Mechanistic Habitat Modeling with Multi-Model Climate Ensembles

A practical guide to preparing and using Global Climate Model output for mechanistically modeling habitat, as exemplified by the modeling of the fundamental niche of a pagophilic Pinniped species through 2100.

Masters project submitted in partial fulfillment of the requirements for the Master of Environmental Management degree in the Nicholas School of the Environment of Duke University, May 2013.

Author: ***Hunter Jones*** *Adviser:* ***Dr. David Johnston***

Abstract

Projections of future Sea Ice Concentration (SIC) were prepared using a 13-member ensemble of climate model output from the Coupled Model Inter-comparison Project (CMIP5). Three climate change scenarios (RCP 2.6, RCP 6.0, RCP 8.5), corresponding to low, moderate, and high climate change possibilities, were used to generate these projections for known Harp Seal whelping locations. The projections were splined and statistically downscaled via the CCAFS Delta method using satellite-derived observations from the National Sea Ice Data Center (NSIDC) to prepare a spatial representation of sea ice decline through the year 2100.

Multi-Model Ensemble projections of the mean sea ice concentration anomaly for Harp Seal whelping locations under the moderate and high climate change scenarios (RCP 6.0 and RCP 8.5) show a decline of 10% to 40% by 2100 from a modern baseline climatology (average of SIC, 1988 - 2005) while sea ice concentrations under the low climate change scenario remain fairly stable. Projected year-over-year sea ice concentration variability decreases with time through 2100, but uncertainty in the prediction (model spread) increases. The general decline in sea ice projected by climate models is detrimental to Harp Seal survival, but the effect of the decreased year-over-year variability is less certain.

Contents

Abstract *1*

PREFACE **3**

INTRODUCTION **6**

RESEARCH METHODS **9**

MECHANISTIC HABITAT MODELING FOR PINNIPEDS 9
CASE STUDY: MODELING SEA ICE FOR HARP SEALS 11
OBTAINING GLOBAL CLIMATE MODEL OUTPUT 13
UNDERSTANDING A MODEL OUTPUT AND ENSEMBLES 16
The Coupled Model Intercomparison Project *16*
Scenarios: The Representative Concentration Pathways *16*
Model Differences *18*
Choosing Appropriate Model Output *20*
PREPARING A MULTI-MODEL ENSEMBLE 23
NetCDF Tools: NCO, CDO, Panoply, and Python *23*
DOWNSCALING FOR REGIONAL ANALYSIS 25

RESEARCH RESULTS AND DISCUSSION **28**

CONCLUSION **36**

REFERENCES **37**

APPENDIX I – EXAMPLE PYTHON SCRIPTS AND CLIMATE UTILITY COMMANDS 40
A. Concatenating multiple NetCDF climate records *40*
B. Averaging across hyperslabs to create climatologies *40*
C. Creating the Anomaly File *41*
D. Regridding the Anomaly File *41*
E. Spline Interpolating the Climate Model Centroids to Raster Surfaces with Python *41*
F. Creating the ensemble with Python *47*
APPENDIX II – A PRIMER ON NETCDF AND MULTIDIMENSIONAL DATA IN ARCMAP 49
ArcMap Multidimensional Tools *50*
Making a NetCDF Feature Layer *50*
Choosing the Climatological Variables for Import *52*
Setting the Dimensions for Importing Climatological Variables *52*
Working with EdGCM data in ArcMap *54*
APPENDIX III: EXAMPLE NETCDF OUTPUT 55

Preface

As an undergraduate student studying Computer Information Systems, I found the apparent juvenescence of the study of computers, even in 2003, peculiarly fascinating. While the "mechanical calculator" had been invented by Blaise Pascal in the 1640s, and the concept for a more sophisticated "analytical engine" had been put forth by Charles Babbage in the 1830s, it wasn't until the 1940s that true computers (with memory, processing, and programmable input) were capable of being built. Thus, modern computing wasn't really born until the middle of the last century.

This is unlike many other scientific fields that have heroes stretching back through many centuries to the renaissance, and some further. Copernicus published his helio-centric astronomical theory in 1543, the same year as Vesalius' presentation of his work on the structure and function of the human body. Astronomy, medicine, and many other branches of science saw some of their most notable advances in the distant past, by wizened men who need only be referenced by last name, inaccessibly couched in prior eras and cultures that would seem to bear little resemblance to those of the present.

While the history of these fields lends them an air of sophistication, it would also seem to cast a quality of derivativeness upon more recent achievements in these fields. The science of computers is most certainly built upon advances in many other branches of science, but that it is a youthful field with some of its forefathers living in the present time, makes it feel more approachable, dynamic, and exciting. John von Neumann developed, in in 1945, the modern computing architecture we still use today. John Backus developed FORTRAN, an ancient (in the computer science time-scale) programming language in 1954 that is still used today in some of the most sophisticated applications of computer science (such as climate modeling). Dennis Ritchie and Ken Thompson developed the C language and the Unix operating system in the 1960s, and the now ubiquitous World Wide Web that runs on Unix computers wasn't developed and tested until the early 1990s. Futurists would deem

this the era of the computer and the Internet; this technology has been rapidly transforming our daily lives and our methods in all branches of science.

But this period of time is known by another name as well - a term that is defined by the rapid advancement of technology preceding computers and the Internet, but that is an essential precedent to this technology as the Internet is an essential artifact of it. Eugene Stoermer called the present era in geologic time the Anthropocene, and it is roughly defined as the period of time during which humans have had a significant influence on the Earth at a global scale. This time period was initiated by the industrial revolution, the creation of the steam engine, and the use of hydrocarbons as a concentrated mobile energy source, and continues today in the form of automobiles, advanced manufacturing techniques, and anything that runs on electricity such as a computer. It has led to a rescaling of the planet in our minds. Thanks to abundant hydrocarbons and subsequent technological advancements such as the World Wide Web, we can fly halfway around the world in under a day, and have remote and synchronous communication with anybody, anywhere.

The term "Anthropocene" often evokes another sentiment: concern. As humans now have the ability to affect global-scale changes in the Earth's basic systems through advanced agricultural techniques, combustion of hydrocarbons, and the global economy. We also have a heightened awareness of our abilities and effects on a global scale, and thanks to satellite technology and computer modeling, a new way to observe our impact on the planet. Just as with nascence of computer science, scientific sub-disciplines borne of this new awareness of the functioning of the Earth and our leading role in the present geological epoch now seem approachable, dynamic, and exciting. Though humans have been observers of nature for millennia, we now have the power (quite literally) to understand how it functions and to dramatically influence its course.

As a Master's student studying the natural sciences, I am again fascinated by their juvenescence. It was only slightly more than a decade ago that ecology became a named branch of science thanks to Ernst Haeckel, and it was merely a half-century earlier that Darwin penned the Origin of the Species – shaping the field of biology and other related disciplines. In 1917 Joseph Grinnell developed the concept of an

ecological niche, initiating ecology's present focus on habitat and enabling a more advanced appreciation of Earth's biodiversity (a term which wasn't coined until the latter half of the 20th century). In conjunction with global climate change (and certainly intimately tied to it), biodiversity is another incredibly important topic in the Anthropocene as humans are thought to be causing a mass extinction of proportions rarely seen on this planet (and only recently known of).

The abstract and cold science of computers is coming of age at the same time as many of the sub-disciplines of the natural sciences, and advancements in either field owe a debt to the to the other. Many of the important methods in the natural sciences, such as habitat and climate modeling, rely on the rapid processing abilities of computers, and advancement in computer science is partially driven by their needs. It is no coincidence that John Von Neumann, creator of modern computer architecture, was also an important contributor to the study of fluid dynamics and weather systems. Our new understanding of natural systems is tied to our understanding of computer systems, and our ability to project the future climate and ecosystems of the earth and their response to our actions is a product of the rapidly growing speed of computer processing.

The subject of this paper, and the importance of the findings that could be generated from this kind of modeling, derives from the co-evolution of the computational and natural sciences and their contribution to our appreciation of our impact on the Earth system and to the resolution of today's global-scale problems of anthropogenic climate change and biodiversity loss. The modeling of future habitat for species based on the output from global climate models fulfills at least two important needs. First, it advances our knowledge of the natural sciences and better allows us to understand how various species on the planet have adapted and may adapt to climactic changes. Second, but just as important, it can help build political will towards and help to shape the policy governing our actions so that we can preserve the climate and biodiversity of the earth that existed before the Anthropocene.

Introduction

Lewis Richardson, a mathematician and physicist born in the 1890s, dreamt up the idea of the modern weather forecast in 1922; weather forecasting was in its infancy at this time and was based largely on basic atmospheric measurements and the drawing of rudimentary maps. Richardson's vision of many computers running in parallel, solving basic physics equations, and rapidly sharing information with each other in each time step to evolve a global forecast was incredibly prescient not only because this is precisely how modern weather and climate models work[1], but also because the computers he was envisioning in his then fantastical vision were human beings (Edwards 2010). It would be nearly 20 more years before modern computers were built, and many more years before they were capable of the sort of information sharing and rapid data processing required to generate a global forecast on a computer at a rate faster than the weather was evolving in real time.

