respectively (thin gray lines), as described in Section 3.2.6. I compare 1000 of these fits to the full daily-resolution output fit (thick black line).

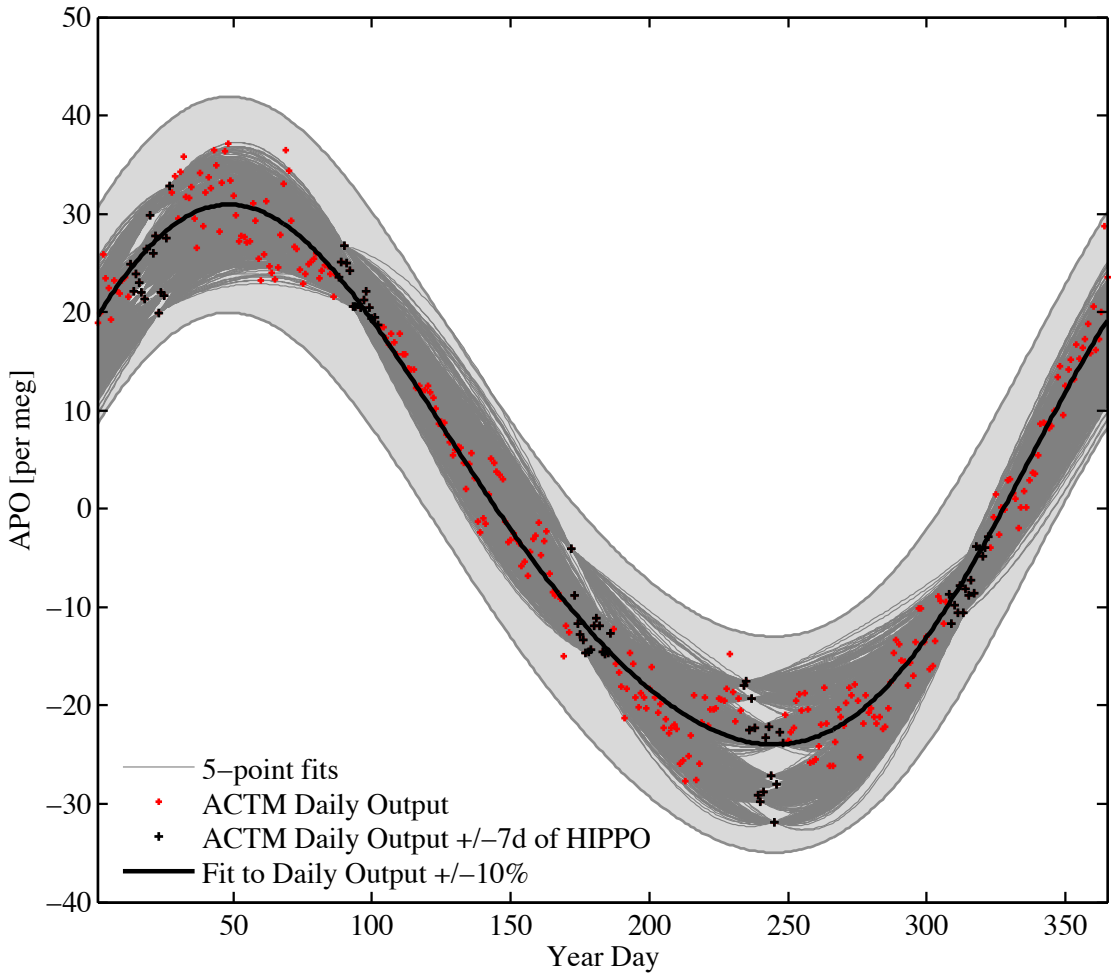

Figure 3.11. 2-harmonic fits (thin gray lines) to 5 points, randomly chosen from the 2-week windows surrounding each Southern Ocean flight from ACTM+GK01 output. Daily resolution output, from which these values are taken, are shown as small red crosses, while daily resolution values that fall within ±7 days of HIPPO mission dates are shown as small black crosses. The fit to the entire daily resolution output is shown as a dark black line with a shaded ±10% error bar on either side. Units: [per meg].

Results for 2-harmonic fits are shown in **Figure 3.11**. Fits generally perform well with amplitudes not exceeding ±10% of the fit to the full-resolution output. Fits for the 1-harmonic case, not shown, agree even more closely, to within ±5% of the daily resolution fit.

Figure 3.12 and **Figure 3.13** show the binned frequency of the timing of the 2-harmonic peak, and of the peak-to-peak amplitude, respectively. Daily resolution peak

timing, YD48, is early relative to the mean fit by 5 days, but falls within the standard deviation of ±6.3 days, or 1.7% of a year. The amplitude of the seasonal cycle has a mean of 54.9 per meg, mere hundredths of a per meg different from the daily fit, with a standard deviation of ±4.2 per meg, or 7.6% of the daily fit. Results from a 1-H fit, not shown, suggest sparse data have the potential to capture the timing and amplitude of the first harmonic of the seasonal cycle even better, with an amplitude uncertainty of ±2.7 per meg, and a phasing uncertainty of ±4.1 days relative to a fit to daily resolution model output.

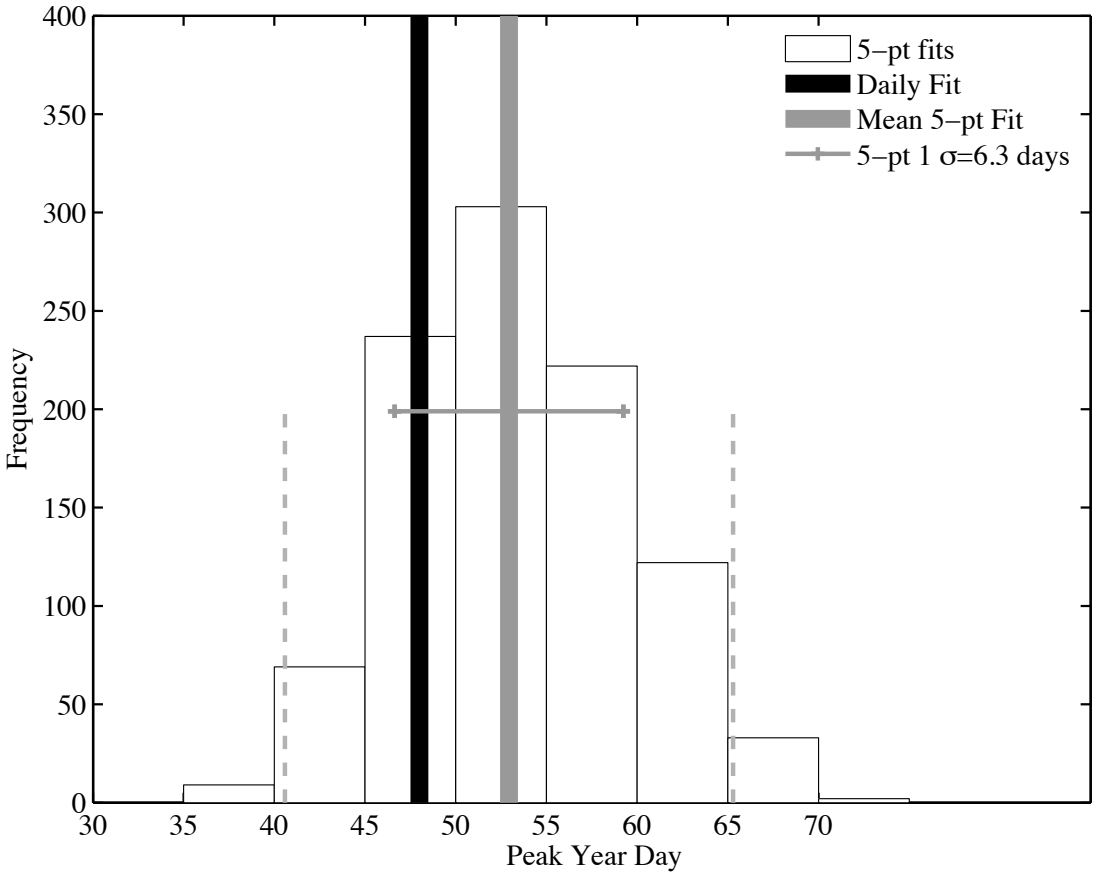

Figure 3.12. 5-day binned peak year days (white bars) from 1000 2-harmonic 5-pt fits to ACTM output. The mean of the 5-pt fits is shown in gray, with horizontal error bar. Peak year day of the fit to daily values (all 365 days) is shown in black.

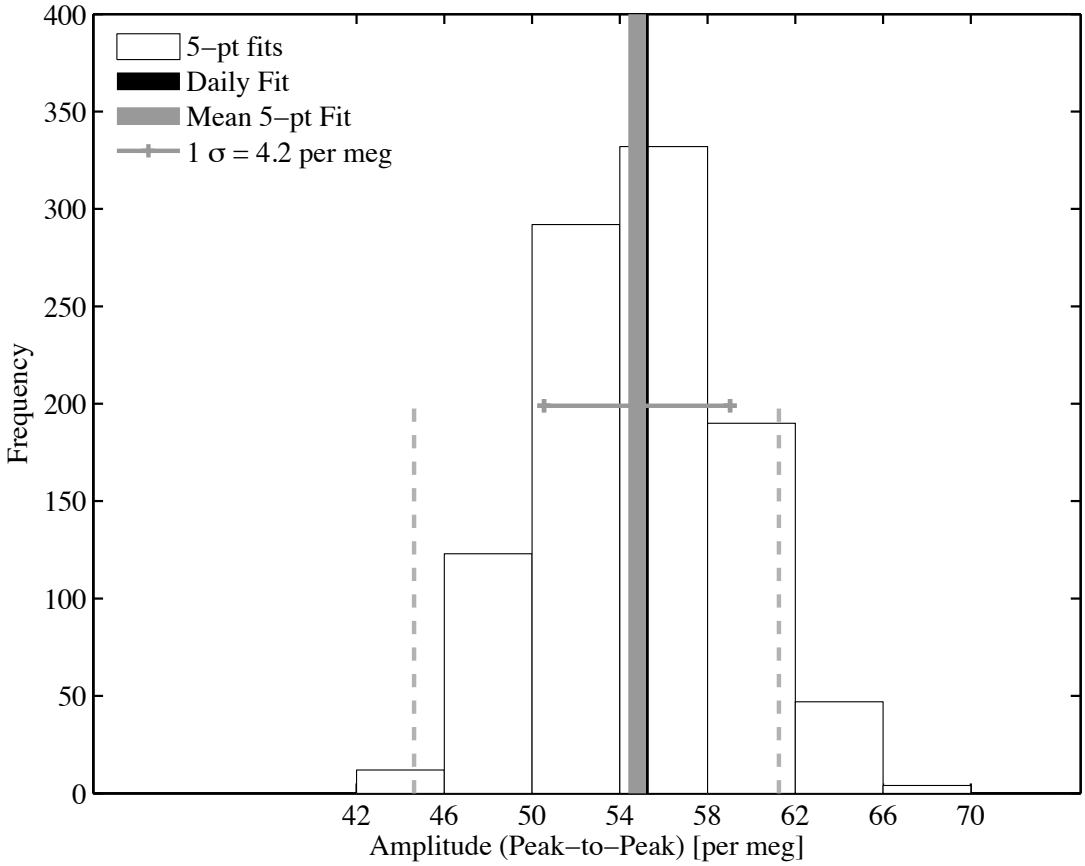

Figure 3.13. 4-per meg binned amplitudes (white bars) from 1000 2-harmonic 5-pt fits. The mean of the 5-pt fits is shown in gray, with horizontal error bar. Amplitude of the fit to daily values (all 365 days) is shown in black. Units: [per meg].

These values represent a rough impression of the initial viability of the 1- and 2-harmonic 5-pt fitting technique given synoptic-scale variability. I return to a notionally similar technique in Section 3.6.5 of this chapter to assess the *specific* fit uncertainties in our HIPPO curtain average seasonal amplitude and phasing after I have adjusted individual HIPPO curtain average values and assessed their full error bars. These later fit uncertainties are likely to be more conservative because fitting algorithms have more latitude when the uncertainty associated with each individual point is larger—i.e. when all sources of error (interannual and synoptic variability, instrument precision, correction uncertainty, etc.) are summed in quadrature.

3.5.4 Compensating for Ar/N_2 seasonal bias introduced to the curtain average

In the process of correcting the O_2/N_2 data for the related scatter seen in the Ar/N_2 data (Section 2.6), I removed the seasonal Ar/N_2 signal—which Keeling et al. [2004] have shown to exist at surface stations due to temperature-driven solubility changes—from both datasets. I reintroduce the seasonality that has been removed from the APO Southern Ocean meridional curtain average by computing the expected Ar/N_2 seasonal bias. I start with climatological mean Ar/N_2 records at Cape Grim Observatory and Palmer Station, which our station records report to have seasonal amplitudes of roughly 10 and 20 per meg, respectively. I take the average of these two measurements as a rough estimate of the mean seasonal cycle at the surface over the Southern Ocean box, so roughly 15 per meg. The seasonality of Ar/N_2 higher up in the atmosphere is not characterized in literature, so I estimate an atmospheric weighted average value instead, using the observations above, that the seasonal signal of an air-sea gas exchange tracer integrated vertically to 300 mb is approximately 60% the seasonal cycle measured at the sea surface (see **Figure 3.9**). This suggests an integrated seasonal cycle of Ar/N_2 over the Southern Ocean of approximately 9 per meg (i.e. ±4.5 per meg). Scaling this down by the ratio of the fractionation of Ar/N_2 to O_2/N_2 (3.77) used when correcting the data (see Chapter 2), I calculate a scaling adjustment of 1.06, i.e. that an upward adjustment of 6%, or 2.4 per meg, is needed in the APO amplitude. To correct for the underestimation of the APO amplitude, I adjust each of the curtain averages by an amount proportional to their deviation from the annual mean (i.e. values near the peak and trough will be adjusted more than those near the mean). These adjustments are shown in **Table 3.4**. I assume these adjustments have an uncertainty of 100%, shown in **Table 3.5** as "Ar Err".

3.5.5 Assessing sampling biases and uncertainty

I also use ACTM-GK01 model output to assess two potential biases associated with my sampling methodology and their respective uncertainties. The first is due to variable meteorological conditions on day-to-week timescales: each HIPPO Southern Ocean cross-section captures a several-hour "snapshot" of the atmosphere above the Southern Ocean near the dateline which may fail to capture the mean meteorological conditions in the days leading up to and following sampling. I refer to this as the "synoptic sampling bias". Because the NCEP reanalysis product captures the wind conditions on the actual day of the model output, it is more likely to capture the location of pressure fronts than model output with recycled winds or winds from different years than the observations. The second component is the "spatial sampling bias", which results from an incomplete characterization of the full atmospheric slice along 180°W. This is due to the lack of data from areas missed by the aircraft or instrument, and to the fact that the aircraft was not flying exactly along 180°W.

I evaluate synoptic and spatial bias with the transport model output that is most directly tied to data—a forward simulation of dissolved climatology fluxes by the ACTM transport model, which are driven with reanalyzed winds that cover the timespan of the HIPPO missions. ACTM has a finer spatial resolution than TM3 (**Table 3.3**) and also captures the magnitude of the CGO and PSA seasonal cycle relative to the weighted average the best, suggesting that it is capturing the vertical and horizontal transport better than TM3 (see **Table 3.13**).

I first estimate the synoptic bias by computing the ACTM APO curtain average for daily output gridded data between 2009 and 2011 (the latter portion of a much longer

run) for the dateline between 65˚S and 45˚S, and then compute a two-harmonic fit to the same data with a linear trend component. For each HIPPO sampling day, I calculate the difference between the modeled two-harmonic fit and the model output for that particular day, which I take as a measure of the synoptic measurement bias for each HIPPO mission. This measure technically also includes the interannual transport uncertainty due to differences in winds on a given day from year to year, but does not include interannual flux differences since the flux values are recycled year to year in the ACTM GK01 run. I apply this estimated bias as an adjustment to the AO2 curtain averages. **Figure 3.14** shows the five HIPPO offsets, both in the context of the full signal (top panel), and with the seasonal mean signal and trend removed (bottom panel). The absolute value of these offsets is between 0 and 3 per meg.

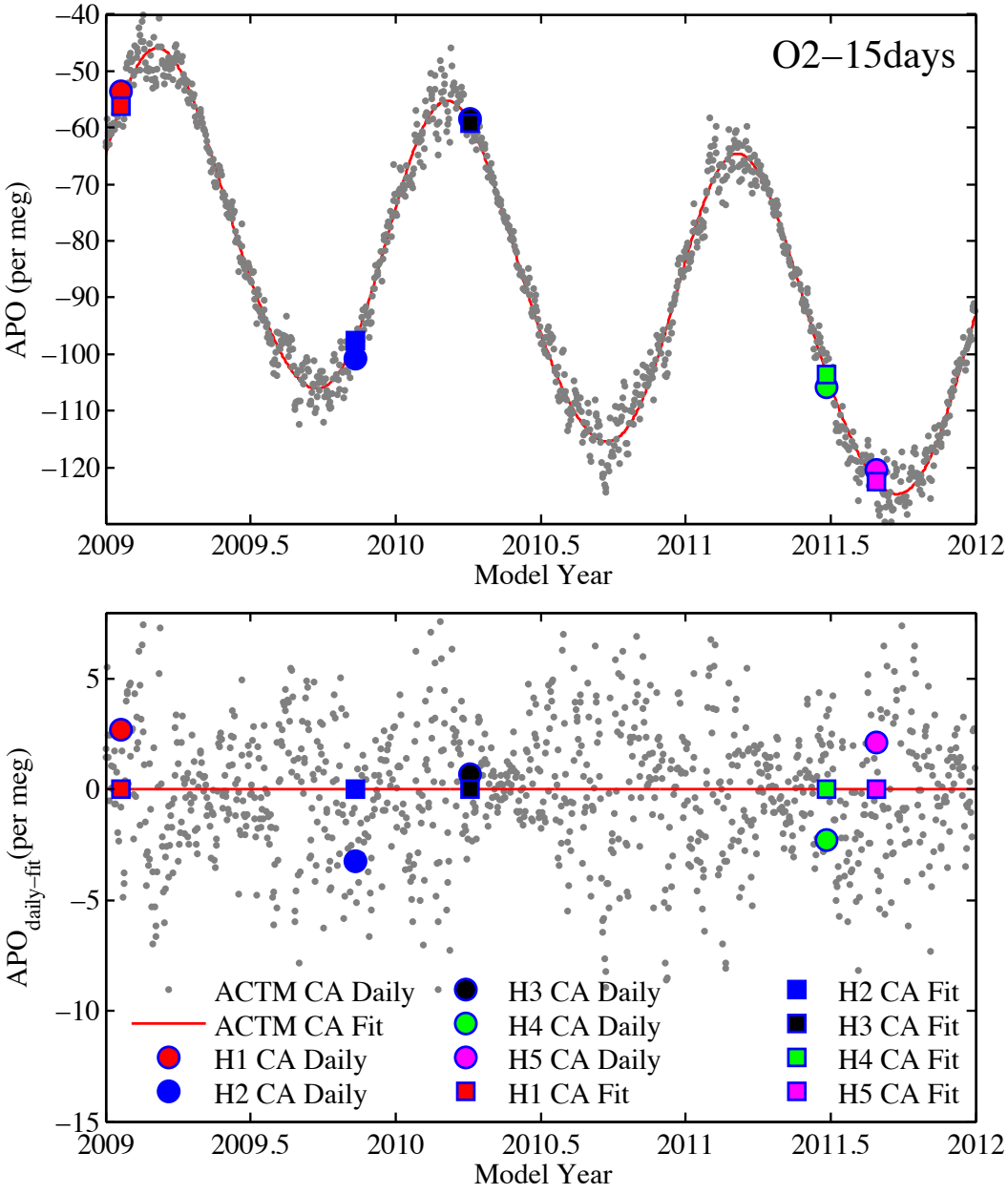

Figure 3.14. Top Panel: Expected daily APO values for the Southern Ocean Curtain Average, using ACTM and the closest matching dissolved climatology O_2 flux delay of 2 weeks (see Section 3.8). Gray points: CA for daily values Jan '09 - Dec '11. Red curve: 2-harmonic fit to these data. Large circles: CA value from daily ACTM output on exact dates of the five HIPPO SO flights. Large squares: CA for the same date, but for the *fit* to the daily ACTM output. Bottom Panel: Same, but with fit to ACTM daily output removed, showing the daily residuals. The difference between the synoptic value (circle) and fit value (square) for a given day gives an estimate of the synoptic bias due to local meteorological conditions on the days of the five SO flights. Units: [per meg].

Applying modeled synoptic offsets to the dataset assumes that the GK01 fluxes are roughly correct in amplitude, but my initial observations of the curtain average amplitude suggest that the amplitude of the fluxes is likely to be 1.3-1.4x too great. Because the magnitude of the scatter around a mean is likely to be proportional to the magnitude of the signal that creates the scatter in this context, I choose, conservatively, to scale down the synoptic bias and uncertainty estimates by the reciprocal of 1.3. Adjusting accordingly, I estimate offsets of -2.1, +2.5, -0.5, +1.8 and -1.6 per meg respectively. I assess the uncertainty of each of these adjustments as the scatter in the ±7 days around each HIPPO point (again divided by 1.3), giving uncertainties of: 2.1, 1.3, 1.1, 1.7, and 2.2 per meg. These adjustments lead to only a small change in the amplitude of the 1- and 2-harmonic fits, of approximately 1 per meg, or 2.5% of the final amplitude.

I evaluate these synoptic bias adjustments and their uncertainty using GK01 O_2 fluxes that have been delayed by 2 weeks. The decision to use delayed fluxes stems from observations in literature [*Manizza et al.*, 2012] and in this study that the phase of the O_2 fluxes in the Garcia and Keeling study is shifted 2-3 weeks early. I explain exactly how I chose a 2-week phase delay in Section 3.8, but here it merits mentioning that this choice could, under the right conditions, have an observable effect on the calculated synoptic offset. This could conceivably be observable if a pressure front were to bring air from a grid box in which a "spring bloom" had just started to occur in one phasing scenario, but which hadn't yet begun in another phasing scenario. However, my results indicate that for this study, such a choice is ultimately a very small contributor to the calculated synoptic offset. I show three phasing scenarios in **Figure 3.14** and Appendix 1 Figures A1-08 and A1-09, in which the O_2 fluxes are delayed 15 days, 22 days and 0 days (i.e. the

original phasing), respectively. In each case, O_2 mixture values are run through the ACTM transport model independently by Dr. Prabir Patra, using phase-shifted flux boundary conditions. APO is, again, calculated independently, as is the 2-harmonic fit to the APO daily output. In all three cases, the sign of each offset is consistent. The relative magnitude between HIPPO campaigns is also quite consistent. These two agreements lead me to believe that the calculated bias in the 2-week delayed scenario is largely representative of synoptic meteorological conditions, not the specific timing of the underlying fluxes.

Dr. Britton Stephens has evaluated the spatial sampling bias in our measurements by comparing two forms of ACTM model output—one produced at full model resolution along 180°W, and another produced along the HIPPO flight tracks at local noon on the day of each HIPPO Southern Ocean flight. (As before, this analysis uses 2-week delayed O_2 fields.) This allows me to estimate and correct for the sampling bias that results from incomplete characterization of the APO field along the meridian. He interpolates and extrapolates the flight track output in 3 different ways: 1) as described above with interpolation and extrapolation followed by masking for regions greater than 4° latitude or 25 mb from an actual observations, 2) no interpolation or extrapolation, only using grid boxes with actual observations, and 3) the full interpolation and extrapolation with no masking. He then compares these to the curtain average for the full model slice along 180°W. The results for the first scenario are shown in **Figure 3.15**. The difference values for the limited extrapolation case (described above) is my best estimate of the sampling bias that results from incomplete characterization of the APO field along the meridian. Scaled down by a factor of 1.3 to reflect the too-large ACTM-GK01 amplitude, this gives

spatial biases of -0.8, -0.2, +1, +4.8, and +2.6 per meg, for HIPPO1 to HIPPO5, respectively. This approach accounts for interpolation bias due to missing sections of the full slice in HIPPO3 and HIPPO4, and bias from not sampling exactly at 180°W. As with the previous adjustment, this adjustment is quite small relative to the amplitude of the seasonal cycle.

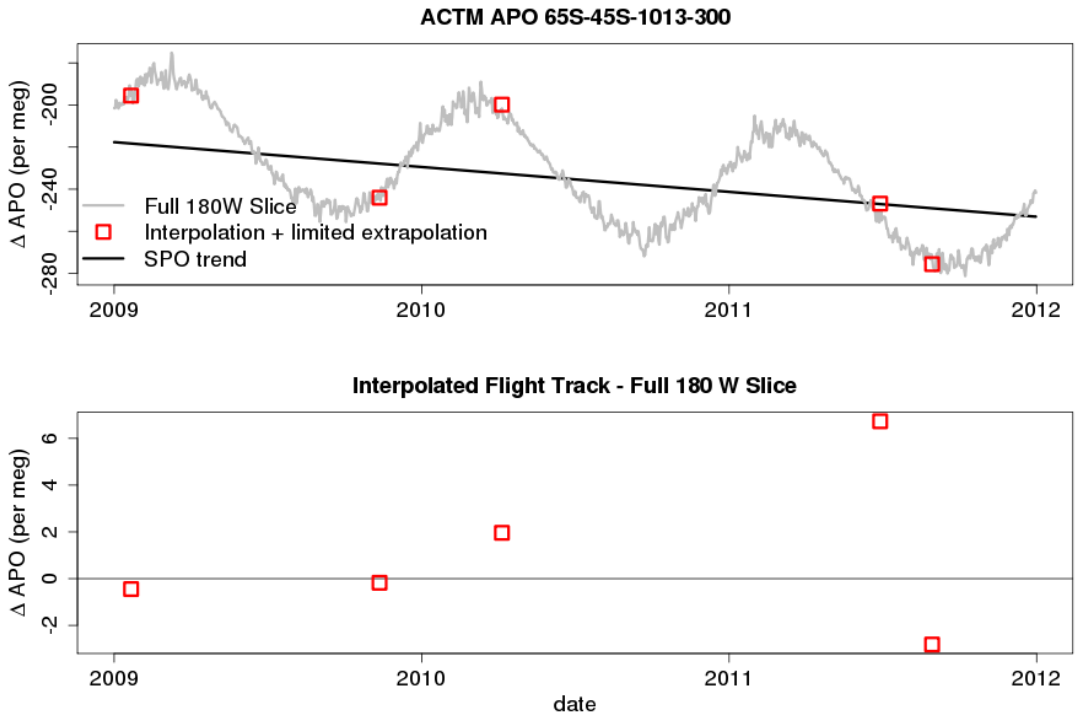

Figure 3.15. TOP PANEL: Comparison of ACTM+GK01 curtain averages computed at full resolution along 180°W (gray curve) with curtain averages computed from simulated flight-track output (red squares), also from ACTM+GK01. Flight track curtain averages are computed using exactly the same method applied to the HIPPO APO data, with interpolation between flight tracks, and a limited extrapolation for sections of flights missing data. BOTTOM PANEL: The same figure, but with the gray line removed to show residuals. I use these values to establish the spatial bias due to 1) incomplete characterization of the meridional slice and 2) the fact that the flights were not flown exactly along 180°W. Units: [per meg]. Figure: B. Stephens.

The ACTM-based adjustments account, in principle, for variability in the curtain averages driven by atmospheric transport variability on both synoptic and interannual time frames in the presence of air-sea fluxes that repeat exactly each year. These adjustments do not account, however, for possible interannual variations in the APO

fields due to interannual variations in air-sea fluxes. Though I do not have any way to directly apply adjustments for this additional effect because I do not have models that represent interannual variations in flux fields between 2009 and 2011, I am able to assess the potential magnitude of this effect by assessing interannual variations in the amplitude of the seasonal cycles at Southern Ocean sites, which I explain below.

3.6 Data-based assessments of uncertainty, bias and trends

Station measurements provide a spatially limited, but direct measure of interannual variability that model output may not reproduce accurately. I consult data from Southern Ocean sampling stations to assess how the seasonal cycle differs in magnitude from year to year. The variation in these values reflects a combination of synoptic and interannual atmospheric transport and air-sea flux differences, so they provide an upper bound for the interannual error due to air-sea fluxes alone. I also consult station data to determine whether the eruption of the Puyehue-Cordón Caulle volcano complex contributed to local CO_2 concentrations at a measureable scale. Finally, I briefly discuss analysis of the uncertainty associated with the scatter around each flask value from correcting a scaled down version of the Ar/N_2 scatter out of the O_2/N_2 values (a different source of Ar/N_2-based uncertainty than that mentioned in Section 3.5.4). I investigate the impact such an uncertainty has when compounded over a set of ~30 flasks used in each Southern Ocean curtain average value.

3.6.1 Assessing atmospheric interannual variability

I calculate the interannual variability of atmospheric conditions at Cape Grim, Palmer Station and South Pole. I choose to do this to effectively cancel out any local effects that might increase APO around one single station, while reducing it at another nearby station for a given year. I start by removing a linear trend from the monthly mean records at Cape Grim, Palmer Station, and South Pole between 1997 and 2012 and then fit 2 harmonics to the individual years' residuals (seasonal+interannual components). I compute the average peak-to-peak amplitude of each year, as seen in **Figure 3.16**, and report the standard deviation of the mean of the three amplitudes over the 16 years. The standard deviation of the 16 cross-site means yields a 1σ value of 7.1 per meg. Expressed as a percent of the amplitude of the average peak-to-peak amplitude (67 per meg), this is 10.5%. For the curtain average, the seasonal cycle is 37.8 per meg, so my uncertainty estimate for the error for the curtain average amplitude is 3.9 per meg 1σ. However, amplitude uncertainty is larger than the uncertainty associated with each individual point, with an individual point's uncertainty at most equal to the amplitude uncertainty multiplied by the square route of two based on Monte Carlo estimates shown in Section 3.6.5. Accordingly, I take this value, 2.9 per meg, as a conservative uncertainty estimate for each of the five HIPPO missions.

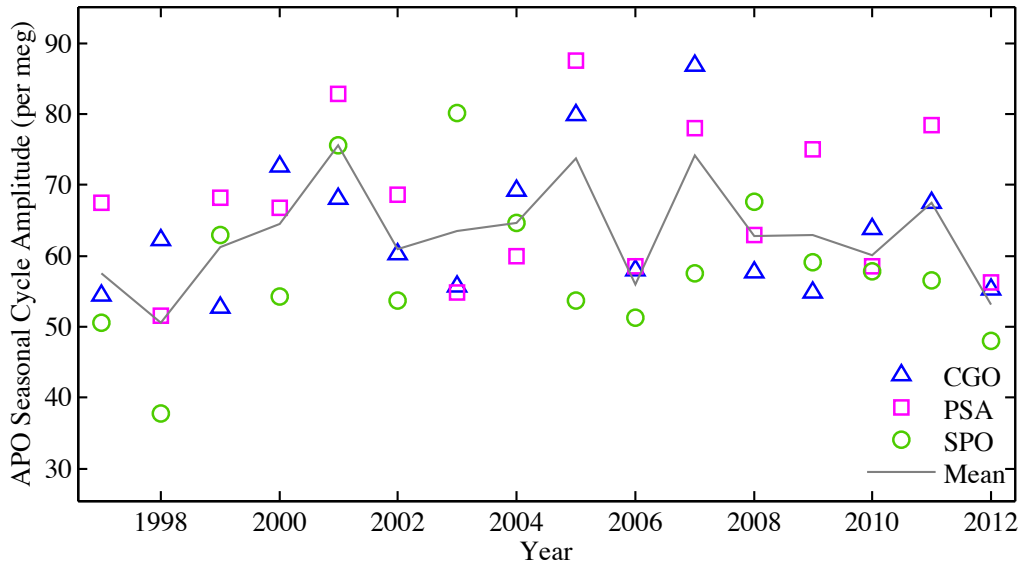

Figure 3.16. Yearly 2-harmonic fitted seasonal peak-to-peak amplitudes at Cape Grim, Palmer and South Pole, with the yearly mean of the three stations shown in gray. Units: [per meg].

3.6.2 Assessing volcanic contribution to HIPPO4 and HIPPO5

The June 3, 2011 eruption of the Puyehue-Cordón Caulle volcano complex (hereafter PCC), and the ash cloud that spread from it, posed a potential threat to the engines of the GV aircraft. As a result, the HIPPO4 Southern Ocean flight turned around at 57°S, instead of 67°S to avoid the ash plume. An estimate of the spatial bias that resulted from this necessity is found in the previous section, but an estimate of the contribution of volcanic CO_2 to the Southern Ocean curtain average of APO requires separate attention. I assess this in two ways—by considering the contribution of large eruptions to atmospheric CO_2 in recent years, and by consulting recent Southern Ocean time series records.

I am unable to find direct studies of the CO_2 loading to the atmosphere during the PCC eruption, so I determine a first-order estimate of this value from studies of similar

eruptions. Gerlach [2011] examined the eruptions of Mount St. Helens in 1980 and Mount Pinatubo in 1991, concluding that the two eruptions were responsible for producing 0.01 Pg and 0.05 Pg, of CO_2, respectively. The Mount Saint Helens eruption has a Volcanic Explosivity Index [*Newhall and Self*, 1982] of 5 (>1 km^3 tephra emitted), while Pinatubo has a VEI of 6 (>10 km^3 tephra emitted); Puyehue-Cordón Caulle, however, has a VEI of 4 (>0.1 km^3 tephra emitted) [*NOAA NGDC*, 2014]. If I assume a nominally consistent emissions ratio for these eruptions, the PCC likely emitted on order 1 teragram, or $2.3*10^{10}$ moles, of CO_2 to the atmosphere. Given a global atmospheric volume of $5.1*10^{21}$ g ($1.76*10^{20}$ mol), this corresponds to a perturbation of 0.13 parts per billion. In a worst-case scenario, if I assume perfect mixture in only the zone of interest, extending down from PCC (40°S) to the bottom of our box (65°S)—roughly 13% of the surface area of the earth—over the weeks between the eruption and HIPPO4, the volcano would have contributed approximately 1 ppb of CO_2 to the local atmosphere, well below the resolving capability of our instruments.

Station records from Cape Grim, Palmer Station and South Pole seen in **Figure 3.17** also show no increase in CO_2, or the seasonality of CO_2 after June 3, 2011. Accordingly, I do not correct the data for volcanic influence.

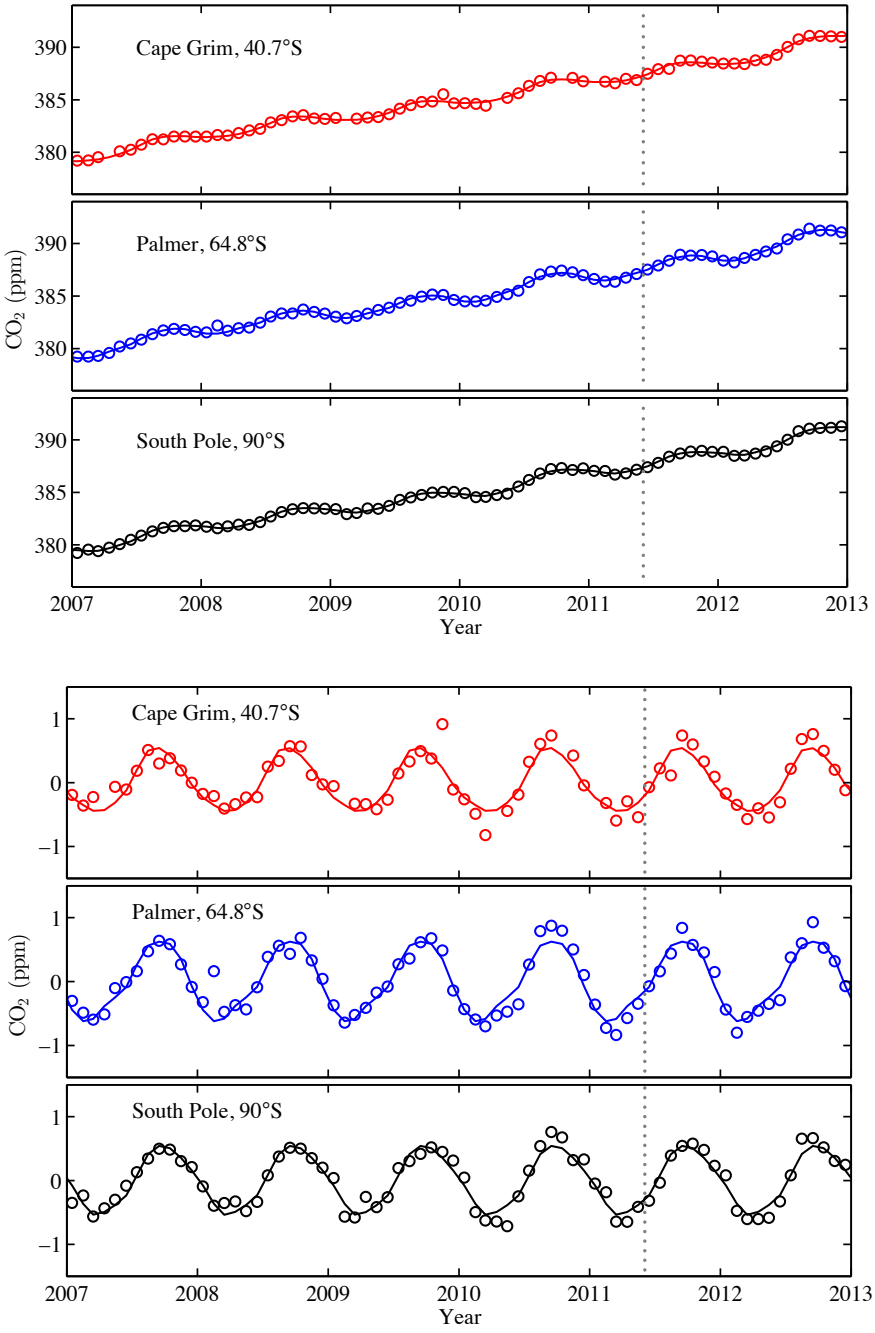

Figure 3.17. Top 3 Panels: CO_2 concentrations for Cape Grim, Palmer and South Pole between 2007 and 2013. Monthly average concentrations are shown as open-faced circles, while a 4-harmonic stiff spline fit is shown as a solid line. The timing of the Puyehue-Cordón Caulle eruption—June 3, 2011—is shown by vertical gray dots. Bottom 3 Panels: The same plots, with the stiff spline removed to show seasonal residuals, which do not show any noticeable change in the seasonal CO_2 anomaly after the PCC eruption. Units: [ppm].

3.6.3 Estimating uncertainty introduced by removing Ar/N_2 scatter from O_2/N_2 values

Each flask in each Southern Ocean curtain average has an individual uncertainty due to the imprecision of the O_2/N_2 measurement on the ISOPRIME mass spectrometer and the uncertainty introduced to the flask value by the removal of a scaled-down version of the Ar/N_2 scatter (see Chapter 2). Based on replicate agreement for flasks collected at La Jolla, the precision for an individual flask on the ISOPRIME is conservatively ±3 per meg for O_2/N_2 and the scatter around the mean in the Ar/N_2 data is ±22 per meg. Scaling the Ar/N_2 precision down by 1/3.77 and adding this in quadrature to the O_2/N_2 value yields an overall imprecision per flask of ±6.5 per meg. In order to estimate the uncertainty of the curtain average, I select the values for HIPPO1 and add a randomly-determined error value, whose 1 sigma is 6.5 per meg, to each flask value, and recomputed the curtain average value 1000 times. The resulting standard deviation of this curtain average value, seen below in **Figure 3.18**, is 1.2 per meg, so this is the value I carry through to the uncertainty estimates in **Table 3.5**.

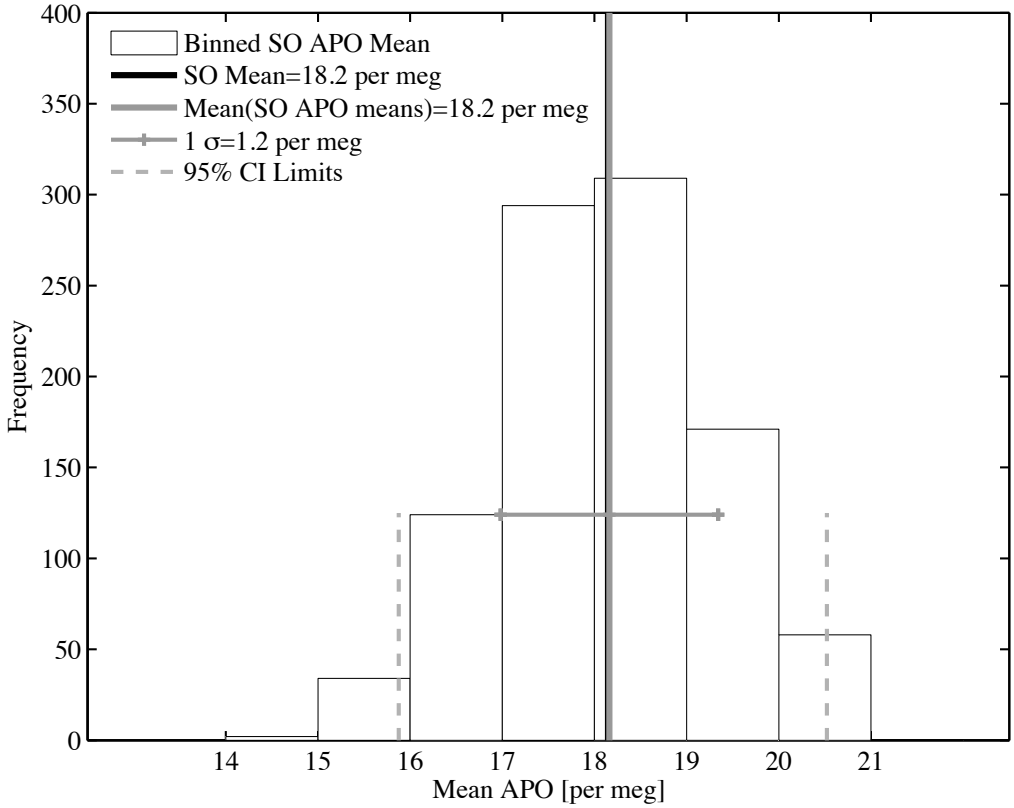

Figure 3.18. 1-per meg binned mean of detrended HIPPO1 curtain averages, as simulated 1000 times by adding an uncertainty to each flask value whose 1 sigma is the quadrature sum mentioned in the text above, of 6.5 per meg. The standard deviation of these 1000 simulations is 1.2 per meg.

3.6.4 The adjusted curtain average value: bias and error totals

It is important to reiterate that the sum of the biases associated with the curtain average adjustment is relatively small, as is the error bar assessed for each curtain average. **Figure 3.19** below repeats **Figure 3.8**, showing the unadjusted curtain average values as light gray circles with colored outlines, but also showing in bolder colored circles with a black dashed 2-harmonic fit line the adjusted values of the curtain average results. This fit has a seasonal cycle of 37.8 per meg, only 2.8 per meg less than the original fit to the unadjusted values suggested. No adjustment exceeds 3.8 per meg in total, and no error bar exceeds 6.6 per meg—roughly 18% of the observed amplitude—

with the average of these being about 13%. The total bias and error budgets for the curtain averages can be found below in **Table 3.4** and **Table 3.5**, respectively.

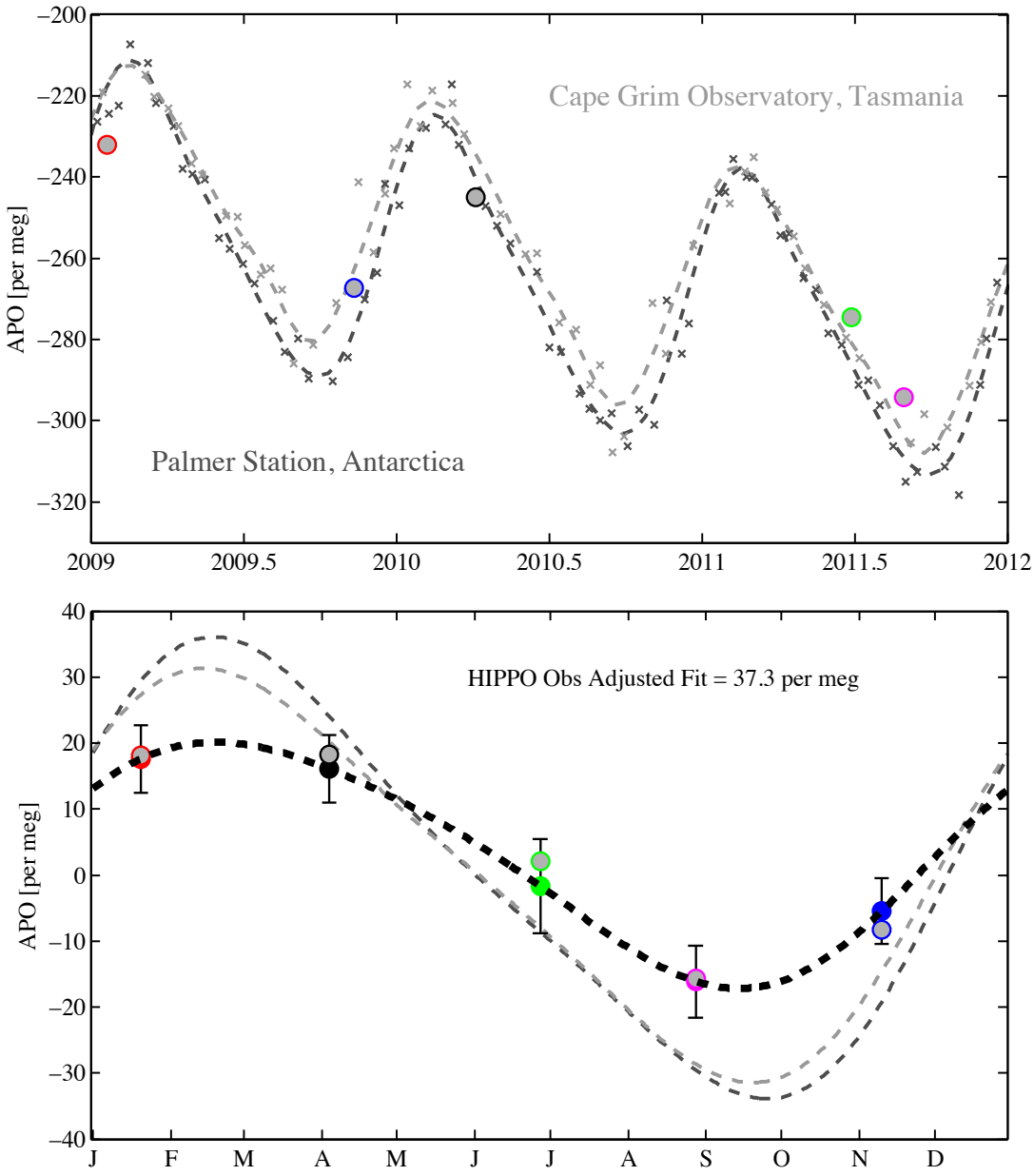

Figure 3.19. A repeat of Figure 3.8, with the top panel the same, but with both the original detrended values and the adjusted detrended values in the lower plot. Unadjusted values are still shown with colored outlines, while the adjusted values are filled colored circles (in order, HIPPO1 (red), 3 (black), 4 (green), 5 (magenta), and 2 (blue). The 2-harmonic fit to the adjusted values has an amplitude of 37.8 per meg, only 2.8 per meg lower than the original amplitude.