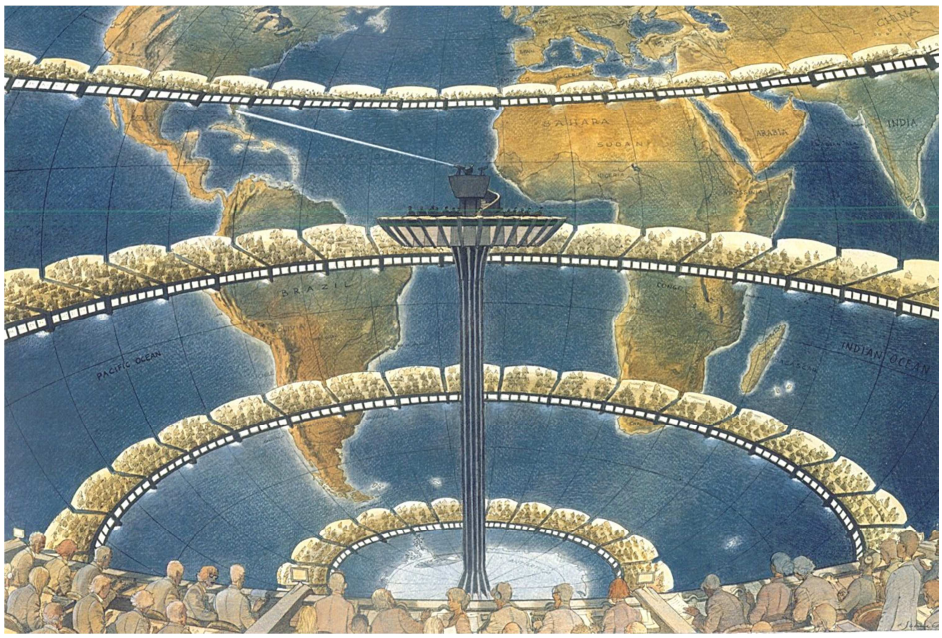


Figure 1. Francois Schuiten's representation of Richardson's "Forecast Factory". Human "computers" would sit in a spherical room in front of the grid cell they were responsible for calculating the weather of, and a conductor in the centered tower would keep time. (University College Dublin, 2006)

[1] Care should be taken not to conflate weather and climate, but they do share this history and are indeed related.

Ninety years later, though still requiring of great improvement, climate modeling is practiced in research centers all over the globe, and results from climate models are an important component of a vast amount of present-day scientific literature and popular news stories. Bodies exist to discuss and disseminate the important findings coming from these models (such as the International Panel on Climate Change), and scandals worthy of the –gate suffix have drawn unwanted attention to the human side of climate modeling (Climategate of 2009). Indeed, climate modeling is an important component of our lives whether we are aware of it or not. It helps us to explain the history of the earth, and to project the future of the climate system, and as such, the future of all species inhabiting the planet.

Projective habitat modeling is one new application of climate models that has only recently become possible due to new methods in species distribution modeling and the availability of high-quality scientific outputs from climate modeling centers. When first developed in the early 1900s, species distribution models were qualitative and imprecise, but they became more quantitative in the '70s and '80s (Elith & Leathwick 2009). It was at that point in time, and continuing to the present, that correlative habitat models became the dominant form seen in scientific publications. The basic logic behind them is simple: for a given species, a number of environmental variables such as sea surface temperature, bottom-type, and chlorophyll concentration (for a marine species) are independent variables with which correlations are drawn to the dependent variable – the presence (and often, absence) of an individual of the target population. If it is suspected that a particular species of fish prefers rocky bottoms, for example, this is included in the model and the degree to which the presence of rocky bottoms, in conjunction with other variables, predicts the presence of the fish determines the goodness of fit of the model.

These models are subject to a number of constraints and drawbacks, however, and a good amount of criticism has been lobbed at them. They have been faulted for of being poorly linked to theory, subject to autocorrelation (Elith and Leathwick, 2009), and too reliant on assumptions and indirect effects that are not explicitly accounted for in the model (Kearney and Porter, 2009). These limitations

become even more pronounced in the face of ecology's latest modeling challenge: global climate change. SDMs are not known to be reliable for projecting future distributions (when species equilibrium is disturbed), in new locations, with completely new environmental factors – all well outside of the range of training data used to fit the models (Elith and Leathwick, 2009).

In addition to these issues with SDMs, each species presents its own modeling complexities, with the modeling of marine mammal species being a particularly difficult endeavor (Pompa et al, 2011). More specifically, pinnipeds (seals, walruses, etc…), are exceptionally difficult to model because they can have rich life histories that span many environments (terrestrial, marine, cryospheric), great annual migratory distances (Laidre et al, 2008), complex intra- and inter-specific interactions, and because they are intelligent mesoconsumers in their ecosystems (Heithaus et al, 2008). At the same time, it is of great importance that the impact of climate change on Pinnipeds be studied, because their overlap with and connection to human populations – due to hunting, indirect fisheries interactions, their status as charismatic megafauna, and their special protection in the U.S. under the Marine Mammal Protection Act (MMPA) - presages increased future conflicts and more challenging management scenarios under a rapidly changing climate. New methods for modeling Pinniped distribution and abundance are of critical importance, then, for informing the latest conservation management programs.

Research Methods

Mechanistic Habitat Modeling for Pinnipeds

Recent advances in the theory of species distribution modeling have created new, relatively underexplored methods for projecting the distribution of species under climate change scenarios. One promising new method, mechanistic modeling, predicates species distribution on inherent physiological constraints that are specific to the given species (Dormann, 2012). A species' Thermal Neutral Zone (TNZ), which is the range of temperatures suitable for an endotherm, is one such constraint which can be applied to surface temperature projections to predict future range. An important distinction between conventional correlative SDMs and mechanistic or process-based SDMs is that the former estimates the 'realized niche' of the species while the latter estimates the 'fundamental niche' (Wiens et al, 2009). Mechanistic models can show the absolute ranges inhabitable by a species under any scenario (withstanding genetic variation) while correlative models can only show the likely distribution of a species under conditions that are similar to those used to build the model. Correlative models assume that the prediction space is within roughly the same time period and ecosystem as the input data used to train the model. Correlative and deterministic models can be combined, and the intersection of their predicted distributions may more closely map the projected 'realized niche' for a future climate scenario.

Hybrid mechanistic/correlative models are rarely used, not only because they are a recent development, but also because the physiological parameters they require are not mapped for most species and these models are more difficult to construct (Kearney and Porter, 2009). In fact, very few papers have been published on this approach, and none have demonstrated its use in mammals – much less mammals whose life histories can span multiple ecosystems and include migrations of thousands of miles per annum. That said, the Thermal Neutral Zones for several species of Pinniped (e.g. Harbor Seals, Hawaiian Monk Seals) have been studied, and

are available for use in mechanistic models. Given the aforementioned complexities inherent in modeling the response of Pinnipeds to climate change, this presents an opportunity for the development of new distribution models that overcome several of the standing limitations – perhaps creating the most reliable projections to date for Pinnipeds, and demonstrating – for the first time – a robust technique for projecting marine mammal distribution under anthropogenic climate change.

The seals for which TNZ information is available also represent a great amount of intra-order diversity in life histories. The Harbor Seal is widespread and not threatened by extinction. Quite the contrary, this seal is considered a nuisance by many along the east and west coasts of the United States because they are suspected of interfering with fisheries (Lunneryd, 2001), they are thought to attract sharks (Nickisch, 2012), and their strict protection under the MMPA means that they cannot be harassed in any way, so their growth is unchecked. Harbor seals are highly loyal to their chosen haul-out spots (Burns, 2002), but perhaps flexible enough to adapt at the pace of climate change. Their birthing time may vary by latitude, which is an added sensitivity to climate change (Temte, 1994).

The Hawaiian Monk Seal, for which the TNZ is pending publishing in an upcoming paper (Williams et al, 2011), is under quite different circumstances. It is one of two Pinnipeds in the U.S. listed as endangered under the Endangered Species Act (Read & Wade 2000), and its range is limited to the Hawaiian archipelago. Sea level rise is a pressing concern for this species, as many of the tiny islands it hauls out on will be submerged due to rising sea levels (Antonelis et al, 2006). Additionally, entanglement in discarded fishing gear, predation, juvenile-adult conflict (Lowry et al, 2011), and adverse interactions with humans on the Main Hawaiian Islands contributes to the poor status of this seal. Mechanistic modeling could be used to inform unusual approaches to management of this species, such as transport to new islands or sand replenishment on existing islands, which may be required to prevent its extirpation.

While these two examples may seem to represent a great deal of variation, much more exists in sub-order Pinnipedia. Some seals are pagophilic (ice-associated) and require sea ice for critical life-history stages such as breeding

(Friedlander et al, 2010). Others, such as freshwater Baikal and Lac De Loups Marins seals are thought to be constrained to freshwater lakes (Smith, 1997). This means they are physiologically limited not only by their TNZ, but also their locomotive capabilities. If climate change significantly alters the hydrological structure of the lake systems these seals depend on, it could severely contract their range.

Case Study: Modeling Sea Ice for Harp Seals

To demonstrate another form of mechanistic modeling, the life history requirements of some pinnipeds can be used as constraints not unlike those of the Thermal Neutral Zones of all animals. Harp Seals (*Pagophilus groenlandicus*), for example, require sea ice as a temporary whelping (birthing) and weaning substrate for their pups. In late February and early March, female harp seals haul out on sea ice and synchronously give birth to their pups (young of the year, known as whitecoats) throughout a roughly two week period (Johnston et al. 2005 & 2012). After whelping, the mothers then spend no more than two more weeks weaning their pups, after which they abandon them and return to the water to begin mating.

Figure 2. A harp seal whitecoat basks in the sun on pack ice in the North Atlantic. This pup was likely photographed within a few weeks of birth, and is in the process of being weaned from its mother. It will convert the fatty milk it receives from its mother into the muscle and other tissues it will need when it enters the cold waters for the first time and must begin foraging on its own.

These seals use sea ice as a substrate for whelping because it is relatively safe from predators, unlike coastal beaches where whitecoats would be much less camouflaged and much easier prey for predators. In fact, in years when sea ice is not readily available these seals have been known to spontaneously abort their pregnancies or to give birth in water rather than hauling out on land. Thus, though

sea ice is highly variable year-over-year, and while it does put the seals at risk of slipping into the water or being crushed between ice floes, it is the preferred location for these seals to pup and wean. This makes it a very important constraining factor in the life histories of these animals.

Another important consideration for the mechanistic modeling of harp seal habitat based on sea ice availability is the habitual return of this species to the same whelping location every year. There are 4 generally accepted whelping locations for harp seals in the North Atlantic, shown in Figure 3 (Friedlander et al. 2010). If sea ice is low off the East coast of Canada, for example, these seals are not known to relocate to the ice off of Greenland or in the White Sea. There is very little mixing between these subpopulations of *Pagophilus groenlandicus* so their ability to adapt to variation in ice cover across the north Atlantic is reduced.

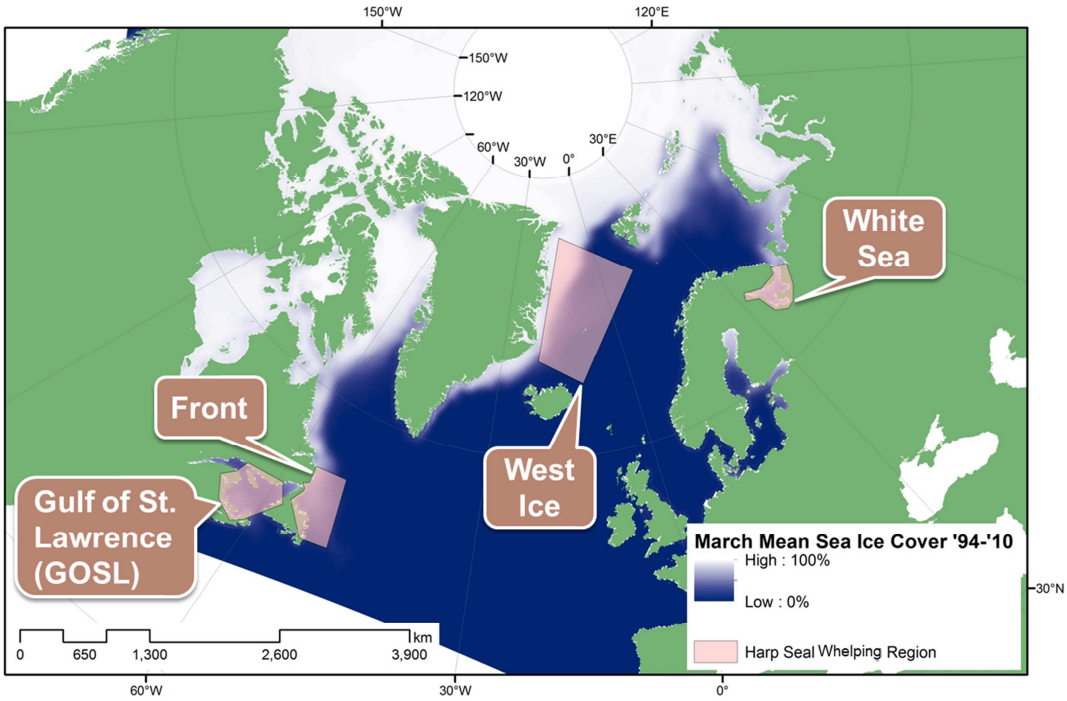


Figure 3. Harp Seal Whelping Locations in the North Atlantic

Given all of these constraints, sea ice concentration is a good mechanistic modeling variable which can be used to project future viable harp seal habitat. Certainly, if sea ice variability were to increase or if sea ice were to vanish entirely in some of these regions the effect on young of the year mortality, and thus on overall

population stability, would be negative. As such, this case study will focus on the use of sea ice concentration projections to determine how harp seals may fare at the end of this century compared to present. Global climate model output will be used to obtain projections of sea ice, and this will be combined with proxy data on seal mortality under differential ice conditions to determine how future ice conditions will affect this seal.

Obtaining Global Climate Model Output

In order to perform mechanistic habitat modeling to project future habitat and species distribution, realistic high-resolution climate model output is needed for variables that could determine a species' fundamental niche and other distribution-constraining characteristics. Sophisticated global climate models are developed by no fewer than 27 modeling centers around the world, each likely with many instances of their models, and each with their own strengths and weaknesses. For example, the NASA Goddard Institute for Space Studies produced Models I, II, III, and E, with Model E having multiple instances due to coupling with different ocean components. Most modeling centers make their model code and some of the output from the models freely available for others to use as well.

Visiting the web site of one or more climate modeling institutes to pull down their code and run one of their models would require a level of programming skill and raw computing power (and patience) not readily available to most practitioners of the natural sciences. Furthermore, running just one model one time is not a sufficient method of generating climate projections because any one model is subject to a great deal of error. Fortunately, repositories of model output have been made available, such as the Earth System Grid (Williams et al, 2009), which is where all model output for this study was obtained. This facilitates the task of deciding which models to use by allowing a large set of them to be obtained and analyzed, but some basic theory on uncertainty is important to further understand how using many models constrains uncertainty.

Hawkins & Sutton (2009) analyzed a set of climate models from many different institutes and applied statistical methods to tease apart the different types

of uncertainty inherent in each models' predictions. They found that there are primarily 3 types of uncertainty present in climate models, and that the fractional uncertainty (the ratio of projected uncertainty to the mean change in temperature) attributable to each type is influenced by several things, including the variable being modeled, the size of the area being modeled, the lead time used to spin up the model (and the data used to initialize it), and the size of the climatology (the period of years averaged together to form each data point). The three types of uncertainty in climate models are:

- Internal variability (of the climate system – such as El Niño)
- Model uncertainty (due to the various ways models are coded)
- Scenario uncertainty (due to uncertainty of future GHG emissions)

The contribution of these uncertainty components to the total varies through the time of the climate model projection, and an interesting phenomenon is observable when the total fractional uncertainty is analyzed: it reaches a minimum well into the climate model run (around 40 years in for a global analysis, and around 80 years in when focused on Greenland). Thus, while total uncertainty rises consistently over time from the beginning of the climate model run, fractional uncertainty does not.

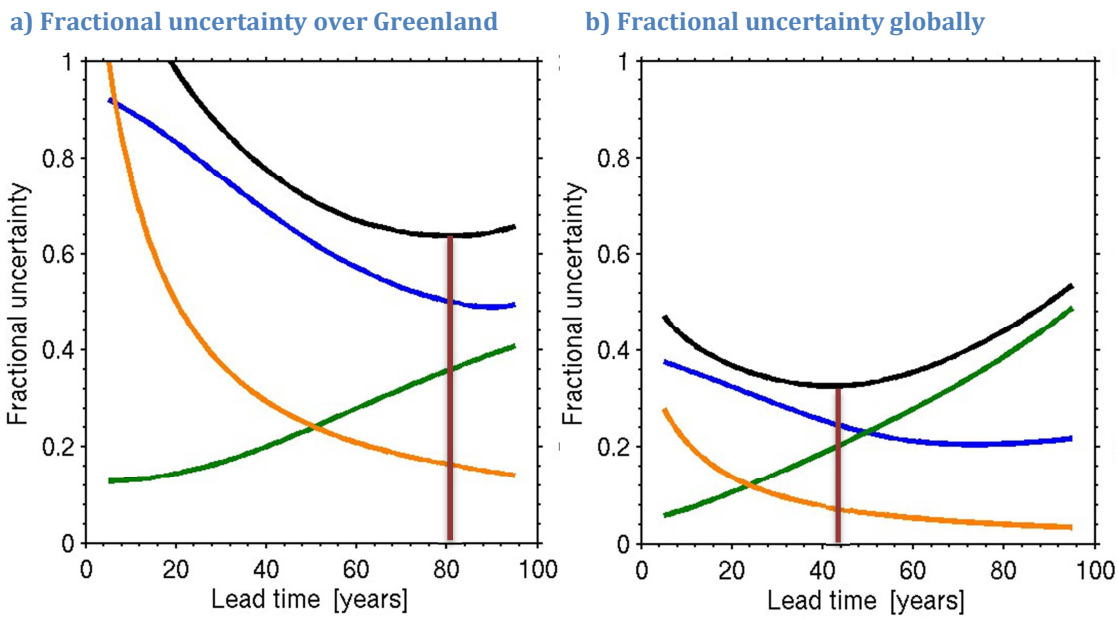


Figure 4. The three types of uncertainty are graphed for Greenland and globally (orange = internal; blue = model; green = scenario) as well as the total fractional uncertainty, based on a 12-year climatology.

The total fractional uncertainty minimum (approximated by the red lines) is dependent upon many factors, but is clearly influenced by the area being modeled (and its size). Charts generated using the authors' web tool at <http://climate.ncas.ac.uk/research/uncertainty/>.

This suggests that, until the minimum point, mean temperature is increasing faster than total uncertainty, and that per predicted degree increase in temperature, the uncertainty is decreasing until well into the model run.