Table 3.4. Southern Ocean curtain average values for the five HIPPO missions. "Raw val" reflects the combined AO2-MEDUSA APO product as produced by the flask analysis/adjustment code, the AO2 code, and with anchoring of AO2 to MEDUSA (Chapter 2). ADJUSTMENTS: "Dtr Val" reflects the value when the SPO linear trend is removed from the raw value (Section 3.2.7). "Met adj." reflects the adjustment needed to remove synoptic+interannual meteorological effects to produce a "mean" climatology (Section 3.5.5); "Spat" reflects the adjustment needed to account for a combination of flight track interpolation and deviation from 180°W (Section 3.5.5); "Ar Adj" reflects the adjustment to restore the portion of the seasonal signal removed during the Ar correction process (Section 3.5.4). "SO CA Adj Val" reflects the final Southern Ocean curtain average value. "SO CA 1H Norm Adj Val" is the "SO CA Adj Val" minus 2.7 per meg which, when a 1-Harmonic fit is applied to the values, reflects a zero mean offset in the Y direction—i.e. this column represents the normalized value of the adjusted curtain average value. "SO CA 2H Norm Adj Val" is the "SO CA Adj Val" minus 2.8 per meg, the equivalent Y adjustment value for a 2-harmonic fit. These last two columns are probably the most useful number for researchers interested in looking at the data. Units: [per meg].

HIPPO Miss.	South. Ocean Flight Date	Year Day	Raw Val	Dtr. Val	Met Adj	Spat Adj	Ar Adj	SO CA Adj Val	SO CA 1H Norm Adj Val	SO CA 2H Norm Adj Val
1	Jan 20, 2009	20	-232.1	18.1	-2.1	0.3	1.2	**17.4**	**14.7**	**14.8**
2	Nov 11, 2009	315	-267.3	-8.3	2.5	0.1	-0.7	**-6.4**	**-9.2**	**-9.1**
3	Apr 5, 2010	95	-245.0	18.3	-0.5	-1.5	0.4	**16.6**	**13.9**	**14.0**
4	Jun 28, 2011	179	-274.5	2.1	1.8	-5.2	-0.1	**-1.4**	**-4.1**	**-4.0**
5	Aug 29, 2011	241	-294.1	-15.7	-1.6	2.2	-0.8	**-15.9**	**-18.7**	**-18.6**

Table 3.5. Estimates of error for the five HIPPO curtain average values. "Met Err" reflects the individual errors assessed for the meteorological synoptic and inter-annual adjustment, which reflects the standard deviation of the difference between daily and fit values for the 7 days before and after the HIPPO points; "Flux IAV Err" represents a scaled down estimate of surface station interannual flux variability, reflecting the fact that I do not attempt to adjust the HIPPO values for interannual flux variability (Section 3.6.1); "Spat Err" reflects the error associated with the spatial adjustment—here 100% the adjustment value carried over as an uncertainty estimate; "Inst Drift Err" reflects the long-term stability of calibration gases used to make the O_2/N_2 and CO_2 measurements in the lab at Scripps; "Ar Err" is the error associated with the Ar/N_2 seasonal adjustment. "Comp Err" is the error associated with the per flask precision in O_2/N_2 and Ar/N_2 recognizing that each CA averages roughly 30 flasks (Section 3.6.3). 1 sigma error reflects the combination in quadrature of the error estimates. Units: [per meg].

HIPPO Mission Number	Southern Ocean Flight Date	Year Day	Met Err	Flux IAV Err	Spat Err	Instr Drift Err	Comp Err	Ar Err	1σ
HIPPO1	Jan 20, 2009	20	2.1	2.9	0.3	2	1.2	1.2	4.4
HIPPO2	Nov 11, 2009	315	1.3	2.9	0.1	2	1.2	0.7	4.0
HIPPO3	Apr 5, 2010	95	1.1	2.9	1.5	2	1.2	0.4	4.2
HIPPO4	Jun 28, 2011	179	1.7	2.9	5.2	2	1.2	0.1	6.6
HIPPO5	Aug 29, 2011	241	2.3	2.9	2.2	2	1.2	0.8	4.9

3.6.5 Uncertainty in harmonic fit parameters

In addition to comparing models to the individual HIPPO curtain average points, I am interested in comparing model estimates of key parameters such as peak-to-peak amplitude and timing of peak and trough of APO cycles. These metrics can be assessed separately for either 1-harmonic or 2-harmonic fits to the 5 HIPPO data points, although for a 1 harmonic fit the peak and trough are not independent (i.e. are fixed half a year apart). There are thus six potentially useful metrics: two amplitude metrics (one each for 1- and 2-harmonic fits) and three phasing metrics (one for 1-harmonic fits and two for 2-harmonic fits), and the length of time between the peak and trough (for 2-harmonic fits only), which is a measure of the asymmetry of the cycle.

To assess uncertainty in these harmonic fit parameters, I use a Monte-Carlo approach in which I select a random point for each HIPPO campaign, using a normal distribution with standard deviations based on error bars assessed in the previous section. I then repeat this selection process 1000 times, generating 1000 sets of the 5 HIPPO points. These 1000 replicates are then used to for either 1-harmonic or 2-harmonic fits allowing confidence limits to be placed on the fitted metrics.

Results from 1-harmonic fits are shown in **Figure 3.20**, with 95% confidence interval limits assessed for each day individually shown as thin, dark gray dashed lines. Distributions of the fitted parameters for the 1 harmonic fits are shown in **Figure 3.21** and **Figure 3.22**.

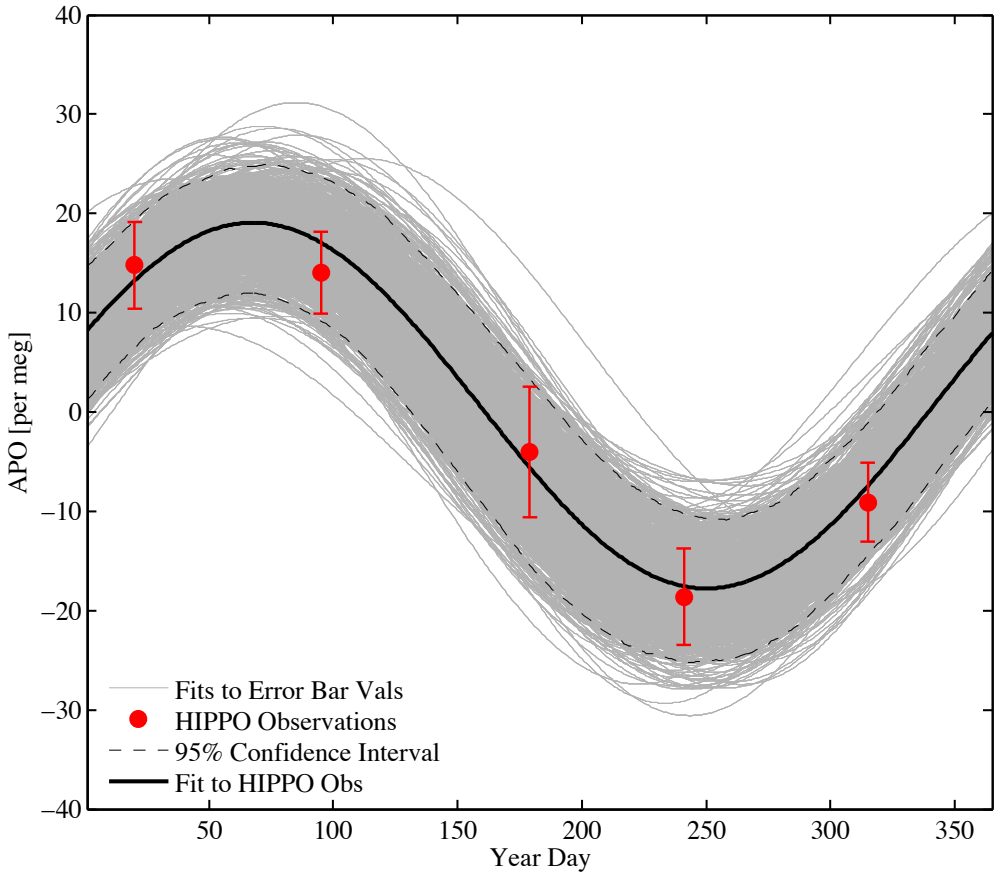

Figure 3.20. 1-Harmonic fit to the five HIPPO data points (red dots with error bars), and 1000 Monte Carlo simulations fitted to randomly-chosen, normally-distributed points whose 1 sigma deviation from the fitted line is equal to the 1 sigma of the HIPPO error bars.

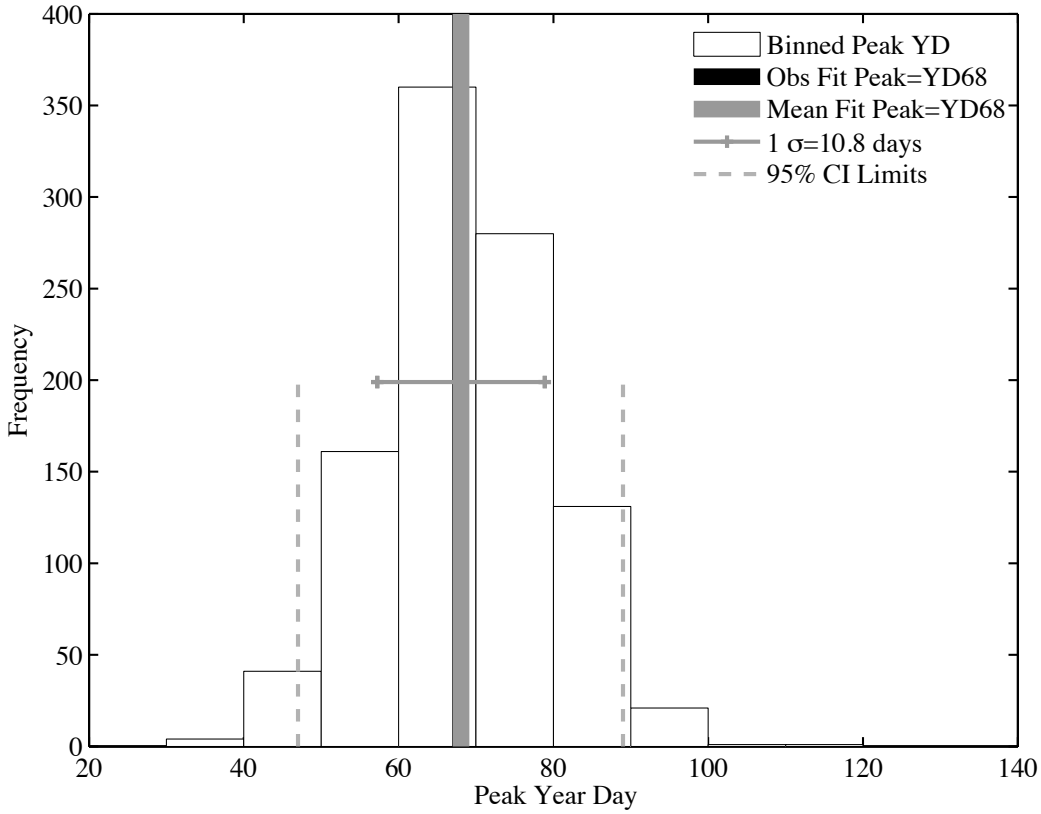

Figure 3.21. Peak timing for the 1000 Monte Carlo simulations. The mean of the 1-Harmonic fit to the data is shown in black, obscured by the gray bar above it, which shows the mean of all 1000 1-Harmonic fits. The standard deviation of the simulation peaks is shown as an error bar on the mean, while 95% confidence limits are shown as dashed vertical lines.

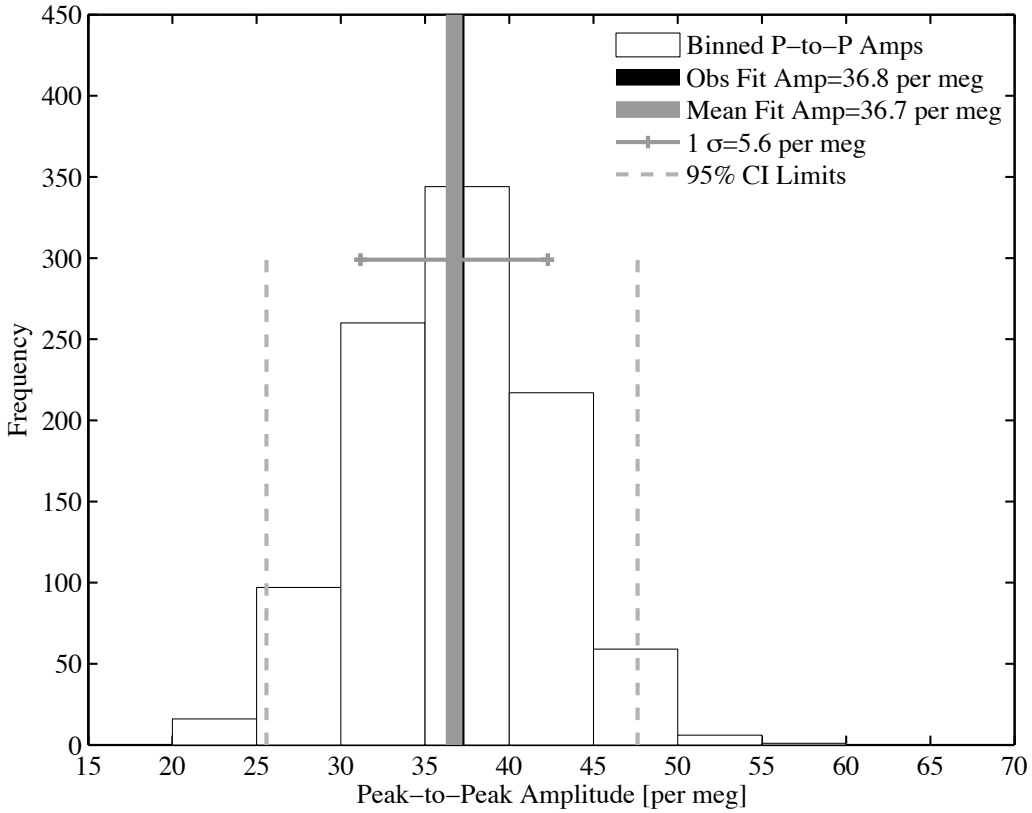

Figure 3.22. Amplitude for the 1000 Monte Carlo simulations. The mean of the 1-Harmonic fit to the data is shown in black, obscured by the gray bar above it, which shows the mean of all 1000 1-Harmonic fits. The standard deviation of the simulation amplitudes is shown as an error bar on the mean, while 95% confidence limits are shown as dashed vertical lines.

These results give 1-harmonic error estimates of ±5.6 per meg for the peak-to-peak amplitude and ±10.8 days for the timing of the peak. The mean of the peak day distribution is exactly the same as the fit to the HIPPO curtain average values (Year Day 68), while the mean of the fitted amplitudes is almost exactly the same, at 36.7 instead of 36.4 per meg, well within the 1 sigma, suggesting that a normal distribution is appropriate under these circumstances.

Monte Carlo runs for the 2 harmonic fit scenario are shown in **Figure 3.23**. While the estimate of the 2-harmonic fit amplitude appears relatively similar to that in the 1-H

scenario, the timing of the peak and trough in the 2-harmonic scenario are considerably less well constrained.

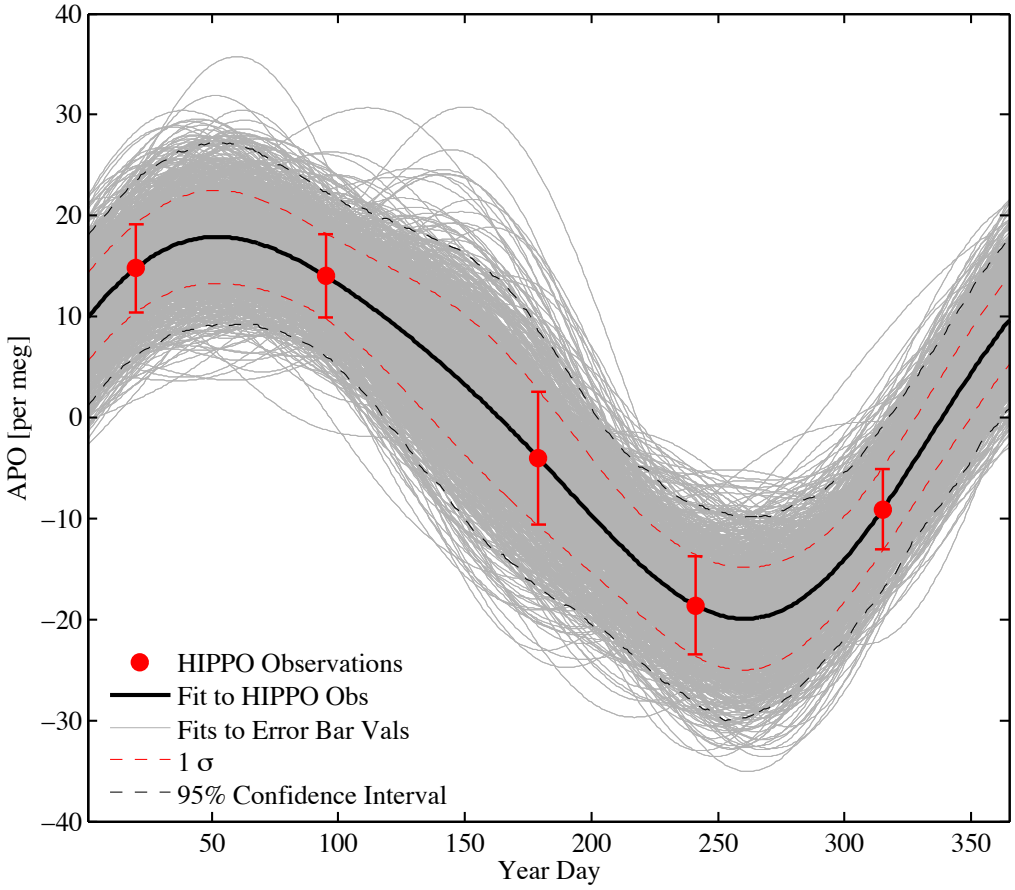

Figure 3.23. Monte Carlo method for 2-harmonic fit scenario, showing 1000 simulations of the HIPPO APO curtain average using randomly-chosen, normally-distributed values whose standard deviation in each of the five cases is the error bar on that individual point. Red points with error bars show the adjusted curtain average values, while the thick black line shows the 2-harmonic fit to these values. Thin gray lines show the Monte Carlo simulations; a thin dashed red dashed line shows the standard deviation of the 1000 fits assessed for each day of the year individually, while a thin dashed gray line shows the 95% confidence limits for the same.

The large uncertainty in the timing of peak and trough results from some of the fits with two harmonics being relatively unconstrained between the HIPPO points, especially near the peak. Error bars on HIPPO1 and HIPPO3 points allow the second harmonic component to vary the timing of the peak quite wildly from between these points to between the HIPPO3 and HIPPO4 points, a feature not supported by station

records, but nonetheless a feature that cannot be ruled out on the basis of HIPPO observations alone. This uncertainty in peak and trough timing leads to an even larger uncertainty in the difference (delta, in days) between the trough and peak timing. This value, of ±36 days, is shown in **Figure 3.24**, below.

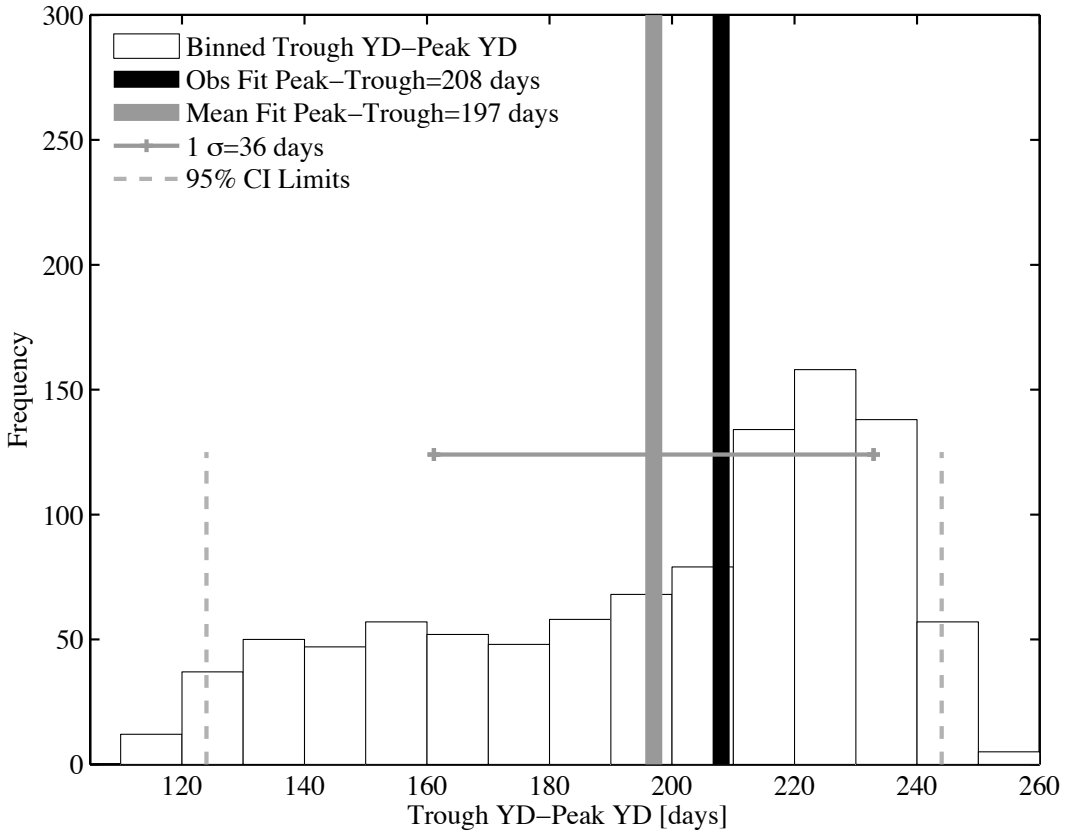

Figure 3.24. An estimate of seasonal asymmetry uncertainty, as calculated from 1000 Monte Carlo simulations, using a 2-harmonic fit approach. Asymmetry is shown as the timing of the seasonal trough year day minus the timing of the seasonal peak year day, with a value of ~182 being perfectly symmetrical with respect to time (i.e. 365/2).

I summarize the remaining 2-harmonic uncertainty estimates among those already mentioned in **Table 3.6**, below to save space. As appropriate, I carry the 1- and 2-harmonic uncertainty estimates over to 1- and 2-harmonic comparisons of data and models.

Table 3.6. Summary of 1- and 2-harmonic metrics discussed above, showing (in order) the 1- and 2-harmonic amplitude, 1- and 2-harmonic peak year day, the 1- and 2-harmonic trough year day, and the 2-harmonic trough-peak difference in days (essentially temporal asymmetry). The first row shows each metric as computed using just the 1- or 2-harmonic fit to the adjusted HIPPO MEDUSA curtain average observations. The second row shows the mean value of the metric as computed from 1000 simulations using the technique described above. The third row shows the uncertainty associated with each of these metrics.

	1-H Amp (per meg)	2-H Amp (per meg)	1-H Peak (YD)	2-H Peak (YD)	1-H Trough (YD)	2-H Trough (YD)	2-H Trough-Peak (days)
Fit to Obs. Value	36.4	37.8	68	52	250	260	208
Mean Fit	36.7	39.8	68	60	251	257	197
1 σ	±5.6	±5.9	±10.8	±27	±10.8	±20	±36

3.7 Model-observation comparisons

The fundamental questions of this study are many: 1) How well do we understand the biogeochemical controls on carbon in the Southern Ocean? 2) How does the amplitude of the seasonal cycle at the surface compare to the amplitude of the seasonal cycle of the entire atmospheric column? 3) What are the seasonal fluxes of APO over the Southern Ocean? 4) How well do ocean models reproduce these atmospheric seasonal cycles? Two *practical* questions that arise from these are: 1) What is the ultimate interpretive power of the HIPPO dataset and the curtain average approach? And 2) How can these be presented in a fashion that is useful for the modeling community? In the following section I present comparisons of model and observed curtain averages. I highlight differences between the various ocean model/dissolved climatology fluxes. Preliminary statements about the differences between atmospheric model runs with

common dissolved climatology fluxes are also possible, although I include only a brief investigation of atmospheric transport model differences here.

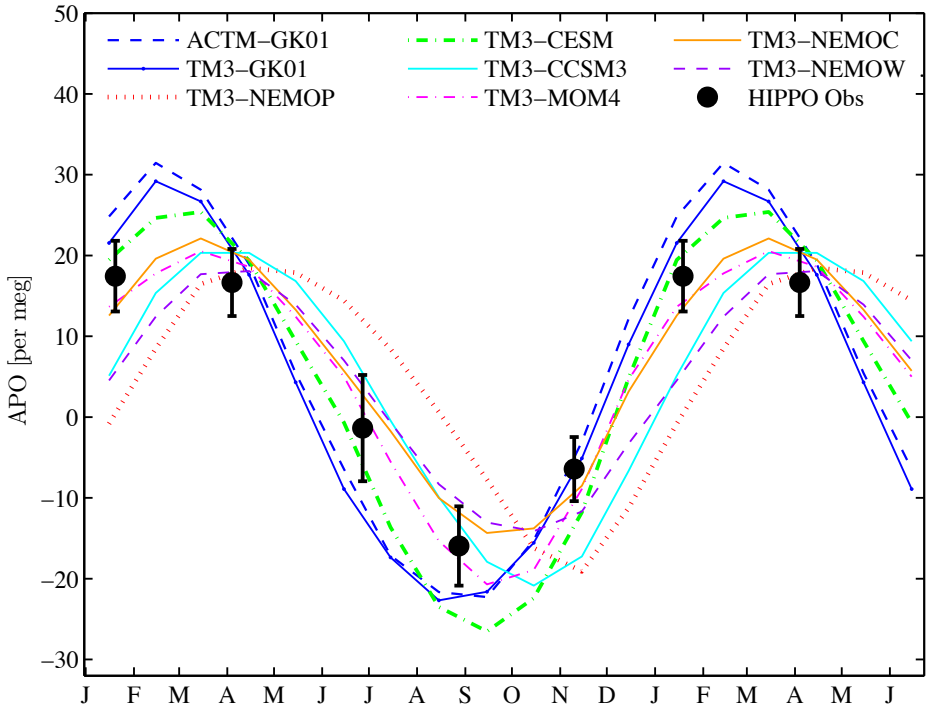

Figure 3.25. 18-month comparison of model output and data (last two points are repeats of first three). Model output is shown as monthly means for consistency between runs, while data are shown with error bars from Table 3.6. Models and data are detrended with their respective South Pole records. Units: [per meg].

Figure 3.25 shows a comparison of detrended monthly mean model output curtain averages and HIPPO curtain averages without harmonic fitting. Some models appear to disagree with data strongly in phase, but with no fitting it is difficult to assess whether models and data disagree or agree on amplitude. For those interested in seeing how all model and data runs compare side-by-side for 1- and 2-harmonic cases, these figures can be found in the first appendix as A1-11 and A1-12. **Figure 3.26** and **Figure 3.27** show a slightly different version of the figures in the appendix, with the model results shown separately, each in their own panel, and with all contributing tracers alongside APO.

Models' APO fits are shown as dash-dot colored lines, while HIPPO observations are shown as solid black lines with shaded gray 1-sigma error bands and with thin gray lines to show 95% confidence limits. The HIPPO shaded error band (1-sigma) notionally represents the uncertainty in comparing non-climatological HIPPO data with climatological mean conditions represented by an average of several years of model output. Stated in a different way, the shaded error bar shows the quadrature sum of instrumental drift uncertainty, Ar/N_2 correction uncertainty, and the error associated with synoptic and spatial bias adjustments (all mentioned variously in Section 3.5). The 95% confidence limits show the limits of the middle 95% of the 1- and 2-harmonic Monte Carlo simulations (i.e. with top 25 and bottom 25 simulations removed), so these limits are relative to the mean seasonal cycle of all 1000 simulations, not the fit to the five HIPPO values. Also shown in **Figure 3.26** and **Figure 3.27** are modeled curtain average contributions from N_2 alone (thick dashed colored line), and O_2 alone (thin solid colored line) and CO_2 (thin dashed colored line). The modeled APO results include simulations based on air-sea fluxes from six ocean models, as well as based on GK01 dissolved climatology flux run [*Blaine*, 2005; *Garcia and Keeling*, 2001; *Gloor et al.*, 2001; *Gruber et al.*, 2001; *Takahashi et al.*, 2009]. Most runs use the TM3 transport model except the ACTM model run with dissolved climatology fluxes shown in the upper left panel.

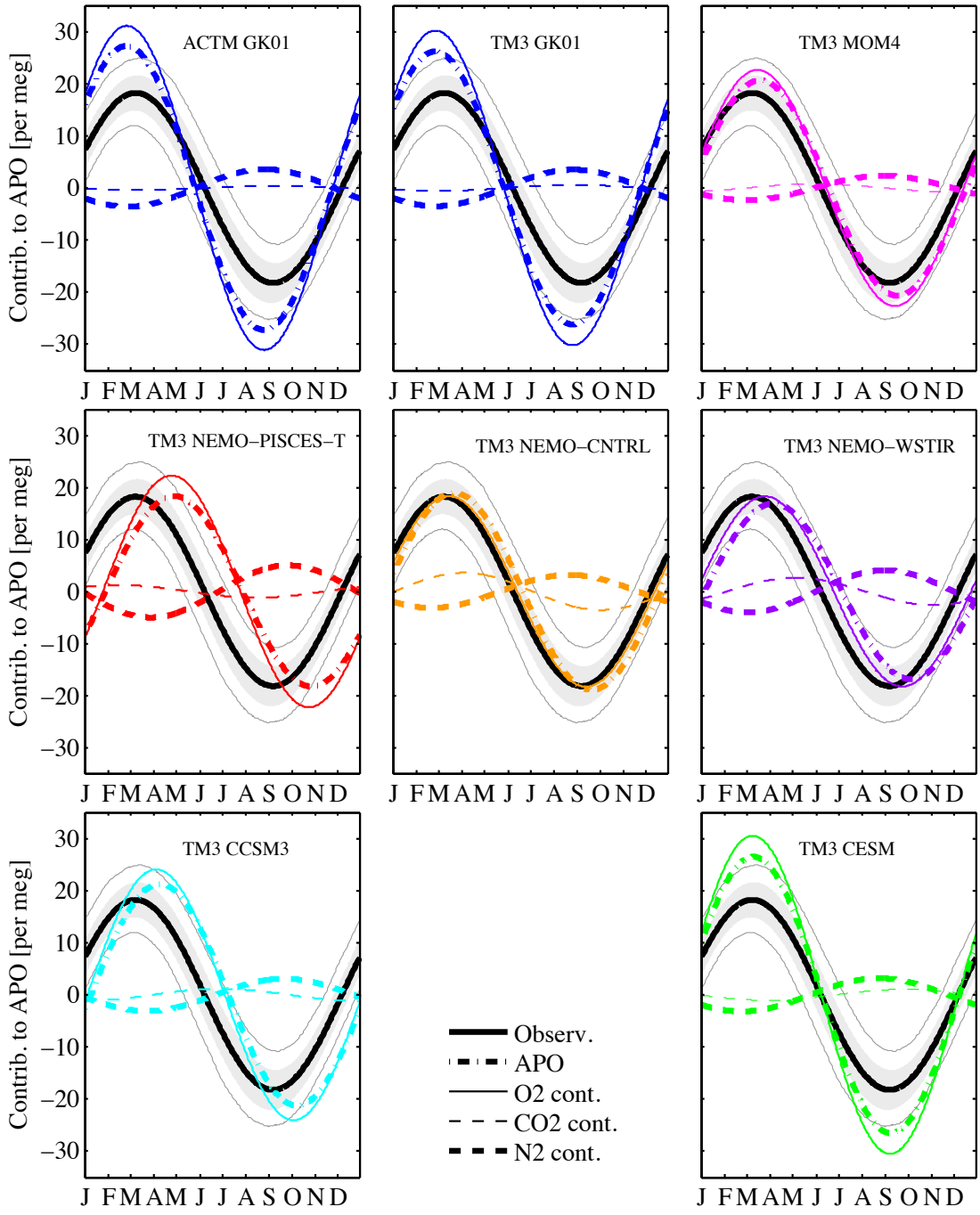

Figure 3.26. Fitted 1-harmonic curtain average seasonal cycles of APO (thick dash-dot colored line), air-sea components of O_2 (thin solid colored line), contribution from N_2 (thick dashed colored line) and CO_2 (thin dashed colored line) for the six ocean model flux runs and two dissolved climatology flux runs compared against a 1-harmonic fit to the HIPPO curtain averages (thick black line). The 1σ distribution around the observed fit is shown with gray shading, while the 95% confidence limits are shown as thin light gray lines. Here, all model tracers are plotted relative to a zero mean value, for intercomparability. Units: [per meg cont.].

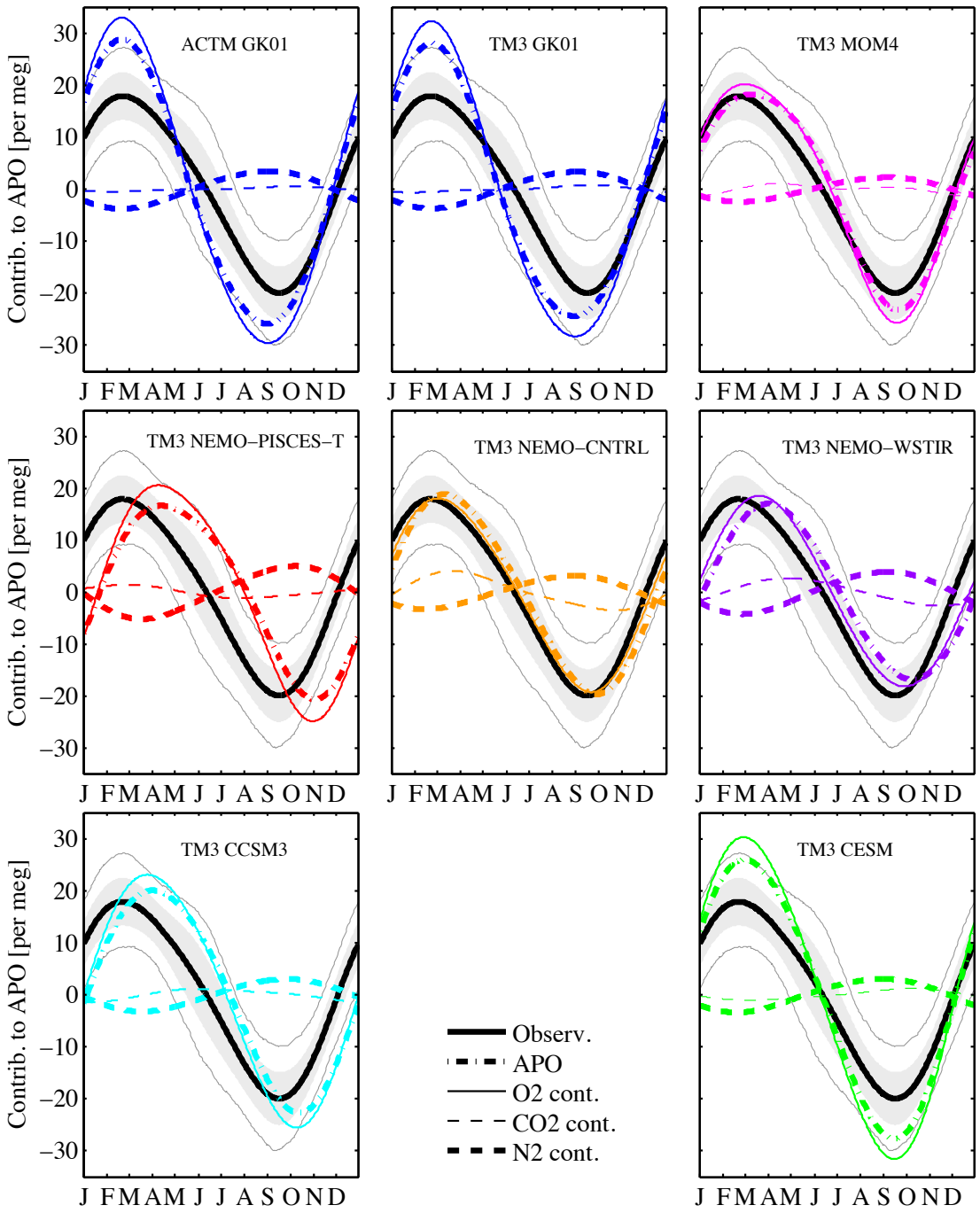

Figure 3.27. Fitted 2-harmonic curtain average seasonal cycles of APO (thick dash-dot colored line), air-sea components of O_2 (thin solid colored line), contribution from N_2 (thick dashed colored line) and CO_2 (thin dashed colored line) for the six ocean model flux runs and two dissolved climatology flux runs compared against a 2-harmonic fit to the HIPPO curtain averages (thick black line). The 1σ distribution around the observed fit is shown with gray shading, while the 95% confidence limits are shown as thin light gray lines. Here, all model tracers are plotted relative to a zero mean value, for intercomparability. Units: [per meg cont.].

As expected from the technique, the two GK01 dissolved climatology runs, seen apart from the other results in **Figure 3.28**, are very similar in shape despite the fact that they disagree strongly on the amplitude of the seasonal cycle at surface stations (see **Figure 3.34**).

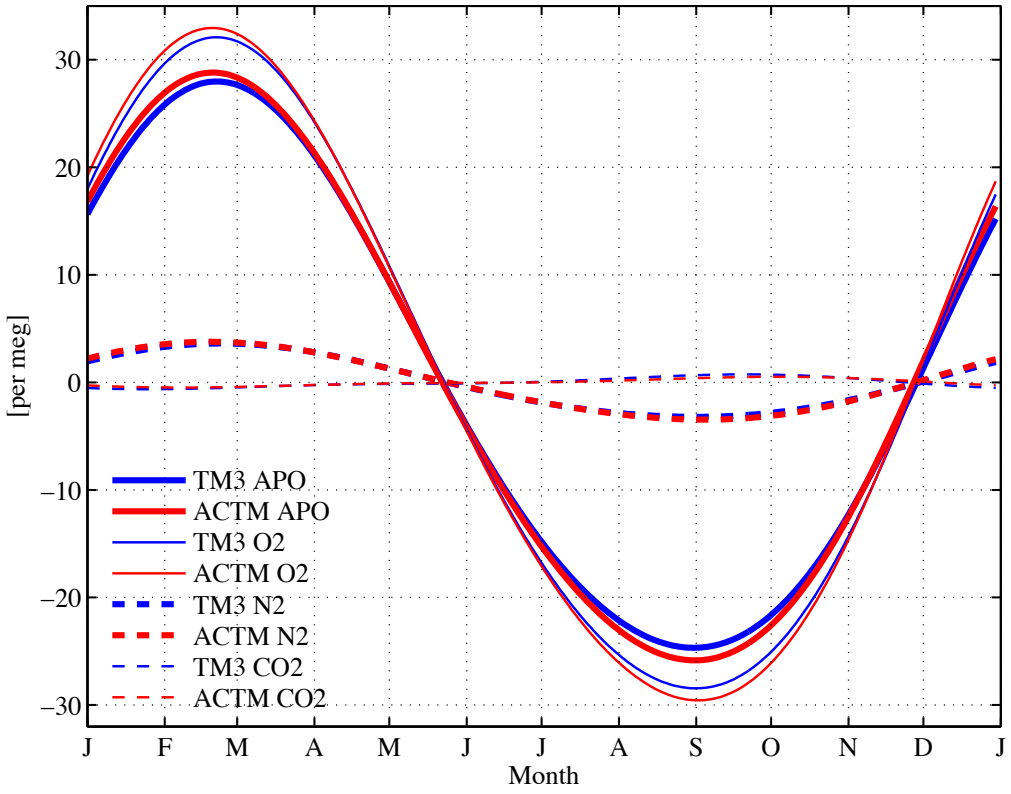

Figure 3.28. Comparison of 2-harmonic fits to individual tracers from ACTM and TM3 GK01 runs, showing APO and its constituents. APO seasonal signals (solid thick lines) are very similar for the two transport models, supporting the premise of the curtain average method. Units: [per meg cont.]

Relative to observations, for 2-harmonic fits, the dissolved climatology runs are close on the timing of the peak, but miss the trough by approximately 3 weeks. For 1-harmonic fits, the dissolved climatologies are roughly 2 weeks early at peak and trough. These observations are broadly consistent with observations by Manizza et al. [2012]. This study, itself, broadly agrees with Garcia and Keeling's [2001] own prediction that

the GK01 O_2 fields, derived by correlation with heat fluxes, would anticipate observations by 2-3 weeks due to lack of a reliable estimator of mixed layer equilibration time. GK01 runs with different atmospheric transport models agree on the amplitude and phase well, showing that surface fluxes are primarily responsible for differences in timing and amplitude with observations.

To establish good quantitative measures, I assess how well fits to models reproduce fits to observed values with the six metrics mentioned above: amplitude of 1- and 2-harmonic fits, timing of peak for 1- and 2-harmonic fits, timing of trough for 2-harmonic fits, and trough year day minus peak year day for 2-harmonic fits. This last value, essentially temporal asymmetry, with ~182 days showing no asymmetry at all, is necessarily the same for 1-harmonic fits, so I do not show this as a seventh metric. I present these values graphically in **Figure 3.29**, tabulate them in **Table 3.7** and **Table 3.8**, and specify which models fall within 1 sigma and 95% confidence levels in the following paragraphs.

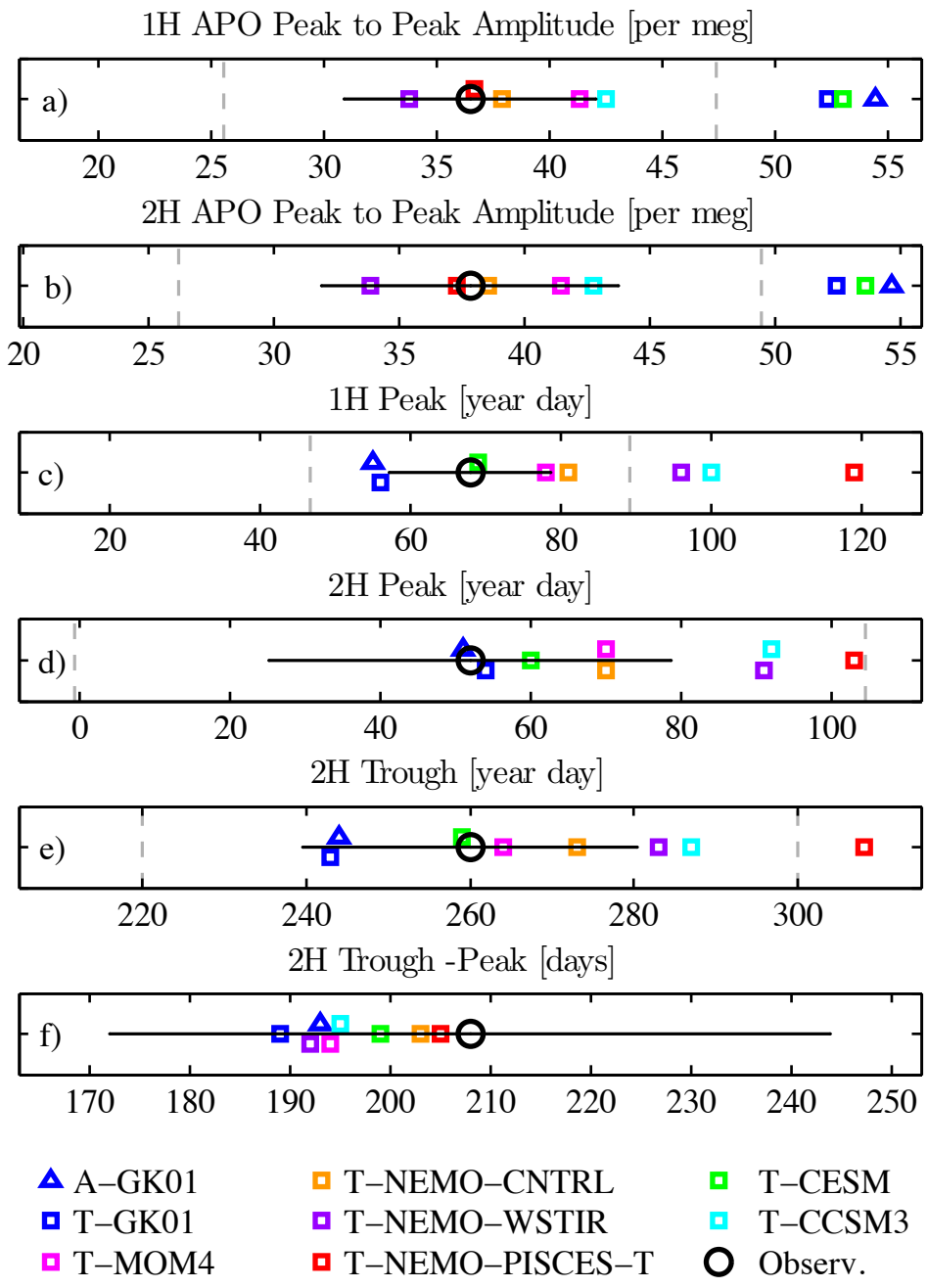

Figure 3.29. Models (A=ACTM; T=TM3) and HIPPO observations compared for each of the six metrics mentioned in the previous paragraph. Observation error bars reflect error values for 1- and 2-harmonic scenarios calculated in the previous section, and tabulated in Table 3.7 and Table 3.8. Uncertainty is shown as a horizontal black line centering around the observed value, while 95% confidence limits are shown as vertical dashed gray lines for all panels but f), in which all models fall within the 1-sigma error bar.