In order to separate the different types of uncertainty in climate models, Hawkins & Sutton relied on a Multi-Model Ensemble (MME) – the compilation of outputs from many models and model runs to form one large dataset. MMEs are very important in the world of climate modeling because "combining models generally increases the skill, reliability, and consistency of model forecasts" (Tebaldi & Knutti 2007). These ensembles do not just include output from the models of many different modeling houses, but also multiple different runs of each model under slightly different conditions. The output of each model run is tagged with an "ensemble member" descriptor that includes the realization (r – differing initial conditions), initialization method (i – differing observational inputs to start the model run), and the perturbed physics number (p – different parameter settings for the run) (Taylor et al. 2011). The NASA GISS E2R model has 98 permutations of these members, and assuming this model and institution is typical, a rough estimate of the total number of runs that could be considered for inclusion in an ensemble is over 13,000 (98 members * 5 models * 27 modeling centers).

While no ensemble includes anywhere close to thousands of members, it is typical for ensembles to comprise 4 – 30 models with one or more different realizations from each. Fortunately, collecting output from each of these models has been made drastically easier by placing them on the Earth System Grid (totaling many petabytes of data), under a common naming scheme (Taylor et al. 2011), and in a common format and convention (Eaton et al. 2011) used for most multidimensional scientific data pertaining to climate: NetCDF Climate and Forecast (CF) Metadata Conventions. Compiling a multi-model ensemble is still a challenge despite this facilitation.

Understanding a Model Output and Ensembles

The Coupled Model Intercomparison Project

In addition to making available much of their climate model output in a common repository, most modeling centers have agreed to participate in coordinated modeling experiments so that their model outputs are more easily compared and more useful in ensembles. The Coupled Model Intercomparison Project, Phase 5 (CMIP5) is the largest of such experiments, and models adhering to the CMIP protocol are used in the preparation of IPCC assessment reports (with the 5th Working Group Report on the Physical Science Basis due to be released in 2013).

The models are run under prescribed conditions to assert as much control and standardization as possible, so as to reduce model variability and exaggerate the proportion of uncertainty coming from the different formulations of the climate models. Each of the aforementioned realizations is prescribed by the CMIP project, starting a model at a slightly different point in time with fresh observations, and creating multiple "climate trajectories" with each realization run for the same model, the spread of which is considered representative of all possible outcomes and hopefully capturing internal climate variability (Taylor et al. 2012). These trajectories differ in large part as a result of multi-year internal climate stochasticity from phenomena such as El Niño.

Scenarios: The Representative Concentration Pathways

The Representative Concentration Pathways (RCPs) are a set of climate change scenarios developed for a wide variety of uses – from determining the forcings to be used in climate model runs to informing economic and policy models. Importantly, they are based on "concentration pathways" rather than "emissions pathways" and each scenario corresponds to an expected radiative forcing by 2100.

In essence, this is a way of backing into climate modeling scenarios – first specifying the forcing and then building a scenario that will lead to this level of forcing. The four RCP scenarios used in climate modeling represent a synthesis of literature at the time of their development (which is why they are considered

representative) – literature that explores the possibility of radiative forcing levels as low as 2.6 Wm-2 and as high as 8.5 Wm-2 (see Figure 5).

These scenarios are intended to provide high spatial resolution (.5 x .5 degree) projections of GHG emissions, air pollutants, and land use/cover changes (LULUCF) which would result in each 2100 radiative forcing scenario (Van Vuuren et al. 2011). Each of the four scenarios was developed independently, so while they are meant to be internally consistent, they may not be completely consistent across the suite of the four scenarios. As can be seen in Figure 6, the area of agricultural land used in the scenarios does not increase with radiative forcing.

Overview of representative concentration pathways (RCPs)

	Description[a]	Publication—IA Model
RCP8.5	Rising radiative forcing pathway leading to 8.5 W/m^2 (~1370 ppm CO_2 eq) by 2100.	(Riahi et al. 2007)—MESSAGE
RCP6	Stabilization without overshoot pathway to 6 W/m^2 (~850 ppm CO_2 eq) at stabilization after 2100	(Fujino et al. 2006; Hijioka et al. 2008)—AIM
RCP4.5	Stabilization without overshoot pathway to 4.5 W/m^2 (~650 ppm CO_2 eq) at stabilization after 2100	(Clarke et al. 2007; Smith and Wigley 2006; Wise et al. 2009)—GCAM
RCP2.6	Peak in radiative forcing at ~3 W/m^2 (~490 ppm CO_2 eq) before 2100 and then decline (the selected pathway declines to 2.6 W/m^2 by 2100).	(Van Vuuren et al., 2007a; van Vuuren et al. 2006)—IMAGE

[a] Approximate radiative forcing levels were defined as ±5% of the stated level in W/m^2 relative to pre-industrial levels. Radiative forcing values include the net effect of all anthropogenic GHGs and other forcing agents

Figure 5. The representative concentration pathways as described by Van Vuuren et al. (2011).

Main characteristics of each RCP

Scenario Component	RCP2.6	RCP4.5	RCP6	RCP8.5
Greenhouse gas emissions	Very low	Medium-low mitigation Very low baseline	Medium baseline; high mitigation	High baseline
Agricultural area	Medium for cropland and pasture	Very low for both cropland and pasture	Medium for cropland but very low for pasture (total low)	Medium for both cropland and pasture
Air pollution	Medium-Low	Medium	Medium	Medium-high

Figure 6. The main characteristics of the RCPs as presented by Van Vuuren et al. (2011).

All of these scenarios are intended to be possible future pathways of emissions and LULUCF, and no degree of likelihood is assigned to the scenarios. Because it is impossible to predict which scenario is most likely (leading to such a large and increasing share of scenario uncertainty discussed by Hawkins & Sutton (2009) and visible in Figure 4) it may be difficult investigators preparing impact studies to decide which scenario's model output to use. We used output from 3 of the 4 scenarios (skipping RCP 4.5) because they captured the high and low scenarios plus one of the intermediate scenarios.

Model Differences

Each climate model is considered a unique effort by distinct teams to represent the physics of the internal climate system in such a way as to enable projection of future states of the climate system (complete with many climatic variables) at some point in the future. Not only is every mode different in terms of how it is assembled, coupled, and run, but also in how the output is prepared. In theory, this level of independence between modeling groups is a better scientific practice as each group is trying to replicate the true climate system while being uninfluenced by the efforts of others, and as such, are essentially conducting independent experiments.

In actuality, the models do experience a great amount of overlap due to several unavoidable factors. The climate modeling community is small, and academic, and as such they share their findings and new ideas with each other at conferences and in peer-reviewed journals. Researchers leave one lab to work for another, and often bring components of their code with them to their new labs. Masson and Knutti (2011) analyzed the overlap in institutions and their models to determine if an evenly weighted multi-model ensemble was truly evenly weighted, or whether the inclusion of several models from different modeling houses with a great amount of overlap lead to an uneven weighting in an ensemble mean. They found that, indeed, a genealogy of climate models could be constructed that would show several models to be similar even if they were the product of different modeling groups.

For example, the GISS and CGCM models were found to be somewhat related, as were the HadGEM and CSIRO models, suggesting that they may share similar components such as the code for how the atmosphere works or general ocean circulation. Masson and Knutti conclude that "there is no evidence for multiple classes of models, different mutually exclusive theories or philosophies in how to build a model… [instead the] climate model landscape rather resembles an evolutionary process" (2011). This follows from the idea that these models are developed over time via academic collaboration in the same community, and given that the fundamentals of the earth system components being modeled are not subject to dispute (which would indicate philosophical differences). Indeed, even such decisions as which programming language to use are fairly consistent across modeling groups (with Fortran being the lingua franca in climate modeling).

In addition to differences in the implementation of physics and the coupling of model components, climate modeling groups may also run their climate models using different grids (varying in projection, datum, and spatial resolution) at different temporal resolutions, and using different variable naming schemes. All of these differences may seem trivial compared to decisions concerning the implementation of climate-related physics and coupling mechanisms, but they do create data management and interpretation issues when preparing MMEs that

deserves specific mention. All of these differences must be remediated before an MME can be prepared.

The first step in preparing climate model output from different modeling houses for use in MMEs is performed by the modeling centers themselves as a requirement for including their model in the CMIP experiments. Each model is processed using CMOR (Climate Model Output Rewriter) which standardizes the files so that they are CF compliant NetCDF files (Taylor et al, 2004). In essence, this means that the variables are all standardized and the files are "self-describing" such that there is enough metadata in them to understand the structure of the file and the naming schemes in place. This allows users of the climate model output to interpret model output and assemble ensembles without questioning fundamental data assumptions (units used, meaning of the variable, etc…).

After climate model output is prepared using CMOR it is still possible for it to be in any projection and using any datum and spatial resolution. As such, additional tools must be used to prepare climate model output from various centers for further analysis. The process for doing this will be covered shortly, but first it may be useful to consider how climate models can be chosen for inclusion in an MME before going through the tedious task of preparing the data files.

Choosing Appropriate Model Output

Multi-Model Ensembles means are considered better projections of climate than any one model output (or any one ensemble from a set of realizations from one model) (Hawkins & Sutton 2009), but when it comes to deciding how many models should be included in an ensemble more is not necessarily better. The multidimensional data from climate models is becoming larger as model resolution improves over time, and so if all 60 models submitted to the CMIP5 project (from 27 modeling centers) were to be included, the data processing task would take a great deal of time. The models need to be interpolated to a common projection and resolution, as well as interpreted because they may employ different variable names or slightly different units (while still adhering to the CF convention).

Secondly, the ability of each model to realistically represent the nearly 500 variables being modeled in the CMIP5 project in all regions that may be studied is inconsistent. Some models are known to be good at representing sea ice while some are not – as a result of many possible differences such as spatial resolution, how the different physical realms are coupled in the models, etc… Knowing which models are good at modeling the variables desired for analysis is a good first step towards narrowing down the possible models for inclusion, but few papers have been written on the estimation of a specific variable across all CMIP5 models (except for the most common variables), and none focus exactly on the region being studied in this paper. At the same time, performing this analysis would be tedious and involved beyond the point of prudence for a paper not intended to delve into model comparison. Thus, another approach must be used for the sake of efficiency.

I chose to take a 2-step approach to determining which models should be included in an analysis of SIC – first surveying literature to determine which models are most frequently used to project SIC, and second analyzing the ability of the chosen subset of models to predict observed values during the historical control run of the climate model (during the same period: 1988 – 2005). The final set of models is thus backed up by other peer reviewed studies and by an analysis of the predictive abilities of the models against observations for the variable we're modeling (SIC) in the regions our study is focused on (harp seal breeding locations).

The literature review was conducted using those papers categorized in the CMIP database as sea ice papers[2]. Of the papers listed, those which were published (or submitted and available for early review) were reviewed, and a tally was kept of the models used in each paper (excluding papers written on the performance of only one model). Models were ranked according to the frequency of their use in papers, and those models lacking all three RCP scenarios we want to examine in our analysis were removed. We eliminated those models which were not used in at least half of the intermodal comparison papers, bringing the total number of models under consideration to 13. The initial selection set is shown in Table 1.

[2] http://cmip.llnl.gov/cmip5/publications/variable?exp=sea%20ice%20area%20fraction

Table 1. The initial set of models chosen from the literature review of CMIP5 SIC papers are shown, with those eliminated in the second selection step in red.

Model Name	Spatial Res. (lat x lon)	Avg. Anomaly in Breeding Regions.
NorESM1-M	384 x 320	23.10%
LASG-IAP FGOALS-s2	196 x 360	25.04%
NorESM1-ME	384 x 320	25.95%
NCAR CCSM4	384 x 320	27.88%
IPSL-CM5A-LR	149 x 182	27.93%
MIROC-ESM	192 x 256	28.42%
FIO-ESM	384 x 320	29.46%
MIROC-ESM-CHEM	192 x 256	29.50%
MOHC HadGEM2-ES	216 x 360	32.36%
CSIRO-Mk3.6.0	96 x 192	34.51%
GISS-E2-R	90 x 144	37.26%
GFDL-ESM2G	210 x 360	45.78%
MRI-CGCM3	210 x 360	50.28%

The initial set of models was further reduced by analyzing the performance of each model against observations acquired from the NSIDC. The historical control run output for each model under each RCP scenario was splined (using the same parameters as would be later used for the delta downscaling technique) to a continuous raster surface with the same cell size as the observations. The absolute value of the local (per cell) difference between these two raster surfaces was calculated in ArcMap 10.1, and the average of this per-cell anomaly was computed for each of the breeding location polygons. The result of this analysis is shown in Figure 7.

Once graphed, a cutoff value had to be determined to rule models out of the final set. As the average anomaly across all models was ~32% SIC (the orange bar in Figure 7), this value was chosen as the cutoff. Thus, any model that performed worse than the ensemble mean anomaly from observations was eliminated from the final ensemble. Since the median was ~30% SIC, more than half of the models made the cut, and a final set of 8 out of the 14 were chosen to be part of the final multi-model ensemble.

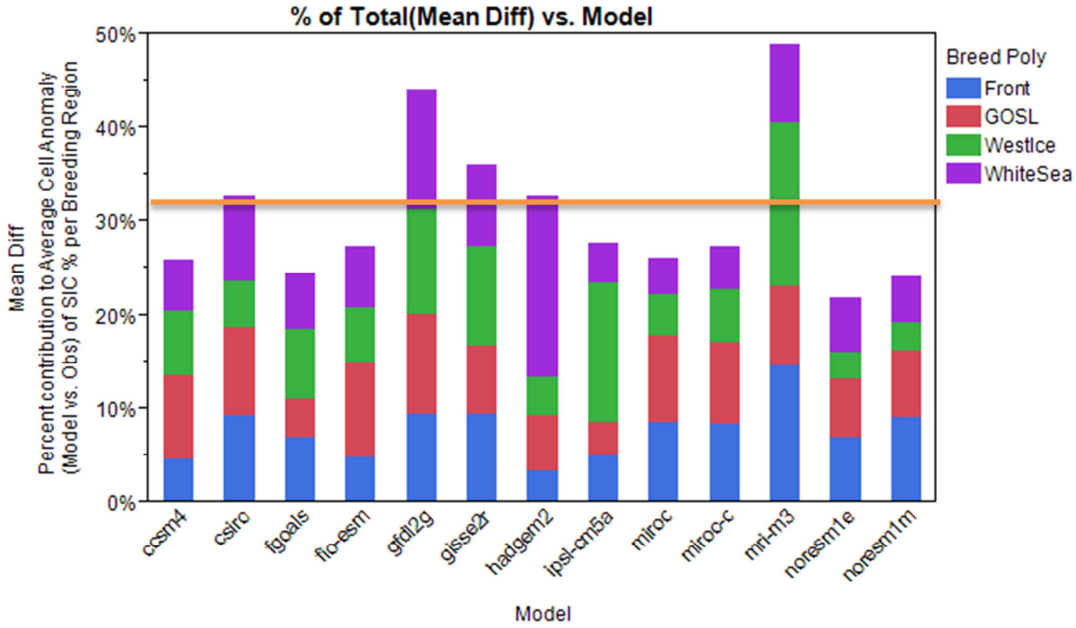


Figure 7. An analysis of the models chosen in the literature review step. The height of each bar represents the average of the SIC anomaly across all the four breeding locations. The colored proportions of each bar represent the contribution of each breeding location to the total anomaly for that model. The orange line is the inclusion cutoff (32%); any model extending past this was excluded.

However, after discussions with climate modelers at NASA GISS, it was concluded that using a slightly larger set of models (returning to the initial set of 13 from the literature review) would be a better ensemble. This is largely because eliminating models based on their ability to predict past climate is not a fair way to assess their capabilities. There is too much internal variability in the climate system and the models may be unfairly penalized for getting the timing of events wrong (like ENSO) even while they, on climatological timescales of 20 to 30 year averages, perform very well compared to historical observations.

Preparing a Multi-Model Ensemble

NetCDF Tools: NCO, CDO, Panoply, and Python

Preparation of a multi-model ensemble is an exercise in data massaging. The output from several climate modeling centers must be obtained from the Earth System Grid, reprojected into a common projection with a common datum and spatial resolution, and summarized as desired for analysis into appropriate spatial

regions and time slices. The details of this process are demonstrated in the appendices to this document, but a brief summary of this process is included here.

All output from the CMIP5 experiment is available on the Earth System Grid (Williams et al, 2009) in NetCDF format, and can be retrieved using WGET scripts[3]. Once obtained, the individual NetCDF files may represent fragments of the desired data range under investigation, so multiple files may need to be combined to generate one file per model, scenario, and realization such that all years to be considered (1850 – 2005 for historical scenarios and 2006 – 2100 or 2300 for the RCPs) will be in one combined file. This can be done using ncrcat (record concatenator), one of the many NetCDF NCO[4] utilities available.

Once a consistent set of files is prepared, they must all be regridded so the cell centroids from all outputs line up to the exact same latitude and longitude pairs, which allows the models to be mixed together and compared. This necessitates some resampling, and is performed using CDO[5] (nearest neighbor interpolation) to resample all model output to the highest spatial resolution in the set.

Next, a set of NCO tools is used including ncra (record average), ncbo (binary operator), and ncks (Kitchen Sink – miscellaneous, subsetting) to find averages and anomalies from baselines, and to spatially subset the data into the whelping regions under investigation. It should be noted at this point that the White Sea whelping location was not included in this analysis because there were not enough data points to get a statistically sound projection of sea ice, and also, because the depth of the sea is not great enough to allow the climate models to accurately model circulation, and thus sea ice.

These NetCDF files were periodically reviewed in Panoply[6], a display application prepared by NASA GISS, to ensure that they were being accurately prepared at each step. They were also further processed in ArcMap using Python scripts (see appendix for examples) in order to downscale the model output and prepare maps; these steps are covered next.

[3] http://www.gnu.org/software/wget/
[4] http://nco.sourceforge.net/
[5] https://code.zmaw.de/projects/cdo
[6] http://www.giss.nasa.gov/tools/panoply/

Downscaling for Regional Analysis

Downscaling climate model output, which can exist at spatial resolutions as course as 100x100 km per grid cell or worse, is a useful approach to regional climate modeling (where much smaller-scale phenomena such as hydrology or habitat availability) are being modeled. Though it should go without saying, downscaling does *not* enhance the reliability of climate model outputs and may lend false precision to an analysis relying on downscaled data. However, it is a worthwhile practice because it can result in a reasonably good estimate of local-to-regional climate change impacts[7].

An important quality of climate models related to regional downscaling is the spatial uniformity of the climate change signal visible even at small scales in contrast to the noise of the unforced climate system (what the climate would be like without Anthropogenic GHG emissions and land use changes). While climate variability is relatively high on local-to-regional scales, on a global scale the variability is dampened by local variations offsetting each other. However, due to the uniformity of the climate change signal (the modeled climate system with anthropogenic forcings minus the unforced system) on all scales, carefully performed downscaling can generate useful and reliable information (Taylor et al. 2013). This is especially true when the downscaling method includes use of high-resolution observations in the study region and a good MME rather than runs from a single model.

One such downscaling approach, though simple to understand and execute, is the CCAFS Delta Downscaling method (Ramirez-Villegas & Jarvis 2010). The basic steps are described here (with additional details in Appendix I):

[7] The type of downscaling discussed here is 'statistical downscaling' in contrast to 'dynamical downscaling'. Statistical downscaling takes completed global climate model runs and applies them to static high-resolution observations. Dynamical downscaling uses global climate model output as an input to a regional climate model, which is then run for the study area.