1-Harmonic Amplitude:

A 1-harmonic fit to the observations has a peak-to-peak amplitude of 36.4 ±5.6 per meg (±15%). Models variously under/overstate this value by -7%/+49%. Of the eight runs, MOM4, NEMO-CNTRL, NEMO-PISCES-T, and NEMO-WSTIR fall within 1 sigma, while CCSM3 falls within 95% confidence limits. CESM, ACTM-GK01, and TM3-GK01 fall outside 95% confidence limits.

2-Harmonic Amplitude:

A 2-harmonic fit to observations has a peak-to-peak amplitude of 37.8 per meg ±5.9 per meg (±17%). Models variously under/overstate this value by -10%/45%, with the majority overstating the amplitude relative to observations. Of the eight runs, CCSM3, MOM4, NEMO-PISCES-T, NEMO-CNTRL, and NEMO-WSTIR fall within 1 sigma, while ACTM-GK01, TM3-GK01, and CESM fall outside 95% confidence limits. The dissolved climatology runs overstate the seasonal cycle similarly, as might be expected with the curtain average technique. Interestingly, CCSM3 (+13%) and CESM (+41%), though related models, disagree strongly.

1-Harmonic Peak Timing:

Modeling the peak of the seasonal signal is potentially quite difficult because it requires adequately representing the slowing of productivity and solubility signals, a process that may cause a wide peak. A 1-harmonic fit to observations indicates a peak at March 9, year day 68 ±10.8 days (±3% of the seasonal cycle). Here I define positive (+)

as leading the phase of the observations, and negative (-) as lagging. Models variously lead/lag this value by 13/-51 days, i.e. 24 Feb to 29 Apr. Of the eight runs, CESM and MOM4 fall within the 1 sigma, while ACTM-GK01 and TM3-GK01 fall within 95% confidence limits. CCSM3, NEMO-PISCES-T, and NEMO-WSTIR fall outside 95% confidence limits.

2-Harmonic Peak Timing:

A 2-harmonic fit to observations indicates a peak at 21 February, year day 52 ±27 days (±7.4% of the seasonal cycle). This error bar is substantially wider than the value calculated for the 1-harmonic fit because the 2-harmonic fit has an additional degree of freedom that allows the model peaks and troughs to move forward or backward relative to each other. 2-harmonic fits to model output also fall quite widely, differing at the peak by 52 days, spanning the time from year day 51 to 103 (20 Feb to 13 Apr). Five models fall within the peak timing uncertainty: ACTM and TM3 dissolved climatology runs, CESM , MOM4 and NEMO-CNTRL. NEMO-WSTIR, CCSM, and NEMO-PISCES-T lag the observed peak considerably, but fall within 95% confidence limits.

2-Harmonic Trough Timing:

In theory, modeling of the trough timing can be easier, as it correlates not with the relaxation of a function (process) as signals slow down, but with the commencement of signals (i.e. solubility and photosynthesis). This is not obvious in the first harmonic, as the uncertainty at the peak and trough is effectively identical due to the fitting process. However, in the case of a 2-harmonic fit, this is more apparent, though the uncertainties

are again large. Observations place the trough timing at 17 September, year day 260 ±20 days (5.5% of the seasonal cycle). The smaller uncertainty in this value also has to do with the placement of the HIPPO points—here a single point, HIPPO5, is near the seasonal minimum, instead of two points, suggesting that the minimum falls somewhere close to this value. In the case of the peak, two similar APO values allow for solutions on either side of these two points, causing a larger uncertainty. Five models fall within 1 sigma: the two dissolved climatology runs, CESM, MOM4 and NEMO-CNTRL. CESM falls closest to observations here, while the two dissolved climatology runs fall near the edge of the error bar. Though they appear to have captured the peak relatively closely, they are further from the observations at the trough, with TM3 and ACTM leading the trough by +17 and +16 days, respectively. NEMO-WSTIR and CCSM3 fall within 95% confidence limits, and only NEMO-PISCES-T is outside of these limits, with timing that effectively mirrors its peak, lagging by -48 days (YD 308).

Seasonal Asymmetry:

The 2-harmonic fit to observations suggests an asymmetrical seasonal cycle, with 57% of the year (208 days) falling between the peak and trough, as defined by the 2-harmonic fit. Though some models, notably the GK01 runs, are only 2-3 percentage points from being seasonally symmetrical, all models satisfy this constraint because the 2-harmonic fit procedure allows a very large 1-sigma of 36 days on this difference, effectively the quadrature sum of the peak and trough uncertainties. Though I believe this uncertainty estimate is overly conservative, considering previous knowledge of seasonal cycles at the surface and at elevation, I cannot constrain this number any more using only

HIPPO observations. Some models are, of course, closer to observations than others. NEMO-PISCES-T is the closest model, with 205 days (56%) between the peak and the trough after it. NEMO-CNTRL is next, at 203 (56%), then CESM at 199 days (55%), CCSM3 at 195 days (53%), MOM4 at 194 days (53%), ACTM-GK01 at 193 days (53%). NEMO-WSTIR at 192 (53%), and TM3-GK01 at 189 days (52%).

Of the six metrics, NEMO-PISCES-T and CESM are closest twice, while ACTM-GK01 is closest once. However, the consistent closest performer throughout the four metrics is MOM4 (within 1-σ 5 times), with NEMO-CNTRL a close second (within 1-σ 4 times) (see **Figure 3.29**).

Table 3.7. Summary of relevant statistics for modeled and observed (AO2/MEDUSA) Southern Ocean curtain average 1-harmonic fits. Amplitudes of APO are in per meg, while amplitudes of oceanic O_2, N_2 and CO_2 are given in both [per meg] and [ppm]. All maxima and minima are given in year day. Units: [per meg] and [ppm].

	—	APO	—	—	O_2	—	—	N_2	—	—	CO_2	—
Model Run or Observ.	Amp [per meg]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD
ACTM GK01	54.4	55	238	62.2 / 13.0	56	238	7.2 / 5.6	57	240	0.8/ 0.2	253	71
TM3 GK01	52.3	56	239	60.4 / 12.7	56	239	7.0 / 5.5	57	240	1.1 / 0.3	244	61
TM3 CCSM3	42.5	100	282	48.1 / 10.1	95	277	6.2 / 4.8	83	265	2.3 / 0.5	180	363
TM3 CESM	53.0	69	252	61.0 / 12.8	68	251	6.4 / 5	54	236	2.1 / 0.5	270	88
TM3 MOM4	41.3	78	261	45.3 / 9.5	75	257	4.7 / 3.7	60	243	1.5 / 0.3	152	335
TM3 NEMO-PISCES-T	36.6	119	301	44.5 / 9.3	115	298	10.0 / 7.8	88	270	2.4 / 0.6	46	228
TM3 NEMO-CNTRL	37.9	81	264	36.6 / 7.7	74	257	6.3 / 4.9	52	234	7.2 / 1.7	94	277
TM3 NEMO-WSTIR	33.8	96	278	36.6 / 7.7	85	267	8.1 / 6.3	62	244	5.2 / 1.2	126	309
Observ.	36.4 ±5.6	68 ±10.8	250 ±10.8	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 3.8. Summary of relevant statistics for modeled and observed (AO2/MEDUSA) Southern Ocean curtain average 2-harmonic fits. Amplitudes of APO are in per meg, while amplitudes of oceanic O_2, N_2 and CO_2 are given in both [per meg] and [ppm]. All maxima and minima are given in year day. Units: [per meg] and [ppm].

	—	APO	—	—	O_2	—	—	N_2	—	—	CO_2	—
Model Run or Observ.	Amp [per meg]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD	Amp [per meg]/ [ppm]	Max YD	Min YD
ACTM GK01	54.7	51	244	62.5 / 13.1	51	246	7.2 / 5.6	51	249	1 / 0.2	278	39
TM3 GK01	52.4	54	243	60.6 / 12.6	53	244	7.1 / 5.4	53	247	1.4 / 0.3	265	26
TM3 CCSM3	42.8	92	287	48.6 / 10.2	85	284	6.2 / 4.8	78	271	2.5 / 0.6	150	10
TM3 CESM	53.6	60	259	61.8 / 12.9	59	259	6.4 / 5	49	244	2.2 / 0.5	284	64
TM3 MOM4	41.5	70	264	45.9 / 9.6	60	263	4.7 / 3.7	55	250	2.4 / 0.6	103	357
TM3 NEMO-PISCES-T	37.3	103	308	45.4 / 9.5	100	305	10.2 / 8	77	282	2.6 / 0.6	59	206
TM3 NEMO-CNTRL	38.5	70	273	37.3 / 7.8	62	266	6.4 / 5	44	245	7.6 / 1.8	81	299
TM3 NEMO-WSTIR	33.9	91	283	36.7 / 7.7	80	273	8.1 / 6.3	58	251	5.3 / 1.2	116	319
Observ.	37.8 ±5.9	52 ±27	260 ±20	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 3.9. Summary of O_2 relative to N_2 peak timing lags, N_2/O_2 amplitude ratios, and over- or under-estimation of curtain average APO amplitude relative to observations, for all model runs' 1-harmonic fits. By convention here, Peak O_2-N_2 Lag=Year Day(O_2_peak)-Year Day(N_2_peak). Though lag is not a direct indicator of the biological component of the O_2 signal, decoupling in the timing of the O_2 and N_2 fields suggests the O_2 signal is subject to a biological forcing that the N_2 signal is not. Likewise, low O_2-N_2 lag values broadly suggest strongly coupled solubility and biology responses, or a large solubility forcing contribution to the O_2 signal. N_2/O_2 amplitude is given as the ratio in ppm (mole fraction) of the N_2 and O_2 seasonal amplitudes, noted here for convention with Manizza et al. [2012]. The final column shows the overestimate or underestimate in percent of the APO seasonal cycle of each model relative to the observed amplitude. The over/underestimate percentage is calculated from the observed APO, which has units of per meg.

Model Run/ Data	Peak O_2–N_2 Lag (days)	N_2/O_2 Amp [ppm/ppm]	Under- or over-estimate of obs. APO CA (%)
ACTM GK01	1	0.43	+49.3
TM3 GK01	1	0.43	+43.6
TM3 CCSM3	12	0.48	+16.6
TM3 CESM	14	0.39	+45.3
TM3 MOM4	15	0.39	+13.3
TM3 NEMO-PISCES-T	27	0.84	0.5
TM3 NEMO-CNTRL	22	0.64	+3.9
TM3 NEMO-WSTIR	23	0.82	-7.4

Table 3.10. Summary of O_2 relative to N_2 lags, APO trough-peak spacing, N_2/O_2 amplitude ratios, and over- or under-estimation of curtain average APO amplitude relative to observations for all model runs' 2-harmonic fits. By convention here, Peak O_2-N_2 Lag=Year Day(O_2_peak)-Year Day(N_2_peak) and Trough O_2-N_2 Lag=Year Day(O_2_trough)-Year Day(N_2_trough). Because these are 2-harmonic runs, the peak and trough are not phase-locked, so I show the difference between trough and peak year day as: APO Trough-Peak Spacing=Year Day(APO_trough)-Year Day(APO_peak). Though lag is not a direct indicator of the biological component of the O_2 signal, decoupling in the timing of the O_2 and N_2 fields suggests the O_2 signal is subject to a biological forcing that the N_2 signal is not. Likewise, low O_2-N_2 lag values broadly suggest strongly coupled solubility and biology responses, or a large solubility forcing contribution to the O_2 signal. N_2/O_2 amplitude is given as the ratio in ppm (mole fraction) of the N_2 and O_2 seasonal amplitudes, noted here for convention with Manizza et al. [2012]. The final column shows the overestimate or underestimate in percent of the APO seasonal cycle of each model relative to the observed amplitude. The over/underestimate percentage is calculated from the observed APO, which has units of per meg.

Model Run/ Data	Peak O_2–N_2 Lag (days)	Trough O_2–N_2 Lag (days)	APO Trough-Peak Spacing (days)	APO Trough-Peak Spacing (year fraction)	N_2/O_2 Amp [ppm/ppm]	Under- or over-estimate of obs. APO CA (%)
ACTM GK01	0	3	193	0.53	0.43	+44.5
TM3 GK01	0	3	189	0.52	0.43	+38.6
TM3 CCSM3	7	13	195	0.53	0.47	+13.0
TM3 CESM	10	15	199	0.55	0.39	+41.6
TM3 MOM4	5	13	194	0.53	0.39	+9.6
TM3 NEMO-PISCES-T	23	23	205	0.56	0.84	-1.4
TM3 NEMO-CNTRL	22	21	203	0.56	0.64	-1.8
TM3 NEMO-WSTIR	22	22	192	0.53	0.82	-10.4
Observ.	n/a	n/a	208±36	0.57±0.1	n/a	n/a

3.7.1 Biological and physical forcings

Figure 3.27 collectively showed the relative contribution of the tracer elements to the APO signal. Though I am unable to directly deconstruct the HIPPO APO signal into its oceanic O_2, N_2 and CO_2 components, looking at model components gives a sense for how biological and physical forcings influence the model seasonal signals. In **Figure 3.30**, I show each tracer in units of [ppm] (equivalents, when each tracer is run within a model context as a trace gas) with axes normalized to the minimum and maximum of the all the models' simulations of the individual tracer. In the fourth panel, I show the combined APO signal in units of [per meg] as per Eq. 3.3. I show these for the 2-harmonic fit scenario only, as the 1-harmonic scenario is visually very similar.

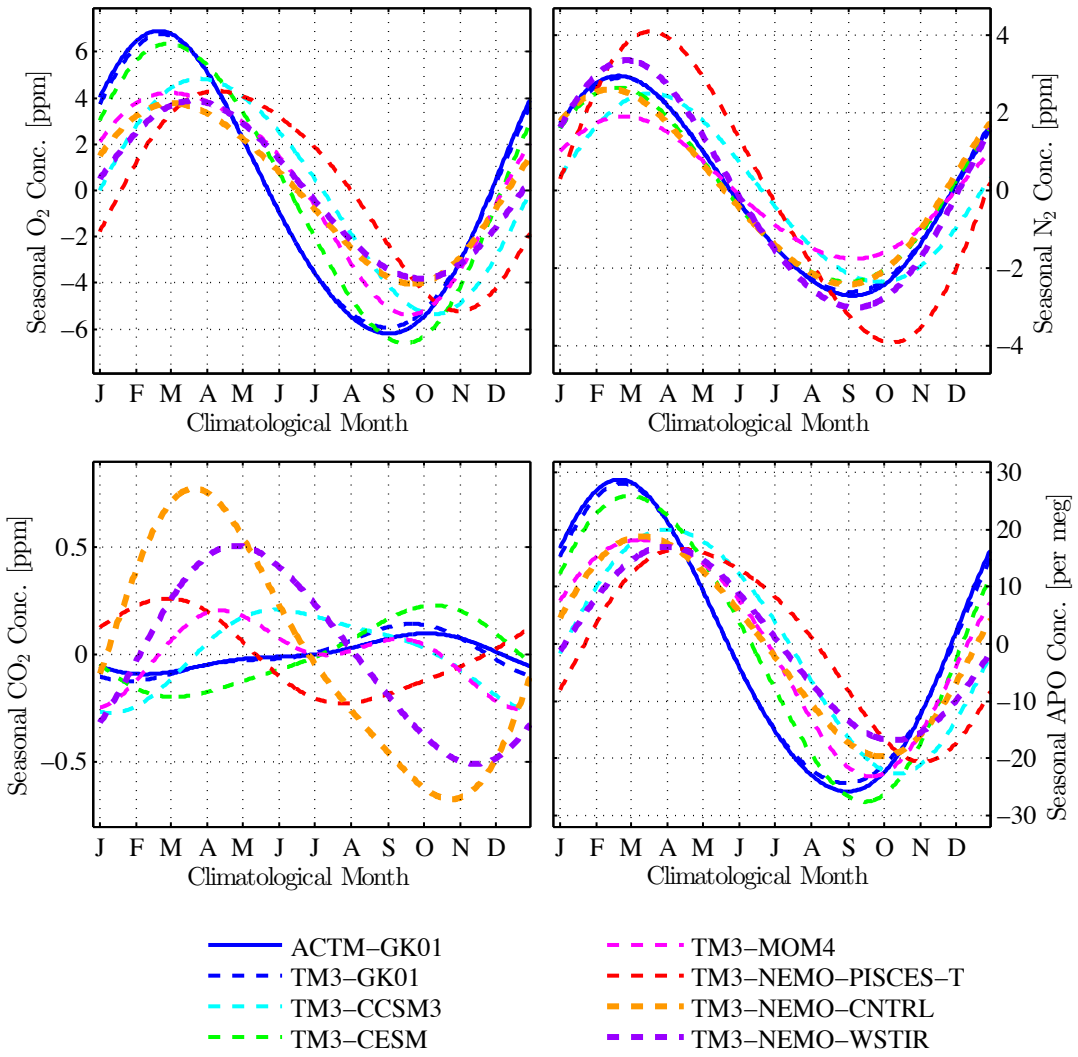

Figure 3.30. 1-harmonic fits to monthly mean seasonal O_2, N_2, CO_2, and APO concentration for the seven forward simulations for the Southern Ocean slice in ppm. Units: [ppm], [per meg], as marked.

Figure 3.27 showed that O_2 is by far the dominant contributor to the APO seasonal signal. **Figure 3.30** adds to this, showing just how similar the shapes and relative model-to-model amplitudes of the O_2 and APO signals are. Also apparent is that while models agree nominally on the timing of the N_2 and O_2 fluxes within a two-month window, they disagree hugely on the CO_2 fluxes due to the need to handle interfering solubility and biology signals. Also apparent is that NEMO-PISCES-T is a major outlier in N_2 relative

to other models: its N_2 fields are both later and larger than other models'. However, its O_2 component is fairly normal for amplitude, though still "late" relative to observations. The large NEMO-PISCES-T N_2 fluxes partly explain the relatively small APO amplitude in **Figure 3.27**, since the N_2 component counteracts the O_2 component in the O_2/N_2 (and APO) calculation—i.e. the larger the N_2 fluxes, the more reduced in amplitude the APO seasonal cycle will appear. The large N_2 fluxes suggest that the model is overestimating the seasonal variation in heat fluxes relative to the others, and that the timing of these fluxes is delayed—two different but possibly related physical model features/problems. However, simply reducing the magnitude of the heat flux seasonal cycle would also reduce the amplitude of the atmospheric O_2 fluxes, suggesting that a single solution is not sufficient to address all of the differences between NEMO-PISCES-T and observations. Furthermore, the delay in O_2 fluxes relative to N_2 fluxes is the largest of the five models by about a week and a half with a $\Delta(O_2\_peak\text{-}N_2\_peak)$ of 27 days and 23 days for 1- and 2-harmonic fits, respectively. This could suggest a different phase in the onset of photosynthesis (biological O_2) in the PISCES-T biogeochemistry model formulation.

The NEMO-PISCES-T run is well-known because a prominent paper associated with it [*Le Quéré et al.*, 2007] appeared to show a slowing in the rate of increase in the Southern Ocean CO_2 sink in recent decades. Observations suggest that increases in latitudinal surface temperature gradients plus decreases in stratospheric ozone have led to an observed increase and poleward shift in the westerlies around Antarctica. Such a shift is expected to enhance northward Ekman transport around Antarctica, increasing SO meridional overturning in the Antarctic Divergence. Le Quéré et al. [2007] argued that this, in turn, leads to outgassing of DIC-rich deep waters, partially offsetting the

increasing sink expected from anthropogenically-induced atmospheric CO_2 rise over the same time period. A subsequent critique of the paper [*Zickfeld et al.*, 2008] has suggested that enhanced Ekman transport would increase the rate at which sinking surface waters produced mode and intermediate waters, entraining additional anthropogenic CO_2. Zickfeld et al. [2008] showed that in their model this counteracting flux overwhelmed the outgassing effect, leading to a net increase in the efficiency of the SO anthropogenic CO_2 sink. A second critique, by Law et al. [2008], also argued that LeQuéré et al. [2007] included Southern Ocean station data from Amsterdam and Ascension Islands (AMS and ASC) that hadn't been filtered properly for high CO_2 outliers previous to 1991. These potentially bad data caused an overly negative station trend, and led to atmospheric inversion results which suggest spurious trends in air-sea CO_2 fluxes.

Though these studies concern themselves primarily with long-term changes, the NEMO-PISCES-T results shown here may well shed some light on why the model predicts a slowing in the rate of increase in the Southern Ocean sink that the Zickfeld et al. [2008] model does not. If the NEMO-PISCES-T solubility component is as far off in the Southern Ocean as it appears to be (lagging other models by several weeks for N_2), solubility forcings may be causing anomalous biological component feedbacks, encouraging in- or out-gassing at the wrong times of year, which could cause anomalously timed sinks and sources of anthropogenic CO_2 in a forward simulation. NEMO-PISCES-T's summer outgassing of CO_2 is significantly out of phase with the dissolved climatology fluxes (**Figure 3.30**, third panel; [*Takahashi et al.*, 2009]), a feature that may be explained by the large heat flux implied by the N_2 signal. Furthermore, the NEMO-PISCES-T O_2 fields trend upward strongly with time (rather than downward as

would be expected from observations), a feature typically associated with a run that has not had time to reach steady state. This has both consequences for interannual signals, and potentially for seasonal signals, as the release of O_2 from over-saturated deep waters should happen predominantly in the winter, when deep water is ventilated by the atmosphere. This signal may well be responsible for the very wide peak in atmospheric O_2 seen in the top left panel of **Figure 3.30** if some additional spurious release of O_2 coincides with the expected ingassing due to ventilation dynamics.

Though I cannot be sure how the features I observe in this study may be related to the controversial CO_2 sink saturation proposed by LeQuéré et al. [2007], repeating their wind-forcing analysis with some of the better-performing runs from this study would provide an added dimension to the discussion of the saturation of the Southern Ocean CO_2 sink, one that could potentially confirm whether models are getting Southern Ocean physics and air-sea fluxes accurately enough to make future projections.

Though CCSM3 and CESM come from the same model family (the ocean component of CESM is the latest NCAR ocean model, while CCSM3 is the penultimate version), the two models disagree greatly on the magnitude and timing of the O_2 and N_2 fields. The two models' heat flux magnitudes appear to be similar, but the onset of austral spring productivity in CCSM3 (peak YD 271) is 27 days later then CESM (YD 244), a significant difference. The large difference in the amplitude of the CESM O_2 fluxes (61.8 per meg) relative to CCSM3 (48.6 per meg), despite the similarity between the two models' N_2 flux amplitudes, also suggests that the biogeochemistry component of the CESM model O_2 fluxes is much greater than its predecessor. The lag in the O_2 fluxes of both models relative to their N_2 fluxes, however, is fairly similar—on average 10 days for

CCSM3 and 14 days for CESM for the 2-harmonic fits (12 and 14, respectively for 1-harmonic fits)—consistent with statements [*M. Long*, pers. comm.] that the biogeochemistry in CCSM4 is not radically different from that in CCSM3.

At this point, though there appear to be notable differences in the biogeochemistry, I cannot attribute the majority of the differences between CCSM3 and CESM to this because changes to the physics of the POP2 model are the predominant changes between the two runs, a point supported by feedback on this dissertation from developers of the NCAR model [M. Long, S. Doney, *pers. comm.*]. If physical changes are, indeed, the predominant driver of differences between the two model versions, processes like outgassing/ventilation dynamics and mixed layer growth might be responsible. The scope of this study doesn't include a more comprehensive analysis of solubility and biological O_2 flux components, so I leave my analysis at this.

MOM4, which approximates the seasonal APO signal well, has O_2 (45.9 per meg) and N_2 (4.7 per meg) amplitudes that are the smallest relative to other models, and yet despite the small N_2 signal (which can cause asymmetry in the APO value if it is large and off-phase from O_2), the O_2 peak is broad and the cycle is asymmetrical. This suggests that the MOM4 springtime biology may constitute a large component of the O_2 signal, since it does not scale only with the solubility, and that the timing of this component is well-spread through the seasonal cycle, not concentrated briefly.

The difference in the NEMO-CNTRL and NEMO-WSTIR runs, at least as I have fit them, is considerable. The wind-stirring parameterization of NEMO-WSTIR [*Rodgers et al.*, 2014] has two principal impacts on the seasonal cycle in mixed layer depth. First, it perturbs the seasonal maximum and seasonal minimum values of mixed layer depth, with

the important result that summer mixed layer depths better resemble ARGO-derived values during summer. With larger summer mixed layer depths for NEMO-WSTIR, both light and nutrient availability are impacted, with the integrated effect that the amplitude of the seasonal cycle in APO driven by the two runs is different. Second, wind stirring perturbs the phasing of the seasonal cycle in mixed layer depth, with an important consequence being seen in the respective timing of re-stratification (spring) and de-stratification (autumn) for the two runs. NEMO-CNTRL re-stratifies approximately 5 weeks earlier than NEMO-WSTIR, and through the joint impact of light and nutrient limitations on productivity, this leads to significant differences in the phasing of air-sea exchanges of O_2 and CO_2, and consequently in the phasing of the seasonal cycle in APO [*K. Rodgers*, pers. comm.].

The findings of this study seem to confirm much of the phenomenological explanation of Rodgers et al. [2014], however whereas Rodgers et al. [2014] find the WSTIR parameterization brings seasonal APO cycles at surface stations in the Southern Ocean into greater agreement with some observations at PSA, this is less apparent in the context of the curtain average, at least using the 1- and 2-harmonic fitting methods of this study. For NEMO-WSTIR, a noticeably larger heat flux (seen by proxy in the N_2 seasonal cycle) is indicative of meaningful changes to the model physics, due to the wind stirring parameterization. O_2 fluxes for CNTRL and WSTIR runs are roughly equivalent on amplitude—consistent with no change to the biogeochemistry itself—but in the case of WSTIR they are delayed, consistent with the predictions of the Rodgers et al. [2014] study.

Lastly, the phasing of the N_2 and O_2 components of the dissolved climatology runs is effectively identical due to the method Garcia and Keeling used to interpolate and extrapolate sparse O_2 observations with heat-flux data. Relative to most other models, the O_2 fluxes also appear quite large.

3.8 Adjusting the O_2 fluxes from Garcia and Keeling [2001]

HIPPO APO data provides a basis for evaluating the accuracy of the southern hemispheric O_2 fluxes estimated by Garcia and Keeling [2001], something of particular worth because I use these fluxes to estimate bias and error in the HIPPO curtain average values. Accordingly, I am interested in using output based in the original fluxes, but that is as close in phase and amplitude to observations as possible before doing so. One way to go about this includes changing the phasing of the O_2 fluxes in the ACTM-GK01 run (which continues with up-to-date reanalysis meteorology through 2012) so as to test the assertion that the phase of the O_2 fluxes is too early. A second way to go about this is to see how much of an adjustment to the amplitude of the fluxes is needed to bring them to the amplitude of the HIPPO observations. This second option reflects recent observations [*Jin et al.*, 2007; *Manizza et al.*, 2012; *Naegler et al.*, 2006] that the amplitude may be too great, and that this may have to do with a scaling factor, known as a_q, used in the Wanninkhof [1992] gas exchange parameterization.

3.8.1 Phase and shape adjustments to Garcia and Keeling [2001]

In their paper, Garcia and Keeling [2001] note that the phase of their O_2 fields might be too early by a few weeks. Without a reliable estimate of the time scale of

mixed-layer dynamics, they chose to simply scale the timing of O_2 fluxes to heat fluxes, effectively allowing the O_2 anomaly to equilibrate instantaneously with the atmosphere. This tends to lead to an APO signal whose O_2 and N_2 components are fairly simultaneous, since the large solubility forcing [*Manizza et al.*, 2012] acts at the same time, and in the same direction for the two species' flux calculations. This suspicion by Garcia and Keeling is preliminarily supported by 1- and 2-harmonic fits to HIPPO curtain average data, which suggest the GK01 fluxes are early with respect to data by 12-13 days. Furthermore, in the case of the 2-harmonic fit, the GK01 fluxes run through both transport models appear to be very temporally symmetrical, something that data (with large uncertainties) and ocean model output suggest may be incorrect. Accordingly, two major questions with regard to the GK01 flux timing are whether they are correct in phase, and whether they are too symmetrical with respect to time.

First, I choose to look at the seasonal "symmetry" of the ACTM-GK01 simulation by delaying the timing of the O_2 and N_2 fluxes in the transport models to determine whether the effectively simultaneous timing of the two flux fields themselves led to an overly symmetrical seasonal cycle. To this end, Dr. Prabir Patra ran two different N_2 flux timings and four different O_2 flux timings through the ACTM model. Together, these offer eight different possible O_2+N_2 phasing combinations. 2-harmonic fits to the individual components are shown in **Figure 3.31** below, as 1-harmonic fits do not allow peak and trough to move independently.

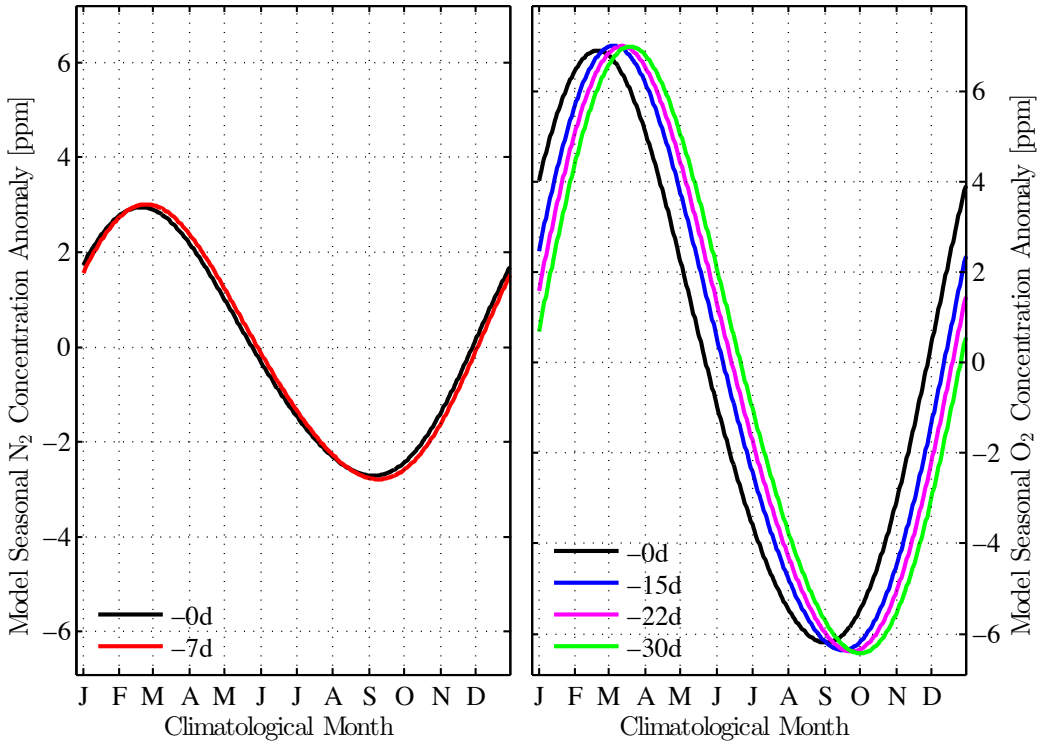

Figure 3.31. Phasing of Blaine [2005] N_2 fluxes (left panel) and Garcia and Keeling '01 O_2 fluxes (right panel). Black lines represent the native timing of the original flux fields, while red, blue, magenta and green represent fluxes whose values are delayed 7, 15, 22, and 30 days (which appears as a forward movement in X). Units: [ppm].

I then compute the 2-harmonic seasonal APO cycle for each of these 8 scenarios to see whether the difference in trough-peak timing, originally 193 days, increases as various O_2 and N_2 delays are implemented. Differences in such a plot are extremely hard to observe, so I present the trough-peak timing difference in the table below.

Table 3.11. Trough-peak difference in days for the eight N_2 and O_2 flux timing scenarios run through ACTM. A 182 (or 183) day difference between trough and peak signifies "symmetry" with respect to time, so seasonal "asymmetry" increases for values as they move away from this number in either direction.

Flux Timing Scenario	*0-Day O_2 Delay*	*15-Day O_2 Delay*	*22-Day O_2 Delay*	*30-Day O_2 Delay*
0-Day N_2 Delay	193	194	195	197
7-day N_2 delay	193	194	195	196

The first thing to note from the table above is that delaying the N_2 fluxes by a week doesn't appreciably increase the seasonal "asymmetry" (trough-peak difference) of the APO signal—in fact in the 30-day O_2 flux field delay scenario, it slightly decreases the asymmetry, so I choose to disregard this option, since I have no physical basis upon which to justify such a change—after all, N_2 is more closely linked to heat fluxes, so allowing it to scale with heat fluxes is fairly reasonable.

The second thing to note is that delaying the O_2 field with respect to the remaining fields does slightly increase the seasonal asymmetry (up to 197 days trough-peak, from 193). These changes don't bring the trough-peak timing difference up to the value predicted by 2-harmonic fits to data—208 days—however. But, since this HIPPO number has a very large uncertainty of ±36 days, it is quite possible that the actual seasonal "asymmetry" is not this pronounced. Nonetheless, it is clear that delaying O_2 fields in time does slightly increase their trough-peak "asymmetry", which has the effect of bringing this metric closer to 2-harmonic fit observations.

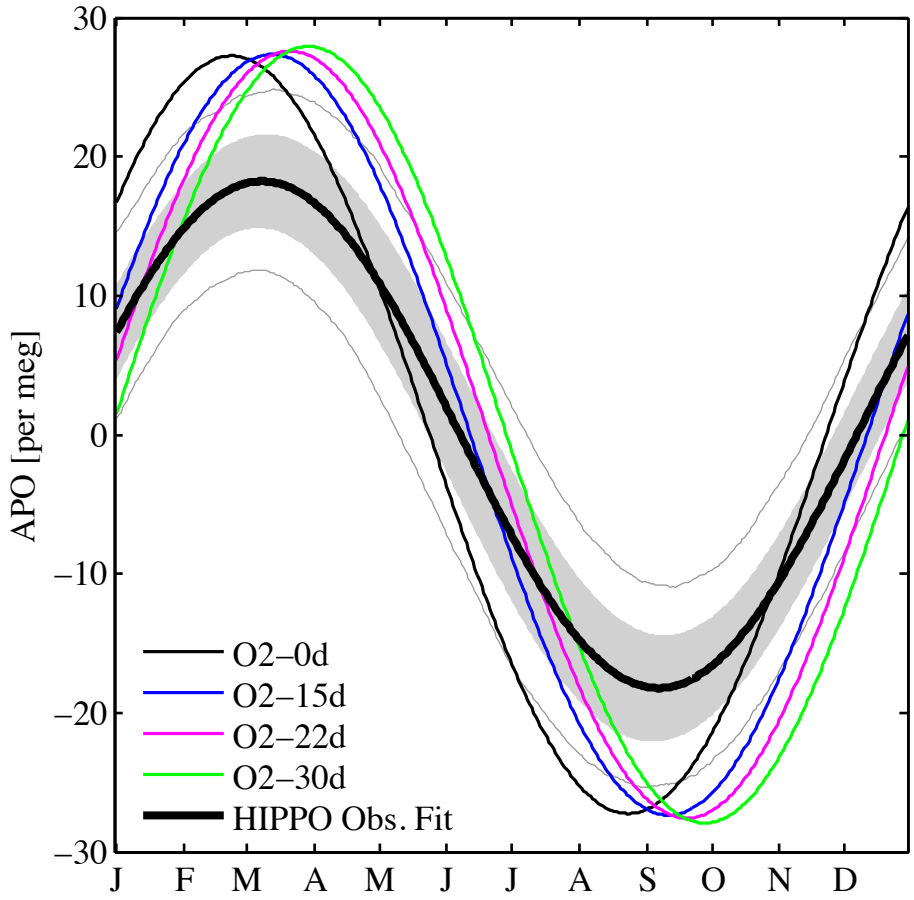

Figure 3.32. Comparison of 1-harmonic fit to observed curtain average (thick black line with shaded gray 1-sigma error bar and light gray 95% confidence limits) against 1-harmonic fits to four APO ACTM-GK01 O_2 flux timing runs for a climatological year. The original GK01 timing (thin black line) is very early in phase relative to observations, while the 15-day delay captures the timing of observations best. Units: [per meg].

The next question is whether delaying GK01 O_2 fields relative to their original timing brings the ACTM-GK01 curtain average APO results closer in phase to observations. Note, though this is fairly apparent from the phasing difference observed in the 1-harmonic fits—ACTM-GK01 leads observations by 12-13 days—this isn't strictly representative of what will happen when the fluxes are delayed and rerun through the model, as non-linear elements (e.g. seasonal changes in meridional transport) might change phasing subtly. However 1-harmonic fits to these reruns, seen in **Figure 3.32** above

as blue (15-day), magenta (22 day), and green (30 day) lines, suggest subtle non-linear components are small because a 2-week (i.e. 15 day) phase delay scenario captures the timing of the HIPPO 1-harmonic fit best. The 2-week delay scenario actually brings the trough to year day 255, while the HIPPO trough is at 250 (see table below), however, this is the closest of the three scenarios, and iterating many times to perfect the exact timing would require substantial additional server time. Accordingly, I choose to use the 2-week delay option when computing bias estimates in Section 3.5.5, and I show the 2-week delayed fluxes (noted) in subsequent figures when appropriate. It is worth noting that an optimal phase shift for the GK01 fluxes is probably around 12-13 days, as seen in the initial 1-harmonic model comparisons.

Table 3.12. Trough timing (year day) of 1-harmonic fits to HIPPO curtain average APO data, and the same for ACTM runs, variously with 0-, 15-, 22-, and 30-day O_2 flux field delays.

HIPPO Trough YD	*ACTM 0-Day O_2 Delay Trough YD*	*ACTM 15-Day O_2 Delay Trough YD*	*ACTM 22-Day O_2 Delay Trough YD*	*ACTM 30-Day O_2 Delay Trough YD*
250±10.8	236	255	263	272

Finally, of scientific note, the 2-harmonic fit to observations, and fits to other model output, suggest preliminarily that the peak of the APO cycle over the Southern Ocean is actually quite a lot wider than the trough, and wider than the GK01 fluxes suggest. This would suggest that though the GK01 fluxes might capture the onset of surface layer photosynthesis in early spring well, they resolve the slowing-down of production and the subsequent slow drawdown of atmospheric O_2 less well. This could be explained by the computational method Garcia and Keeling [2001] used, as this might be expected of a data product whose air-sea exchanges were scaled to heat fluxes. The exact timing of the peak, and its width, is determined not only by heat fluxes, but by the

availability of nutrients in surface waters, something heat flux scaling cannot approximate since nutrients and heat fluxes do not necessarily correlate in time. Further, heat flux scaling might fall short in capturing the nature of ventilation well, since the rate of air-sea exchange associated with ventilation processes is more likely to depend on the large air-sea CO_2 and O_2 gradients when deep water meets atmosphere than on heat flux alone. All in all, though a two-week O_2 flux delay brings the ACTM-GK01 results closest to fits to observations in 1- and 2-harmonic scenarios, the fluxes are still probably not capturing the full picture about competing biological, physical and ventilation signals correctly.

3.8.2 Flux amplitude adjustment to Garcia and Keeling [2001]

Jin et al. [2007] evaluated the Garcia and Keeling [2001] fluxes directly, comparing the seasonal signals at surface stations with model output separated into thermal, biological and ventilation signals. Their conclusion was that the O_2 solubility signal is overstated by about 30%. I cannot distinguish physical and biological forcings in the present analysis, so I am unable to evaluate these components separately. However, a largely unrelated study by Naegler et al. [2006], came to a conclusion that has significance for the GK01 fluxes, though the study concerned CO_2 not O_2. The authors re-evaluated the global gas exchange scaling parameter a_q used to determine piston velocity, in light of recent improvements of the oceanic ^{14}C inventory. New, lower, estimates of the inventory suggest that this empirical scaling parameter was set artificially high in order to increase air-sea fluxes of ^{14}C to replicate the perceived oceanic inventory.

The authors recommend that the present a_q value of 0.39 cm/hr/(m/s)^2, determined by Wanninkhof [1992] should be reduced to a value of 0.32.

Interestingly, such a rescaling would have little effect for ocean models. Though it slows their flux transfer velocity, it doesn't significantly reduce the total value of the fluxes, as these are largely driven by internal ocean processes. However, for Garcia and Keeling [2001], it has a first-order significance because the O_2 flux magnitude was inferred directly from a combination of observed super- or undersaturated water samples and a "known" exchange velocity scaling factor of 0.39 from Wanninkhof [1992]. Reducing this scaling factor to 0.32 reduces their amplitude by 18%. This is an elegant solution, because it has the potential to improve the GK01 fluxes without decreasing the magnitude of the model ocean fluxes, several of which capture the APO amplitude reasonably well.

In **Figure 3.33** below, I show 1-harmonic fits to ACTM-GK01 APO with 15-day delayed O_2 fields, as I have shown above that this timing scenario most closely captures the timing of the observations. I scale the contributing O_2 fluxes (not the APO value) down by a factor of 0.32/0.39 and compare these with the fit to the observed HIPPO data. I also include a second rescaling, of 0.276/0.39, because this rescaling brings the GK01 fluxes into best agreement with the observed amplitude of this study.

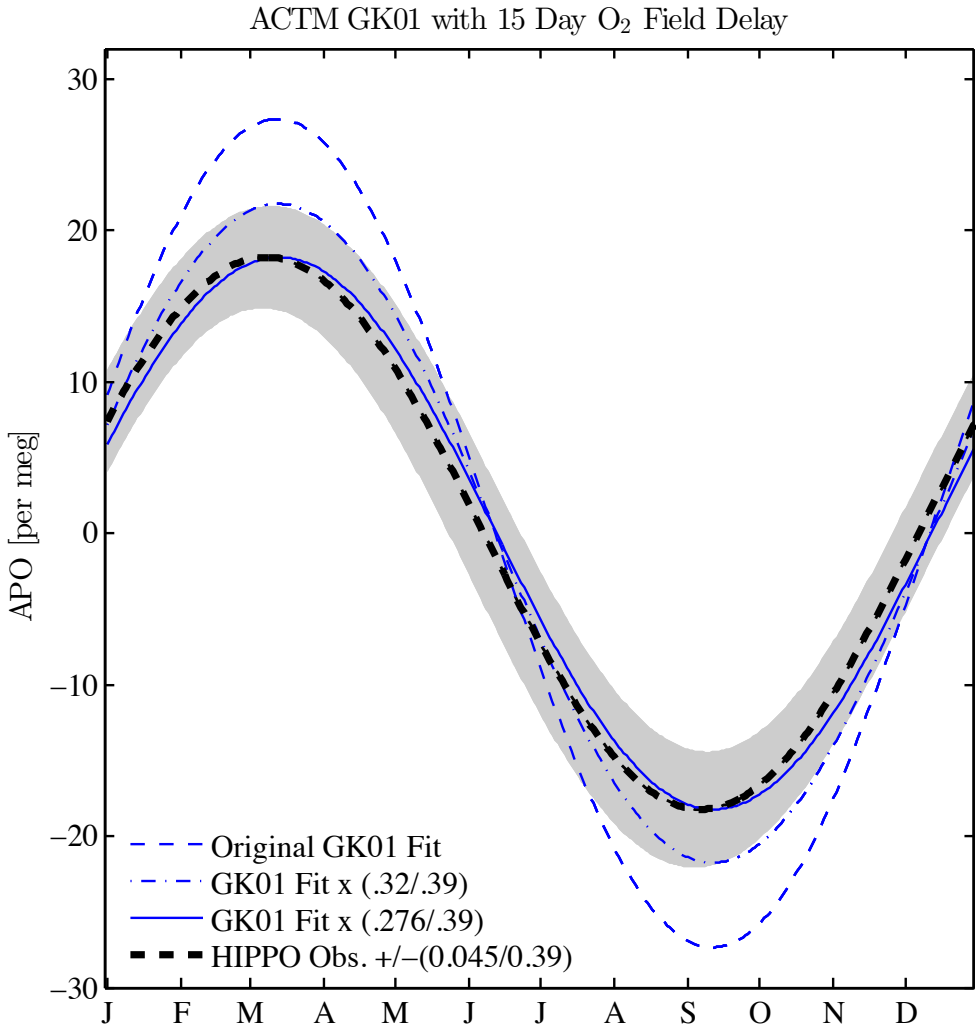

Figure 3.33. Comparison of ACTM GK01 runs using a 15-day O_2 flux delay with 1-harmonic fit to HIPPO observations. The APO cycle given the original unadjusted O_2 fields is shown as a thin dashed blue line. The amplitude implied by the Naegler a_q scaling is shown as a blue dash-dot line, and the amplitude implied by this study is shown as a solid blue line. The fit to observations has a continuous 1-sigma at any point in the fit of about 3.3 per meg (5.6 on the amplitude), which corresponds in this context to adding or subtracting 0.045/0.39 from the global gas exchange scaling parameter a_q.

The Naegler et al. [2006] rescaling significantly improves the fit to observations, producing a seasonal APO cycle with an amplitude of 43.4 per meg, while my empirically chosen value of (0.276/0.39 reduces the amplitude to 36.4 per meg, i.e. the 1-

harmonic fit amplitude observed in this study. In the context of a scaling parameter, this value has an uncertainty of ±0.045, suggesting the scaling parameter could be as low as 0.231 or as high as 0.321, just encompassing the global gas exchange scaling parameter of Naegler, but not the original value of Wanninkhof [1992]. I don't intend to suggest that a value of 0.276 is necessarily the correct scaling factor value, given that other elements of the Wanninkhof [1992] equation might also be scrutinized to bring observations and dissolved climatologies into agreement, however I note the value here as a matter of record, in case if can be of use in future work.

The observation that the Garcia and Keeling [2001] fluxes are too great by ~40-45% is in contrast to the findings of Garcia and Keeling themselves, who present their results alongside the Najjar and Keeling [1997] dissolved climatologies. Garcia and Keeling show their fluxes run through the TM2 transport model with ECMWF86 reanalysis, while Keeling et al. [1998a] show the Najjar and Keeling [1997] fluxes run through the same (Najjar and Keeling [1997] themselves do not). Both Garcia and Keeling [2001] and Keeling et al. [1998a] come to the conclusion that the Najjar and Keeling fluxes are too small over all, with Keeling et al. concluding the fluxes need to be increased by a factor of 1.23 in the Southern Hemisphere to bring them into agreement with station observations. My results suggest that this observation may have relied incorrectly on the performance of TM2, which appears to mix the surface signals out too strongly, as indicated by subsequent studies [*Blaine*, 2005; *Stephens et al.*, 2007].

3.9 Implications for atmospheric transport models

Though this study utilizes only two transport models, it provides a first opportunity to evaluate whether these models are capturing the vertical mixing of seasonal ocean signals properly. One quick first-order estimate of this effect is the degree to which the transport models capture the difference between the amplitude of the seasonal signal for the CA and for nearby level 1 station output.