- Historical control run output and projected global warming scenario output should be obtained from the Earth System Grid for the time period of interest and from the models to be used in the MME.
    - For the example, in this document historical data for the period from 1990 – 2006 and future climate model projections for the period 2006 – 2100 were obtained. This output was from 14 original models, 3 RCP scenarios, and two variables: Sea Ice Area Fraction (SIC) and Surface Temperature (TS).
- The model output should all be standardized and preprocessed as described in previous sections so that the desired climatologies are produced for analysis.
    - In this document, 17-year climatologies for the mean February and March SIC and every month's TS were prepared as well as historical climatologies for each model ensemble member.
- Each future forecast climatology (RCP scenario) should be differenced with the historical climatology to produce an anomaly file. This anomaly represents the change in the variable mean from the historical period to the future period being forecast. There will be as many anomaly files as there were RCP scenario files.
- Each anomaly file can then be splined to a continuous surface and applied to the observational raster containing the present day climatology (using basic raster algebra – adding the overlapping cells together) to arrive at a forecast future variable value in the same resolution as the observation raster – a downscaled climate projection.

A key component of this methodology is the use of the historical file and the future forecast (RCP scenario) files from the same model ensemble member. This ensures that the historical variable mean and forecast variable means were generated from the same climate model code under the same initial conditions. The only differences, then, are the forcings mandated by the different scenarios. This

ensures that if a model has certain biases (such as it represents the climate as more humid or arid than observations would demonstrate) they exist in both files. When the anomaly is calculated by differencing these two files, the bias is largely canceled out while the climate change signal becomes more pronounced.

Research Results and Discussion

Multi-Model Ensemble projections of the mean sea ice concentration anomaly for harp seal whelping locations under the moderate and high climate change scenarios (RCP 6.0 and RCP 8.5) show a decline of 10% to 40% by 2100 from a modern baseline climatology (average of SIC, 1988 - 2005) while sea ice concentrations under the low climate change scenario remain fairly stable, though still mostly below the baseline (see Figure 9 and Figure 11). This demonstrates that, outside of the most moderate scenario of global warming (RCP 2.6), seals will have to cope with declining sea ice in all whelping locations. As could be expected, there is a positive correlation between the severity of the climate change scenario and the decline in sea ice.

In absolute terms, a 10% to 40% decline in sea ice may mean a decline to 0% ice cover if the baseline ice concentration was already low. To better see this represented, the output from statistical downscaling is represented in Figure 10, which shows spatially the data presented in Figure 9 and Figure 11. In all whelping locations it is clear that as the global climate change scenario increases in severity the sea ice concentration declines predictably. Though the White Sea is shown in Figure 10, it should again be noted that not enough data was available to accurately predict sea ice in this region – rather it is shown for the sake of completeness and because downscaling was spatially applied to the entire study area, even outside of whelping locations.

The change in projected year-over-year sea ice concentration variability with time through 2100 (Figure 12 and Figure 13) was considered in 5-year intervals. These figures were constructed by finding the standard deviation of the 5 year moving average to show how much variability occurred from the two years prior to any given year through the two years following that year. This is one of many possibly ways of exposing temporal variability of sea ice concentration, and these figures were prepared to assess whether interannual fluctuations in sea ice would increase or decrease over time as sea ice declines. Interestingly, while under the

moderate (RCP 2.6) scenario the variability trend remains fairly flat, under the intermediate and high scenarios (RCPs 6.0 and 8.5) variability declines (this is especially visible in Figure 13, which is simplified to just show the average variability metric for the MME, and includes a linear trend line.

Uncertainty in the prediction of sea ice concentration (model spread) increases over time, as would be expected given the increasing scenario uncertainty and model uncertainty. This can clearly be seen in Figure 9 as the red bars indicating 1 standard deviation from the MME mean expand over time due to model predictions becoming more and more spread out at each year. This increase in uncertainty is representative of the likely spread of outcomes due to nuances in climate model physics, coupling, and the fact that anthropogenic emissions become more difficult to predict on longer timescales. While scenario uncertainty is difficult to address, model uncertainty may decline with each new CMIP project as models increase in their predictive capabilities.

The general decline in sea ice projected by climate models is undoubtedly more of a hindrance to Harp Seal survival than a beneficial feature of the climate system in 2100, but the extent to which it hinders their survival must still be considered. Unfortunately, population estimates for these seals are not frequently undertaken, and an understanding of trends in pup mortality is lacking. As such, proxy measures relating seal survival to sea ice trends must be used. Two such measures exist: landed catches (the number of seals pulled in from the annual seal hunts) and strandings (the number of seals found on coastlines that are either dead or nearly so).

Use of both of these proxies is demonstrated in a 2012 paper by Johnston et al., showing that there is a negative relationship between strandings and sea ice, such that decreased sea ice leads to more strandings, and thus a higher mortality rate for harp seal pups. Also, it was shown that landed catches of harp seals declines to zero at sea ice concentrations of 10% or lower, suggesting that so few seals survived in such light ice years that no pups were able to be taken. Though these proxies would need to be strengthened before making deterministic statements about how sea ice decline will eventually lead to a decline in harp seal numbers in

these whelping locations, a clear connection can be seen in both of these proxies, and this suggests that this method of mechanistic modeling, conducted with better observations of seal populations, could lead to fairly accurate forecasts of seal survival and whelping location.

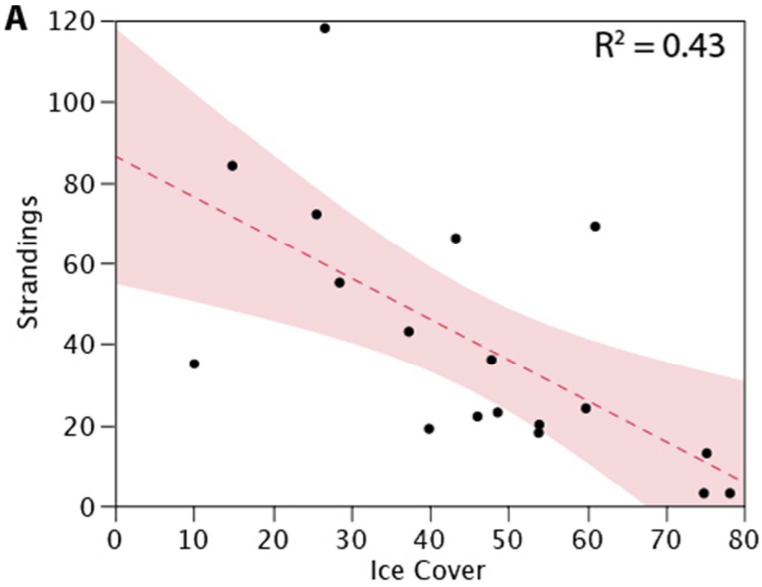


Figure 8. Strandings of young-of-the-year harp seals are plotted against ice cover from records collected during the period 1979 – 2011, and a linear trend line is fit to show the general relationship between these variables. As ice cover increases, strandings decrease, suggesting that lower ice years lead to higher mortality in newborn harp seal pups. Figure from Johnston et al, 2012)

The effect of the decreased year-over-year variability is less certain still; it is possible that as the sea ice variability decreases, seals could benefit from predictable ice conditions by learning over time that ice conditions will be fairly consistent, allowing them to make longer-term choices about where to whelp (though there is no evidence that they can or will do this). On the other hand, increased consistency in ice cover is not helpful against a trend of sea ice decline because the absence of ice is the factor that causes increased seal mortality, not variability. Seals have already adapted to current levels of variability (Johnston et al. 2005), so if anything this finding suggests that changing variability will not be an important factor in their survival through 2100.

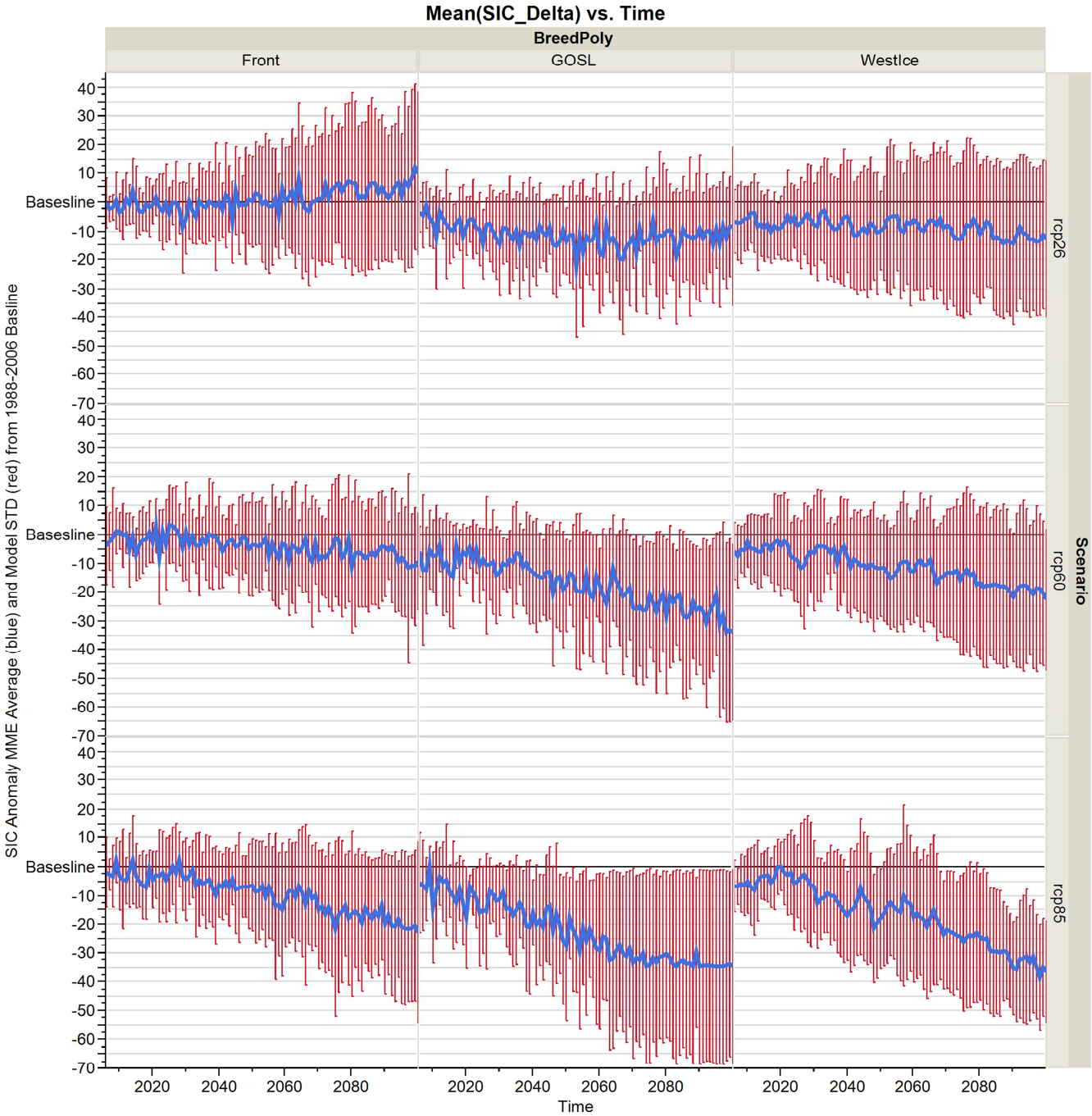


Figure 9. The Multi-Model Ensemble anomaly mean is plotted (blue) against error bars (red) representing 1 standard deviation from the MME mean for each year, based on the output from 13 different climate model inputs. This is done for each whelping location and RCP scenario. The standard deviation increases over time due to the sources of uncertainty discussed in the previous section (see Figure 4).

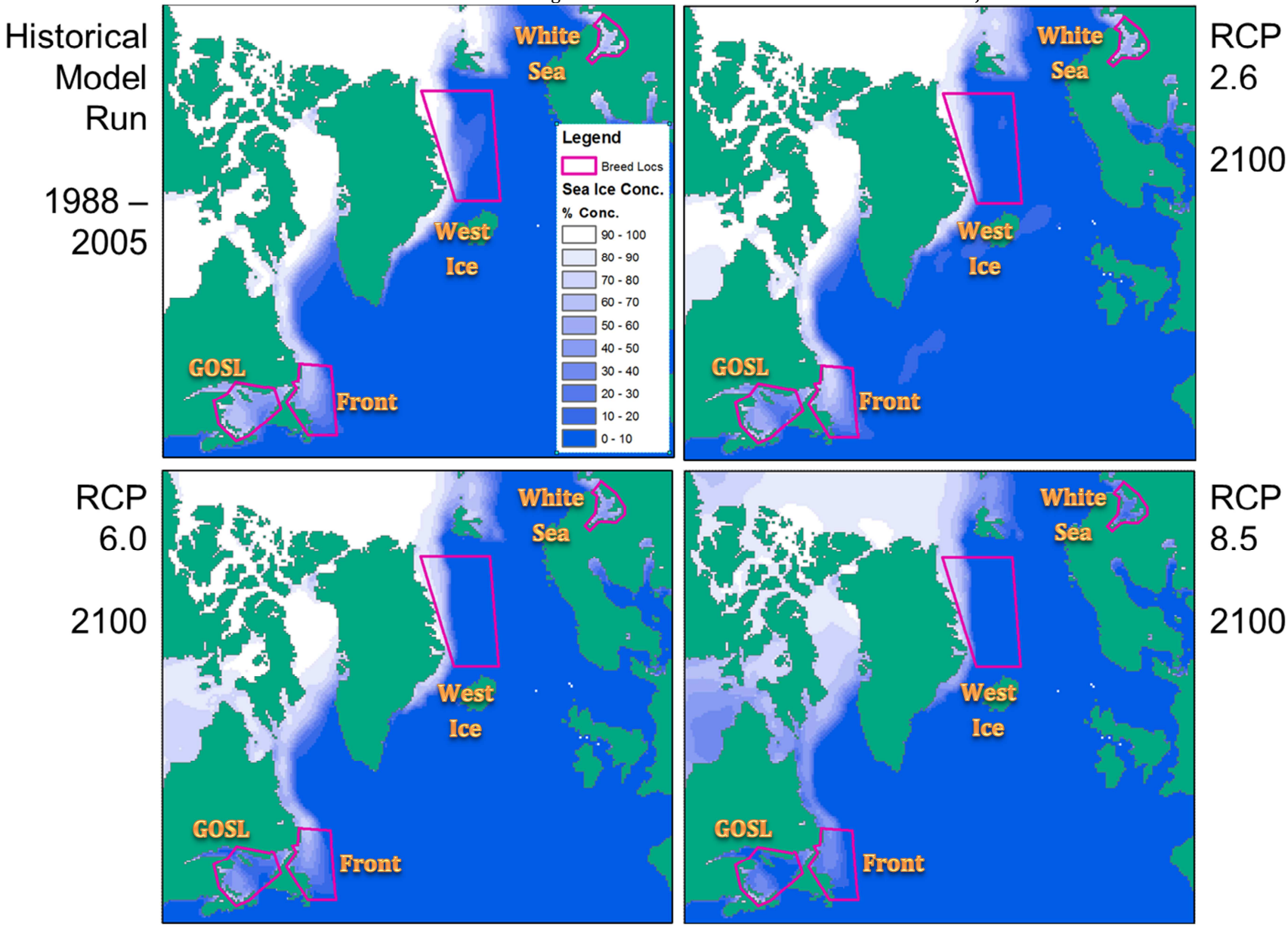


Figure 10. A spatial representation of sea ice concentration in 2100 under several RCP scenarios using the Delta Downscaling method on the 13 model MME mean projection of sea ice concentration. Though a general trend of reduced ice in the sub-Arctic and Arctic is visible, the story for each whelping location is more nuanced. It is only under the harshest scenario that significant declines are visible in all polygons.

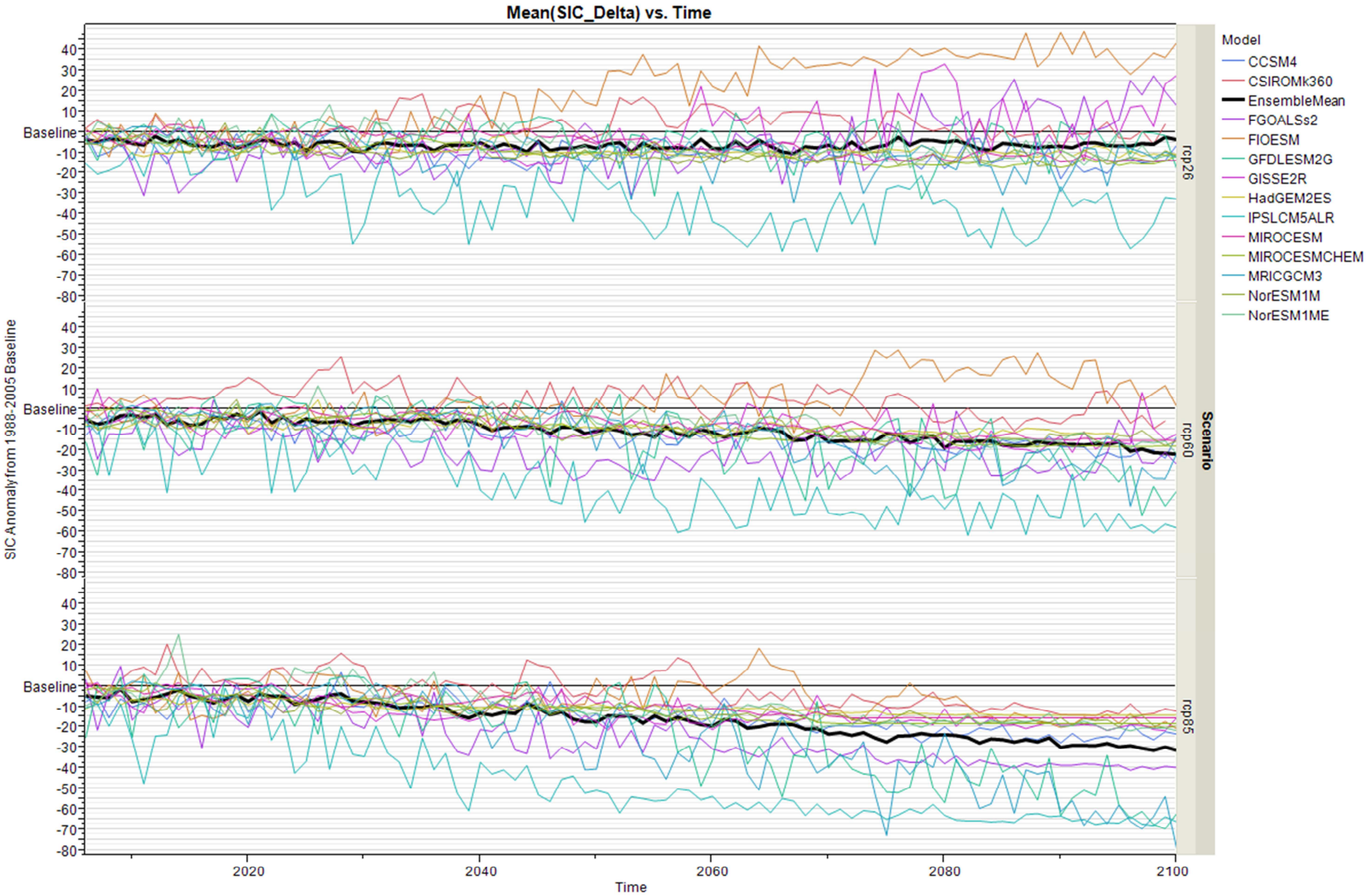


Figure 11. The ensemble mean SIC anomaly from the 1988-2005 baseline under each RCP scenario is plotted against the 13 individual model anomalies from the same baseline. The output from all three harp seal whelping polygons is collapsed to allow for the individual model output to be more perceptible.