Since the 1-harmonic fits to data have the lowest uncertainties in this study, I evaluate 1-harmonic surface station data only. I follow the convention of the TransCom3 Study Revised Protocol Appendix B [*Gurney et al.*, 2003; *Gurney et al.*, 2000] in extracting station output from gridded output. This convention differs for each location, depending on whether a station typically falls within a land or ocean box, and depending on prevailing winds around that station. For Cape Grim Observatory, I extract not at the exact location of the station, but at the closest grid point to the southwest. This measure is taken because the exact location of CGO tends to be considered by most transport models as lying at the coast of a land box, which causes the station to show a mixture of land and ocean signals. In reality, however, station "baseline" data are only taken when winds are coming from between 190° and 280° [*Law et al.*, 2010], so comparing data to a nearby model grid box that doesn't have a large land signal is a better representation of the sampled site conditions. For Palmer, local terrestrial fluxes are small, and no sample grid point direction is specified by the TransCom protocol, so I simply output model results at the closest ocean grid point. **Figure 3.34** shows the seasonal cycle of the curtain average of each ocean model alongside 1-harmonic fits to the weekly-resolution seasonal cycle output at Cape Grim Observatory and Palmer Station.

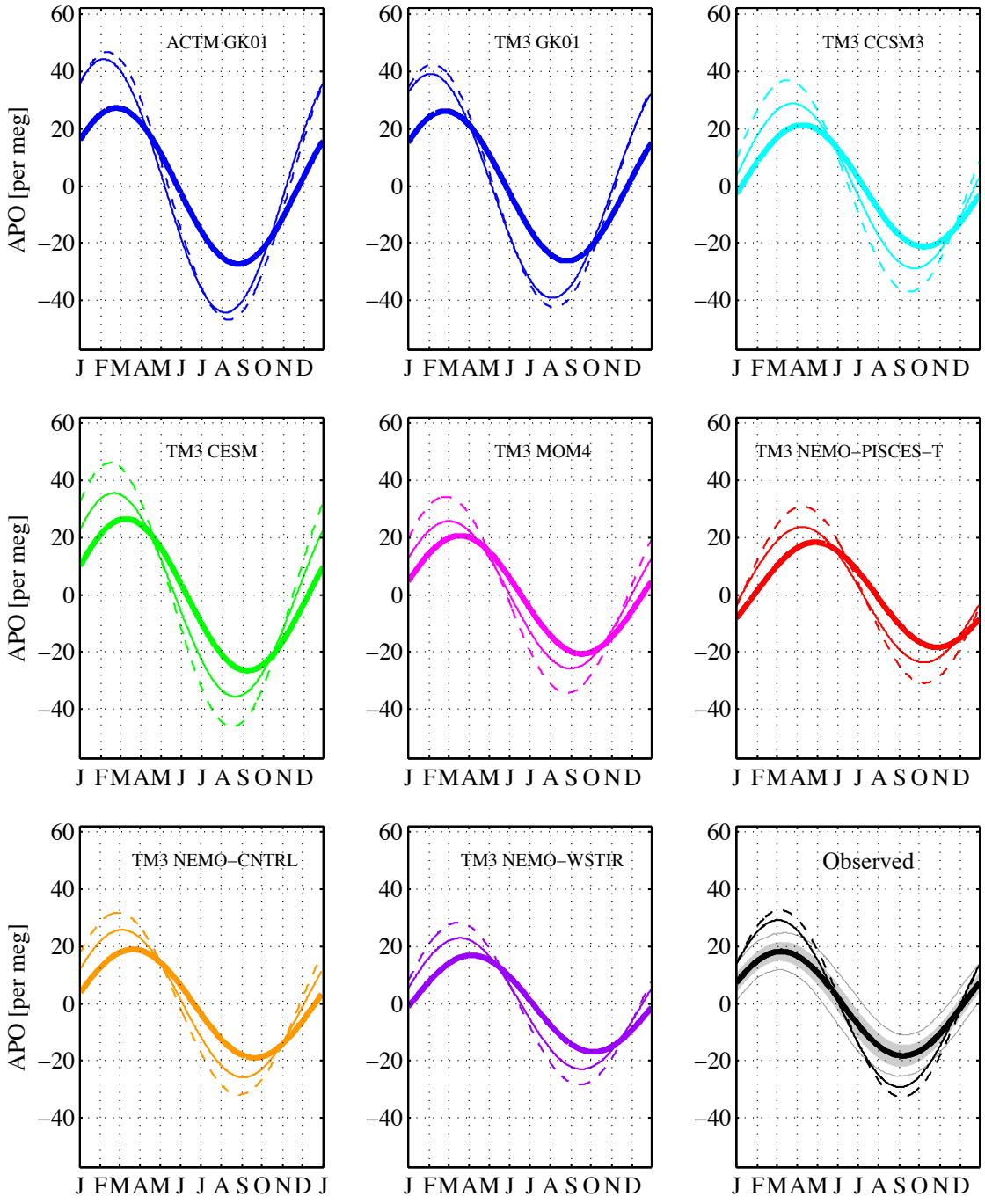

Figure 3.34. PSA (thin dashed), CGO (thin line), and curtain average (thick line) 1-harmonic fits. The 1-harmonic fit to observations in the final panel shows the 1-sigma error bar as a light gray shaded error bar, and 95% confidence limits as light gray thin lines, both computed from 1-harmonic Monte Carlo simulations, as shown above.

Of all the runs, the two dissolved climatologies capture the similarity of the seasonal signal's amplitude at CGO and PSA best. All other models predict a CGO signal whose amplitude is roughly half way between the PSA and CA amplitudes. Interestingly, though the CAs for the two dissolved climatology runs are almost identical in amplitude, the station records are not, with TM3 about 10-15% smaller than the equivalent CGO and PSA records in ACTM, suggesting that TM3 tends to vertically mix the APO signal more than ACTM does (this relationship also holds true when CGO is outputted at its true geographic point). This particular observation is a notable contrast to Stephens et al. [1998], who found that TM3 tended to trap signals more strongly than its predecessor, TM2. A future comparison with TM2 might reveal whether the TM2 model in fact had an even stronger tendency to mix vertically over the Southern Ocean than TM3, and by association ACTM.

ACTM also tends to capture the ratio of the CA amplitude to surface station amplitude best. ACTM predicts a CGO/CA amplitude ratio of 1.62 relative to the observed 1.6, and a PSA/CA amplitude ratio of 1.71 to the observed 1.79. These results are shown in the table below, and pressure/latitude contour plots of each model at the five HIPPO points can be found in appendix Figure A1-08.

Table 3.13. Amplitudes from 1 harmonic fits to seasonal curtain average, CGO (40.7°S, 144.7°E) and PSA (64.8°S, 64.1°W), as determined from model runs and observations. TM3 simulations are output at 44°S, 140°E, for CGO and 63.2°S, 65°W for PSA. ACTM is output at 43°S, 143°E for CGO and 65.6°S, 64.7°W for PSA. Units: [per meg].

Model Run, Data	CA Amp [per meg]	CGO Amp [per meg]	PSA Amp [per meg]	CGO/ CA	PSA/ CA	CGO/ PSA
ACTM GK01	54.4	88.4	93.5	1.62	1.71	0.95
TM3 GK01	52.3	78.2	84.7	1.49	1.62	0.92
TM3 CCSM3	42.5	57.6	73.9	1.36	1.74	0.78
TM3 CESM	53.0	71.0	92.1	1.34	1.74	0.77
TM3 MOM4	41.3	51.4	68.5	1.25	1.66	0.75
TM3 NEMO-PISCES-T	36.6	47.3	61.6	1.29	1.68	0.77
NEMO-CNTRL	37.8	51.5	63.4	1.36	1.68	0.81
NEMO-WSTIR	33.8	45.7	56.5	1.36	1.67	0.81
Observed	36.4±5.6	58.2	65.1	1.60	1.79	0.89

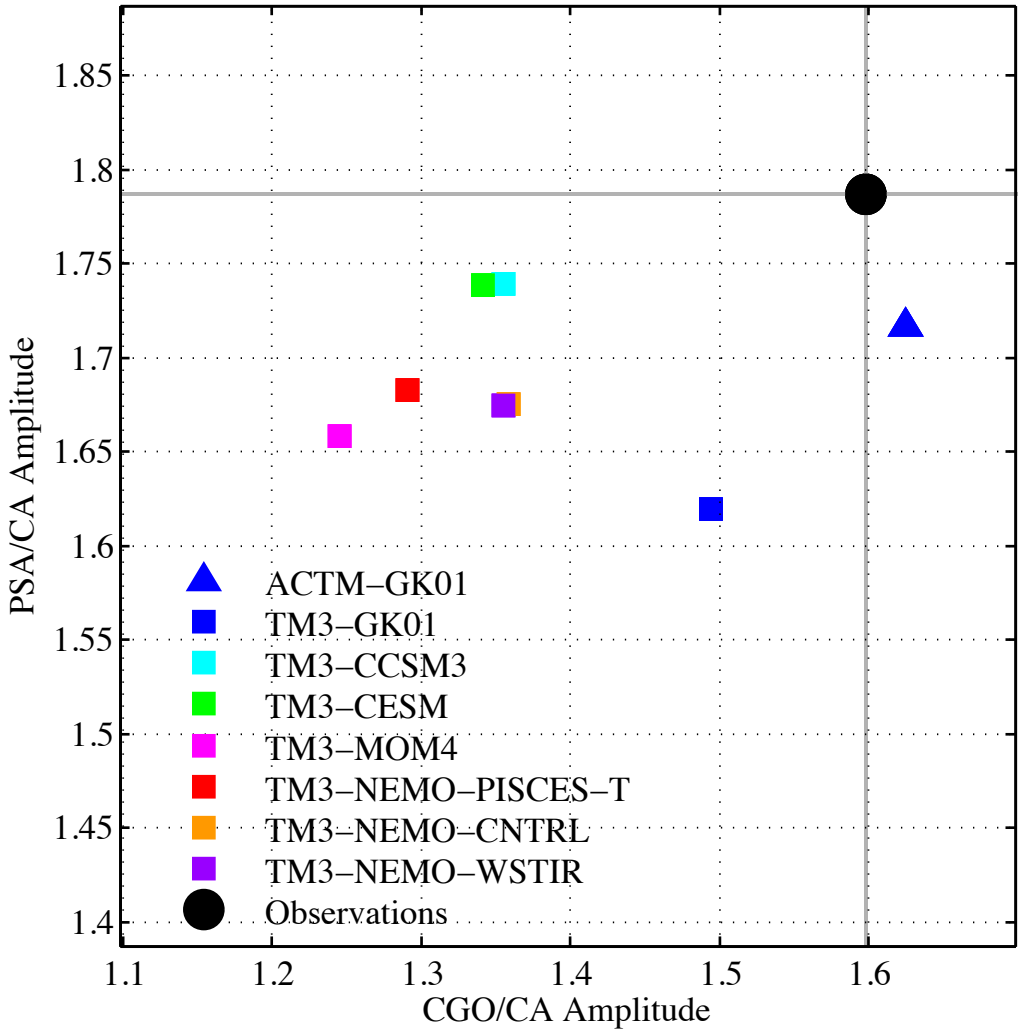

Figure 3.35. 1-harmonic CGO/CA amplitude ratio plotted against PSA/CA amplitude ratio for each of the eight model combinations and observations, with the observed relationships shown centrally in black, with gray lines of "constant ratio" extending up/down and left/right.

Figure 3.35 shows that all models underestimate the CGO/CA relationship relative to observations except for the ACTM-GK01 run. This suggests that vertical mixing at this latitude (~41°S) in TM3 may be too strong, mixing into the vertical column a signal that should be more trapped. The same holds true for all PSA/CA values, suggesting that both models may be slightly overestimating vertical transport at this latitude. For now, I leave

further observations of these features for future studies, as the complexity of this problem is beyond the scope of this chapter.

3.10 Further analysis of NEMO model runs

Here I provide further analysis of differences between the NEMO-CNTRL and NEMO-WSTIR results against APO observations in light of the station record comparisons in Section 3.9. It is worth considering the Rodgers et al. [2014] methodology in slightly more detail first, however. Rodgers et al. interpolate CGO and PSA data and model output to weekly resolution using the seasonal trend decomposition by Loess (STL) fitting algorithm of Cleveland et al. [1990] and compare these results in Figure 9, panels e and f, of their paper. In 9f, the WSTIR run performs better than the CNTRL run, capturing the amplitude and phasing of the seasonal cycle better, while in 9e, the CNTRL run performs better in comparison to CGO, capturing the phase and amplitude of the seasonal cycle better. **Figure 3.36** shows the performance of the two models at PSA, CGO and over the curtain average compared against my observations and fitting techniques. I do this to see whether my methods confirm the findings of the authors: that the WSTIR wind stirring parameterization improves agreement of models with observed APO at Palmer Station. Here again, all fits are 1-harmonic for consistency with the previous plot, but now for consistency with Rodgers et al., PSA and CGO station fits are performed over detrended yearly output between 1999 and 2004.

In order to investigate whether the choice of grid point for the station output has a considerable influence on the predicted phase and amplitude, I plot each model

simulation at its interpolated geographical point (solid line) and at the closest grid point by the methods of the Rodgers et al. [2014] paper (dashed line). In station and curtain average cases, the phasing of the NEMO-CNTRL seasonal cycle is earlier than the equivalent WSTIR cycle because the wind stirring parameterization delays the onset of stratification in the Southern Ocean, in agreement with Rodgers et al. [2014], whose Figure 9 panels e and f shows the wind stirring parameterization delays and dampens the seasonal APO cycles at PSA and CGO.

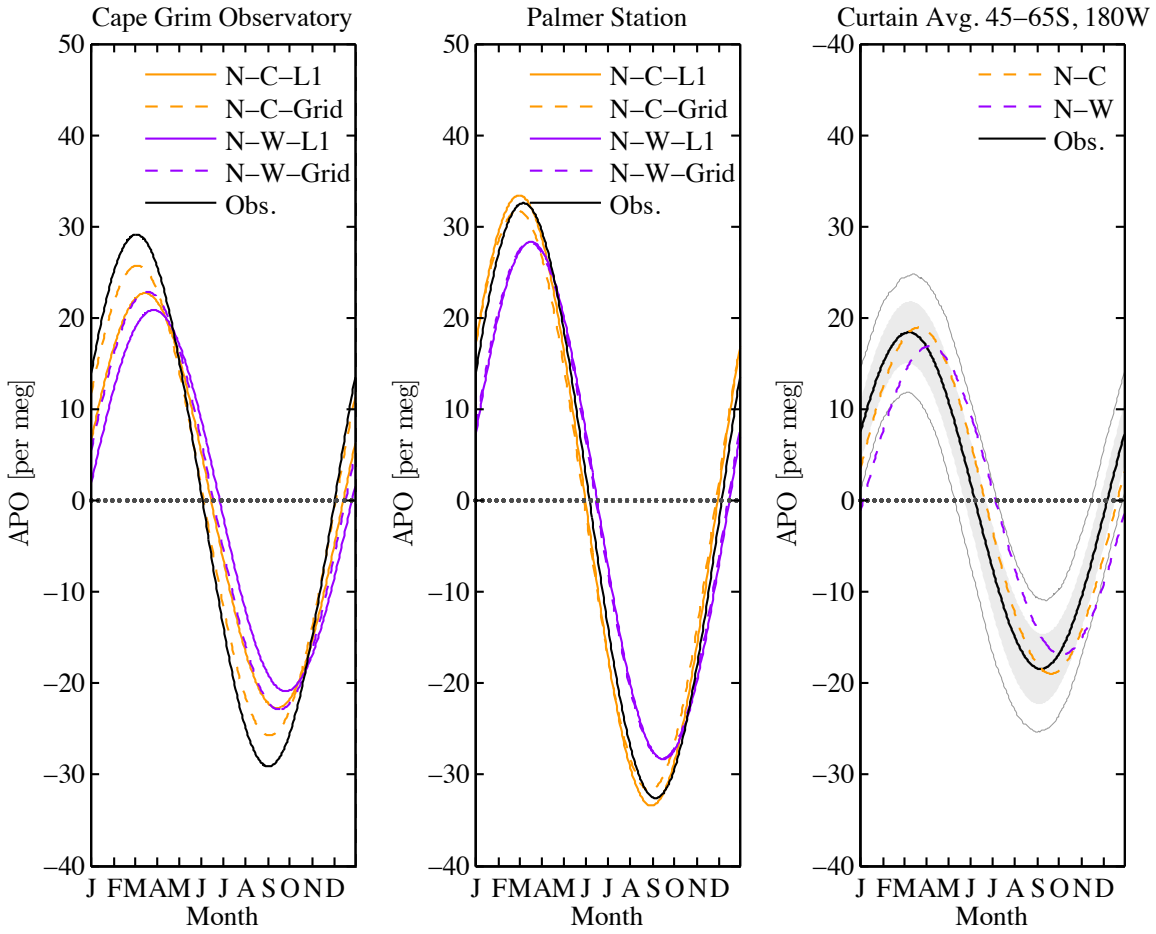

Figure 3.36. A comparison of NEMO-CTNRL and NEMO-WSTIR runs with station data and the curtain average. To investigate whether the choice of grid point is responsible for discrepancies between Rodgers et al.'s [2014] and this study's findings at surface stations, I have outputted both models at their lowest level, interpolated to their geographical location (solid line, "L1" in legends) and at their closest grid box (dashed line, "Grid" in legends), as per Rodgers et al.'s protocol discussed in the previous section. In all cases considered, NEMO-CNTRL appears to be the better fit. Units: [per meg].

Table 3.14. 1-harmonic fit amplitude of HIPPO, NEMO-CNTRL and NEMO-WSTIR observations/output for curtain average, CGO (at grid point), CGO (at station location), PSA (at grid point), PSA (at station location). For model output "at station location" refers to the value interpolated to the location of the station from gridded output. In each case, the observed amplitude from HIPPO or station is in bold and italics. The closest match to this value from the four NEMO-CNTRL and NEMO-WSTIR simulations is noted in bold. In all three cases, NEMO-CNTRL comes closer to observations. I don't have a direct measure of station first harmonic amplitude uncertainty, but they are likely to be very small (on order 1-2 per meg) given the large number of years used to evaluate first harmonic timing. Units: [per meg].

	Curtain Average Amplitude (per meg)	CGO Gridded Amplitude (per meg)	CGO Station Amplitude (per meg)	PSA Gridded Amplitude (per meg)	PSA Station Amplitude (per meg)
Observations	***36.4 ±5.6***	*n/a*	***58.2***	*n/a*	***65.1***
NEMO-CNTRL	**37.9**	**51.5**	45.4	63.5	**66.8**
NEMO-WSTIR	33.8	45.8	41.7	56.5	56.6

Table 3.15. 1-harmonic fit trough timing (year days) of HIPPO, NEMO-CNTRL and NEMO-WSTIR observations/output for curtain average, CGO (at grid point), CGO (at station location), PSA (at grid point), PSA (at station location). For model output "at station location" refers to the value interpolated to the location of the station from gridded output. In each case, the observed amplitude from HIPPO or station is in bold and italics. The closest match to this value from the four NEMO-CNTRL and NEMO-WSTIR simulations is noted in bold. In all three cases, NEMO-CNTRL comes closer to observations. I don't have a direct measure of station first harmonic peak and trough uncertainty, but they are likely to be very small (on order a day or two) given the large number of years used to evaluate first harmonic timing. Units: [per meg].

	Curtain Average Trough YD	CGO Gridded Trough YD	CGO Station Trough YD	PSA Gridded Trough YD	PSA Station Trough YD
Observations	***250 ±10.8***	*n/a*	***246***	*n/a*	***249***
NEMO-CNTRL	**264**	**247**	258	241	**243**
NEMO-WSTIR	278	261	269	258	260

However, this adjustment increases the phase discrepancy with the curtain average APO cycle as measured by the 1-harmonic fit. The amplitude and trough year day calculated from 1-harmonic fits to observations is noted in italicized bold font in

Table 3.14 and **Table 3.15**. For each of the six metrics (amplitude and trough timing for the curtain average, CGO and PSA), the closest value from NEMO output is marked in bold. In all six cases, this value comes from NEMO-CNTRL output, suggesting that the CNTRL run performs better at Cape Grim and Palmer Station, and also over a larger slice of the atmosphere near the dateline. It bears noting that some of the NEMO-WSTIR values do fall within the error bars of the data, including the curtain average amplitude, suggesting that by some criteria, it is possible the WSTIR run is performing better, and that the uncertainty in my measurements simply obscures this.

Some of the disagreement between this study and Rodgers et al. [2014]—over whether WSTIR improves agreement with observations' phase and amplitude at PSA—stems from my decision to fit all tracers with one harmonic to minimize phasing and amplitude errors associated with fitting five sparse data points. As mentioned above, Rodgers et al. [2014] compare APO output and data using the STL method, which preserves seasonal asymmetry more than harmonic fits. My approach admittedly simplifies the comparison, looking for only the phasing and amplitude of the fundamental frequency, however, but I have applied it using the same fitting method for data and output.

To shed further light on the question of whether WSTIR improves agreement with output at PSA, I show a comparison of 1999-2004 CGO and PSA observations alongside the two NEMO models' daily output for the same years. This comparison, seen in **Figure 3.37**, gives a sense for the timing on a year-to-year basis that fits to data slightly obscure. I show the comparison of station observations and model output as produced at the grid point to the Southwest of CGO (top) and PSA (bottom), according to the protocol of the

Rodgers et al. [2014] paper, and for consistency, model output and station data are detrended by a linear fit to the interannual component of their respective South Pole records.

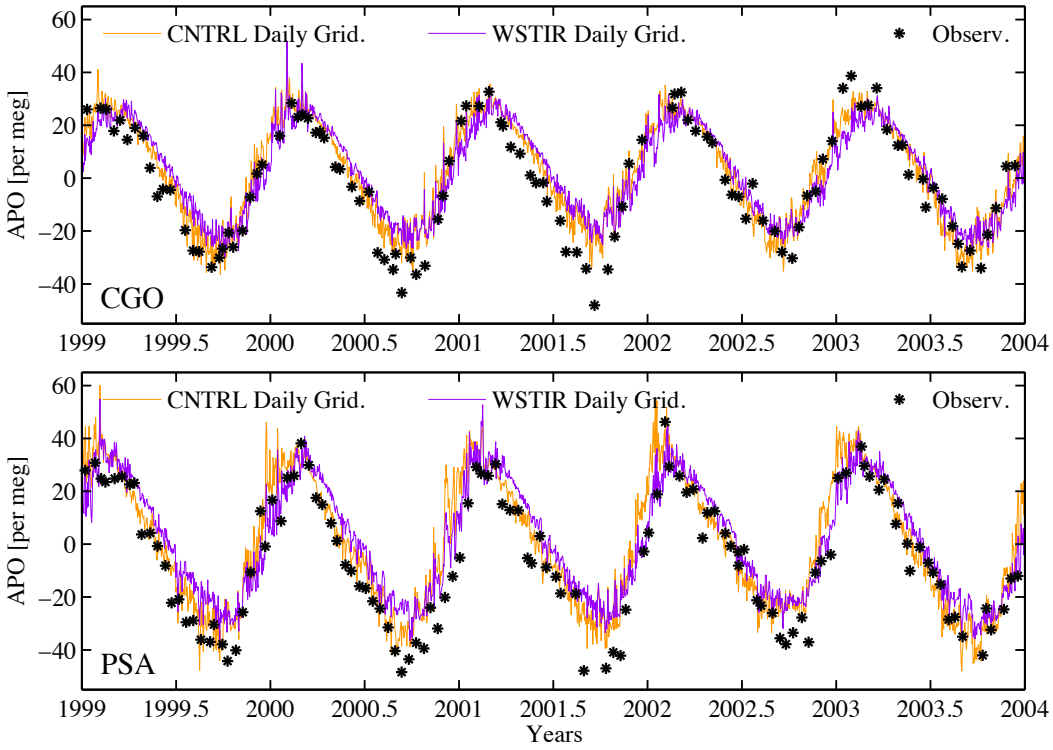

Figure 3.37. Comparison of triplicate flask mean APO observations (black asterisks), NEMO-CNTRL (orange), and NEMO-WSTIR (purple) between 1999 and 2004. These are the five years that go into calculating the mean seasonal cycle for model output. Units: [per meg].

This comparison shows that the timing of the Cape Grim peak is closer in the CNTRL run (orange), but that the timing of the PSA peak seems closer in the WSTIR run (purple). In both cases, however, CNTRL captures the seasonal amplitude more closely. As different fitting techniques emphasize different features in the data, the ranking of the model performance may depend on fitting technique. Rogers et al.'s [2014] conclusion that NEMO-WSTIR is the better performer based on Loess fitting and our conclusion that

NEMO-CNTRL performs better based on one-harmonic fits may simply be emphasizing different features in the data.

In terms of real-world phenomenology and how the Rodgers et al. [2014] runs have captured these, it is worth considering these final few points. Rodgers et al. have seen differences in the robustness of the WSTIR run match against observations for different regions. The closer agreement of WSTIR at PSA data may reflect the close proximity of the Antarctic divergence where oxygen-depleted Circumpolar Deep Water (CDW) sustains some of the sharpest upper-ocean vertical gradients in O_2 found over the Southern Ocean, and the seasonal cycle of entrainment and detrainment may be expected to be of critical importance. Nearer to the subtropical front, closer to the equatorward end of the curtain average, subduction is active, and vertical gradients in O_2 are weaker over the planetary boundary layer, such that the physical controls in surface conditions through solubility should be expected to find more pronounced expression. Given these differences, the application of the wind-stirring parameterization may be expected to impact the seasonal cycle in O_2 fluxes differently in these regions. The heterogeneity in the mechanisms controlling the seasonal cycle in O_2 fluxes is interesting but complicated [*K. Rodgers*, pers. comm.]. The great virtue of the curtain average method is that it provides an overarching framework of large-scale seasonality variability that will be critical in attribution of mechanisms over the diverse biogeochemical regimes of the Southern Ocean.

3.11 Conclusions and further work

I have collected measurements of atmospheric potential oxygen at five points in the seasonal cycle over a vertical slice above the Southern Ocean near the dateline, whose weighted average value I refer to as a "curtain average," and I have calculated the seasonal amplitude and timing of the first and second harmonic fits to this signal. I calculate the curtain average as a means toward quantifying the seasonal APO flux into and out of the Southern Ocean, but also as a metric for evaluating how well ocean models are capturing this seasonal cycle, so I compare HIPPO observations with output from two transport models and seven sets of ocean flux estimates.

To assess the validity of this method, I have first shown that the curtain average reduces disagreement between the transport models using the same ocean model boundary conditions by comparing CAs of TransCom model output against station records of the same. In order to show that the five points in time are sufficient to constrain the seasonal cycle, I have selected five points at random from model output over many iterations and shown that a scenario in which typical synoptic-scale variation is preserved captures the phase of the cycle to within ±4.1 days and ±2.7 per meg for the 1-harmonic fit and to within ±6.3 days (1σ), and the amplitude to within ±4.2 per meg for the second harmonic fit.

In order to assess bias associated with the interpolation method, I have compared gridded ACTM model output with output averaged only along the flight tracks for areas actually covered by the data, and with a limited interpolation and extrapolation scheme. In order to assess bias associated with synoptic-scale variability, I have compared curtain averages performed on daily-resolution output to the seasonally fitted value for those

days. For the most representative data-based APO seasonal fluxes, I have chosen to adjust the timing and amplitude of the Garcia and Keeling O_2 fluxes, because comparison with curtain averages and station data, and recent studies, suggest these anticipate the phasing of observations by about two weeks, and a recently revised gas-exchange parameter suggests the amplitude is too large.

I have also shown that interannual variability in station measurements, which is a combination of interannual flux variability and interannual meteorological variability, is reasonably low, and have incorporated this value into the uncertainty estimate associated with the observations. Furthermore, I have shown that potential effects from nearby volcanic sources are statistically negligible and unnoticeable at sampling stations, thus unlikely to affect the curtain average values at all.

With these adjustments for temporal and spatial sampling biases, as estimated with model output and observations at surface stations, I have determined that my data are sufficient to constrain the first harmonic of the seasonal cycle of APO over this portion of the Southern Ocean to within a little over a week, and the second harmonic to within a month. This allows me to evaluate models and dissolved climatologies without the vertical mixing uncertainty introduced by the transport models used to circulate these ocean signals through the atmosphere. I find that a 1-harmonic fit to the seasonal cycle of the meridionally integrated APO over the Southern Ocean slice (65 to 45°S, along 180W, between the surface and 300 mb), has an amplitude of 36.4 ±5.6 per meg, with peak and trough at YD 68 and 250, each with an uncertainty of ±10.8 days. I find that a 2-harmonic fit to the same has a mean seasonal amplitude of 37.8 ±5.9 per meg, with seasonal peak and trough at YD52 ±27 days and YD260 ±20 days. This suggests a strongly

asymmetrical seasonal cycle with 208 ±36 days falling between the peak and subsequent trough, roughly 57 ±10% of the year, though this last number is of limited interpretational value because of the large error bar associated with the 2-harmonic fit used to determine it. I conclude that the high uncertainties in the 2-harmonic fit are the result of the added degree of freedom that allows the second harmonic component to shift peak and trough timing independently, producing sometimes unrealistic seasonal cycles when a Monte Carlo method is used to estimate error. Though these outlier Monte Carlo fits are largely implausible given our understanding of station records, I cannot rule them out. Accordingly, the interpretive power of the 1-harmonic fits is stronger, and the interpretive value of the 2-harmonic seasonal asymmetry measurement (trough day-peak day) is less so.

Given these conclusions, I limit the majority of my interpretation to 1-harmonic fits. I find 1-harmonic fits to dissolved climatology estimates of Southern Ocean seasonal APO tend to anticipate the observed cycle at its trough by 12-13 days and overestimate its amplitude by 43% (TM3) and 49% (ACTM). The CESM ocean model also appears to overestimate the seasonal cycle by 45%. NEMO-PISCES-T is closest on amplitude, underestimating it by only 1%, but gets the phasing late by well over a month (51 days), a feature that may reflect on its ability to accurately capture the secular trend in anthropogenic CO_2 uptake. NEMO-CNTRL is a close second for amplitude (+3.9%). NEMO-CNTRL and MOM4 are the closest models to the observed CA APO cycle in phase and shape when all six 1- and 2-harmonic metrics are considered in concert.

Interestingly, the two generations of the NCAR models: CCSM3 and CESM disagree strongly with each other on phase and amplitude. In phase, CESM is close to

observations (-1 day), while CCSM3 is approximately a month late (-32 days), while CCSM3 is closer on amplitude (+17%), and CESM overestimates greatly (+45%). Though the implication of this might initially be that the biogeochemistry is the root cause of the change, the large modifications to the POP2 physical oceanography are a significant potential cause, possibly implicating changes to overturning/ventilation dynamics, and mixed layer growth.

I note also that NEMO-CNTRL and NEMO-WSTIR runs are quite different, as might be expected from observations in the Rodgers et al. [2014] paper. I find that for the curtain average, and for both CGO and PSA, the NEMO-CNTRL run is closer to the observations than NEMO-WSTIR in both amplitude and phasing of the first harmonic, and that the backwards phase shift (delay) brought on by the introduction of the wind-stirring parameter has brought NEMO-WSTIR further from, not closer to observations' timing when run in conjunction with TM3. Though some of the offset between the observed and modeled curtain average may be due to error in, or to the representativeness of, the airborne data, the agreement of the NEMO-CNTRL output at Cape Grim and Palmer Station seem to confirm that the CNTRL run represents the seasonal APO cycle better for amplitude and phase, at least given the fitting methods of this study. Though for the curtain average, it seems clear that the CNTRL run is the better performer, further investigation of the bias introduced by fitting methods is needed to resolve this issue at surface stations. Despite these findings, it is important to remember that neither the CNTRL or WSTIR run were intended by the authors to be representations of state estimates, but of sensitivity to changes to mixed layer dynamics.

I also conclude that Garcia and Keeling [2001] O_2 fields require several improvements/adjustments. First, the asymmetry of the APO curtain averages generated by Garcia and Keeling O_2 fluxes and Blaine's N_2 fluxes cannot be significantly changed simply by adjusting the timing of the Blaine [2005] N_2 fluxes, though a small increase in the timing difference between trough and peak (from 193 to 197 days) can be achieved by delaying the O_2 fields. This adjustment tends to bring the APO cycle's temporal "asymmetry" into slightly better agreement with other models and the 2-harmonic fit to observations, however the error bar on the observations' seasonal "asymmetry" is too large to know whether this is truly a feature of the data

I conclude that, of four phasing scenarios run in the ACTM model using the Garcia and Keeling O_2 fluxes, a 2-week delay (15 days) brings the 1-harmonic fit to ACTM GK01 output closest to the phasing of the observations. A slightly smaller adjustment of 12-13 days, as noted from comparisons of the phasing of the 1-harmonic fits to ACTM GK01 and the HIPPO curtain average, is probably the optimal adjustment for the O_2 fields, however due to server time limitations, I am presently unable to confirm this number.

I also conclude that rescaling the Garcia and Keeling [2001] fluxes based on the re-evaluated global gas exchange scaling parameter of Naegler et al. [2006], 0.32, substantially improves the amplitude of the simulated 1-harmonic APO seasonal cycle, though our, admittedly limited, observations suggest an even smaller a_q value of 0.276 ±0.045. The Naegler et al. [2006] scaling factor does just fall within our uncertainty estimate, while the original Wanninkhof [1992] value of 0.39 does not. I suggest that, in the Garcia and Keeling [2001] paper itself, initial amplitude agreements at surface

stations between data and GK01 fluxes may have been due to overly-strong vertical mixing in the TM2 atmospheric transport model used by the authors.

In sum, I conclude that a thorough re-evaluation of the Garcia and Keeling [2001] fluxes is necessary to 1) incorporate new gas-exchange estimates as suggested by Naegler et al. [2006] and this study, 2) improve the phasing of the fields, and 3) to evaluate whether the preliminary observations of this study are correct that the GK01 fields are overly symmetrical with respect to time. The addition of 13 years'-worth of subsequent sea-surface O_2 samples is promising in this regard.

I also conclude that the amplitude of the first harmonic fit to the seasonal curtain average, as I have described it (to 300mb, roughly the tropopause), constitutes approximately 60% of the seasonal cycle observed at surface stations at either edge of its latitudinal extent (63% for CGO and 56% for PSA observations). Few model runs capture both of these relationships closely, with the ACTM-GK01 closest at 60.8% and 57.3%, respectively. TM3 model runs typically perform less well, slightly underestimating the amplitude of PSA relative to the CA, and significantly underestimating CGO relative to the CA, perhaps due to vertical mixing problems at the latitude of CGO.

This study is the first to introduce the curtain average metric. A great deal of additional work could be done to investigate the accuracy of transport models when run with ocean fluxes that appear to reproduce observed seasonal CA amplitudes well. Additional work could also look at vertical and horizontal gradients, and station-station gradients, as HIPPO output provides the first opportunity to measure these values simultaneously alongside station measurements, with a single sampling system. Finally, curtain average work need not be limited to the Southern Ocean. This study could be

replicated in other zones to assess whether models that perform well over the Southern Ocean continue to do so elsewhere.

4 Seasonality, and meridional and vertical gradients of atmospheric Ar/N_2

4.1 Introduction

Air from the MEDUSA whole-air sampler was analyzed for Ar/N_2 ratios both for practical purposes—to aid in identifying possible O_2/N_2 ratio fractionation—and for scientific purposes—to examine the value of Ar/N_2 in representing environmental processes. This chapter presents a early look at the Ar/N_2 observations from the HIPPO Global campaign, and includes a small literature review in order to place the measurements in scientific perspective.

Atmospheric argon constitutes about 0.9% of the dry atmosphere. Changes in its concentration in the atmosphere on a seasonal basis are exceedingly small. Against a relatively large background concentration of Ar, small changes are hard to measure, so a ratiometric mass spectrometer method [*R. F. Keeling et al.*, 2004] is used to detect very small variations in the ratio of Ar to molecular nitrogen, since N_2 (~78% of dry atmosphere) is a relatively stable background against which to compare Ar values. Using this technique, atmospheric Ar/N_2 ratio changes are reported as deviations of one part in a million relative to a known reference gas, by the same equation used for O_2/N_2 in Chapter 2 (see Eq. 2.1). Accordingly, measurements in this chapter are expressed as $\delta(Ar/N_2)$, with units of per meg.

Studies over the last decade suggest that the atmospheric Ar/N_2 ratio has potential as an inert tracer of air-sea heat exchange and atmospheric mixing [*Battle et al.*, 2003; *Blaine et al.*, 2006; *Cassar et al.*, 2008; *Ishidoya et al.*, 2013; *R. F. Keeling et al.*, 2004].

Because argon is a noble gas, its abundance in the atmosphere is unaffected by biological and chemical processes. N_2 is also very inert and its atmospheric burden is relatively stable since sources and sinks are small relative to its natural abundance. Atmospheric Ar and N_2 abundances are controlled primarily, then, by physical processes such as solubility-driven air-sea exchange. Both gases are more soluble in colder water, though the effect is more pronounced for Ar [*Weiss*, 1970], such that the Ar/N_2 ratio of the overlying atmosphere changes as the gases are taken up or given off by the ocean in response to heat flux: ocean heating will cause the Ar/N_2 ratio in the atmosphere to increase, while ocean cooling will cause it to decrease [*Battle et al.*, 2003; *Blaine*, 2005; *Craig and Wiens*, 1996; *R. F. Keeling et al.*, 2004; *Weiss*, 1970]. Atmospheric mixing will cause the changes induced near the ocean surface to be dispersed throughout the atmosphere.

Measurements of the atmospheric Ar/N_2 ratio at the Earth's surface began around 2002 at nine monitoring stations in the Scripps O_2 Program—Alert, Canada; Cold Bay, Alaska; Cape Grim, Tasmania; Cape Kumukahi and Mauna Loa, Hawaii; La Jolla, California; Palmer Station and South Pole, Antarctica; and American Samoa—and back to 2011 for Barrow, Alaska. These records document seasonal cycles in the atmospheric Ar/N_2 ratio that are strongest at high latitudes, and smaller, but detectable in the tropics [*Battle et al.*, 2003; *R. F. Keeling et al.*, 2004]. These same studies have also shown that these stations experience mean offsets in Ar/N_2 based on their latitude, though the degree to which this offset is real is uncertain—sampler-sampler offsets may be responsible for some of the observed station-station differences.

Further studies also document variations in Ar/N_2 ratio with elevation.

Adachi et al. [2006] observed an enrichment of the ratio under night-time inversion conditions over the Anza Borrego desert which they attributed to a combination of gravimetric and thermal diffusive fractionation. Ishidoya et al. [2013] have documented a clear diminution of the ratio from flasks measured in the stratosphere over Japan, which they attribute to similar effects.

Many factors complicate the sampling, analysis and interpretation of the Ar/N_2 ratio. Measurement reproducibility is relatively poor because small temperature gradients in sampling and analysis apparatus can cause relatively large thermal fractionation of the gas at inlets, flask outlets and T-junctions [*Blaine et al.*, 2006]. Samples collected at ground stations may be subject to station-dependent sampling biases, which can interfere with resolving meridional and zonal gradients [*Battle et al.*, 2003]. Furthermore, the various laboratories which measure the ratio use different reference tanks, and reference scales, hampering intercomparison.

One particularly poorly understood aspect of the geographic variation in the Ar/N_2 signal is its vertical gradient within the free troposphere, which measurements in Chapter 2 of this dissertation suggest to be approximately 2 per meg/km. Tropospheric vertical gradients have been observed in multiple prior airborne measurement campaigns. Steinbach [2010] observed a 30 per meg depletion of the Ar/N_2 ratio at 4 km (7.5 per meg/km) relative to surface values over the Amazon basin during the BARCA experiment, but interpreted this as an inlet artifact. Steinbach suggested that rear-facing inlets used in her study preferentially sample lighter molecules (e.g. N_2, M28) over heavier molecules (e.g. Ar, M40) when the velocity of the sample air passing the inlet is increased (the opposite being true for forward-facing inlets). Since aircraft velocity and

altitude often correlate, the author concluded that increased air speed of the aircraft used for the BARCA study was responsible for the observed depletion at altitude, and cited other recent studies [*Langenfelds*, 2002] that appear to support the author's hypothesis. As only a few research groups measure the atmospheric Ar/N_2 ratio (and do so with various samplers and inlet types), and because aircraft time is limited, studies have yet to be able to rule out the possible presence of a very small gradient in the troposphere.

The HIPPO campaign yielded ~1600 samples, collected by the GV aircraft between 67˚S and 87˚N, from the planetary boundary layer to the lower stratosphere along the middle of the Pacific Ocean. These measurements provide the first opportunity to evaluate large-scale geographical trends and gradients in Ar/N_2 without potential station-dependent sampling biases. The large number of samples makes it possible to statistically resolve small signals that might otherwise not be detectible due to imprecision in the Ar/N_2 measurements.

Chapter 2 outlines corrections to HIPPO O_2/N_2 samples using the observed scatter in the Ar/N_2 signal. This process required quantifying certain natural signals in the Ar/N_2 data so as not to correct them out of the O_2/N_2 data erroneously. I found that the data had a distinct Ar/N_2 depletion signal in the lower stratosphere which correlated well with the depletion of N_2O, and (less closely) with elevation. I found additionally that a gradient exists in the tropospheric data, of -2.21 per meg/km, which I also preserved in the Ar/N_2 data before using Ar/N_2 to correct O_2/N_2 for use in Chapter 3 because it had little or no effect on the ultimate value of the curtain average. In this chapter, I provide a more thorough examination of these gradients, and further investigate the HIPPO Ar/N_2 data

for evidence of seasonal cycles and interhemispheric gradients. In both cases, statistically significant signals appear to be present.

4.2 Methods

Data in this chapter come from two principle sources: the HIPPO MEDUSA dataset, which uses 1.5-liter “B” flasks (seen in **Figure 2.5**), and the Scripps O_2 Program sampling network, which uses 5-liter spherical “round-bottom” flasks [*R. F. Keeling et al.*, 1998b]. I described HIPPO in the introduction to this dissertation, and MEDUSA sampling and analysis methods in Chapter 2. MEDUSA methods for this chapter are no different than those already discussed, and again I consider only Flask Box 2 data, as the scatter in these measurements from sampling artifact is substantially lower than the scatter in Box 1 data.

Scripps O_2 Program sampling methods used for sampling at surface stations are slightly different than those used on HIPPO. Flasks for surface sampling are prepared by initially filling them with dry working tank air. They are then shipped to sampling stations where they are sampled during suitable wind conditions. Sample air is cryogenically cooled by passing it through a trap in a chiller bath. The resulting, dried whole air sample is shipped back to the Scripps Institution of Oceanography, and analyzed in the same lab, but on a different rack than the MEDUSA flasks. Nearly all surface samples are collected in 5-liter round-bottom glass flasks, although "B" flasks have also been collected in parallel with round-bottom flasks at La Jolla for many years.

During round-bottom sample preparation, a small magnetic stir bar is dropped into the flask stem, which is then purged with a tank of background air. When the flask

valve is opened, the magnet is lowered magnetically into the flask. During analysis, the magnet is activated by a stir base beneath it, spinning inside the flask to continuously mix the air to produce the most reproducible "sweep-out" curve [*R. F. Keeling et al.*, 1998b]. Round-bottom flasks are typically sampled in triplicate, and the representative value for a given date is calculated as the mean of these values. If, however, any of the flasks is broken, or deemed to be contaminated by the analysis software, this value is removed, and a mean value is calculated from the remaining flask or flasks. Though round-bottom data for Ar/N_2 extend back to 2002 for many stations, I choose to consider only data taken after 2006 to determine mean conditions because these data were taken with an aspirated inlet to minimize possible inlet fractionation [*Blaine et al.*, 2006].

Though round-bottom flasks are run on a different rack than the MEDUSA flasks, a portion of the round-bottom sample air is directed to the same ISOPRIME mass spectrometer on which MEDUSA flask Ar/N_2 and O_2/N_2 samples are analyzed. Some small difference between the two flask systems is to be expected due to a T-junction in the plumbing that brings calibration gases to the mass spectrometer for the B flask rack.

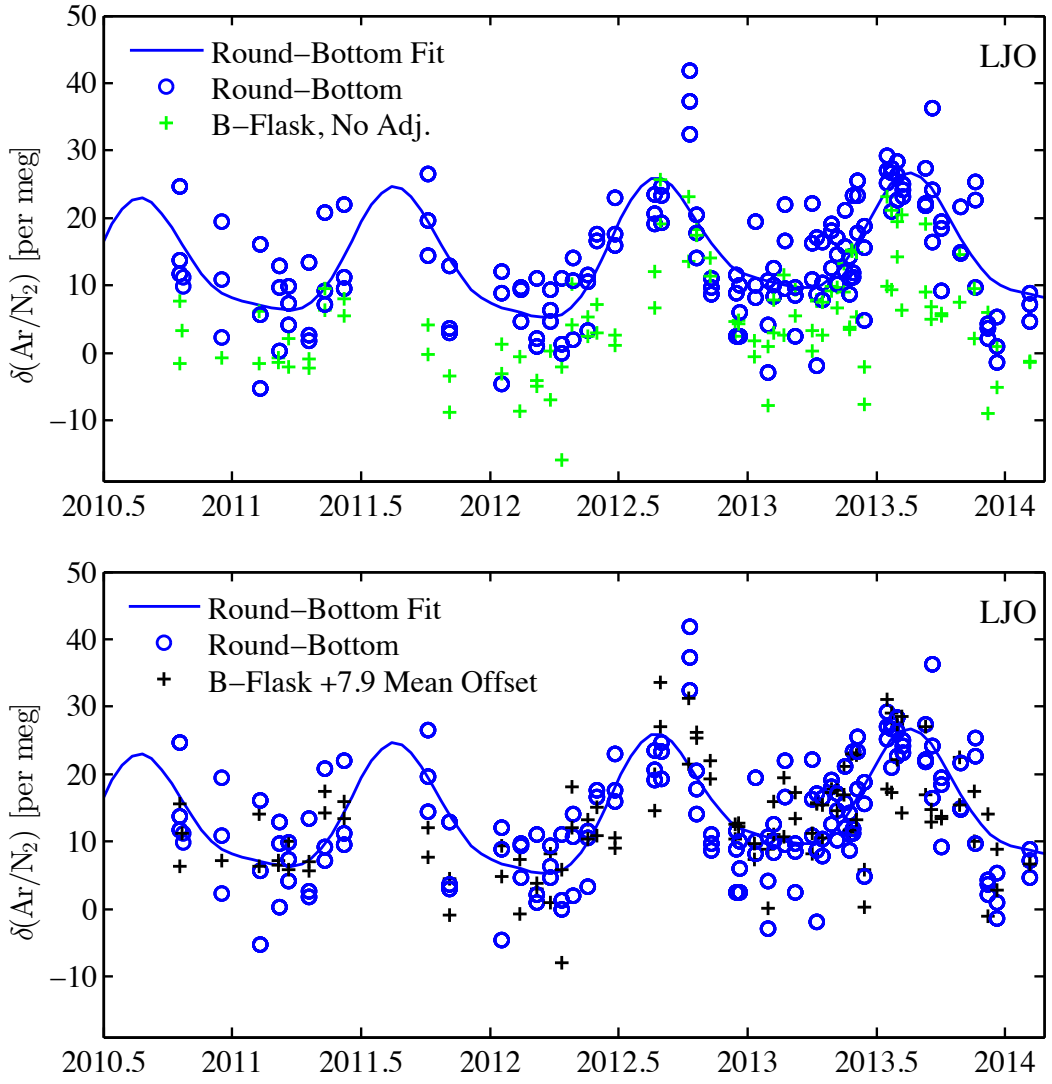

Figure 4.1. Comparison of round-bottom (blue circles) and B (crosses) flask $\delta(Ar/N_2)$, all concurrently sampled at La Jolla, CA. Unadjusted B-flasks (green) in the upper plot are low relative to round bottoms. An adjustment of +7.9 per meg (bottom panel, black) based on the mean offset (RB-B) of the concurrently-sampled flask types, brings them into agreement with round-bottoms. In both panels, a 2-harmonic fit with a stiff spline (blue line) shows the seasonal cycle of the round bottom flasks more clearly.

A comparison of La Jolla round bottom and B-flasks suggests that a fairly consistent offset (round bottom minus B-flask) of +7.9 per meg exists. Because this offset is likely due to diffusive fractionation at a plumbing T junction, this implies that a smaller offset of +2.4 per meg also exists between the round bottom and B-flask O_2/N_2

measurements (using the fractionation scaling relationship of 3.77 from Keeling [2004]), although this is not relevant in the present context.

Importantly, the Ar/N_2 offset observed above is representative of only the offset between flask types using the same sampling and analysis apparatus, but the offset of HIPPO Ar/N_2 data from station round bottoms may be greater or less than this value, because HIPPO flasks were captured by the MEDUSA sampler. I explore this offset further below in Section 4.5.

4.3 Exploring the nature of the tropospheric vertical Ar/N_2 gradient

As noted in the introduction, there are relatively few studies of the vertical structure of Ar/N_2 in the atmosphere [*Adachi et al.*, 2006; *Ishidoya et al.*, 2008; *Ishidoya et al.*, 2013; *Langenfelds*, 2002; *Steinbach*, 2010]. Recent studies by Ishidoya et al. [2013] have focused on quantifying the diminution of the Ar/N_2 ratio by diffusive separation in the stratosphere over Japan, while Adachi et al. [2006] studied the enrichment of the ratio under very specific inversion conditions close to the earth's surface. Langenfelds [2002] observed depletion on order 100-300 per meg of the Ar/N_2 ratio between the surface and 7 km over the Southern Ocean during several research flights, but regarded the gradient as a sign of inlet artifacts alone. He used the Ar/N_2 data variously to disqualify other tracer data, or to correct O_2/N_2 data for inlet effects. Because the goal of Langenfelds's work was not the measurement of the Ar/N_2 gradient, and because conclusions about this data are limited by frequent changes in the configuration of the system's sample inlet, the study sheds little light on the nature or magnitude of the tropospheric Ar/N_2 gradient.

Steinbach's [2010] graduate thesis affords one of the more concentrated efforts to understand vertical gradients in the Ar/N_2 ratio. Steinbach focused largely on profiles of O_2/N_2 over the Amazon jungle during the BARCA experiment. Her analysis of Ar/N_2 gradients was also used to determine the extent of fractionation effects on O_2/N_2 data. Again, the apparent fractionation of O_2/N_2 was determined from Ar/N_2 data, and scaled down using the $Ar:O_2$ fractionation scaling factor of Keeling [2004]. However, in an effort to justify removing vertical Ar/N_2 gradients, Steinbach ran a series of laboratory experiments and modeling simulations to explain the observed fractionation (vertical gradient) mechanistically. Her choice to remove the vertical gradient (on order -30 per meg over 4km) was based on a laboratory experiment in which she exposed an Oxzilla O_2 sensor to a characterized reference gas, and varied the flow rate around the inlet between 19 and 43 m/s (a scaled down representation of aircraft velocities). The slow flow, meant to simulate an aircraft traveling near the earth's surface, showed a higher O_2/N_2 ratio than higher flow, meant to represent the aircraft at elevation.

Steinbach theorized that the sign of the fractionation depended on the orientation of the aircraft sample inlet, with rear-facing inlets showing depletion of Ar/N_2 at altitude, and forward-facing inlets showing enrichment. Such an assertion, however, is only supported by unpublished results from the COBRA field campaign, which used a forward-facing inlet showing a small enrichment (+9 per meg/10 km) in flasks over land during the COBRA campaign [*H. Graven*, pers. comm.]. Steinbach proposes a physical mechanism underlying such enrichment or depletion signals, citing the likelihood that the ratio depends not upon altitude, but upon airspeed of the sample aircraft; aircraft tend to move faster at higher elevation through less dense air, and Steinbach posits that such an

increase of speed leads to undersampling of heavy molecules in rear-facing inlets, and over-sampling in forward-facing inlets. Steinbach doesn't provide a direct in situ measure of the correlation of sample depletion with airspeed, and she admits that the applicability of her results to other studies is debatable, as the inlet tubing used during BARCA was neither pressure nor flow controlled.

The MEDUSA inlet was both pressure and flow controlled, and more importantly was a rear-facing inlet inside of a forward facing shroud (see Figure A1-07), which raises the possibility that Steinbach's conclusions may not match ours. Fortunately, because of the large number of samples we took during HIPPO, and the situational variables provided by the NCAR on-board data system, I am able to regress sample Ar/N_2 against both altitude and airspeed.

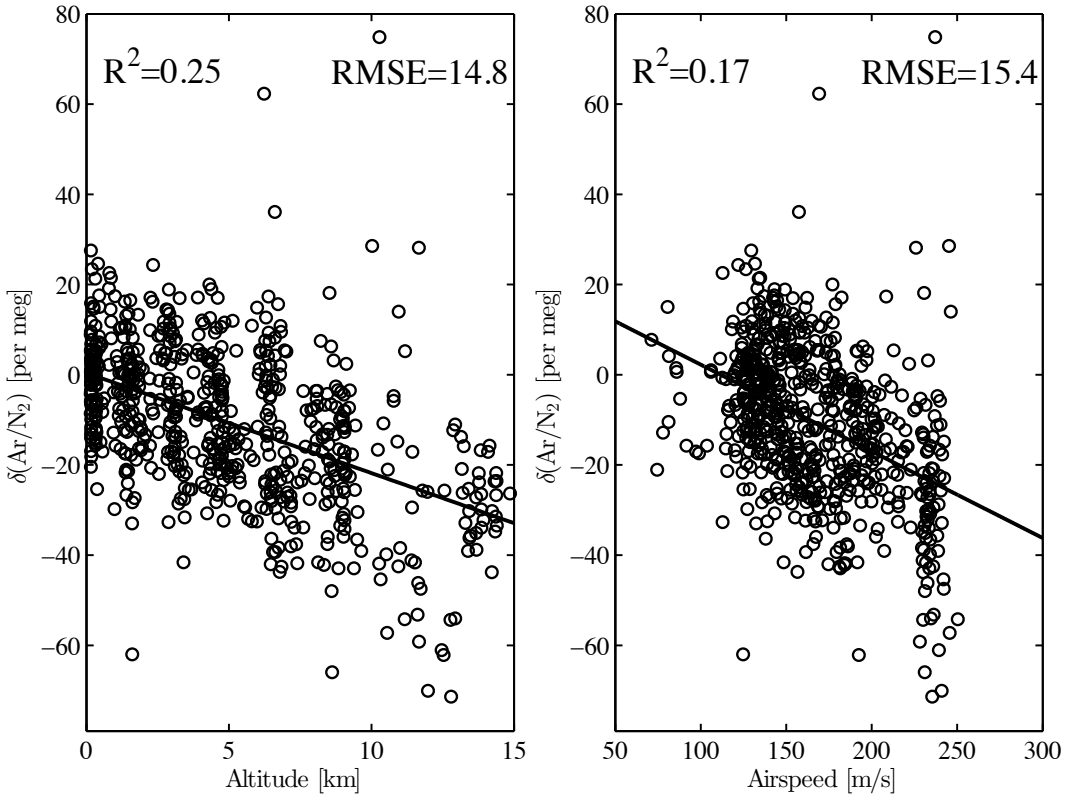

Figure 4.2. Regression of uncorrected Ar/N_2 against altitude (left panel) and airspeed (right panel). The regression against altitude shows a higher r^2 value, and lower root mean square error, suggesting a slightly closer (inverse) relationship.

The results, shown in **Figure 4.2**, suggest both variables scale inversely with the Ar/N_2 ratio, though altitude scales more closely, with a higher coefficient of determination (r^2=0.25) and lower root mean square error (14.7 per meg) than airspeed (r^2=0.17, RMSE=15.4). Preliminarily, this suggests a greater correlation with altitude than airspeed. However, the differences are small, and because airspeed and altitude tend to scale together, the effect of one or the other is difficult to determine independently.

Fortunately, there are cases in which airspeed can be considered to be independent of altitude. In any given elevation band (e.g. 5-6 km), flasks sampled therein were sampled under a variety of associated airspeeds due to slightly different aircraft

operating conditions. In theory, moving up and down the airspeed axis at a given elevation gives an altitude-independent assessment of the effect of airspeed alone. Of course, for any increment of elevation, *dz*, there will be relatively few flask samples, limiting interpretive power. Rather than assess small bands of elevation individually, I instead choose to remove the mean airspeed at all altitudes so that the remaining residuals can be evaluated as a function of airspeed deviations alone.

To do this, I mask out stratospheric data to focus exclusively on the tropospheric vertical gradient, and plot altitude vs. airspeed for all Box 2 Ar/N_2 data (crosses in **Figure 4.3**, with Ar/N_2 value in per meg as color axis for reference). I then fit a regression against these data (blue line) to establish the mean ratio of airspeed to elevation. Because of the properties of the GV avionics system, and the needs of the aircraft, this relationship is largely linear, with airspeed increasing at higher elevations. I then remove this line from the air speed data, giving just the residual values by altitude (circles).

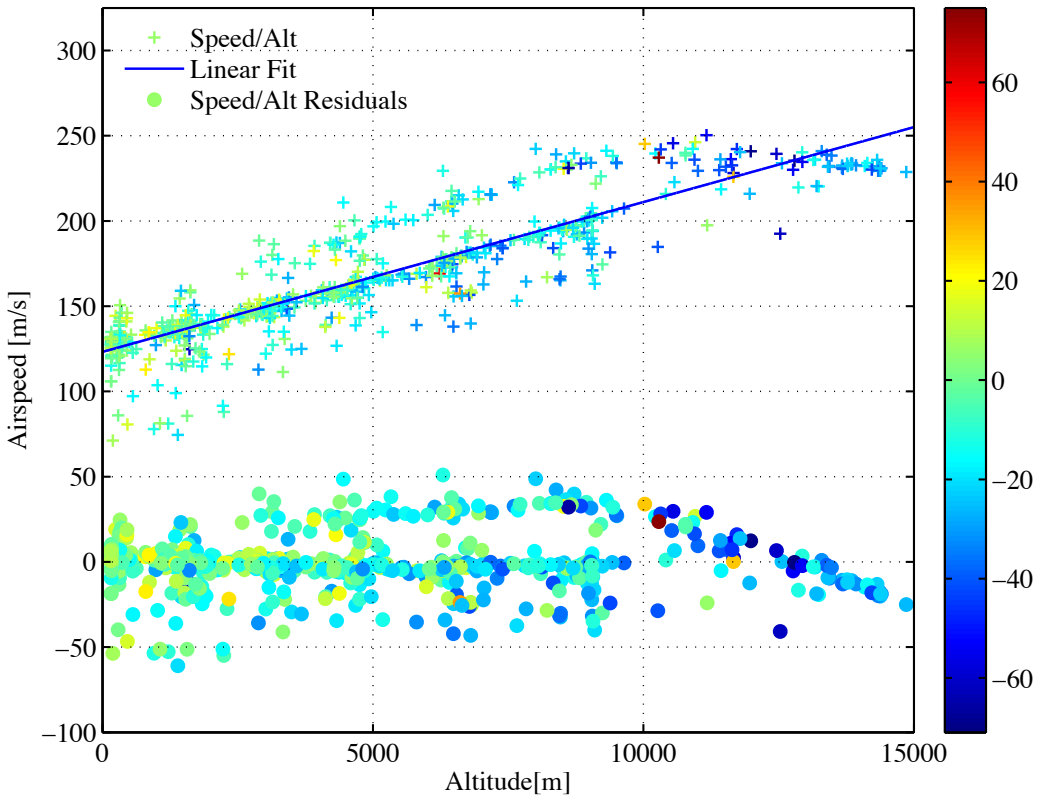

Figure 4.3. The relationship of altitude and airspeed shown from the raw data (crosses), and with the linear fit relationship with elevation (blue line) removed (circles). $\delta(Ar/N_2)$ values in per meg are shown as color for context.

The variation in the residuals is now largely independent of altitude, removing the possibility that an altitude-related phenomenon might underlie any correlations. Positive values (circles above 0) represent instances where the plane's speed exceeded the mean for that elevation. If Steinbach's theory holds correct for HIPPO samples, such instances should lead on average either to depletion in the Ar/N_2 ratio relative to the mean at that elevation because of the preferential sampling of the smaller molecule (N_2) by MEDUSA's rear-facing inlet within the HIMIL (HIAPER Modular InLet), or to enhancement in the Ar/N_2 ratio from the opposite effect at the forward facing inlet to the HIMIL. Similarly, residual airspeed values below zero indicate moments when the plane

travelled more slowly than the mean for that elevation. In such cases, one would expect an enrichment or depletion of the Ar/N_2 ratio relative to the mean value at that elevation. I plot these residuals against Ar/N_2 in **Figure 4.4**.

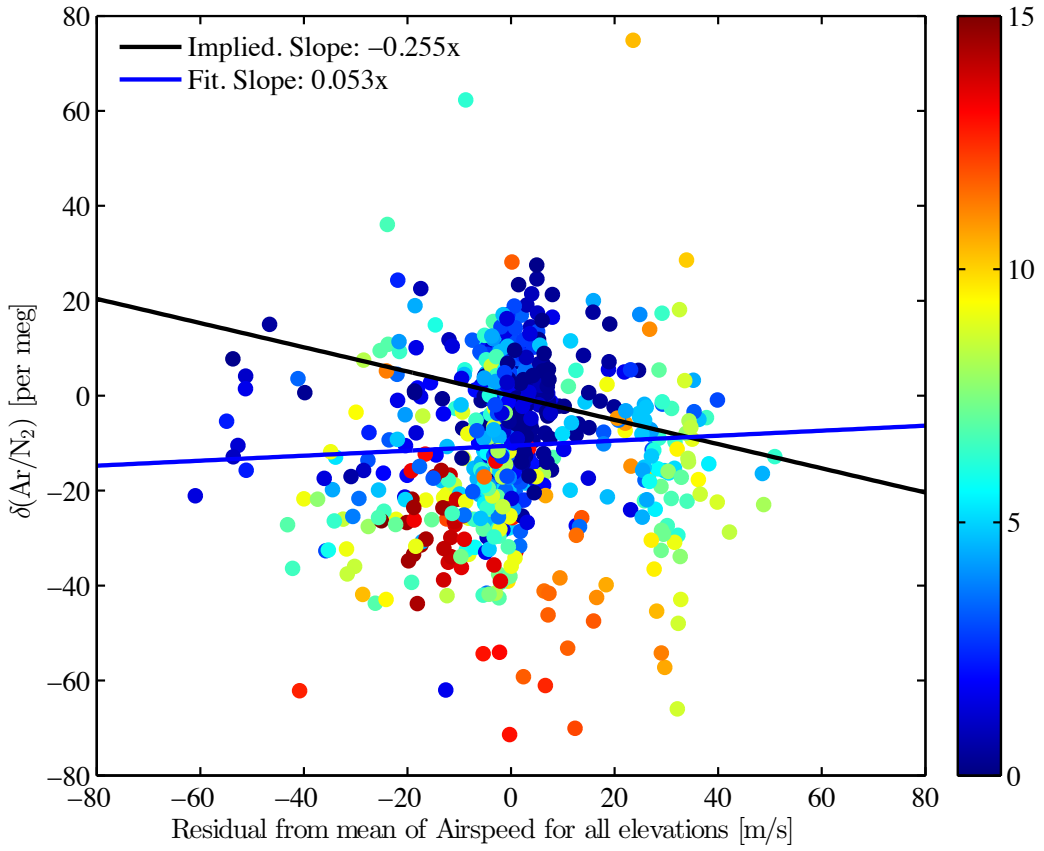

Figure 4.4. Altitude-independent regression of airspeed and $\delta(Ar/N_2)$. The fit to the data is shown in blue, while the relationship implied by the mean changes over the full column (as described below) is shown in black. The color axis shows the altitude of each flask in [km] for reference.

A regression of the two variables suggests a very slight positive correlation (thick blue line) with a slope of 0.053 per meg/(m/s). Such a value is statistically meaningless according to the fit parameters, but qualitatively, this would suggest that depleted $\delta(Ar/N_2)$ values correlated very slightly with slower air speeds.

This raises the question of what the "implied" relationship would be if the observed depletion with elevation *were* due to airspeed. My only way of determining this assumes the depletion/enrichment effect is relatively constant with elevation given a range of speeds, which is probably a fair assumption since the model Steinbach [2010] used suggested the degree of over- or undersampling depends on the mean velocity of the molecule, not the density of the gas. Accordingly, I ascertain this empirically for my flasks: I assume Steinbach's theory—that the depletion with altitude is a function of air speed—and calculate the difference between the mean Ar/N_2 near the Earth's surface and at 10km, and the mean airspeed near the earth's surface, and at 10km. This gives me end points (0.163 per meg, 123.2 m/s) at 0 meters (a slight extrapolation for clarity) and (-22.1 per meg, 210.4 m/s) at 10km. This implies a depletion/airspeed relationship of -0.255 per meg/(m/s). This implied relationship is denoted by a thick black line (arbitrarily centered at 0 per meg and 0 m/s) in **Figure 4.4**. Both statistically, and visually, the slope of this line does not fit the data, suggesting that a single airspeed-dependent inlet fractionation effect cannot explain both the scatter and vertical gradient in the observed Ar/N_2 values, at least in the context of HIPPO flasks.

I also regress Ar/N_2 depletion against ram pressure (seen in Appendix 2, Figures A2-02 through A2-04) because Steinbach [2010] suggests that fractionation at system inlets may be due to the pressure nozzle effect, which itself is caused by pressure gradients normal to the sample streamlines upstream of the inlet. The pressure gradients are induced by ram pressure, and may cause sampled Ar/N_2 to decrease with elevation for rear-facing inlets, and increase with elevation for forward-facing inlets. My regression suggests that there is no statistically-significant correlation with depletion of argon.

Another pertinent consideration in determining the cause of the vertical gradient is whether the gradient is consistent between low and high latitudes. In theory, if convective mixing is failing to completely homogenize the Ar/N_2 ratio in the troposphere, this effect would be more pronounced at higher latitudes where convection is generally less vigorous than in the tropics because of weaker insolation, and thereby heating. Of course, this effect is muddied by the lower tropopause at higher latitudes, which gives the full profiles (troposphere and lower stratosphere) a larger depletion at 15 km than for tropical samples, which are entirely tropospheric because of the limited vertical range of the GV aircraft. **Figure 4.5** shows this stark contrast in the HIPPO Ar/N_2 data, with tropical flasks (23.3°S to 23.3°N) in the upper panel, and mid and high latitudes in the lower.

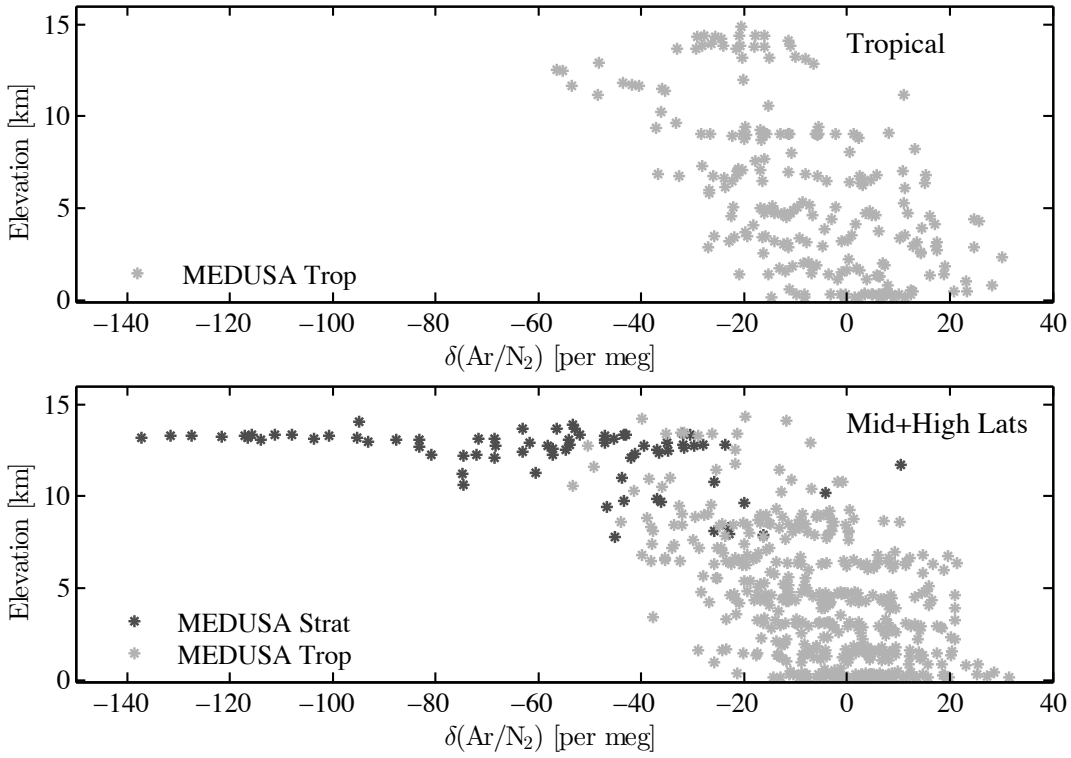

Figure 4.5. Comparison of Flask Box 2 $\delta(Ar/N_2)$ flask profiles for roughly tropical latitudes (23°S-23°N, top panel) and mid-high latitudes (≥23°, bottom panel).

Though the difference is marked for all flasks, it's more subtle for just tropospheric flasks.

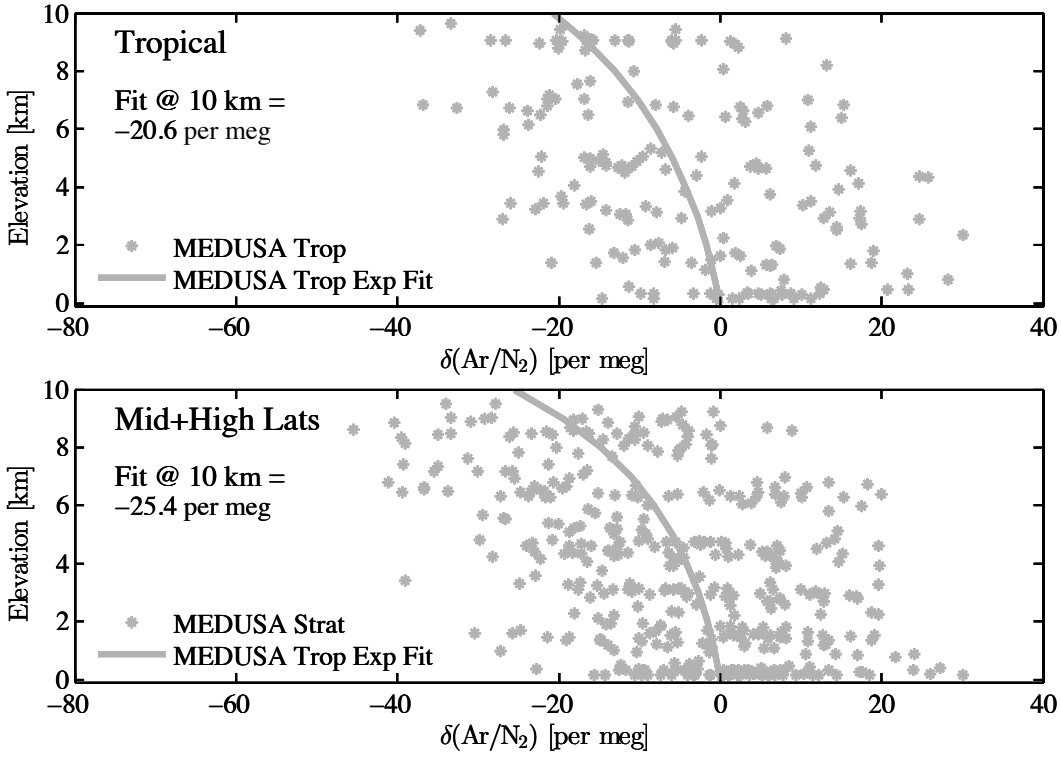

Figure 4.6. Box 2 tropospheric $\delta(Ar/N_2)$ gradient between the surface and 10 km for tropical flasks (top panel, between 23.3°S and 23.3°N) and mid- and high latitude flasks (bottom panel). Note: the mean value for all flasks has been shifted up very slightly for this and the following plots (on order 1-5 per meg, according to the exponential fit parameters) to 0 for each fit to show an offset from a zero starting value since the quantity of interest is the vertical gradient, not the absolute numerical value.

In the tropics, data average -20.6 per meg less than the surface at 10 km, while in the mid and high latitudes, they average -25.4 per meg less than the surface. Though this is by no means conclusive evidence, it supports the theory that convective strength may play a part in mixing out the Ar/N_2 signal in the troposphere. Convective strength can be looked at another way—as the difference between summer and winter conditions. Tropical differences may be small due to consistently warm conditions, so instead I look at just extratropical data. Here, if flasks are divided into two broad categories: those sampled between the first day of (local) spring and the last day of (local) summer, and those sampled between the first day of (local) fall and the last day of (local) winter, the

difference in the vertical gradient may be apparent due to stronger winter convection over the ocean. **Figure 4.7**, below, shows this difference.

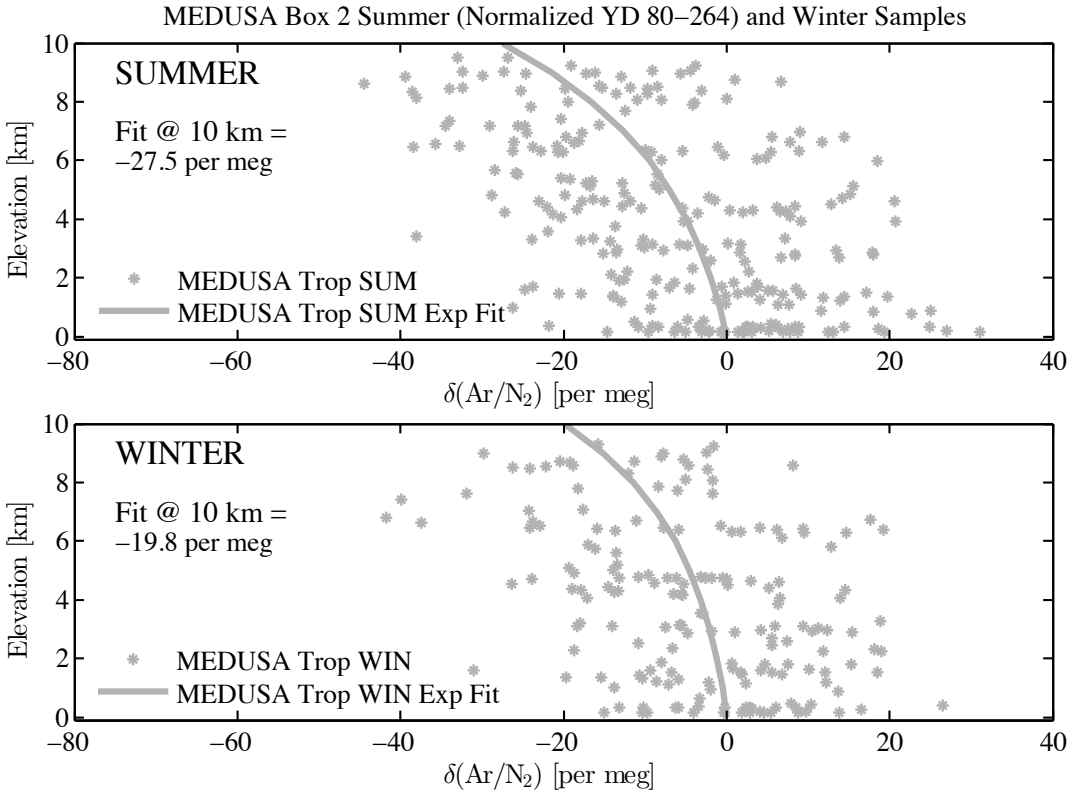

Figure 4.7. Comparison of "summer" and "winter" flasks from MEDUSA box 2. Northern hemisphere flasks between YD80 and 264 (Mar 21 to Sept 21), and southern hemisphere flasks between YD264 and YD80 are shown in the top plot, "summer", while the opposite is true for the bottom plot, "winter". "Winter" conditions show a smaller difference between surface and 10 km, suggesting stronger convection of the atmosphere, consistent with natural winter conditions over the ocean.

The figure shows that the gradient for extratropical flasks in "winter" is smaller (-19.8 per meg/10 km) than for "summer" (-27.5 per meg/10 km), suggesting that stronger convection may be erasing the gradient partly, further support for the vertical gradient being real, or at least partly representative of real differences between summer and winter atmospheric convection regimes.

4.4 Modeling the vertical Ar/N_2 gradient

Atmospheric transport models do not account for the possibility that gravitational separation may very slightly alter the ratio of one gas to another in the atmosphere. Absent a model that directly approximates gravitational separation, I look to a simple 1-dimensional diffusion model used recently by Ishidoya et al. [2013], from the original by Lettau [1951], to see whether small differences in the vertical molecular diffusivity of Ar and N_2 might lead to subtle, predictable changes in the ratio in the troposphere at a scale large enough to explain the vertical gradient seen in the HIPPO flask data. I also look to see whether the same model nominally agrees with the observed stratospheric depletion with and without the fit to the vertical gradient with altitude removed.

I simplify Lettau's equation by removing a source term that doesn't pertain to this study and rewrite it as:

$$C_i = C_{i0} * \exp\left(-\frac{\mu_i}{H}\int_0^z \frac{gT_0}{g_0 T} Q_i dz\right) \quad \textbf{Eq. 4.1}$$

In which the Q_i term represents:

$$Q_i = \frac{D_{iN2}}{D_{iN2} + K_z} \quad \textbf{Eq. 4.2}$$

And in which μ_i represents:

$$\mu_i = \frac{m_i - m_{N2}}{m_{N2}} \quad \textbf{Eq. 4.3}$$

And where H is the scale height of N_2 under the chosen circumstances:

$$H = \frac{k * T_o}{m_{N2} * g_0} \qquad \textbf{Eq. 4.4}$$

For these equations, *k* refers to the Boltzmann constant. C_{i0} and C_i, g_0 and g, and T_0 and T are the number density of the gas *i*, gravity and temperature at the lower (*0*) and upper (*z*) bounds in km, respectively. The terms m_i and m_{N2} refer to atomic or molecular mass of the species in question, here Ar (M40), and of N_2 (M28). D_{iN2} refers to the mutual molecular diffusivity constant of gas *i* with N_2, which here is 0.18, also from Lettau [1951]. The term K_z refers to the eddy-diffusivity coefficient, which changes with height. I take estimates of these values from Massie and Hunten [1981], who provide 6 different sets of estimates of eddy-diffusion coefficients with height from studies conducted between 1975 and 1981 (see **Table 4.2**).

I first plot all six of the models against height to 10 km, with all tropospheric data from 10 km or below. Here, because I show data from all latitudes, I consider an average case of a surface temperature, T_0, of 288K.

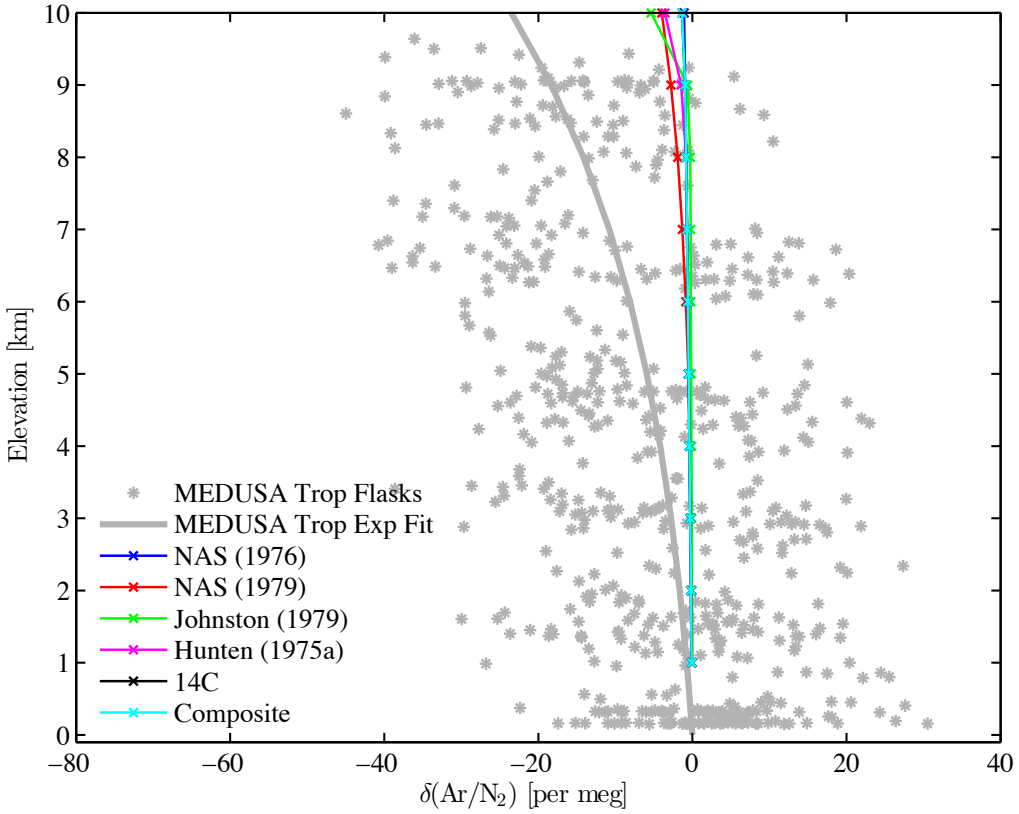

Figure 4.8. Surface to 10 km plot of all HIPPO MEDUSA Box 2 tropospheric flasks (gray stars) without altitude correction, exponential fit to data (bold gray line), and six model calculations. Uncorrected MEDUSA data show a much greater depletion than the model predictions in the troposphere.

Over ten kilometers, the models predict a range of gradients between -1.0 and -5.3 per meg (see, **Table 4.1**, column 3). By contrast, an exponential fit to HIPPO data gives a much larger gradient of -25.4 per meg. Changing the surface temperature to 273K, equivalent to higher latitudes, only slightly increases the magnitude of the gradient, bringing them to between -1.1 and -5.6 per meg, still a factor of 21 to 4 too small. This observation leads to the conclusion that the model and data do not agree, and that this may be due to one of two obvious reasons. Either the data demonstrates an artifact that the model does not, or the model does not accurately represent the degree to which an

influence or influences, be they diffusion, gravitational separation or the mixing of stratospheric air back into the upper troposphere, act upon the vertical gradient.

Our study, however, asks a question that few others have—whether a small vertical gravitation signal in an inert gas can be detected within the troposphere. A second question is whether the Lettau [1951] model is able to reproduce the vertical gradient apparent in the HIPPO stratospheric data, a quantity that is, to some extent, easier to measure due to the larger signal.

Though I've primarily compared stratospheric Ar/N_2 with N_2O (because the height-dependence changes due to tropopause height and N_2O reflects stratospheric air age), the Ar/N_2 gradient does scale with elevation, too. Since the Lettau [1951] model is set up to compare a quantity against altitude, this is the quantity I compare here. Having shown above that the choice of surface temperature within earth's natural ranges has a relatively small impact on the gradient, I choose again to consider an "average" case surface temperature of 288K. I now compare a fit to all data from the troposphere and stratosphere against the same model with six sets of eddy diffusion coefficients. In the top panel of **Figure 4.9**, I show the data without the vertical "tropospheric" gradient removed, and in the bottom panel, I remove the gradient with the linear fit I use in Chapter 2, Eq. 2.8.

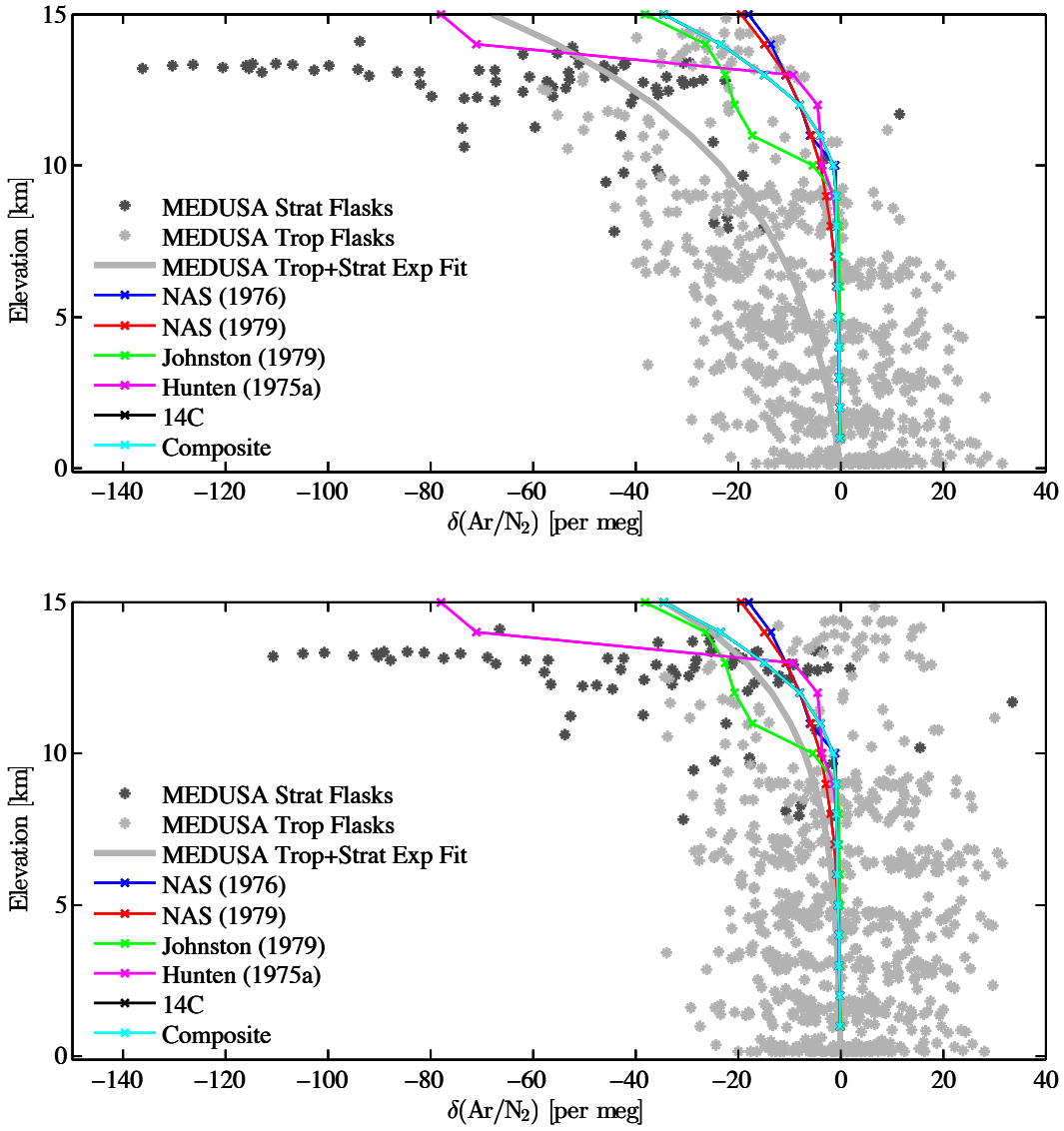

Figure 4.9. For a surface temperature (T_0) of 288K, diffusive separation model by Lettau [1951] for $\delta(Ar/N_2)$, using 6 different sets of eddy-diffusivity coefficient estimates shown variously as colored lines with exes. Top Panel: HIPPO vertical gradient preserved. Lower Panel: HIPPO vertical gradient removed using a linear fit to tropospheric flasks only. Because the composite values draw strongly from the ^{14}C study for lower atmosphere values, the ^{14}C values are completely obscured by the composite curve. MEDUSA data from HIPPO are shown for samples in the troposphere (light gray stars), and for samples in the stratosphere (dark gray stars). A fit to all MEDUSA data is shown as a bold light gray line.

In the upper panel, the fit to the MEDUSA flasks, in light gray, suggests a depletion of -25.4 per meg by 10 km, and of -68.1 per meg by 15 km. In the bottom, with the extrapolated tropospheric gradient removed from all data, the value at 10 km is -7,

and at 15 km is -34.4. The first 15-km fit value of -68.1 per meg is within the boundaries of the outlier model calculation using Hunten [1975a] eddy mixing data, which estimates very low eddy-diffusion coefficients in the lower stratosphere, but is well outside all predictions at 10 km. In the lower panel, the MEDUSA data fit is within the predicted depletion for the remaining models at 15 km, which range from -18 to -38.2 per meg (see column 4, **Table 4.1**), although still slightly low at 10 km. None of the models are able to capture the extreme depletion of a dozen or so flasks sampled at high latitudes. Lowering the surface temperature to 273K, in keeping with the lower annual mean surface temperature at higher latitudes lowers the most extreme case [*Hunten*, 1975a] to -84.1 at 15 km, which still doesn't quite capture the extreme MEDUSA data points.

I conclude that the Lettau [1951] model does not predict the vertical gradient that I observe in tropospheric flasks, even when using eddy diffusion coefficients from the most extreme outlier model run of Hunten [1975a], but that it does capture the average gradient I see higher in the atmospheric column when stratospheric values are factored in. When the tropospheric gradient is removed from the MEDUSA data, the resulting exponential fit falls within the various Massie and Hunten [1981] eddy mixing estimates used by the Lettau [1951] model. I also conclude that, because the eddy diffusion coefficients are provided irrespective of latitude by Massie and Hunten [1981], it is impossible to fine tune the elevation of the tropopause in the model to compare MEDUSA data against a completely equivalent profile for a given latitude. The model using Hunten [1975a], which predicts low-elevation drop-offs in eddy diffusion coefficients (2300 cm^2/s at 14 km), show the sensitivity of the model to stratospheric eddy diffusion coefficients.

Table 4.1. Depletions in the Ar/N_2 ratio predicted for each of the six studies at 10 and 15 km for each of three T_0 scenarios (273, 288, 300K). Depletions are relative to a normalized value of 0 per meg at the surface. The gradients for HIPPO data are reflective of both troposphere and stratosphere flasks, which is why the numbers differ slightly from those in Figure 4.6. For the 273K scenario, only >23° flasks are used; for the 300K scenario, only 23°S-23°N flasks are used; and for the 288K scenario, all flasks are used.

Model Run, Data	T_0=273 Ar/N_2 depletion @ 10km [per meg]	T_0=273 Ar/N_2 depletion @ 15km [per meg]	T_0=288 Ar/N_2 depletion @ 10km [per meg]	T_0=288 Ar/N_2 depletion @ 15km [per meg]	T_0=300 Ar/N_2 depletion @ 10km [per meg]	T_0=300 Ar/N_2 depletion @ 15km [per meg]
NAS [1976]	-1.1	-18.9	-1.0	-18.0	-1.0	-17.2
NAS [1979]	-4.2	-20.6	-4.0	-19.5	-3.8	-18.7
Johnston [1979]	-5.6	-40.3	-5.3	-38.2	-5.1	-36.6
Hunten [1975a]	-3.8	-82.3	-3.6	-78.0	-3.4	-74.9
^{14}C	-1.4	-36.4	-1.3	-34.5	-1.3	-33.1
Composite	-1.4	-36.4	-1.3	-34.5	-1.3	-33.1
HIPPO	-26.4	-87.1	-25.4	-68.1	-14.3	-32.4

Table 4.2. Eddy-diffusion coefficients from 6 studies for elevations between 1 and 16 km. A portion of a larger table, reproduced from Massie and Hunten [1981] Table 3. Units: 10^3 cm^2/s. Tropopause height is indicated by EDC value dropping quickly—most studies place this around 9-10 km.

Elevation (km)	**NAS [1976]**	**NAS [1979]**	**Johnston [1979]**	**Hunten [1975a]**	**^{14}C**	**Composite**
1	100	270	300	100	100	100
2	100	210	300	100	100	100
3	100	160	300	100	100	100
4	100	120	300	100	100	100
5	100	94	300	100	100	100
6	100	70	300	100	100	100
7	100	54	300	100	100	100
8	100	43	300	100	100	100
9	100	34	150	65	100	100
10	100	27	20	30	80	80
11	20	21	7	30	30	30
12	17	17	6.5	30	17	17
13	14	14	6.6	16	10	8
14	12	11	6.2	2.3	7	5.8
15	10	9.2	4.7	2.3	5.2	4.6

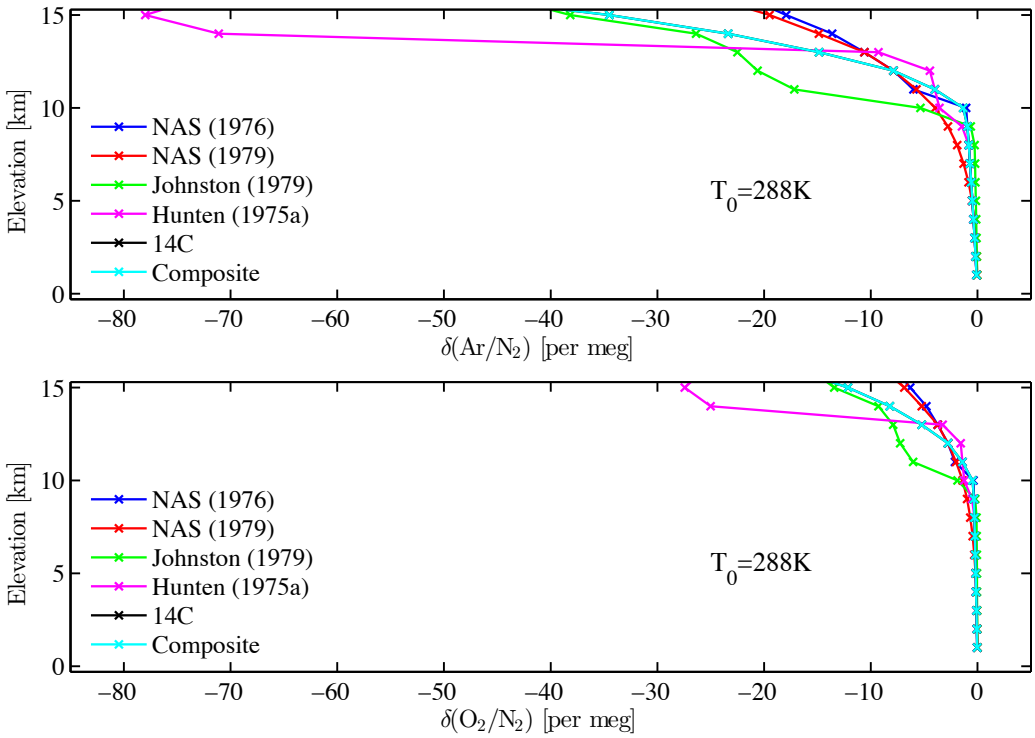

Figure 4.10. $\delta(Ar/N_2)$ (top) and $\delta(O_2/N_2)$ (bottom) depletion with elevation at 288K. Both species show major depletion in the stratosphere when eddy diffusion coefficients drop greatly under different thermodynamic conditions, encouraging gradients to form.

The Lettau [1951] model has an additional use for the purposes of this chapter and the previous chapter—it provides an independent assessment of the ratio of O_2/N_2 depletion to Ar/N_2 depletion with height, a value one might otherwise have assumed to be 1/3, scaling with the relative mass differences between the O_2 and N_2 and Ar and N_2 molecules, $(m_{O2}\text{-}m_{N2})/(m_{Ar}\text{-}m_{N2})$. The Lettau [1951] model confirms this fact partly, suggesting that the depletion in the ratio scales with the term μ_i, but deviates slightly from this value because the mutual molecular diffusivity coefficient for O_2/N_2 is slightly higher than for Ar/N_2—0.19 instead of 0.18, and because the terms are located in the exponent, albeit at extremely small exponent values, where the function approximates a linear relationship. As a result, depletion does not scale exactly as 1/3, but as a 1/2.84, or 0.352

(see **Figure 4.11**). This value is basically constant for all scenarios, regardless of T_0 because μ_i is simply a multiplicative factor, and exponent values are exceedingly small.

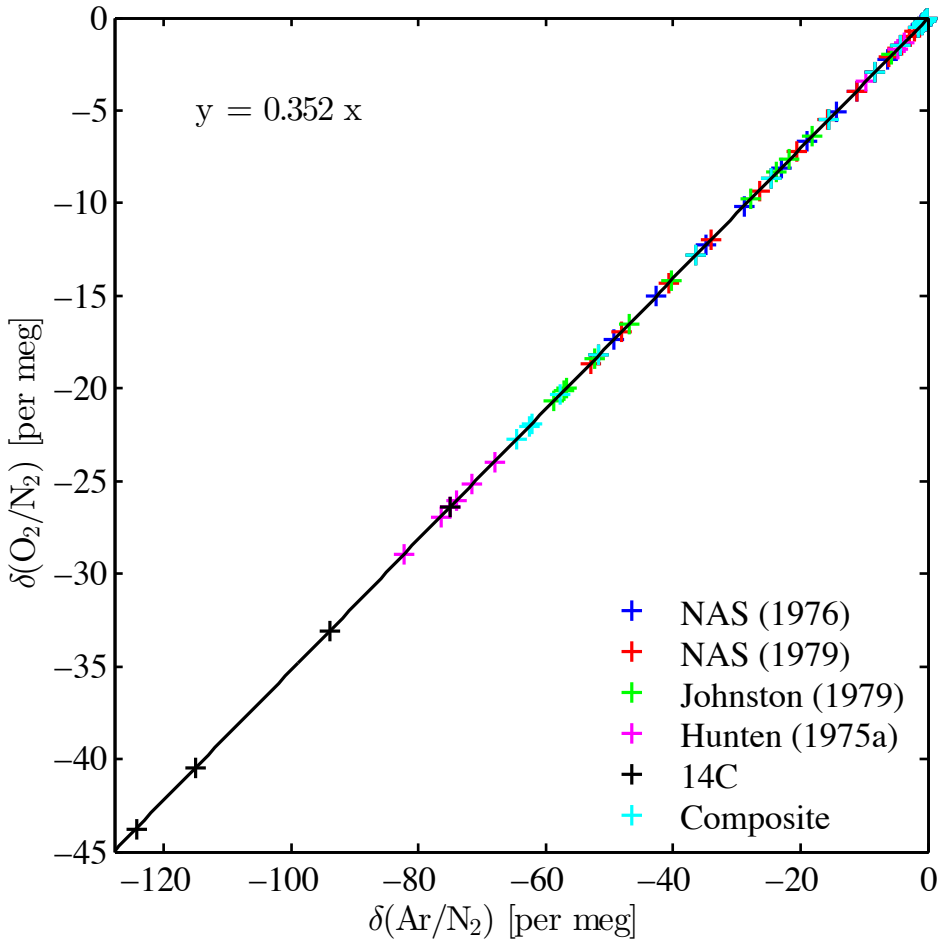

Figure 4.11. For any given T_0 at the surface (here 288k), the relationship between the depletion of O_2/N_2 and Ar/N_2 in per meg in the lower atmosphere is modeled to be a relatively constant 1/2.84, or 0.352. Here I show values up to 20 km for each set of eddy diffusion coefficients, as noted in the legend.

4.5 Seasonal cycles

As I do not fully understand the nature of the tropospheric gradient I've observed, I choose to exclude high-altitude samples for the remainder of the chapter, when considering seasonal cycles and station-station gradients. This reduces possible bias that high-altitude samples might introduce, but it is also a more logical choice of data, since

values from the planetary boundary layer are more representative of surface station values than those above it, which may lag in time from their surface equivalents. For the next section, I choose to include all Ar/N_2 data below 800 mb, as these data show very little diminution in the Ar/N_2 ratio, as seen below in **Figure 4.12**.

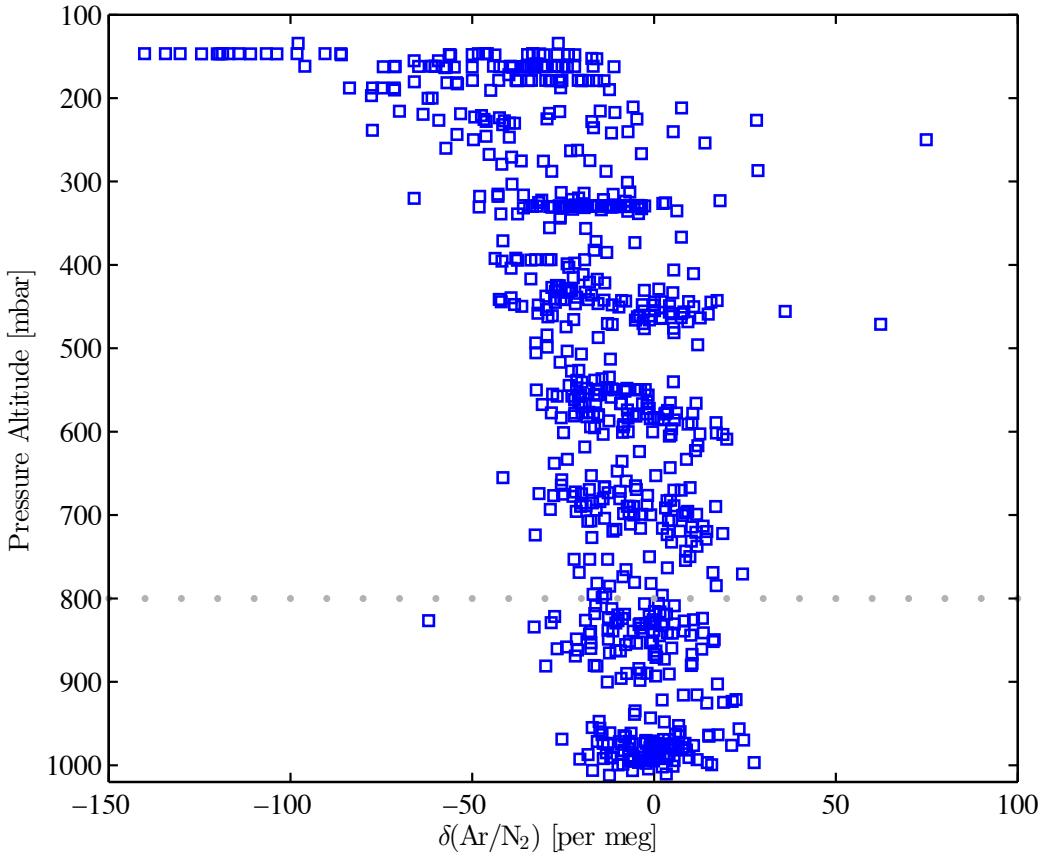

Figure 4.12. Ar/N_2 vs. pressure altitude, showing how the gradient is very small near the surface when plotted against pressure instead of altitude

A simple test of the HIPPO MEDUSA Ar/N_2 data is whether the data roughly reproduce the phase and magnitude of the seasonal cycle that station observations have shown to exist. **Figure 4.13** shows northern (top panel) and southern (bottom panel) hemisphere station records. Southern hemisphere time series in the Scripps O_2 Network show peak-to-peak amplitudes of nominally 10-20 per meg with maxima and minima

typically in austral fall and spring, respectively. Likewise, northern hemisphere stations show a similar range of amplitudes, with maxima (minima) in boreal fall (spring). The two hemispheres are in roughly opposite phase because the seasonal heat fluxes in the two hemispheres are out of phase, and the Ar/N_2 ratio in the atmosphere tends to track heat fluxes closely.

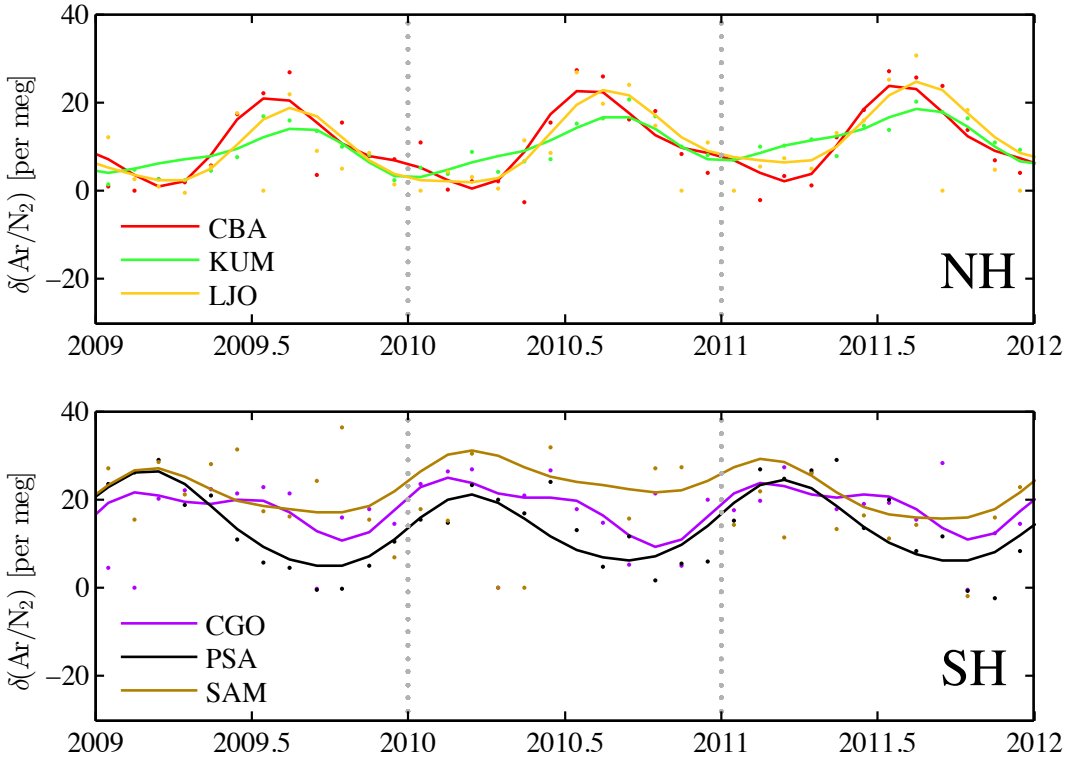

Figure 4.13. Scripps O_2 Program Ar/N_2 data (dots) and 2-harmonic fits (lines) to Scripps O_2 Program Ar/N_2 data at Cold Bay, Alaska (CBA), Cape Kumukahi, Hawaii (KUM), La Jolla, CA (LJO), Cape Grim Observatory, Tasmania (CGO), Palmer Station, Antarctica (PSA), and American Samoa (SAM) between 2009 and 2012. Vertical gray dotted lines at the beginning of each year are included to help the eye distinguish the difference in northern and southern hemisphere phasing.

I attempt to see whether these seasonal cycles are detectable in the HIPPO data. Because the scatter in the HIPPO observations is large, I choose to fit a single function to all data in the Southern Hemisphere and to all data in the Northern Hemisphere. This gives a rough sense for the integrated seasonal cycle over each hemisphere. To have a

sufficient number of points, I choose all MEDUSA Box 2 data below 800 mb, having not corrected for altitude. I exclude results from MEDUSA Box 1 for reasons discussed in Chapter 2, i.e. greater scatter in Ar/N_2 data likely caused by an adverse thermal environment on the airplane. I choose to fit both a single harmonic and a 2-harmonic fit to the data to see which captures the seasonal cycle best. The resulting fits seen in **Figure 4.14** show that northern and southern hemisphere seasonal signals are roughly opposite in phase, with similar 1- and 2 harmonic peak-to-peak amplitudes. Single-harmonic fits have amplitudes of 14.6 (NH) and 13.1 (SH) per meg Ar/N_2, while 2-harmonic fits show amplitudes of 15.5 (NH) and 17.0 (SH).

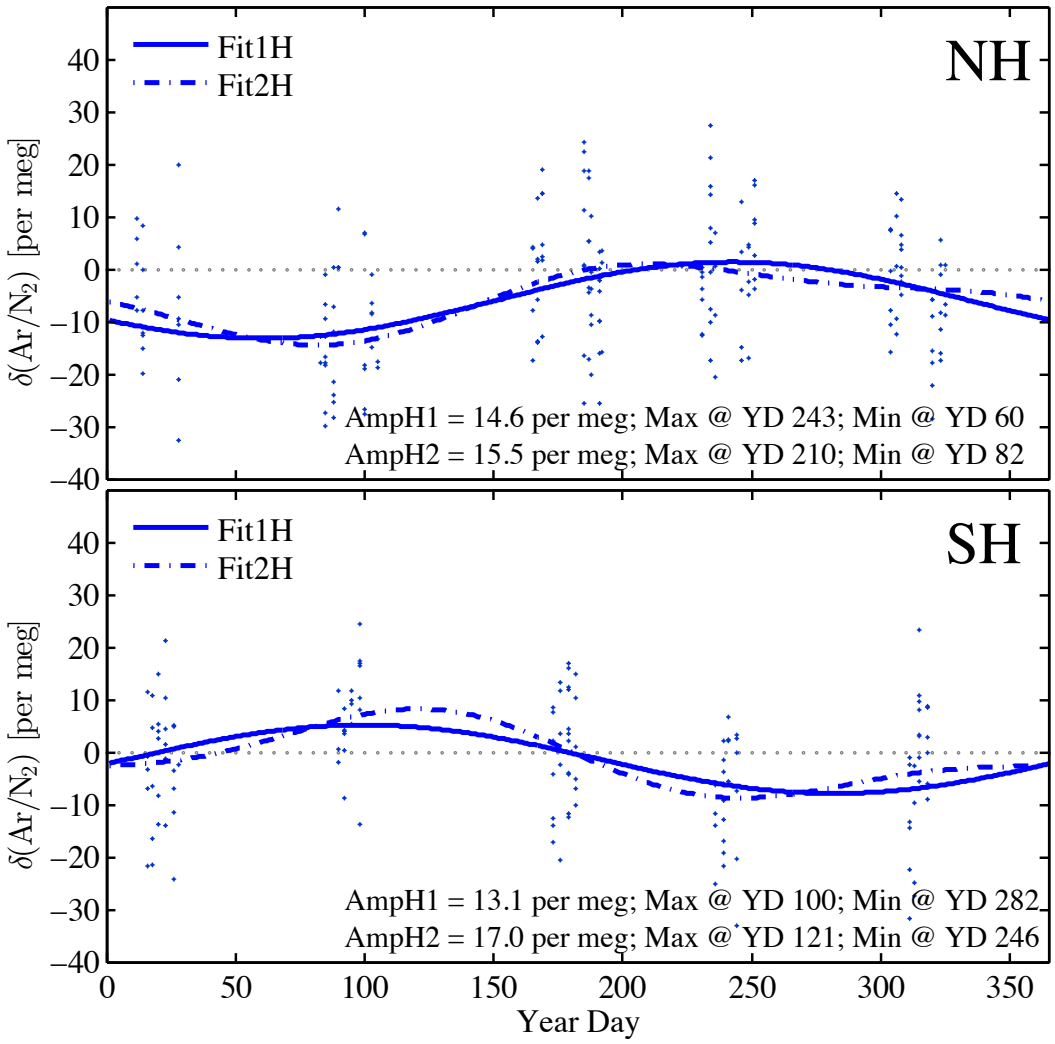

Figure 4.14. Single- (solid line) and 2-harmonic (dash-dot line) fits of lower tropospheric Ar/N_2 in the Northern (upper panel) and Southern (lower panel) hemispheres for all latitudes. Ar/N_2 data, as per other figures, is sourced from Box 2 flasks only. Shown are flasks from all 5 HIPPO missions, below 800 mb. The Northern Hemisphere has roughly twice as many data points because the plane originates from and returns to North America. Gray lines at 0 per meg are shown for clarity. The Northern Hemisphere annual mean is several per meg lower than the Southern Hemisphere annual mean.

The phase of the HIPPO observations appears to match the phase of the station observations in the previous figure reasonably well, with seasonal hemispheric maxima largely in early local autumn. The 2-harmonic fit appears to be a closer fit for all data, though in the southern hemisphere, the fit causes an extreme asymmetry. Additionally, the HIPPO amplitudes fall right in the middle of the ranges observed in the station

records, as might be expected of an average of all latitudes within a hemisphere. Notably, data in the Southern Hemisphere sit slightly below 0, with annual mean values of -1.23 (1H) and -0.93 (2H) per meg, while data in the Northern Hemisphere sit almost entirely below zero, with an annual mean of -5.82 (1 harmonic) or -3.72 (2 harmonics).

4.6 Interhemispheric difference

Having shown that HIPPO observations within a given hemisphere are reasonably similar to station records in phase and amplitude, a second question arises—whether the observed difference in the mean of northern and southern hemisphere HIPPO records is also present between station records. **Figure 4.15** below shows a comparison of HIPPO (thick blue line) northern and southern hemisphere one- and two-harmonic fits, and the same station records (dashed lines here). In this case, I have adjusted all station records down equally to bring them into the same range as the HIPPO Ar/N_2 values I have been looking at throughout this dissertation. Here, in contrast to the 7.9 per meg offset seen between La Jolla round bottom and La Jolla B-flasks, the offset is larger, 14.6 per meg. Further comparisons may reveal why these two values are different—possibly due to additional offsets from the MEDUSA rack—however, at least a portion of this offset is clearly related to analysis offsets between the two flask racks in the Scripps O_2 Program lab.

Offsets between NH and SH sites is now visible, and fairly comparable, in both station and MEDUSA data. Northern hemisphere sites, whose annual mean values are 12.3 (CBA), 11.1 (KUM) and 11.3 (LJO) are now a few per meg below a (normalized) 0,

while southern hemisphere sites whose annual means are 18.7 (CGO), 13.4 (PSA), and 22.1 (SAM) are closer to (normalized) zero, just like the HIPPO observations.

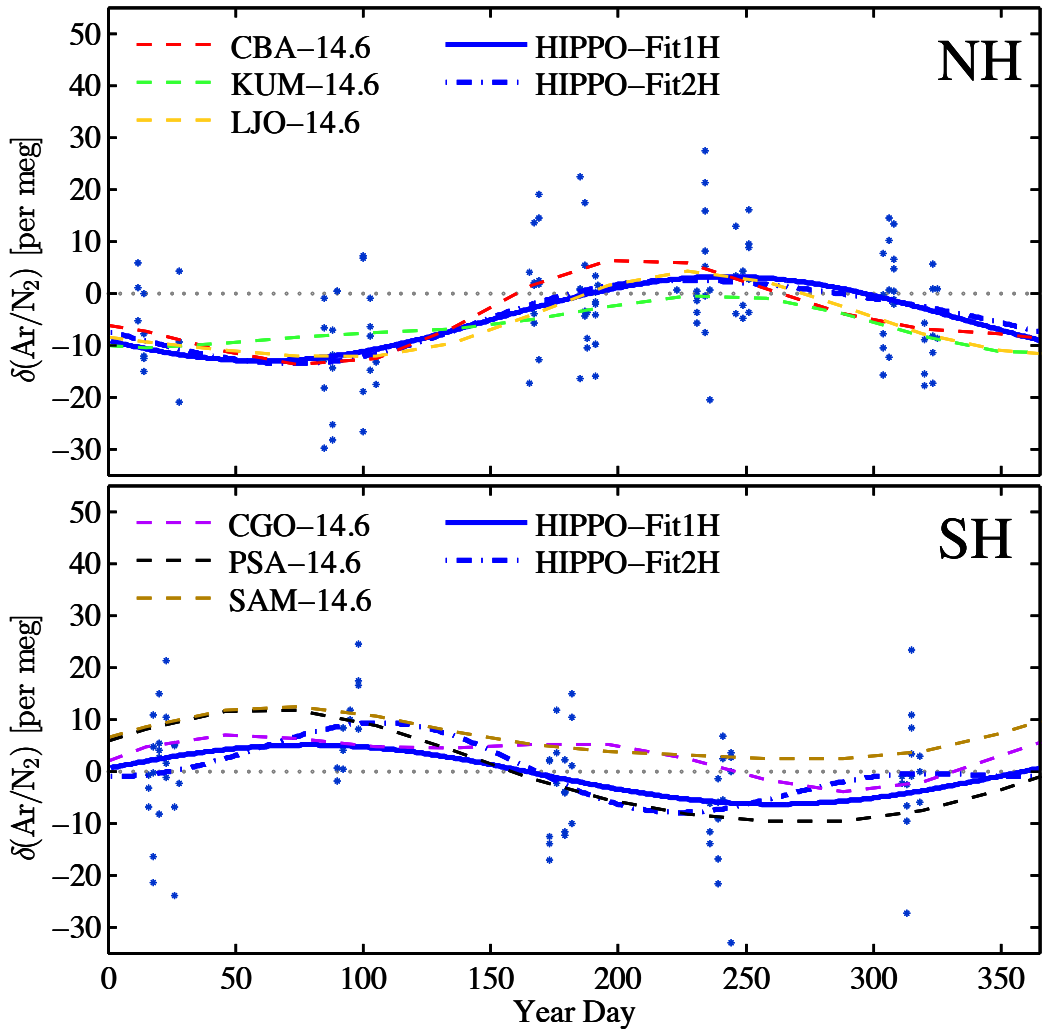

Figure 4.15. Comparisons of 1- and 2-harmonic fits to HIPPO data by hemisphere for all flasks (blue dots) below 800 mb. Station 2-harmonic records are shown alongside the MEDUSA data, and show good agreement with the aircraft flasks.

In the northern hemisphere, the 2-harmonic fit is slightly closer to station observations, while in the southern hemisphere, the single harmonic is closer. Because phase-opposite signals are detectable independently in the northern and southern hemispheres, and because a difference in the annual mean of the cycle is clear, I conclude

that the HIPPO data and station data both have, to an extent limited primarily by scatter, an ability to reproduce absolute interhemispheric differences and seasonal cycles. The HIPPO data, then, have the potential to provide the first measure of the interhemispheric Ar/N_2 difference that is independent of potential sampler-sampler offsets (i.e. that is internally consistent).

This observation is valuable, in that I can now use the observed interhemispheric difference in mean HIPPO values to ascertain whether station-to-station offsets are primarily representative of sampler-sampler offsets, or of true geographic gradients. Because I am now quantifying differences, not absolute numbers, the mean offset between MEDUSA and the station records falls away.

My ability to compare differences is limited primarily by the smaller number of observations in the restricted MEDUSA dataset. Because of this limitation, it would be difficult to fit anything more than simple linear difference (i.e. I do not attempt to characterize possible bulges or dips near the equator). To maximize the number of open-ocean MEDUSA values that can be used to determine this fit, I choose to measure the difference between two distant stations in the Pacific—Cold Bay, Alaska, and Cape Grim, Tasmania. I start by fitting linear trends to the near-surface data from each of the five HIPPO campaigns independently, as shown in **Figure 4.16**.

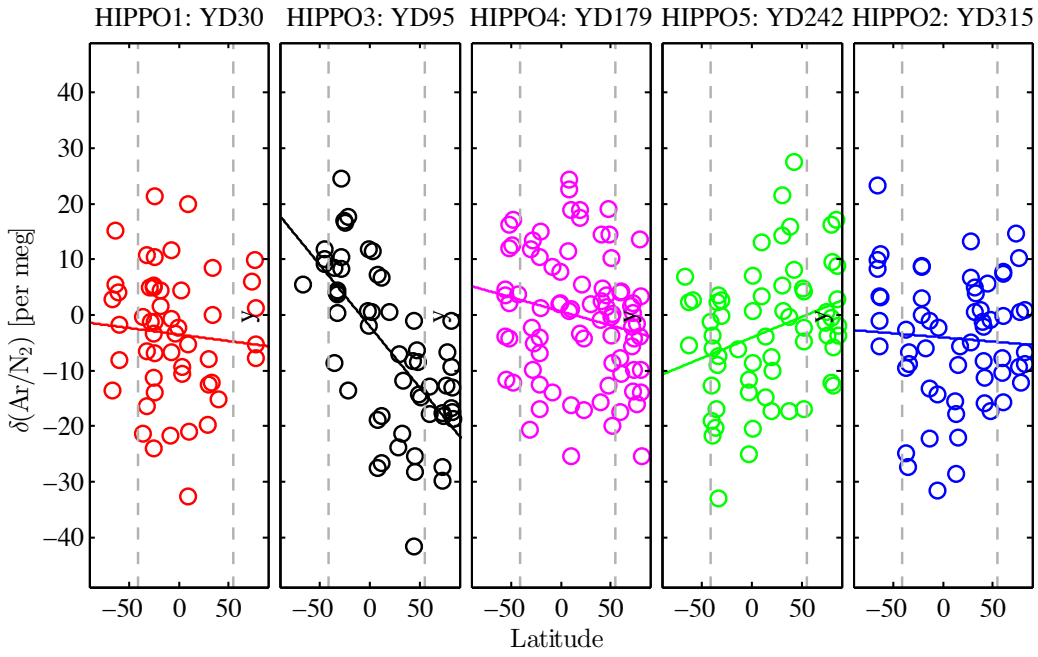

Figure 4.16. Near-surface (below 800 mb) MEDUSA $\delta(Ar/N_2)$ flask measurements for HIPPO1, 3, 4, 5, and 2 respectively against latitude. Linear fits to the data show the gradient observed between northernmost and southernmost data for each of the five campaigns. Fits are extrapolated out to 90˚S and 90˚N for consistency between plots and to aid the eye. Gray vertical dotted lines indicate the latitudes of Cape Grim Observatory (41°S) and Cold Bay, Alaska (55°N), respectively.

I then calculate the difference between the latitude of Cold Bay (~55°N) and Cape Grim (~41°S) for each of these five points in time, and fit two seasonal harmonics to these 5 values, preserving the mean offset. For station data, I do the same, but by necessity this requires a slightly different approach. I first isolate monthly mean Ar/N_2 data from CGO and CBA between 2006 and 2012 (monthly mean values are the means of all samples taken within that particular month). I then subtract the monthly mean values of CGO from the monthly mean values at CBA, and take the average of these seven years' data as the mean value over the 2006-12 period. I represent the standard deviation in these values with the red error bars shown in **Figure 4.17**. I then fit 2 harmonics to the 12 monthly mean values, shown below as a thick red line.

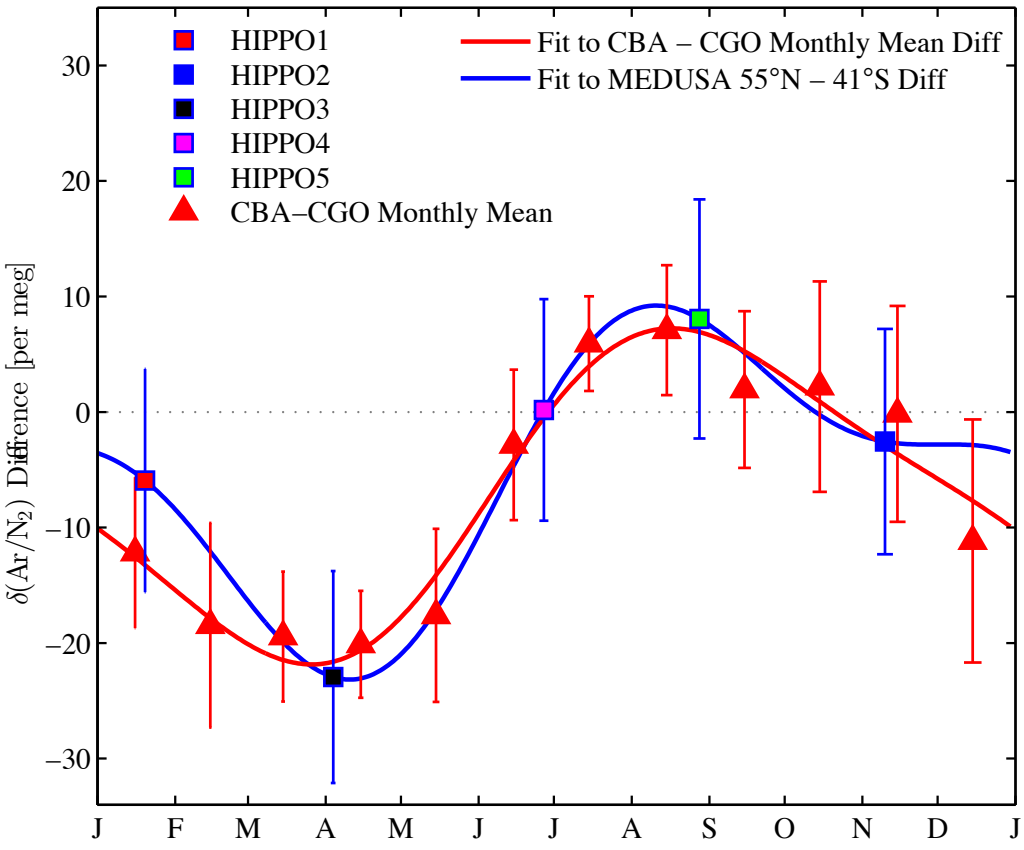

Figure 4.17. Seasonal difference in $\delta(Ar/N_2)$ between Cold Bay, Alaska (55˚N) and Cape Grim, Tasmania (41˚S) as described by 2-harmonic fits to HIPPO data and station data. The difference in HIPPO near-surface (below 800 mb) $\delta(Ar/N_2)$ flask measurements is shown as a colored square with a blue outline, located temporally at the mid-point of each of the five HIPPO missions. The error bar for each value is the RMSE of the linear fit against latitude used to calculate the difference. The thick blue line is the 2-harmonic fit to these five values, preserving a non-zero mean. Station data (red triangles) are computed by subtracting the monthly mean value at Cape Grim from the monthly mean value at Cold Bay for each month between 2006 and 2012. These values are then averaged by month to produce a climatological monthly average for the period, whose standard deviation is the 1σ red error bar. A 2-harmonic fit to these values is shown as a thick red line.

The similarity is quite good. The two methods agree to within 6 days (peak) and 16 days (trough) on phase, to within 3.3 per meg on amplitude, and they agree within 1.4 per meg on the mean annual offset, which station data suggest is -7.0 per meg, and which MEDUSA suggests is -5.6 per meg (lower in northern hemisphere). These results also agree on sign with the mean annual difference computed from modeling efforts by

Keeling et al. [2004], which the authors found to be -3.6 per meg (see **Table 4.3**). Both curves reproduce the seasonal asymmetry of the interhemispheric gradient well, and the offset in y, showing a small ~3 month period in mid-late boreal summer when the interhemispheric gradient (NH-SH) flips positive.

I assess the uncertainty in the station and HIPPO annual mean gradients differently. For HIPPO data, I run 1000 Monte Carlo simulations of the seasonal interhemispheric annual mean difference by selecting 5-point sets at random from within normal distributions around the five HIPPO values, whose 1σ's are the error bars seen in the plot above. I fit two harmonics to these 1000 sets of 5 points and calculate the annual mean offset and standard deviation of these. In the case of the HIPPO values, I find the standard deviation to be ±4.6 per meg.

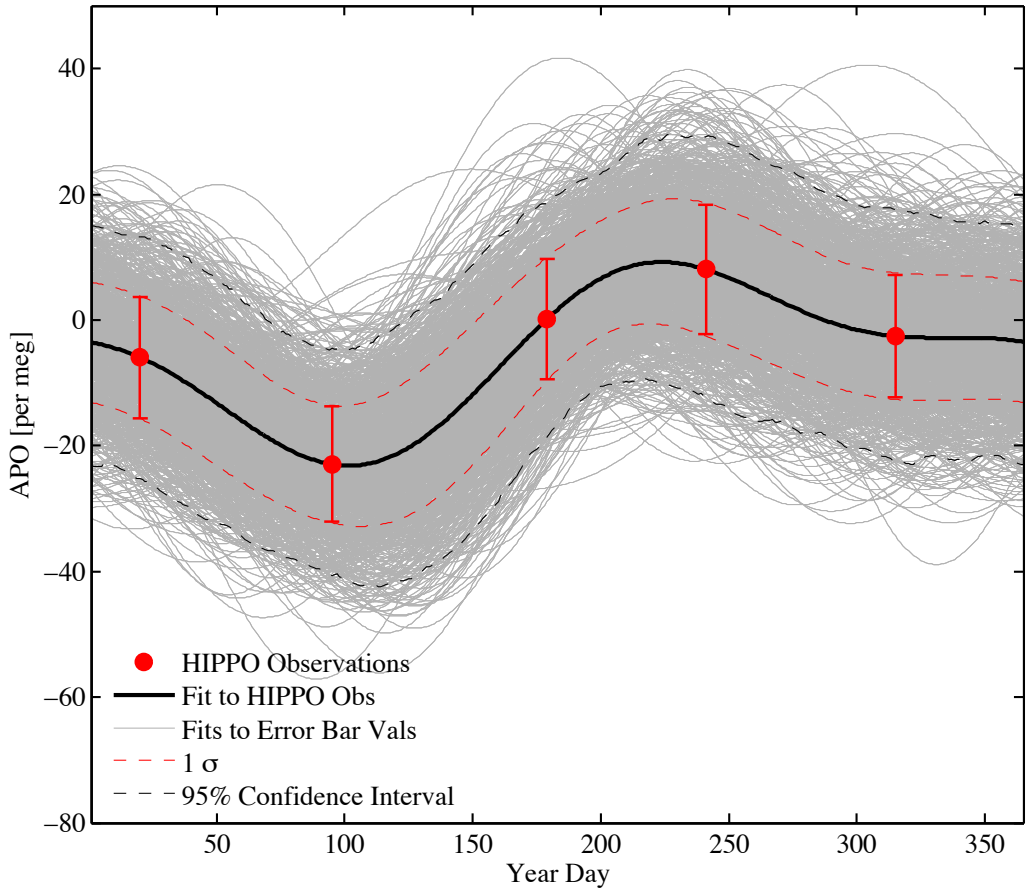

Figure 4.18. 1000 Monte Carlo simulations (thin gray lines) of the seasonal cycle of the interhemispheric Ar/N_2 difference between 55°N and 41°S, using HIPPO data. 55°N-41°S IH differences for each of the 5 HIPPO missions are shown as red circles with 1σ error bars representing the RMSE of the linear fit used to determine the IH difference. A 2-harmonic fit to these values is shown as a thick black line. A continuous 1σ error bar on the fit calculated from the Monte Carlo simulations is shown with a thin dashed red line, while 95% confidence limits are shown with thin dashed dark gray lines. These have been calculated by removing the top and bottom 2.5% of the 1000 runs at each of the 365 days.

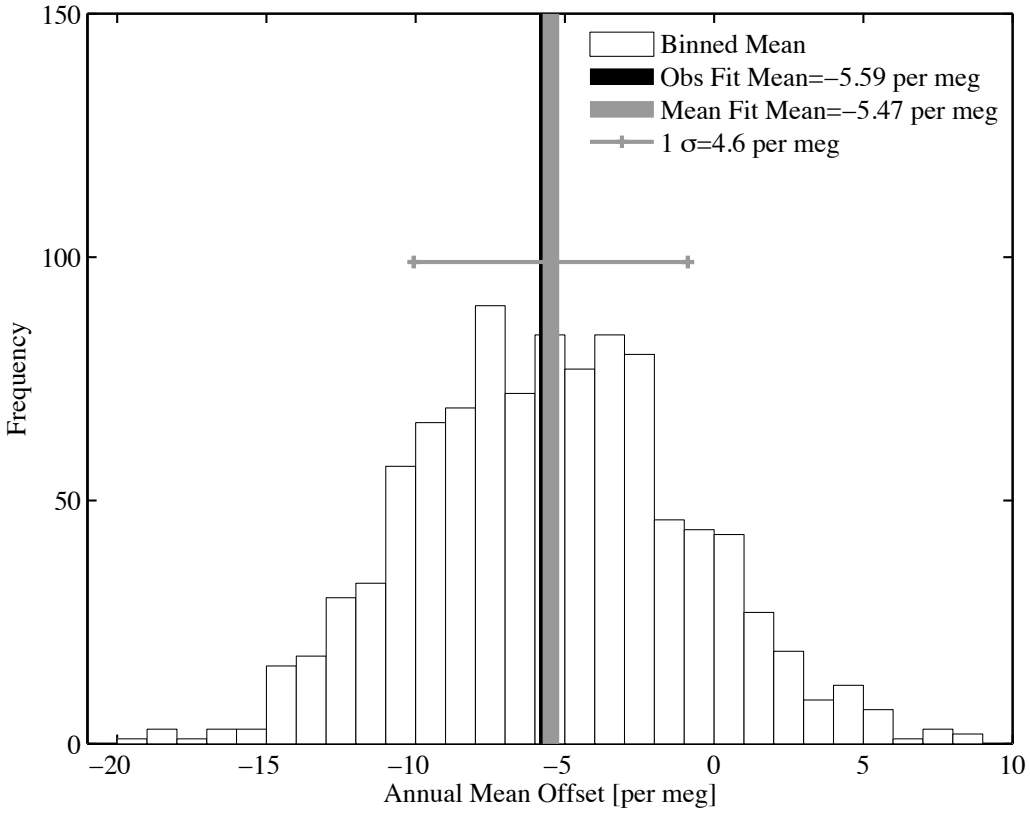

Figure 4.19. 1 per meg binned annual mean offset value (i.e. offset in y) for 1000 Monte Carlo simulations. Fit to observations is shown as a thick black line that is largely obscured by the mean of the 1000 simulations, shown as a thick gray line with a thin gray 1σ error bar. Monte Carlo simulations suggest the uncertainty in the annual mean offset value for the fit to the 5 HIPPO 55°N-41°S interhemispheric Ar/N_2 differences is ± 4.6 per meg.

In the case of station values, I run a similar set of Monte Carlo simulations, but using monthly resolution means of the CBA-CGO difference from 2006-2012. In this case, because 12 data points are provided for fitting, the uncertainty is lower, giving a 1σ of ±2.2 per meg.

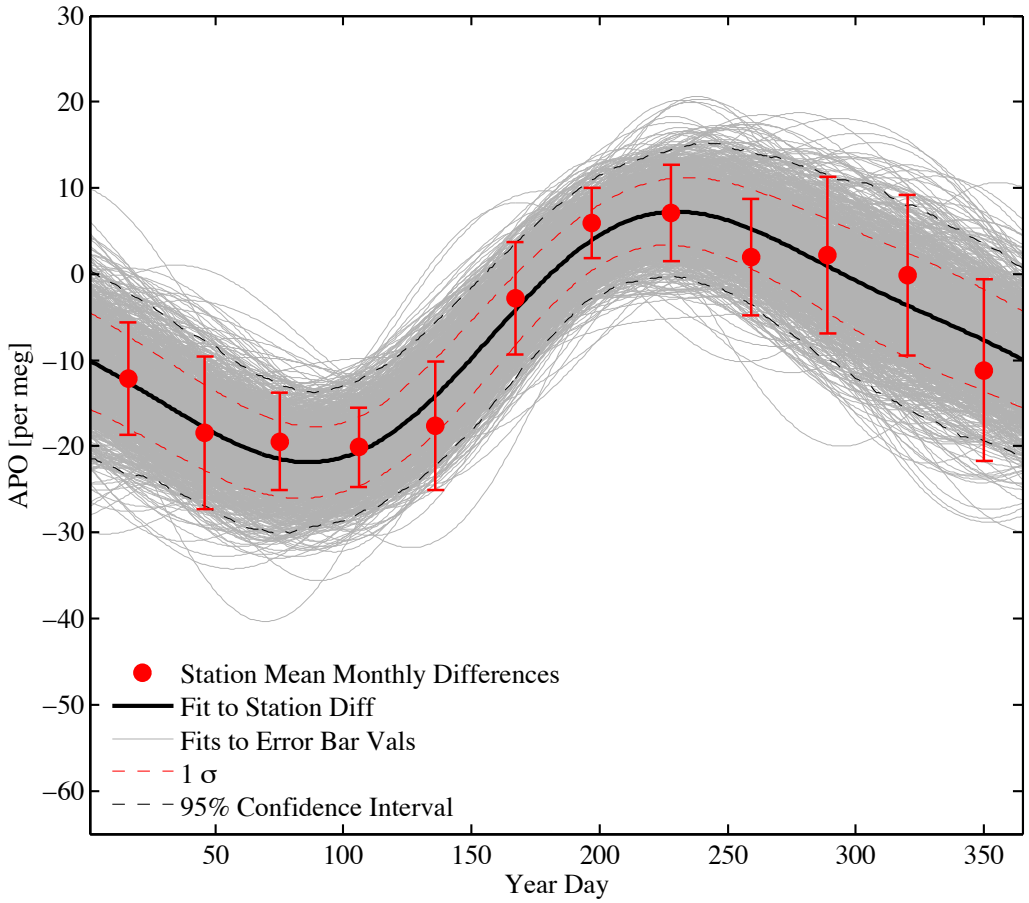

Figure 4.20. 1000 Monte Carlo simulations (thin gray lines) of the seasonal cycle of the interhemispheric Ar/N_2 difference between Cold Bay Station (55°N) and Cape Grim Observatory (41°S), using station data. CBA-CGO IH monthly mean differences are calculated as the mean difference between the two stations' monthly mean values for the period between 2006 and 2012. The standard deviation of these monthly mean values is shown as a red error bar. A 2-harmonic fit to these values is shown as a thick black line. A continuous 1σ error bar on the fit calculated from the Monte Carlo simulations is shown with a thin dashed red line, while 95% confidence limits are shown with thin dashed dark gray lines.

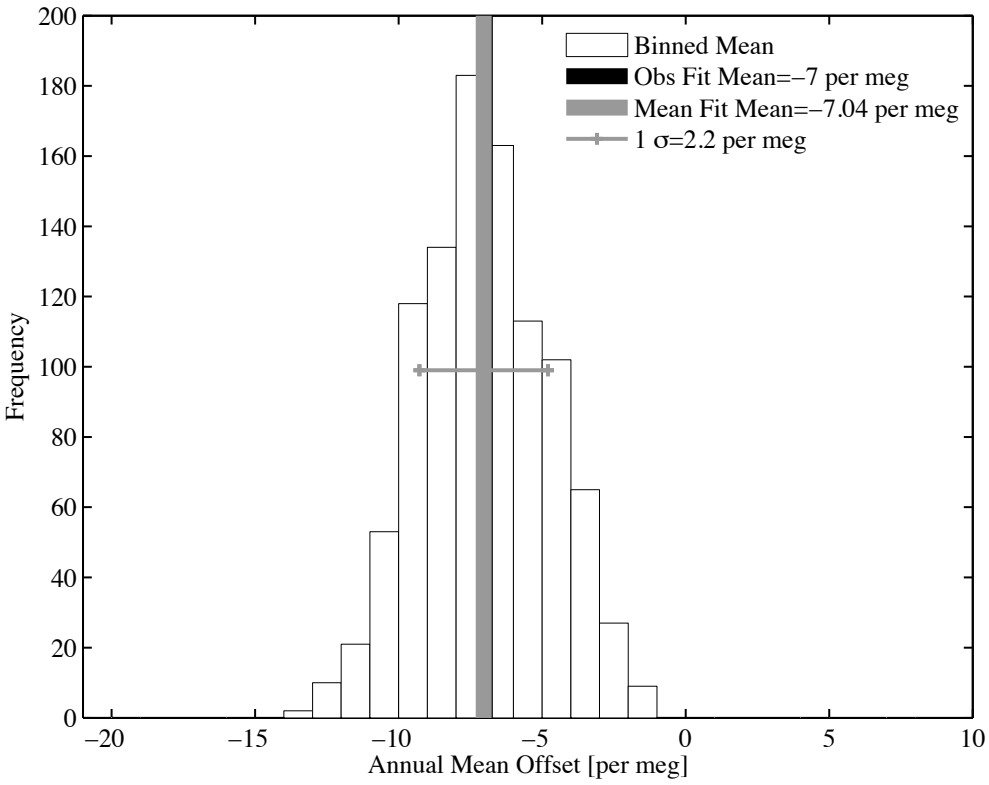

Figure 4.21. 1 per meg binned annual mean offset value (i.e. offset in y) for 1000 Monte Carlo simulations. Fit to station monthly mean differences (2006-2012) between Cold Bay and Cape Grim is shown as a thick black line that is largely obscured by the mean of the 1000 simulations, shown as a thick gray line with a thin gray 1σ error bar. Monte Carlo simulations suggest the uncertainty in the annual mean offset value for the fit to the CBA-CGO interhemispheric Ar/N_2 differences is 2.2 per meg.

In the case of Keeling et al. [2004], I have no estimate of the uncertainty in the annual mean offset, as the value is taken from model output, and as an uncertainty is not provided by the authors. I speculate that the uncertainty in the value may be on order twice the value itself, but I do not include such a number in the table below, as it has no basis in analysis. However, nominally such a number shows a sign (negative) that agrees with the other two measurement-based approaches, and is within the uncertainty of the HIPPO approach.

Table 4.3. Side-by-side comparison of the results of the CBA-CGO gradient fit for station data and 55°N – 41°S HIPPO MEDUSA flask data, and for Keeling et al. 2004. Values are for seasonal 2-harmonic fit amplitude, phase, and mean offset.

Model Run, Data	Amplitude [per meg]	Max YD	Min YD	Max [per meg]	Min [per meg]	Mean [per meg]
Station	29.1	230	86	7.2	-21.8	-7.0 ±2.2
Keeling et al. [2004]	n/a	n/a	n/a	n/a	n/a	-3.60
MEDUSA	32.4	224	102	9.2	-23.2	-5.59 ±4.6

The results show that the mean offset shown in station data fall within the error bars of the HIPPO data, supporting the assertion that, at least in the case of Cape Grim and Cold Bay, station-station offsets are probably primarily the result of true geographical gradients, not station sampler-sampler offsets and that, at least for near-surface flasks, MEDUSA does not suffer from significant latitudinal or seasonally varying sampling artifacts.

One possible explanation for the interhemispheric difference observed in both station and MEDUSA data involves the mechanics of ocean circulation, and particularly the transport of heat between the two hemispheres. North Atlantic Deep Water formation provides a net cold flux southward, while the return flow at surface provides a net warm flux northward [*Crowley*, 1992]. It seems reasonable to suggest that the warming of upwelled deep water in the Southern Hemisphere will preferentially cause Ar to be released to the atmosphere (relative to N_2) while the cooling of warmer surface water to form deep water will involve preferential uptake of Ar from the atmosphere. However, for the purpose of this study, I do not attempt to conclude further about this phenomenon, noting only that this, and further high-resolution meridional Ar/N_2 measurements might

conceivably provide an independent assessment of interhemispheric heat fluxes in the future.

4.7 Conclusion

Over the course of this chapter, I have investigated a number of new Ar/N_2 phenomena which until HIPPO were difficult or impossible to quantify. I have first shown that vertical gradients in the troposphere do not scale with airspeed or ram pressure, two major potential causes of fractionation according to recent studies. Furthermore, a small difference between tropical and extratropical flasks, and a larger difference between extratropical summer and winter flasks suggests that a correlation of depletion at elevation with less vigorous vertical tropospheric mixing is not unreasonable.

I have also shown that the vertical Ar/N_2 gradient in the troposphere is not well captured by the one-dimensional mutual molecular diffusivity equation of Lettau [1951], reproduced by Ishidoya et al. [2013], suggesting either that the model is not appropriate for determining diffusive gradients in the troposphere, or that gradients in the troposphere are due to another phenomenon, either natural or artificial. Vertical gradients in the stratosphere, allowing for differences in tropopause height, however, are reproduced well by the Lettau model forced with six different eddy-diffusion coefficient sets, further suggesting that the Ar/N_2 depletion I observe in the stratosphere is not artifact.

As a matter of interest for future studies of the O_2/N_2 vertical gradient, the model used in this chapter appears to predict a vertical gradient in O_2/N_2 that is roughly 1/2.84 times (0.352x) as strong as the observed gradient in Ar/N_2, a relationship that scales largely with the mass difference of Ar (M40) and O_2 (M32) when compared with N_2 (M28), i.e. 1/3, but which deviates slightly from this ratio due to differences in mutual

molecular diffusivity estimates and the fact that the scaling factors are located in the exponent of the equation.

I have also shown that HIPPO MEDUSA data are quite good at broadly representing the timing and amplitude of the seasonal cycle of Ar/N_2 at low elevations for both northern and southern hemispheres. Comparisons of station seasonal cycles between 2006 and 2012 and HIPPO data raise an important point for the Scripps O_2 Program—that Ar/N_2 measurement offsets exist between B-flasks and round-bottom flasks from station time series. I measure a 7.9 per meg offset between La Jolla round bottom and B-flasks from between 2010 and 2014, while a mean offset of 14.6 ±6.7 (Station-HIPPO) exists between the station round-bottoms and B flasks sampled on MEDUSA based on analysis by Dr. Britton Stephens of each of the 10 northbound and southbound legs of HIPPOs 1-5. The most likely reason for an offset between flask types alone is the T junction which delivers calibration gases to the B-flask rack, as this is a common source of offsets in analysis systems. However, this analysis requires further attention since the offset appears to be larger between station round-bottoms and HIPPO MEDUSA B-flasks which suggests an additional sampler-related offset may exist. Though the station-MEDUSA offset appears fairly constant throughout the HIPPO missions, it is not obvious to what extent sampling system architecture, flow rate, chilling mechanism, sampling temperature, and many other variables may contribute to increasing the baseline analytical offset of 7.9 per meg. I suggest that further investigation is necessary to see whether such an effect is fractionation-related, and whether an O_2/N_2 offset of 4.5 per meg (17/3.77) exists between HIPPO MEDUSA samples and station samples. Resolving

such questions will be important for future work, as aircraft data and station data are synthesized into a large, vertically integrated dataset for Ar/N_2 and O_2/N_2.

I have also shown that the northern hemisphere Ar/N_2 values for stations and HIPPO values are on average lower than those in the southern hemisphere. I have shown that the difference between two distant stations, Cold Bay Alaska (~55°N) and Cape Grim Observatory (~41°S), and its seasonal variability, is reproduced very well in HIPPO flasks, with very similar shapes, phases, amplitudes and annual mean offsets. The annual mean CBA-CGO station-station offset calculated by the two methods, of -7.0 ±2.2 per meg (station) and -5.59 ±4.6 per meg (HIPPO) is quite close, and agrees in sign and order of magnitude with the modeled estimate by Keeling et al. [2004], of -3.60 per meg. This suggests not only that mean offsets between the two stations (and possibly others) are primarily natural, not an artifact of sampler-sampler offsets, but also that there is a steady-state interhemispheric gradient that favors higher Ar/N_2 levels over the more oceanic southern hemisphere, possibly due to the steady-state ocean circulation between the hemispheres that transports heat from south to north.

These final observations could lead to an independent appraisal of interhemispheric oceanic heat transports given further investigation and modeling, so a next line of inquiry lies in modeling Ar/N_2 based on solubility relationships with heat fluxes [*R. F. Keeling and Shertz*, 1992], and based on the recent model-derived N_2:Ar flux scaling factors [*Manizza et al.*, 2012]. Though the Keeling and Shertz [1992] method will allow the calculation of seasonal Ar/N_2 ratios, a more serious investigation into mean conditions, using inversion techniques or possibly previously-calculated mean annual heat flux fields [*Gloor et al.*, 2001], will be required to make updated estimates of sinks

and sources of heat, to determine whether models are able to reproduce the interhemispheric gradient I see in MEDUSA and station data. To a limited extent, the findings from this chapter may offer an independent assessment of whether the Gloor et al. [2001] mean fields are accurately capturing the oceanic heat sinks and sources.

Appendix 1: HIPPO flight maps, supplemental images and figures

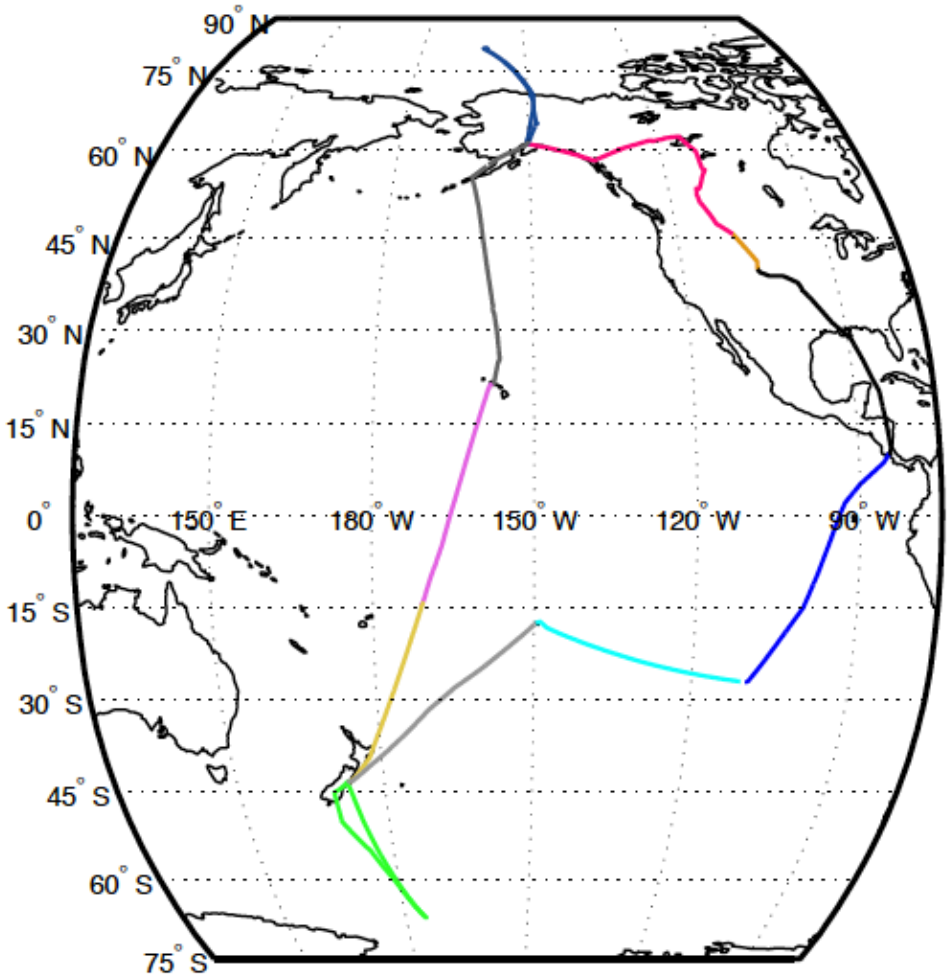

Figure A1-01. HIPPO1 Flight track. Research flights, starting in Colorado are RF01 (orange), RF02 (magenta), RF03 (gray blue), RF04 (dark gray), RF05 (pink), RF06 (mustard), RF07 (green), RF08 (light gray), RF09 (cyan), RF10 (dark blue), RF11 (black).

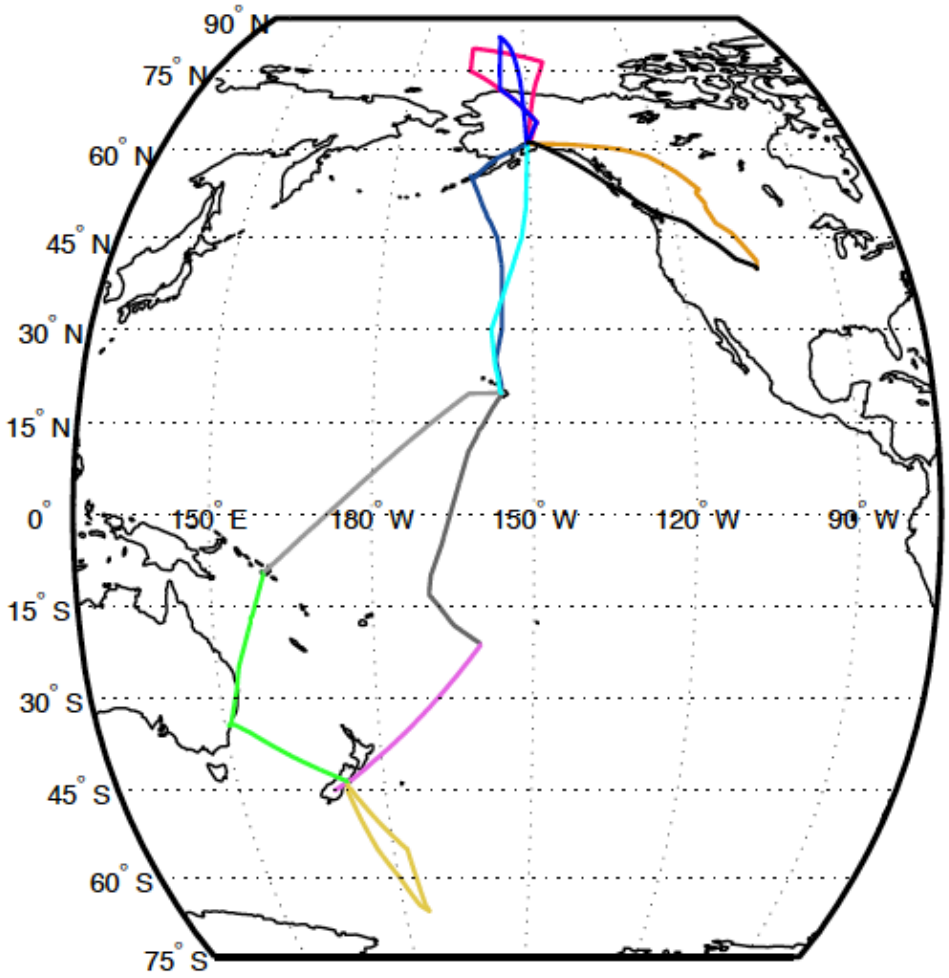

Figure A1-02. HIPPO2 Flight track. Research flights, starting in Colorado are RF01 (orange), RF02 (magenta), RF03 (gray blue), RF04 (dark gray), RF05 (pink), RF06 (mustard), RF07 (green), RF08 (light gray), RF09 (cyan), RF10 (dark blue), RF11 (black).

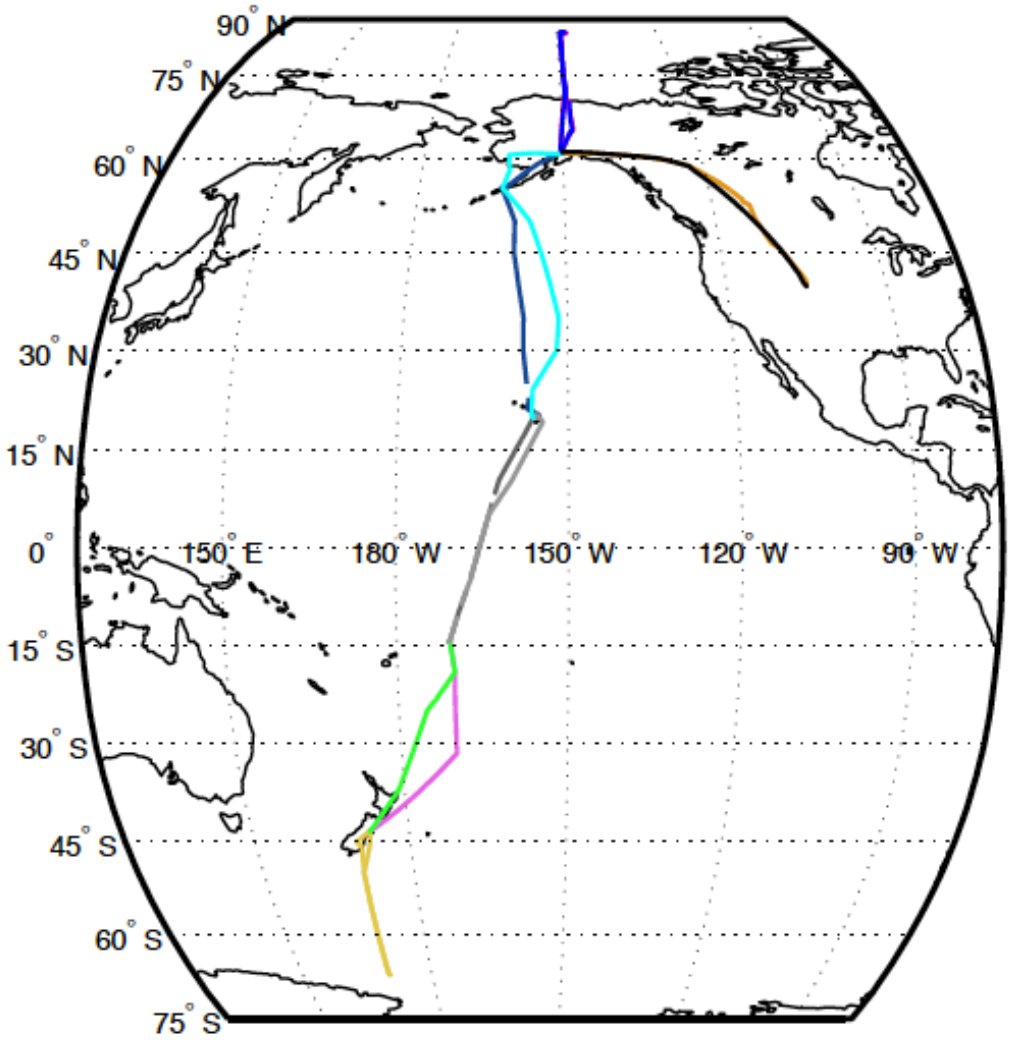

Figure A1-03. HIPPO3 Flight track. Research flights, starting in Colorado are RF01 (orange), RF02 (magenta), RF03 (gray blue), RF04 (dark gray), RF05 (pink), RF06 (mustard), RF07 (green), RF08 (light gray), RF09 (cyan), RF10 (dark blue), RF11 (black).

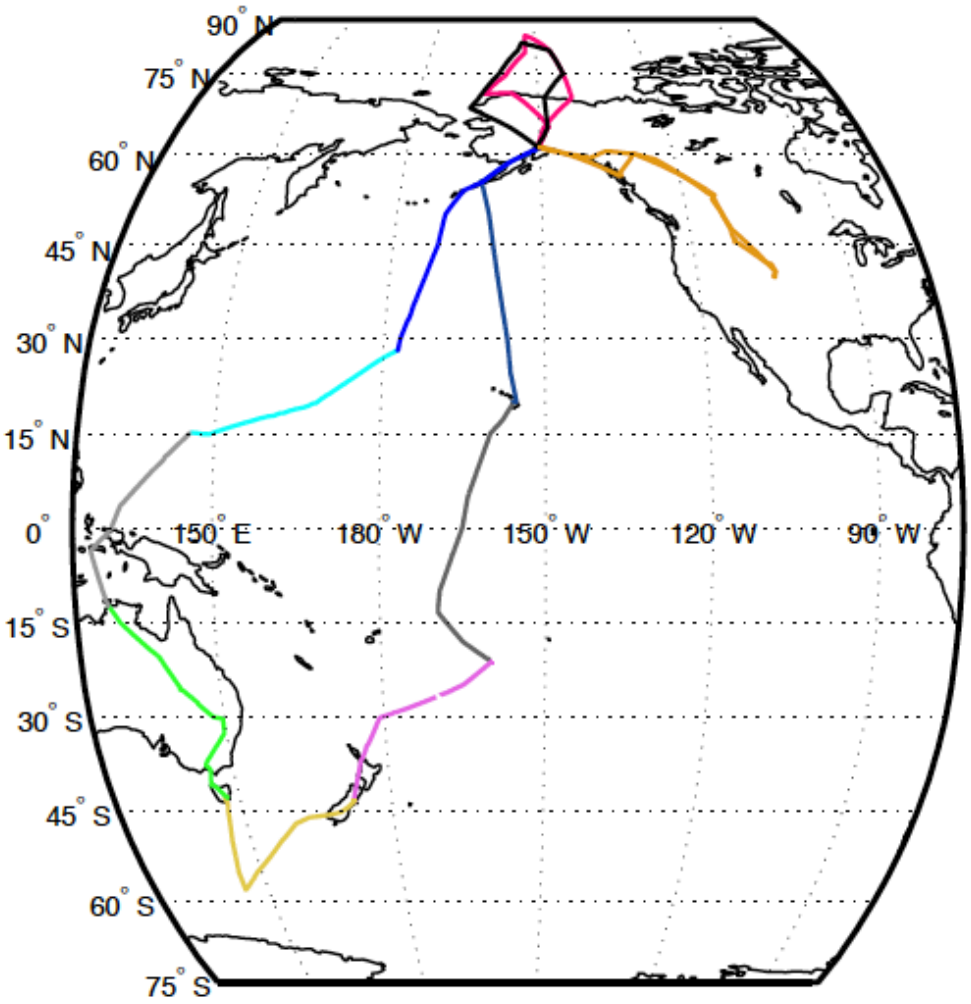

Figure A1-04. HIPPO4 Flight track. Research flights, starting in Colorado are RF01 (orange), RF02 (magenta), RF03 (gray blue), RF04 (dark gray), RF05 (pink), RF06 (mustard), RF07 (green), RF08 (light gray), RF09 (cyan), RF10 (dark blue), RF11 (black), RF12 (orange again).

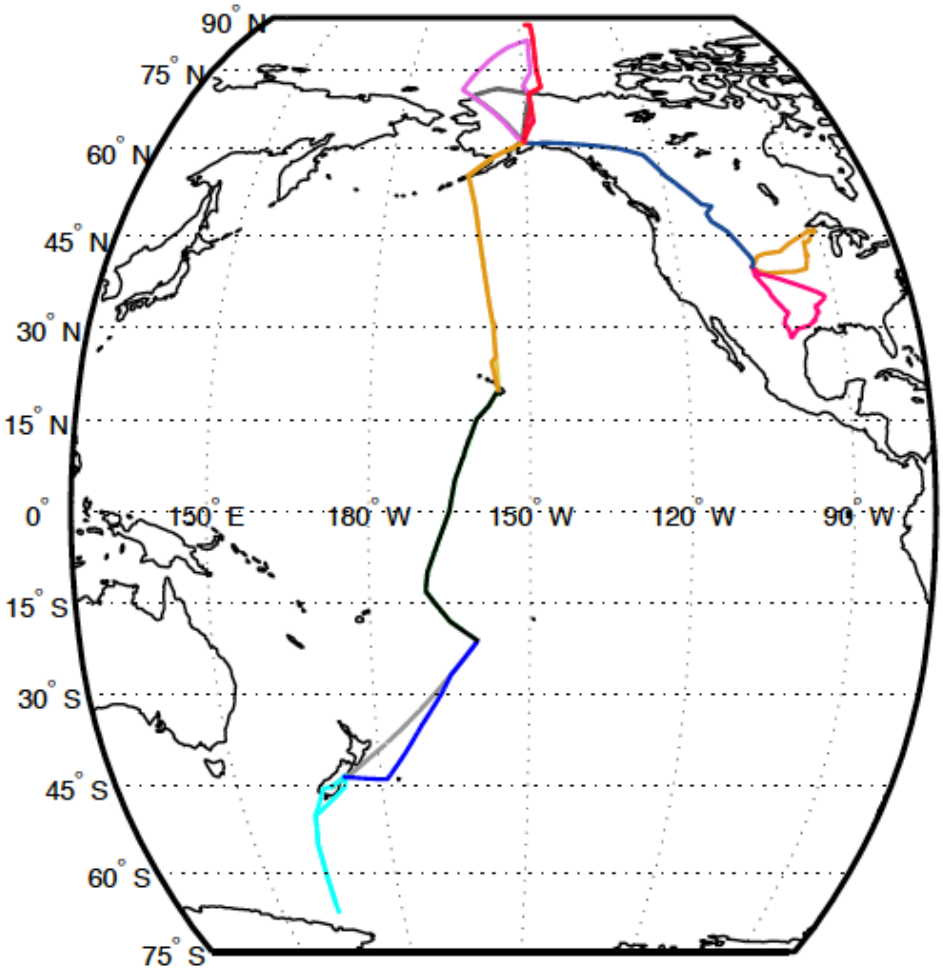

Figure A1-05. HIPPO5 Flight track. Research flights, starting in Colorado are RF01 (orange), RF02 (magenta), RF03 (gray blue), RF04 (dark gray, obscured), RF05 (pink), RF06 (mustard, obscured), RF07 (green, obscured), RF08 (light gray), RF09 (cyan), RF10 (dark blue), RF11 (black), RF12 (orange again), RF13 (magenta again), Ferry Flight 01 (not shown, but would be obscured by RF03).

Figure A1-06. HIMIL, or HIAPER Modular InLet, as seen disassembled in the laboratory. Under operating conditions, the inlet is placed on the belly or top of the aircraft's fuselage, as in figure A1-07. The streamlined top portion lowers down over the inlet lines (seen respectively in stainless steel and copper here, and the panel closes to protect the interior compartment. Air enters the inlet from the top left through a small aperture, and slows down considerably from its outside relative airspeed. Air is then directed into the forward-facing and rear-facing sample inlets. The HIMIL inlet aperture sits 12 inches off the body of the aircraft to avoid air from the aircraft's turbulent boundary layer, which contains vented or leaked cabin air, and the bottom panel of the inlet seals with a Viton O-ring to the fuselage to prevent outward leaks of cabin air into the inlet, which could then make their way into individual instrument sample inlets. Photograph courtesy of Dave Rodgers: https://www.eol.ucar.edu/instruments/hiaper-modular-inlet

Figure A1-07. The MEDUSA HIMIL seen atop the GV fuselage during the HIPPO campaign. Air passed from right to left in the image, entering the small aperture at the front of the red HIMIL. MEDUSA air was picked off from the rear-facing inlet to avoid the introduction of moisture to the sample line. Photograph: B. Stephens.

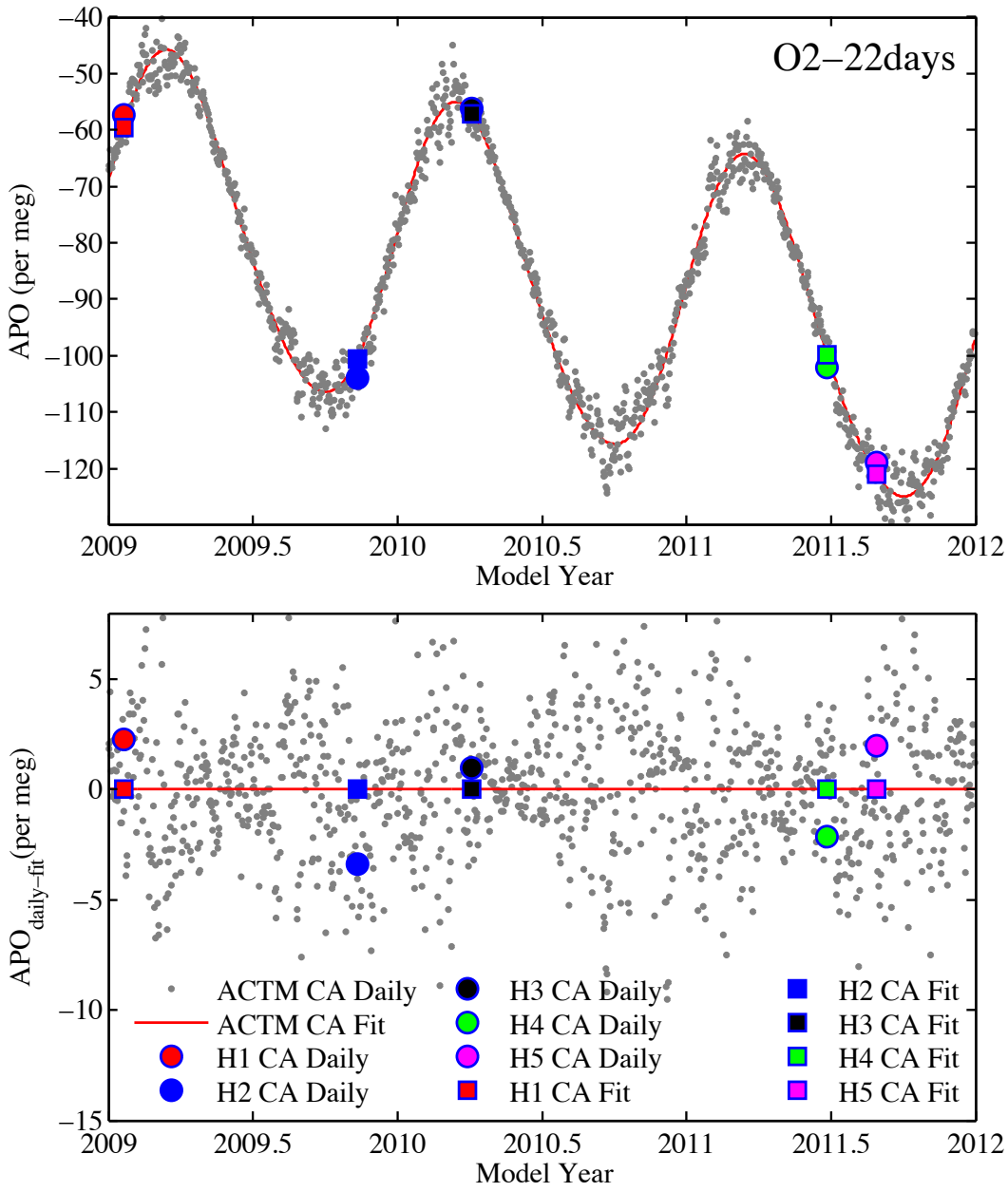

Figure A1-08. The same plot as Figure 3.14, but for 3-week (22 days) delayed O_2 fluxes. Bias estimates for each of the five HIPPO dates are similar to those in the previous plots, with the sign of all five biases the same. Units: [per meg].

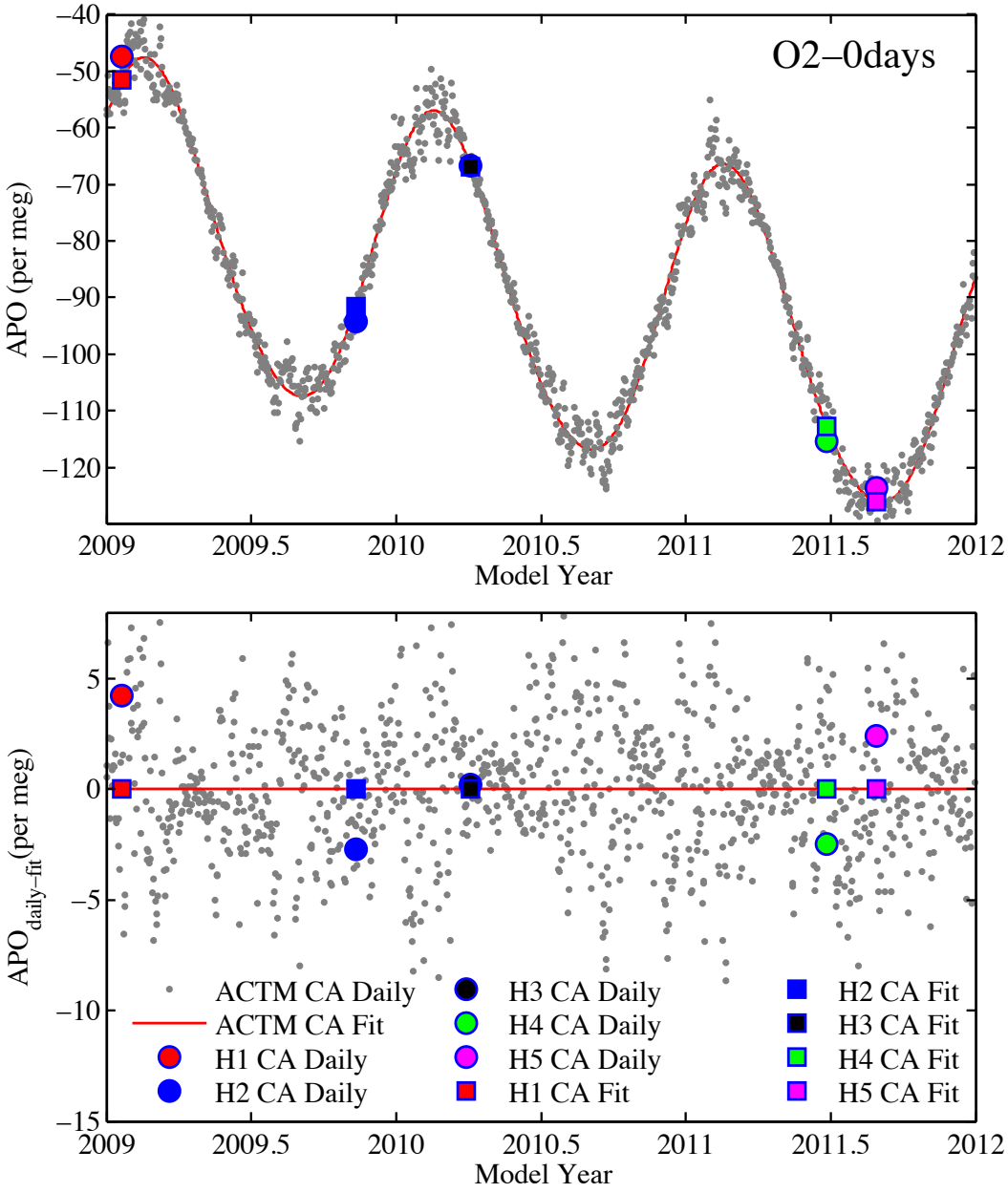

Figure A1-09. The same plot as Figure 3.14 and A1-08 but for the original (non-delayed) GK01 O_2 fluxes. Again, individual bias estimates are similar to those in the previous two plots, and the sign of the bias is consistent. The similarity in the three figures suggests that the meteorological conditions, not the timing of the fluxes, are the primary driver in determining the synoptic correction—consistent with my expectation, and the purpose of the bias estimate. Units: [per meg].

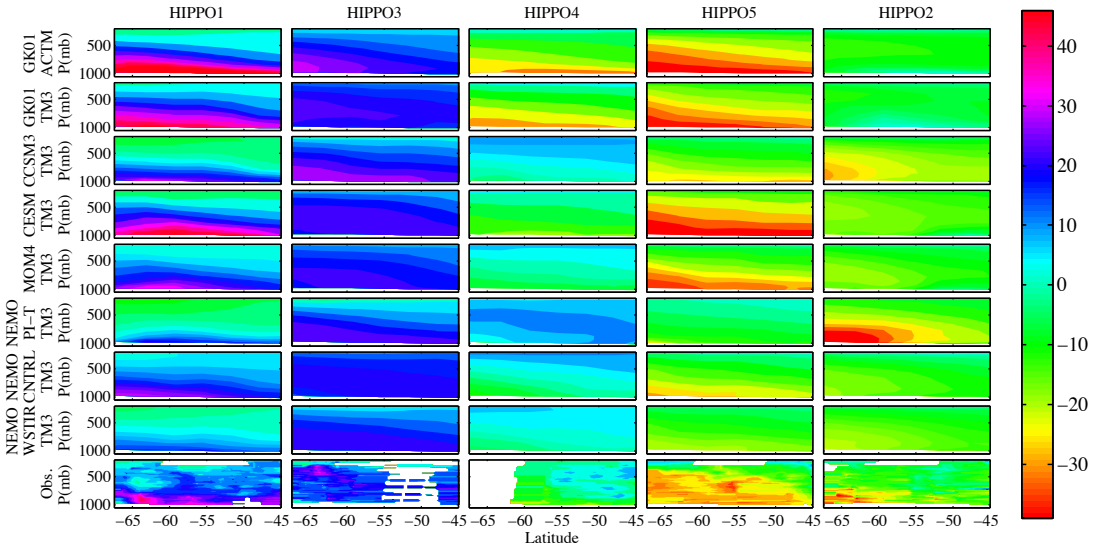

Figure A1-10. Comparison of APO (in per meg) data and model output on the dates of the five HIPPO Southern Ocean flights. Dissolved climatology runs are denoted “GK01” on the left. Output is interpolated to year day of HIPPO flight, and an average of the years between 1999 and 2003 (inclusive) is calculated to provide a climatological mean. All models and data are normalized by a linear fit to South Pole data. For models that did not output daily resolution, values are interpolated from the two closest weekly time stamps.

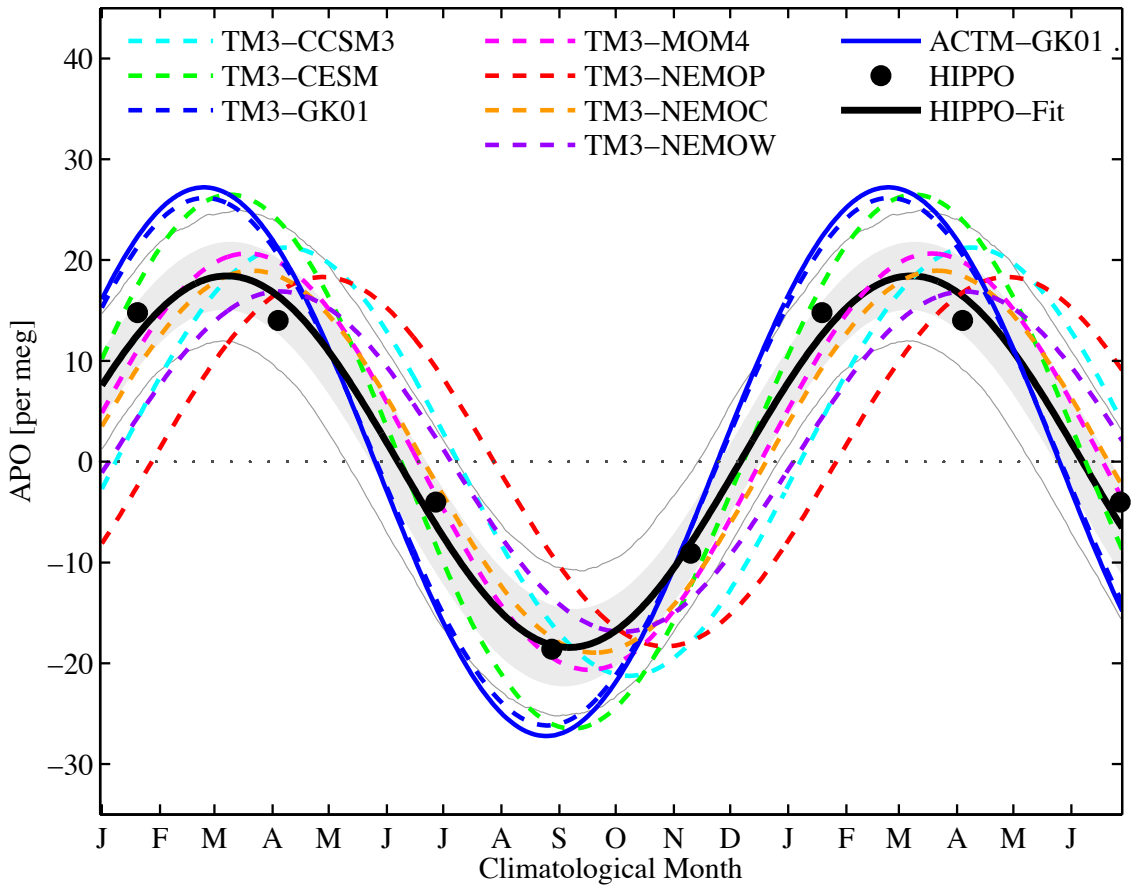

Figure A1-11. 18-month comparison of observed and modeled 1-harmonic Southern Ocean Meridional Curtain Average APO seasonal cycles. Shaded light gray error band represents the 1 sigma error (roughly ±6 per meg) on the continuous HIPPO fit amplitude, as determined by fitting 1 harmonic to 1000 instances of normally distributed randomly-generated points whose 1 sigma are the five HIPPO error bars. Thin light gray lines show 95% confidence limits of the fit. Units: [per meg].

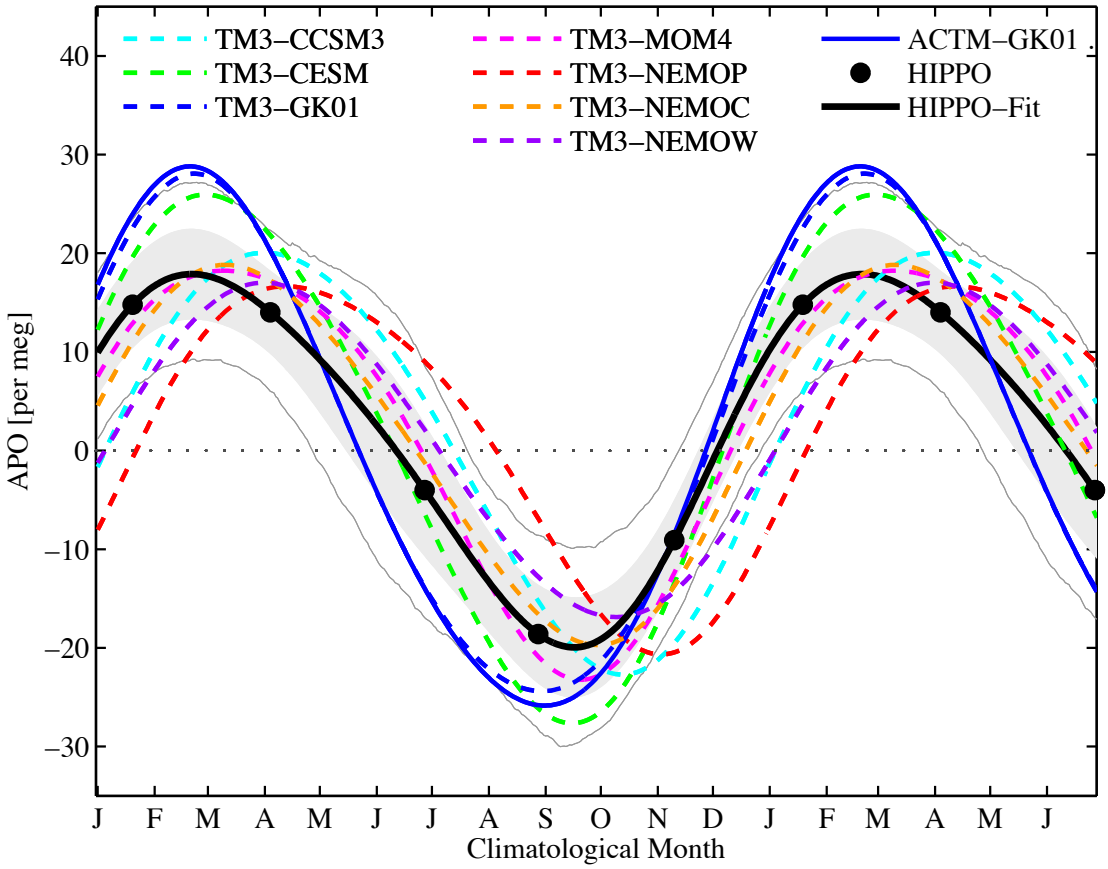

Figure A1-12. 18-month comparison of observed and modeled 2-harmonic Southern Ocean Meridional Curtain Average APO seasonal cycles. Shaded light gray error band represents the 1 sigma error (roughly ±6 per meg) on the continuous HIPPO fit amplitude, as determined by fitting 2 harmonics to 1000 instances of normally distributed randomly-generated points whose 1 sigma are the five HIPPO error bars. Thin light gray lines show 95% confidence limits of the fit. Units: [per meg].Units: [per meg].

Appendix 2: Identifying Ar/N_2 Offsets

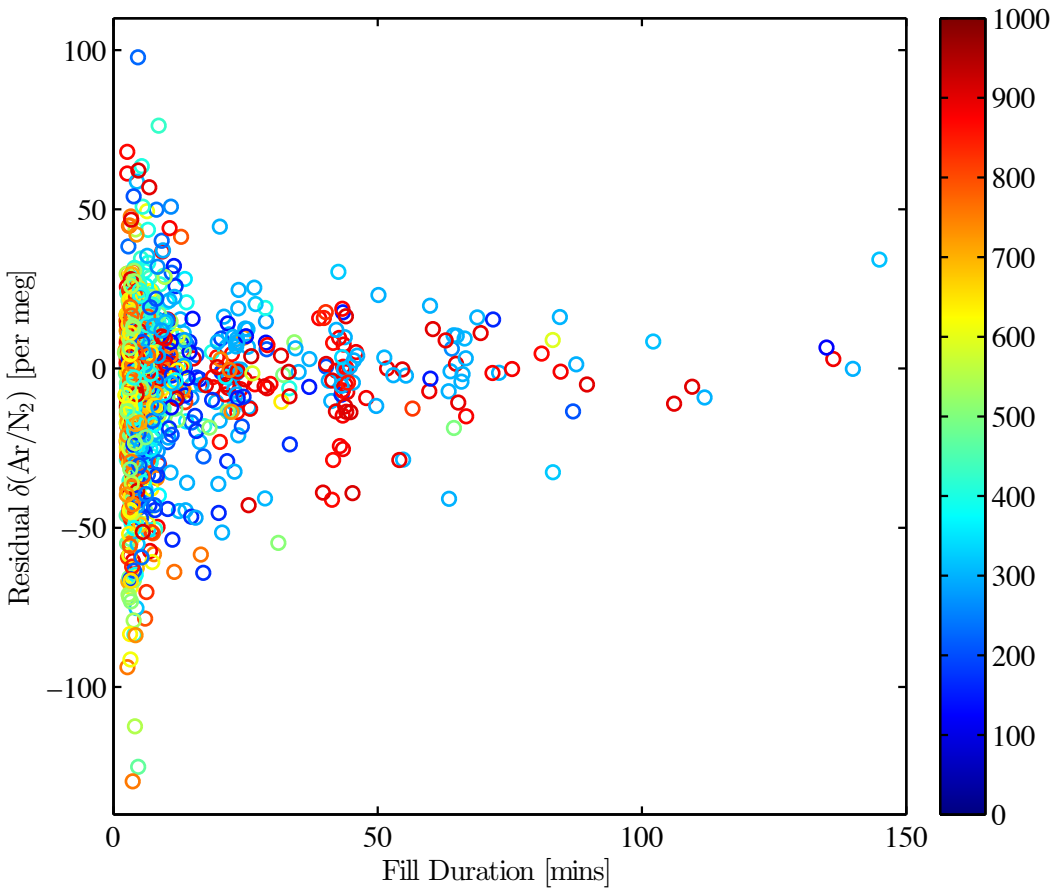

Figure A2-01. Ar/N_2 vs. fill duration. Pressure altitude in mbar (color axis) shown for context. The Ar/N_2 distribution is skewed negative because higher altitude flasks are sampled over a longer period of time at lower flow, and this figure does not attempt to remove the skewness because doing so would likely remove one or more of the processes upon which the Ar/N_2 ratio depends. The even distribution of Ar/N_2 values at longer fill durations suggests there is no dependence of Ar/N_2 on fill time.

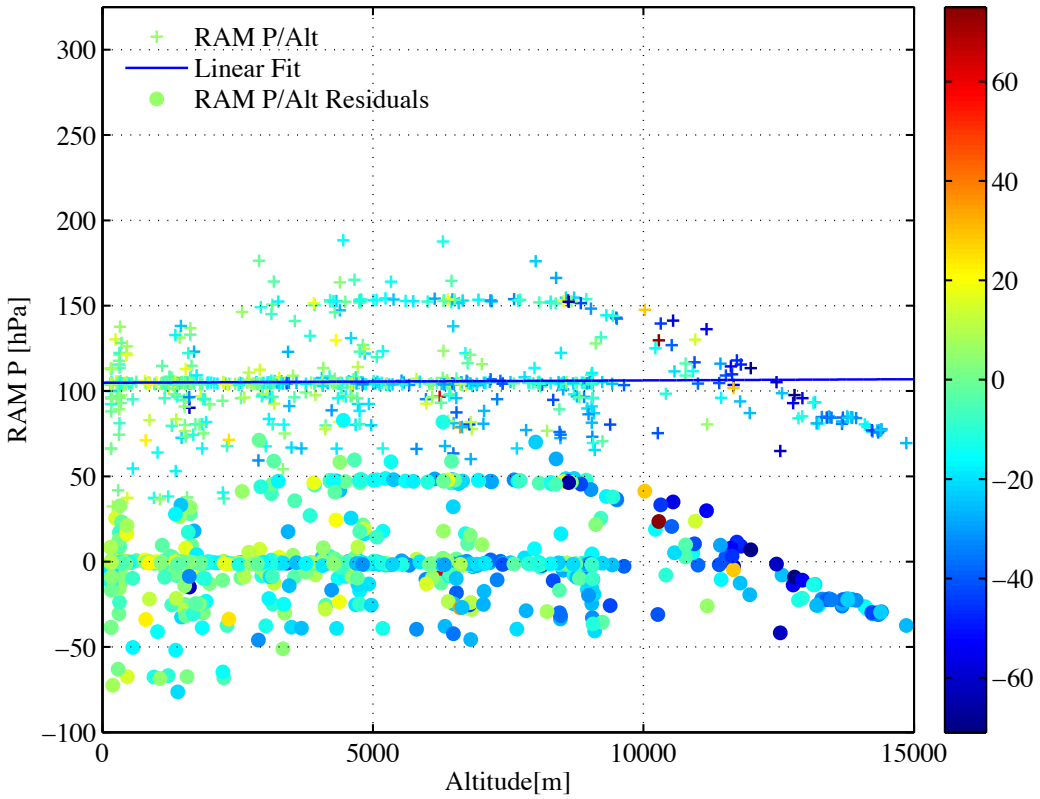

Figure A2-02. The relationship of altitude and ram pressure (hPa, variable QCXC) shown from the raw data (crosses), and with the mean ratio with elevation (blue line) removed (circles). $\delta(Ar/N_2)$ values in per meg are shown as color for context. In this instance, unlike in the relationship of airspeed and altitude, there is very little if any ram pressure correlation with altitude. Nonetheless, I remove the relationship in keeping with the technique in Section 4.3, and plot the residuals below in figure A2-03.

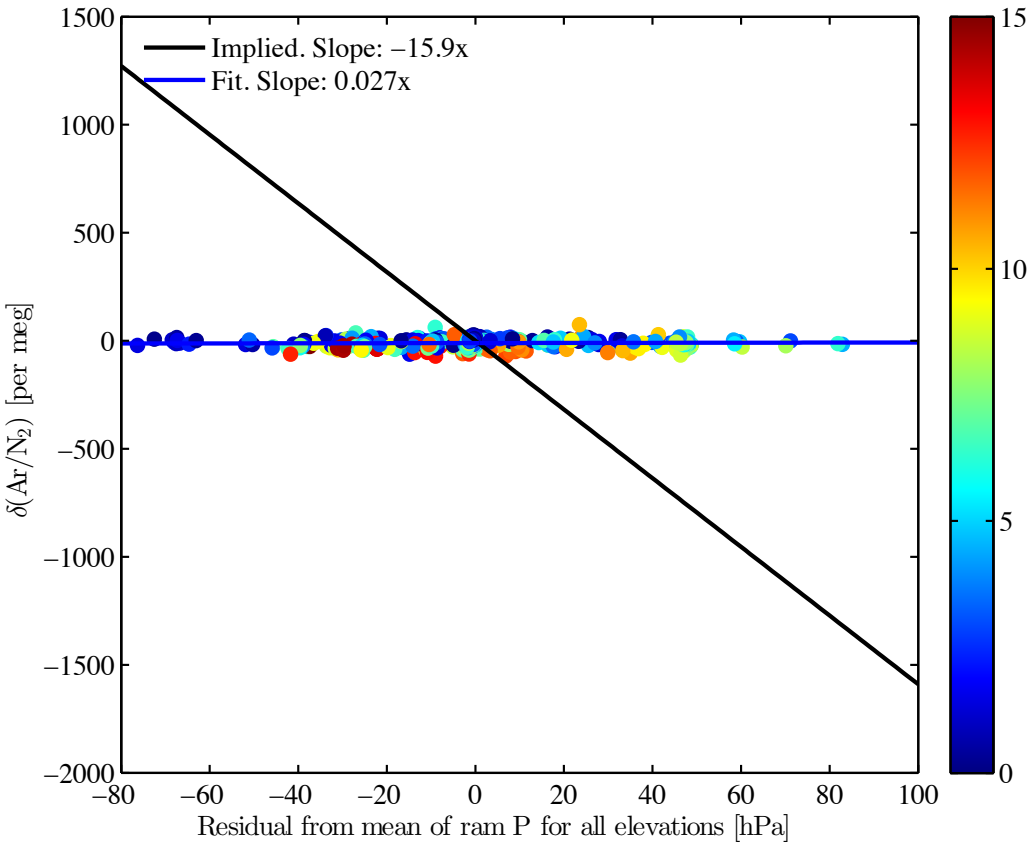

Figure A2-03. Ar/ N_2 plotted against the residuals of ram pressure (hPa) vs. altitude. If depletion correlated with ram pressure, a negative correlation of -15.9 per meg/hPa would be seen based on a similar analysis to that mentioned in Section 4.3 instead of the positive correlation of 0.027 per meg/hPa. Since there is very little correlation of ram pressure with altitude, as seen in the previous plot, this metric is less useful than with airspeed, as seen in Chapter 4. Figure A2-04 below serves as a more useful metric.

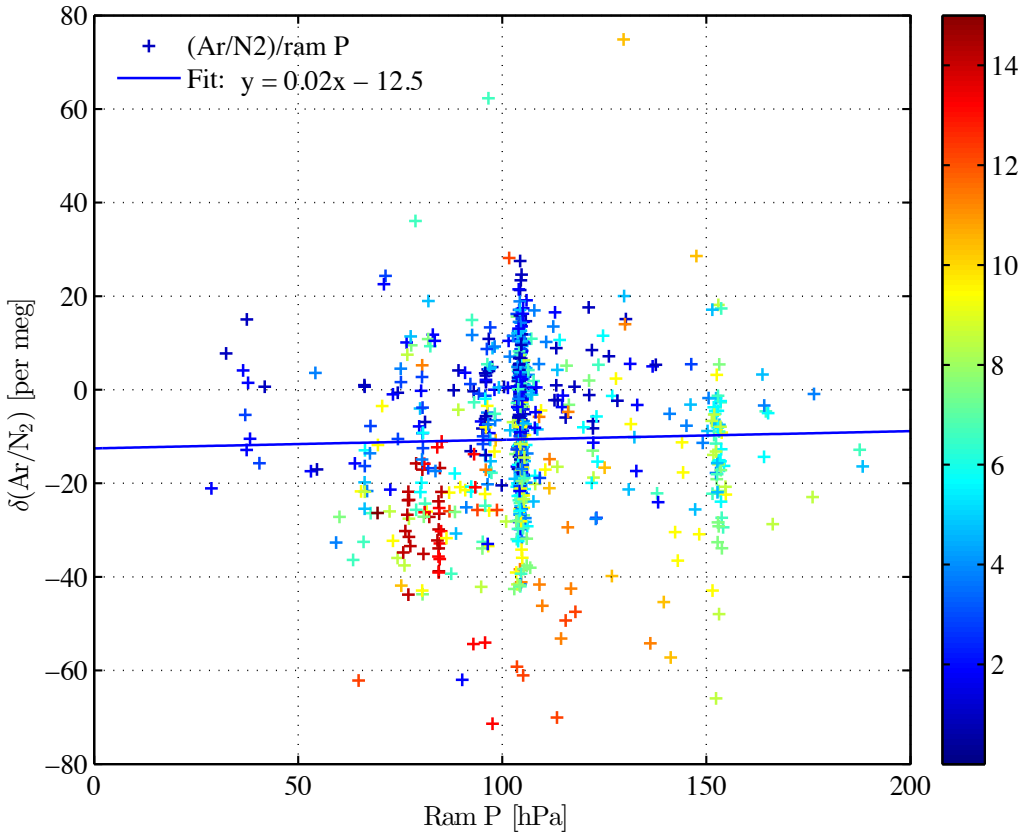

Figure A2-04. Box 2 tropospheric $\delta(Ar/N_2)$ plotted against ram pressure (hPa) with altitude (km) as the color axis. If ram pressure were directly responsible for a depletion in Ar/N_2, the relationship would be anti-correlated. Here I see a small, statistically insignificant positive correlation based on fit parameters, which suggests ram pressure has little effect on Ar/N_2 in the context of the HIPPO missions.

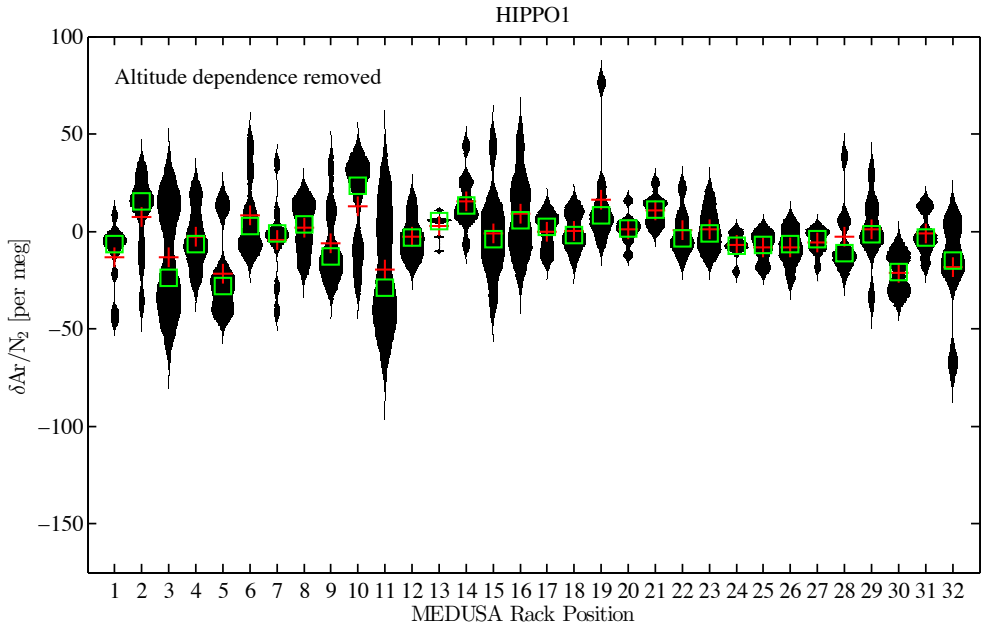

Figure A2-05. Residual scatter in tropospheric $\delta(Ar/N_2)$ by MEDUSA rack position for HIPPO1 with the altitude dependence removed. Mean value of flasks at each position is shown by a red cross, and median value by a green square. Of all five missions, this shows the least consistent depletion signal in early Box 1 flasks, with most odd position means low and evens high. This is the result of the original plumbing scheme in which even positions were plumbed diptube in. This arrangement was changed after HIPPO1 to standardize all flasks as diptube out.

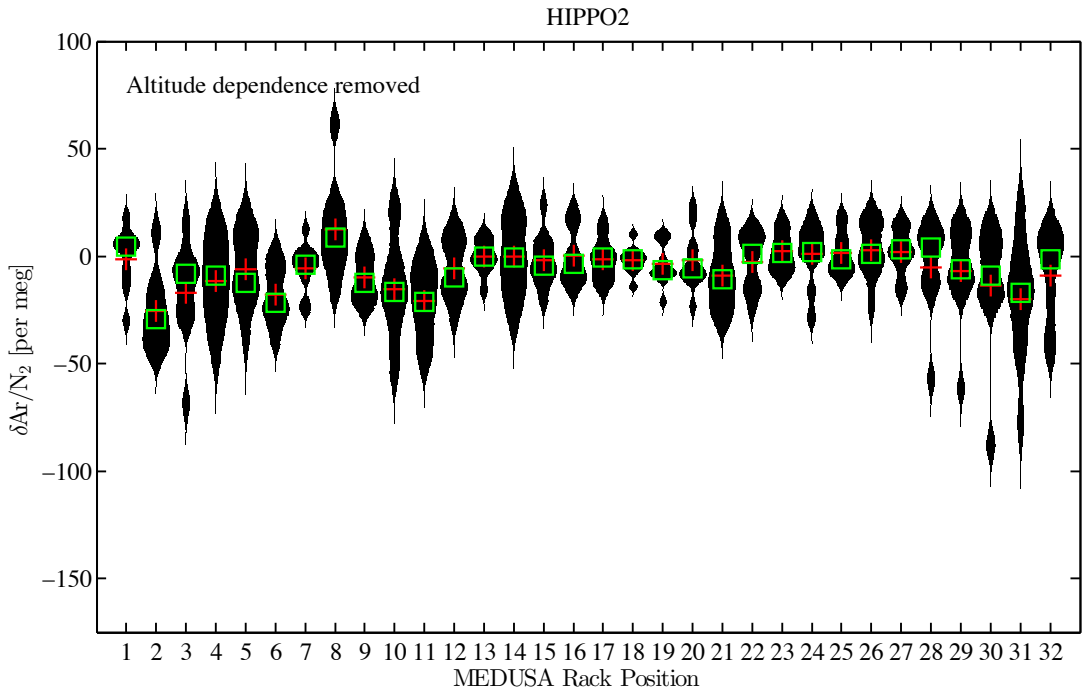

Figure A2-06. Residual scatter in tropospheric $\delta(Ar/N_2)$ by MEDUSA rack position for HIPPO2 with the altitude dependence removed. Mean value of flasks at each position is shown by a red cross, and median value by a green square. The improved plumbing scheme led to a more consistent mean value. By many measures, this was the best performing of the five HIPPO missions for $\delta(Ar/N_2)$.

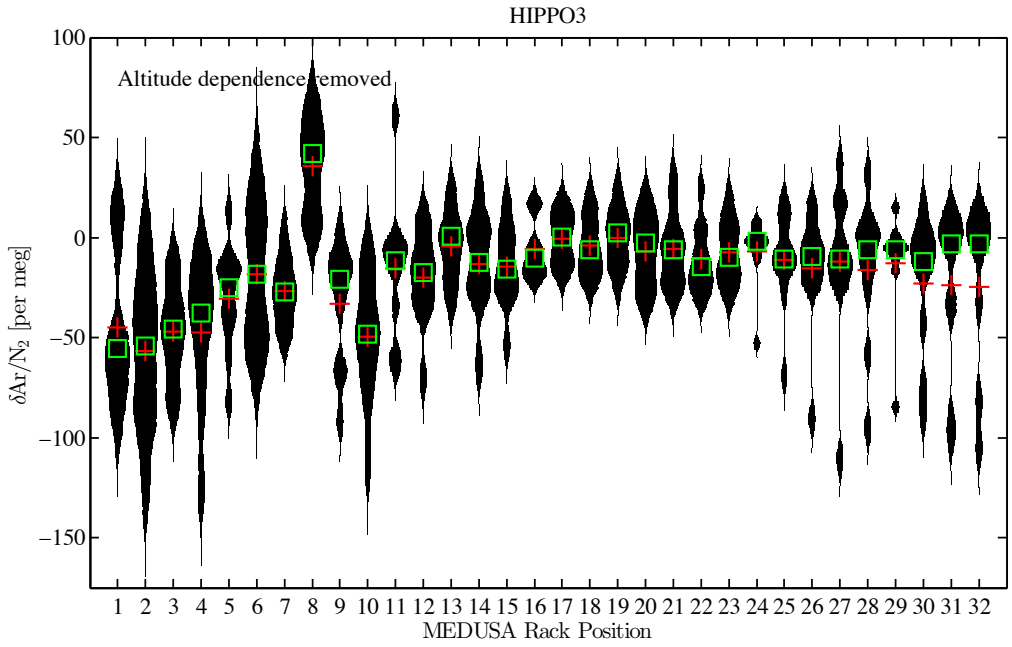

Figure A2-07. Residual scatter in tropospheric $\delta(Ar/N_2)$ by MEDUSA rack position for HIPPO3 with the altitude dependence removed. Mean value of flasks at each position is shown by a red cross, and median value by a green square. The strong depletion in early flasks, a common occurrence in the remaining HIPPO missions, is most pronounced here, and a strong enrichment in position 8 is due to an accidental backwards labeling (and thus plumbing) of the flask, one that persisted through HIPPO4.

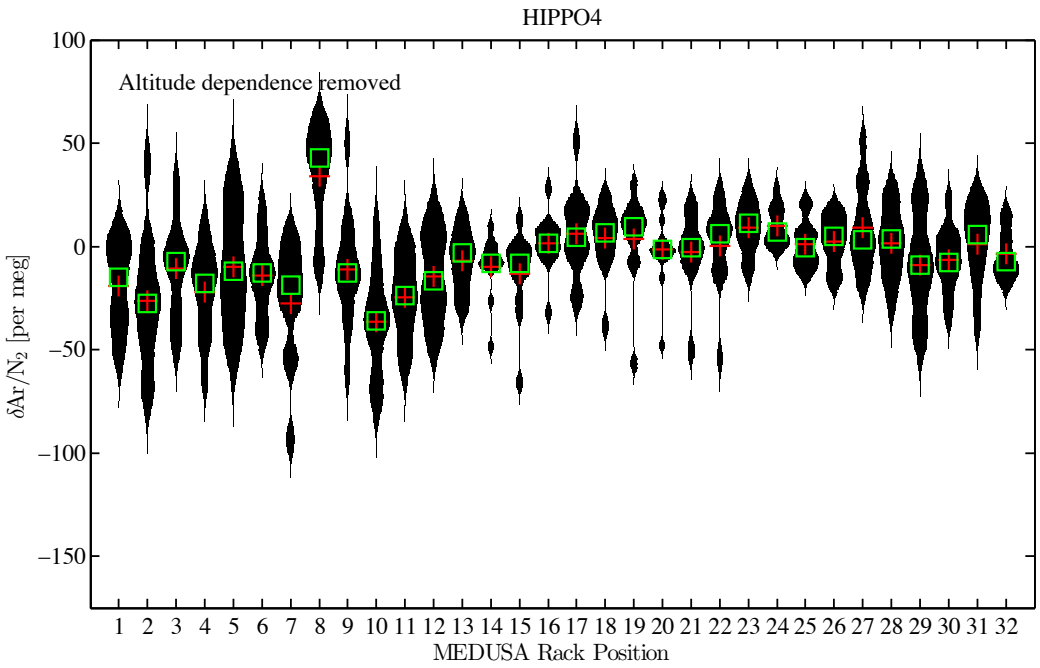

Figure A2-08. Residual scatter in tropospheric $\delta(Ar/N_2)$ by MEDUSA rack position for HIPPO4 with the altitude dependence removed. Mean value of flasks at each position is shown by a red cross, and median value by a green square. This mission is largely consistent with HIPPO3 and 5, but positions 1 and 2 seemed to perform better. This may be due to the policy instituted at the beginning of the mission of closing all flasks as soon as possible after sampling. A strong enrichment in position 8 is due to an accidental backwards labeling (and thus plumbing) of the flask, that began in HIPPO3.

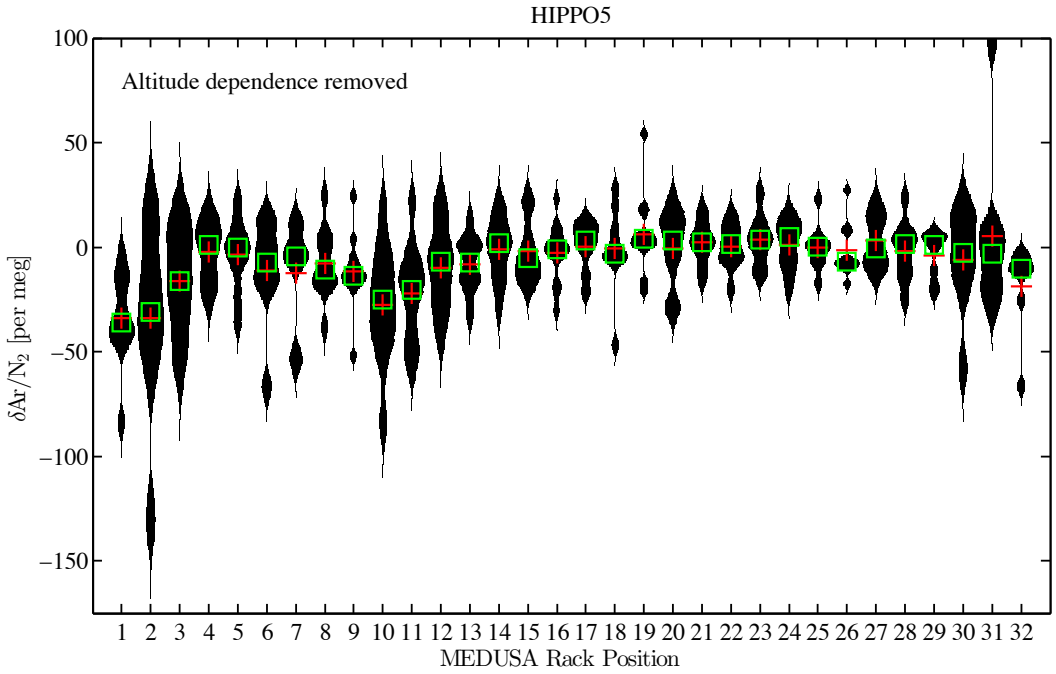

Figure A2-09. Residual scatter in tropospheric $\delta(Ar/N_2)$ by MEDUSA rack position for HIPPO1 with the altitude dependence removed. Mean value of flasks at each position is shown by a red cross, and median value by a green square. Similar again to HIPPOs 3 and 4, with dips centered on positions 1 and 10. This mission also showed an improvement on the extreme depletion seen on early flasks in HIPPO3 likely due to the policy of closing flasks as soon as they were sampled.

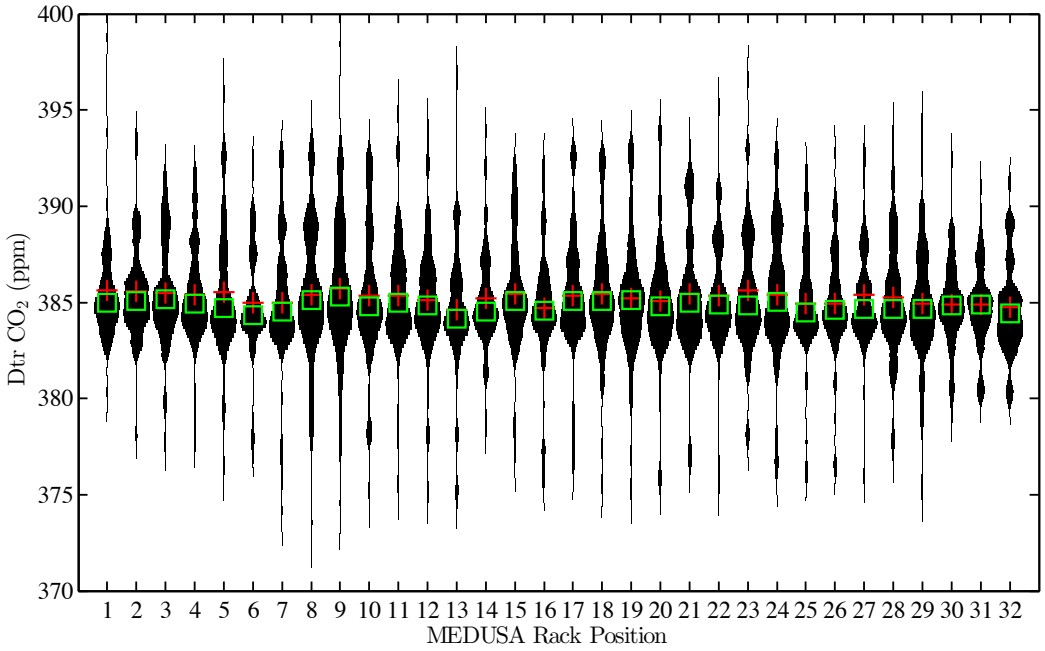

Figure A2-10. Detrended CO_2 values by position. Scatter is relatively consistent by position, and the most-affected positions for $\delta(Ar/N_2)$ (1-4, 9-11) show no systematic bias or scatter, suggesting that early flasks didn't experience cabin leaks.

.

Bibliography

Adachi, Y., K. Kawamura, L. Armi, and R. F. Keeling (2006), Diffusive separation of the lower atmosphere, *Science*, *311*(5766), 1429-1429.

Akima, H. (1978), A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points, *ACM Transactions on Mathematical Software (TOMS)*, *4*(2), 148-159.

Andrews, A., K. Boering, B. Daube, S. Wofsy, M. Loewenstein, H. Jost, J. Podolske, C. Webster, R. Herman, and D. Scott (2001), Mean ages of stratospheric air derived from in situ observations of CO2, CH4, and N2O, *Journal of Geophysical Research: Atmospheres (1984,2012)*, *106*(D23), 32295-32314.

Aumont, O., and L. Bopp (2006), Globalizing results from ocean in situ iron fertilization studies, *Global biogeochemical cycles*, *20*(2).

Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray (2003), An ecosystem model of the global ocean including Fe, Si, P colimitations, *Global biogeochemical cycles*, *17*(2).

Baker, D., R. M. Law, K. Gurney, P. Rayner, P. Peylin, A. Denning, P. Bousquet, L. Bruhwiler, Y. H. Chen, and P. Ciais (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988-2003, *Global biogeochemical cycles*, *20*(1), GB1002.

Battle, M., M. Bender, M. B. Hendricks, D. T. Ho, R. Mika, G. McKinley, S. Ä. Fan, T. W. Blaine, and R. F. Keeling (2003), Measurements and models of the atmospheric Ar/N2 ratio, *Geophysical Research Letters*, *30*(15).

Battle, M., S. Mikaloff-Fletcher, M. L. Bender, R. F. Keeling, A. C. Manning, N. Gruber, P. P. Tans, M. B. Hendricks, D. T. Ho, and C. Simonds (2006), Atmospheric potential oxygen: New observations and their implications for some atmospheric and oceanic models, *Global biogeochemical cycles*, *20*(1).

Bender, M. L., P. P. Tans, J. T. Ellis, J. Orchardo, and K. Habfast (1994), A high precision isotope ratio mass spectrometry method for measuring the O2/N2 ratio of air, *Geochimica et cosmochimica acta*, *58*(21), 4751-4758.

Blaine, T. W. (2005), Continuous measurements of atmospheric Ar/N2 as a tracer of air-sea heat flux: Models, methods, and data, PhD dissertation, 225 pp., Univ. of Calif., San Diego, La Jolla.

Blaine, T. W., R. Keeling, and W. Paplawsky (2006), An improved inlet for precisely measuring the atmospheric Ar/N2 ratio, *Atmos. Chem. Phys*, *6*, 1181-1184.

Brodeau, L., B. Barnier, A.-M. Treguier, T. Penduff, and S. Gulev (2010), An ERA40-based atmospheric forcing for global ocean circulation models, *Ocean Modelling*, *31*(3), 88-104.

Buitenhuis, E., C. Le Quéré, O. Aumont, G. Beaugrand, A. Bunker, A. Hirst, T. Ikeda, T. O'Brien, S. Piontkovski, and D. Straile (2006), Biogeochemical fluxes through mesozooplankton, *Global biogeochemical cycles*, *20*(2).

Cassar, N., G. A. McKinley, M. L. Bender, R. Mika, and M. Battle (2008), An improved comparison of atmospheric Ar/N2 time series and paired ocean‚Äêatmosphere model predictions, *Journal of Geophysical Research: Atmospheres (1984,2012)*, *113*(D21).

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990), STL: A seasonal-trend decomposition procedure based on loess, *Journal of Official Statistics*, *6*(1), 3-73.

Collins, W. D., C. M. Bitz, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, P. Chang, S. C. Doney, J. J. Hack, and T. B. Henderson (2006), The community climate system model version 3 (CCSM3), *Journal of Climate*, *19*(11), 2122-2143.

Craig, H., and R. Wiens (1996), Gravitational enrichment of 84Kr/36Ar ratios in polar ice caps: a measure of firn thickness and accumulation temperature, *Science*, *271*(5256), 1708-1710.

Crowley, T. J. (1992), North Atlantic deep water cools the southern hemisphere, *Paleoceanography*, *7*(4), 489-497.

Doney, S. C., D. M. Glover, and R. G. Najjar (1996), A new coupled, one-dimensional biological-physical model for the upper ocean: Applications to the JGOFS Bermuda Atlantic Time-series Study (BATS) site, *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2), 591-624.

Dunne, J. P., A. Gnanadesikan, J. L. Sarmiento, and R. D. Slater (2010), Technical description of the prototype version (v0) of tracers of phytoplankton with allometric zooplankton (TOPAZ) ocean biogeochemical model as used in the Princeton IFMIP model, *Biogeosciences*, *7*, 3593.

Garcia, H. E., and R. F. Keeling (2001), On the global oxygen anomaly and air-sea flux, *Journal of Geophysical Research. C. Oceans*, *106*, 31155-33166.

Gerlach, T. (2011), Volcanic versus anthropogenic carbon dioxide, *Eos, Transactions American Geophysical Union*, *92*(24), 201-202.

Gibson, J., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano (1997), ECMWF reanalysis project report series, 1. ERA description. European Centre for Medium-Range Weather Forecasts, *Reading/UK, 72pp*.

Gloor, M., N. Gruber, T. M. C. Hughes, and J. L. Sarmiento (2001), Estimating net air-sea fluxes from ocean bulk data: Methodology and application to the heat cycle, *Global biogeochemical cycles*, *15*(4), 767-782.

Griffies, S. M., C. Boning, F. O. Bryan, E. P. Chassignet, R. d. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier, and D. Webb (2000), Developments in ocean climate modelling, *Ocean Modelling*, *2*(3), 123-192.

Griffies, S. M., A. Gnanadesikan, K. W. Dixon, J. Dunne, R. d. Gerdes, M. J. Harrison, A. Rosati, J. Russell, B. L. Samuels, and M. J. Spelman (2005), Formulation of an ocean model for global climate simulations, *Ocean Science*, *1*(1), 45-79.

Gruber, N., M. Gloor, S. M. Fan, and J. L. Sarmiento (2001), Air-sea flux of oxygen estimated from bulk data: Implications for the marine and atmospheric oxygen cycles, *Global biogeochemical cycles*, *15*(4), 783-804.

Guenther, P., A. Bollenbacher, C. Keeling, and E. Stewart (2001), CALIBRATION l\ 1ETHODOLOGY FOR THE SCRIPPS 13c; 12c AND 18 0 I 16 0 STABLE ISOTOPE PROGRAM, 1996-2000.

Gurney, K. R., R. M. Law, A. S. Denning, P. J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y. U. H. A. N. CHEN, P. Ciais, and S. Fan (2003), TransCom 3 CO2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, *Tellus B*, *55*(2), 555-579.

Gurney, K. R., R. M. Law, P. Rayner, and A. Denning (2000), *Transcom 3 experimental protocol*, Department of Atmospheric Science, Colorado State University.

Hamme, R. C., and R. F. Keeling (2008), Ocean ventilation as a driver of interannual variability in atmospheric potential oxygen, *Tellus B*, *60*(5), 706-717, doi:10.1111/j.1600-0889.2008.00376.x.

Hansen, J., M. Sato, and R. Ruedy (1997), Radiative forcing and climate response, *JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES-*, *102*, 6831-6864.

Heimann, M., and S. Körner (2003), *The Global Atmospheric Tracer Model TM3: Model description and user's manual release 3.8 a*, MPI-BGC.

IPCC AR5 WGI (2013), IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change *Rep.*, 1535 pp. pp.

Ishidoya, S., S. Morimoto, S. Sugawara, T. Watai, T. Machida, S. Aoki, T. Nakazawa, and T. Yamanouchi (2008), Gravitational separation suggested by O2/N2, ∂ 15N of N2, ∂ 18O of O2, Ar/N2 observed in the lowermost part of the stratosphere at northern

middle and high latitudes in the early spring of 2002, *Geophysical Research Letters*, *35*(3).

Ishidoya, S., S. Sugawara, S. Morimoto, S. Aoki, T. Nakazawa, H. Honda, and S. Murayama (2013), Gravitational separation in the stratosphere, a new indicator of atmospheric circulation, *Atmospheric Chemistry and Physics*, *13*(17), 8787-8796.

Iwasaki, T., T. Maki, and K. Katayama (1998), Tracer transport model at Japan Meteorological Agency and its application to the ETEX data., *Atmos. Environ.*, *32*, 4285-4295.

Jin, X., R. G. Najjar, F. Louanchi, and S. C. Doney (2007), A modeling study of the seasonal oxygen budget of the global ocean.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter (2002), Ncep-doe amip-ii reanalysis (r-2), *Bulletin of the American Meteorological Society*, *83*(11), 1631-1643.

Keeling, C. D. (1960), The concentration and isotopic abundances of carbon dioxide in the atmosphere, *Tellus*, *12*(2), 200-203.

Keeling, C. D., S. C. Piper, and M. Heimann (1989), A three-dimensional model of atmospheric CO2 transport based on observed winds: 4. Mean annual gradients and interannual variations, *Geophysical Monograph Series*, *55*, 305-363.

Keeling, R. F. (1988), Development of an Interferometric Oxygen Analyzer for Precise, Doctoral thesis.

Keeling, R. F. (2013), bluemoon.ucsd.edu, edited.

Keeling, R. F., T. W. Blaine, B. Paplawsky, L. Katz, C. Atwood, and T. Brockwell (2004), Measurement of changes in atmospheric Ar/N2 ratio using a rapid-switching, single-capillary mass spectrometer system, *Tellus B*, *56*(4), 322-338.

Keeling, R. F., A. C. Manning, E. M. McEvoy, and S. R. Shertz (1998b), Methods for measuring changes in atmospheric O2 concentration and their application in southern hemisphere air, *Journal of geophysical research*, *103*(D3), 3381-3397.

Keeling, R. F., R. P. Najjar, M. L. Bender, and P. P. Tans (1993), What atmospheric oxygen measurements can tell us about the global carbon cycle, *Global biogeochemical cycles*, *7*(1), 37-67.

Keeling, R. F., S. C. Piper, and M. Heimann (1996), Global and hemispheric CO2 sinks deduced from changes in atmospheric O2 concentration, *Nature*, *381*(6579), 218-221.

Keeling, R. F., and S. R. Shertz (1992), Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, *Nature*, *358*(6389), 723-727.

Keeling, R. F., B. B. Stephens, R. G. Najjar, S. C. Doney, D. Archer, and M. Heimann (1998a), Seasonal variations in the atmospheric O2/N2 ratio in relation to the kinetics of air-sea gas exchange, *Global biogeochemical cycles*, *12*(1), 141-163.

Khatiwala, S., F. Primeau, and T. Hall (2009), Reconstruction of the history of anthropogenic CO2 concentrations in the ocean, *Nature*, *462*(7271), 346-349.

Kort, E. A., J. Eluszkiewicz, B. B. Stephens, J. B. Miller, C. Gerbig, T. Nehrkorn, B. C. Daube, J. O. Kaplan, S. Houweling, and S. C. Wofsy (2008), Emissions of CH4 and N2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, *Geophysical Research Letters*, *35*(18), L18808.

Langenfelds, R. L. (2002), Studies of the global carbon cycle using atmospheric oxygen and associated tracers, Doctoral thesis, University of Tasmania.

Large, W., and S. Yeager (2009), The global climatology of an interannually varying air-sea flux data set, *Climate Dynamics*, *33*(2-3), 341-364.

Law, R. M., R. J. Matear, and R. J. Francey (2008), Comment on" Saturation of the Southern Ocean CO2 Sink Due to Recent Climate Change", *Science*, *319*(5863), 570-570.

Law, R. M., and P. J. Rayner (1999), Impacts of seasonal covariance on CO2 inversions, *Global biogeochemical cycles*, *13*(4), 845-856.

Law, R. M., L. P. Steele, P. B. Krummel, and W. Zahorowski (2010), Synoptic variations in atmospheric CO2 at Cape Grim: a model intercomparison, *Tellus B*, *62*(5), 810-820.

Le Quéré, C., M. R. Raupach, J. G. Canadell, and G. Marland (2009), Trends in the sources and sinks of carbon dioxide, *Nature Geoscience*, *2*(12), 831-836.

Le Quéré, C., C. Rödenbeck, E. T. Buitenhuis, T. J. Conway, R. Langenfelds, A. Gomez, C. Labuschagne, M. Ramonet, T. Nakazawa, and N. Metzl (2007), Saturation of the Southern Ocean CO2 sink due to recent climate change, *Science*, *316*(5832), 1735-1738.

Lettau, H. (1951), Diffusion in the upper atmosphere, *Compendium of meteorology*, 320-333.

Levy, H., J. D. Mahlman, and W. J. Moxim (1982), Tropospheric N2O variability, *J. Geophys. Res.*, *87*, 3061-3080.

LiCor (1996), Li-6252 CO2 Analyzer Operating and Service Manual.

Long, M. C., K. Lindsay, S. Peacock, J. K. Moore, and S. C. Doney (2013), Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1 (BGC), *Journal of Climate*(2013).

Madec, G., P. Delecluse, M. Imbard, and C. Lévy (1998), OPA 8.1 ocean general circulation model reference manual, *Note du Pôle de modélisation, Institut Pierre-Simon Laplace*, *11*.

Mahlman, J., and W. Moxim (1978), Tracer simulation using a global general circulation model- Results from a midlatitude instantaneous source experiment, *Journal of the Atmospheric Sciences*, *35*(8), 1340-1374.

Maksyutov, S., and G. Inoue (2000), Vertical profiles of radon and CO2 simulated by the global atmospheric transport model, *CGER report, CGER-I039-2000, CGER, NIES, Japan*, *7*, 39-41.

Manizza, M., R. F. Keeling, and C. D. Nevison (2012), On the processes controlling the seasonal cycles of the air-sea fluxes of O 2 and N 2 O: A modelling study, *Tellus B*, *64*.

Manning, A. C. (2001), *Temporal variability of atmospheric oxygen from both continuous measurements and a flask sampling network: Tools for studying the global carbon cycle*.

Massie, S., and D. Hunten (1981), Stratospheric eddy diffusion coefficients from tracer data, *Journal of Geophysical Research: Oceans (1978,2012)*, *86*(C10), 9859-9868.

Mikaloff Fletcher, S. E., N. Gruber, A. R. Jacobson, S. C. Doney, S. Dutkiewicz, M. Gerber, M. Follows, F. Joos, K. Lindsay, and D. Menemenlis (2006), Inverse estimates of anthropogenic CO2 uptake, transport, and storage by the ocean, *Global biogeochemical cycles*, *20*(2).

Naegler, T., P. Ciais, J. C. Orr, O. Aumont, and C. Rodenbeck (2007), On evaluating ocean models with atmospheric potential oxygen, *Tellus B*, *59*(1), 138-156.

Naegler, T., P. Ciais, K. Rodgers, and I. Levin (2006), Excess radiocarbon constraints on air-sea gas exchange and the uptake of CO2 by the oceans, *Geophysical Research Letters*, *33*(11).

Najjar, R. G., and R. F. Keeling (1997), Analysis of the mean annual cycle of the dissolved oxygen anomaly in the World Ocean, *Journal of marine research*, *55*(1), 117-151.

Najjar, R. G., and R. F. Keeling (2000), Mean annual cycle of the air-sea oxygen flux: A global view, *Global biogeochemical cycles*, *14*(2), 573-584.

Nevison, C. D., N. M. Mahowald, S. C. Doney, I. D. Lima, and N. Cassar (2008), Impact of variable air-sea O2 and CO2 fluxes on atmospheric potential oxygen (APO) and land-ocean carbon sink partitioning, *Biogeosciences*, *5*(3), 875-889.

Newhall, C. G., and S. Self (1982), The Volcanic Explosivity Index (VEI) an estimate of explosive magnitude for historical volcanism, *Journal of Geophysical Research: Oceans (1978,Äì2012)*, *87*(C2), 1231-1238.

NOAA NGDC (2014), Significant Volcanic Eruptions Database, edited, National Geophysical Data Center, NOAA, doi:10.7289/V5JW8BSH.

Olivier, J. G., J. A. Peters, and G. Janssens-Maenhout (2012), *Trends in global co2 emissions 2012 report*, PBL Netherlands Environmental Assessment Agency.

Patra, P. K., S. Maksyutov, M. Ishizawa, T. Nakazawa, T. Takahashi, and J. Ukita (2005), Interannual and decadal changes in the sea-air CO2 flux from atmospheric CO2 inverse modeling, *Global biogeochemical cycles*, *19*(4), GB4013.

Rasch, P., N. Mahowald, and B. Eaton (1997), Yienger, JJ and Levy, H. II. 1995, *Empirical model of*.

Rodgers, K. B., O. Aumont, S. E. Mikaloff Fletcher, Y. Plancherel, L. Bopp, C. de Boyer Montégut, D. Iudicone, R. F. Keeling, G. Madec, and R. Wanninkhof (2014), Strong sensitivity of Southern Ocean carbon uptake and nutrient cycling to wind stirring, *Biogeosciences*, *11*(15), 4077-4098, doi:10.5194/bg-11-4077-2014.

Sabine, C. L., R. A. Feely, N. Gruber, R. M. Key, K. Lee, J. L. Bullister, R. Wanninkhof, C. Wong, D. W. Wallace, and B. Tilbrook (2004), The oceanic sink for anthropogenic CO2, *Science*, *305*(5682), 367-371.

Sabine, C. L., and T. Tanhua (2010), Estimation of anthropogenic CO2 inventories in the ocean, *Annual review of marine science*, *2*, 175-198.

Sallee, J.-B., R. Matear, S. Rintoul, and A. Lenton (2012), The physical transport of anthropogenic carbon into the Southern Ocean interior, paper presented at EGU General Assembly Conference Abstracts.

Sarmiento, J. L., T. M. Hughes, R. J. Stouffer, and S. Manabe (1998), Simulated response of the ocean carbon cycle to anthropogenic climate warming, *Nature*, *393*(6682), 245-249.

Severinghaus, J. P. (1995), Studies of the terrestrial O2 and carbon cycles in sand dune gases and in biospere 2, Doctoral thesis, Columbia University.

Smith, R., P. Jones, B. Briegleb, F. Bryan, G. Danabasoglu, J. Dennis, J. Dukowicz, C. Eden, B. Fox-Kemper, and P. Gent (2010), The parallel ocean program (POP) reference

manual: ocean component of the community climate system model (CCSM), *Los Alamos National Laboratory, LAUR-10-01853*.

Steinbach, J. (2010), Enhancing the Usability of Atmospheric Oxygen Measurements Through Emission Source Characterization and Airborne Measurements, Doctoral thesis, MPI-BGC.

Stephens, B. B., G. Brailsford, A. Gomez, K. Riedel, S. Mikaloff Fletcher, S. Nichol, and M. Manning (2013), Analysis of a 39-year continuous atmospheric CO2 record from Baring Head, New Zealand, *Biogeosciences*, *10*(4), 2683-2697.

Stephens, B. B., K. R. Gurney, P. P. Tans, C. Sweeney, W. Peters, L. Bruhwiler, P. Ciais, M. Ramonet, P. Bousquet, and T. Nakazawa (2007), Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2, *Science*, *316*(5832), 1732-1735.

Stephens, B. B., R. F. Keeling, M. Heimann, K. D. Six, R. Murnane, and K. Caldeira (1998), Testing global ocean carbon cycle models using measurements of atmospheric O2 and CO2 concentration, *Global biogeochemical cycles*, *12*(2), 213-230, doi:10.1029/97gb03500.

Stephens, B. B., R. F. Keeling, and W. J. Paplawsky (2003), Shipboard measurements of atmospheric oxygen using a vacuum-ultraviolet absorption technique, *Tellus B*, *55*(4), 857-878, doi:10.1046/j.1435-6935.2003.00075.x.

Taguchi, S. (1996), A three-dimensional model of atmospheric CO2 transport based on analyzed winds: Model description and simulation results for TRANSCOM, *Journal of geophysical research*, *101*(D10), 15099-15015,15109.

Takahashi, T., S. C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R. A. Feely, and C. Sabine (2002), Global sea-air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects, *Deep Sea Research Part II: Topical Studies in Oceanography*, *49*(9), 1601-1622.

Takahashi, T., S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. Chavez, and C. Sabine (2009), Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8), 554-577.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, *93*(4).

Tohjima, Y. (2000), Method for measuring changes in the atmospheric O2/N2 ratio by a gas chromatograph equipped with a thermal conductivity detector, *Journal of Geophysical Research: Atmospheres (1984,2012)*, *105*(D11), 14575-14584.

Uppala, S. M., P. Kållberg, A. Simmons, U. Andrae, V. Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, and G. Kelly (2005), The ERA-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961-3012.

Wanninkhof, R. (1992), Relationship between wind speed and gas exchange, *J. geophys. Res*, *97*(25), 7373-7382.

Wanninkhof, R., and W. R. McGillis (1999), A cubic relationship between air-sea CO2 exchange and wind speed, *Geophysical Research Letters*, *26*(13), 1889-1892, doi:10.1029/1999gl900363.

Weiss, R. (1970), The solubility of nitrogen, oxygen and argon in water and seawater, paper presented at Deep Sea Research and Oceanographic Abstracts, Elsevier.

Wofsy, S. C. (2011), HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 2073-2086.

Zickfeld, K., J. C. Fyfe, M. Eby, and A. J. Weaver (2008), Comment on "Saturation of the Southern Ocean CO2 Sink Due to Recent Climate Change", *Science*, *319*(5863), 570, doi:10.1126/science.1146886.

Errata:

p. 109, line 10. The text should read: “The **seventh** TM3 run uses…”

p. 137, Figure 3.19. In-figure text should read “HIPPO Obs Adjusted Fit = **37.8** per meg”

p. 147, Figure 3.25, line 1 of caption. The text should read “last two points are repeats of first **two**”