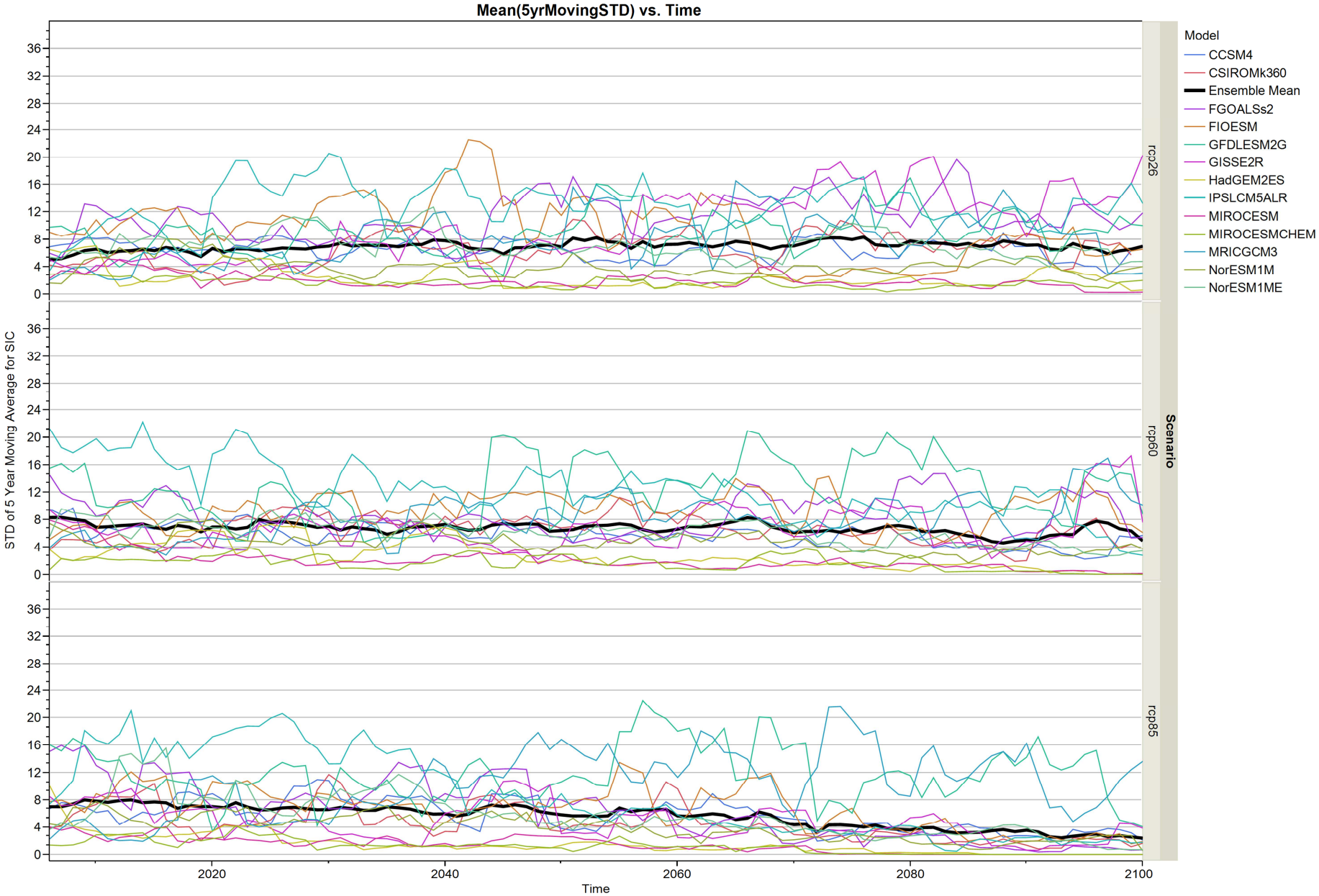


Figure 12. The standard deviation of the 5-year moving average was calculated for each model's prediction of SIC anomaly (as in Figure 11) in order to estimate the year-over-year variability of sea ice concentration in 5 year intervals. In RCPs 2.6 and 6.0 the variability is roughly the same throughout the 21st century but in the most extreme climate change scenario, RCP 8.5, variability gradually decreases, on average, across the whelping locations.

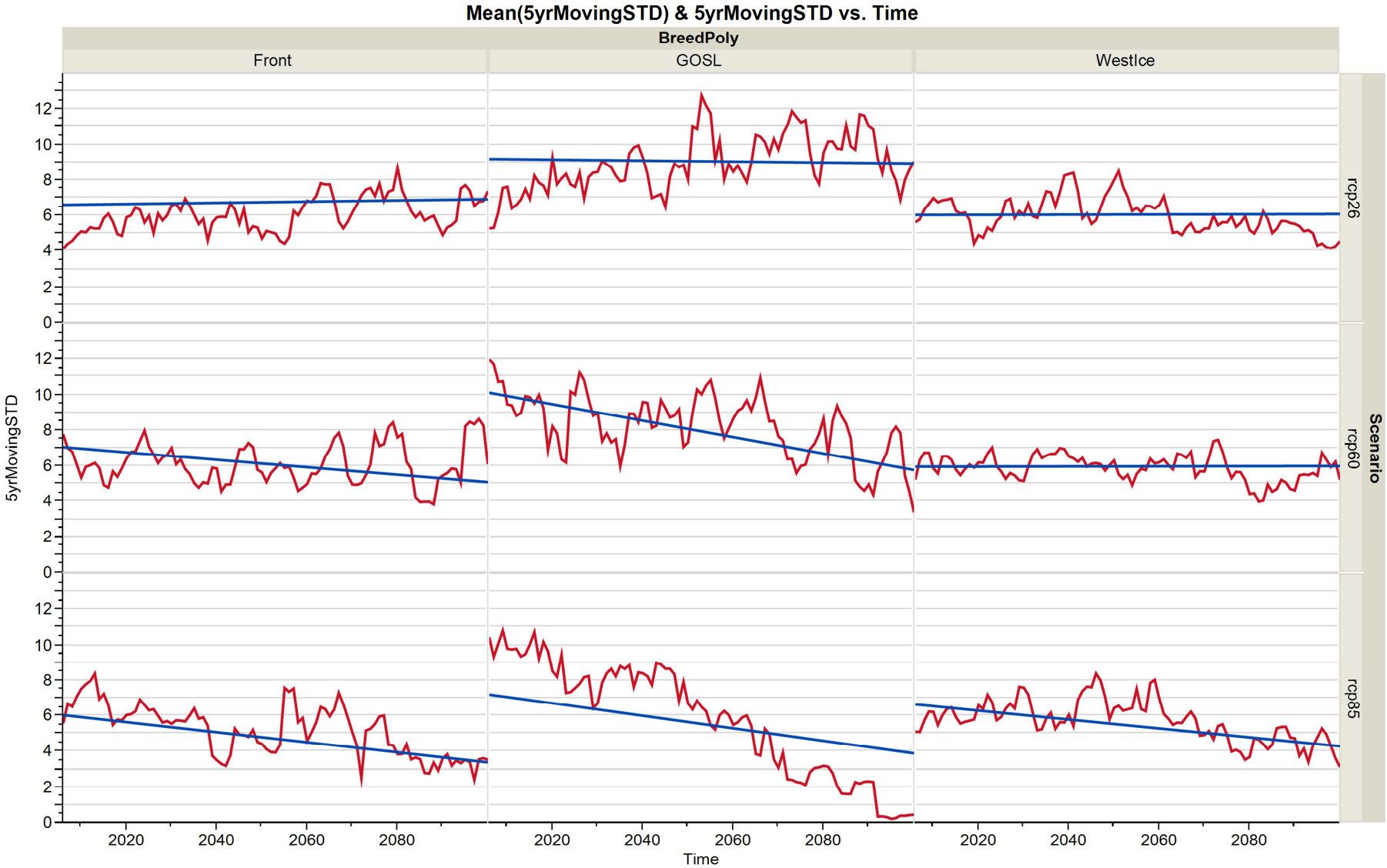


Figure 13. This figure is based on the same data as Figure 12, except broken out further by whelping location, and also averaged across all models so that only the MME mean is shown (red). A linear trend line was fit to each time series to show the general direction of the variance estimate, demonstrating that in most whelping locations and under most scenarios a decrease in variability in sea ice is projected to occur throughout the 21st century.

Conclusion

Mechanistic modeling, as a new form of species distribution modeling, shows great promise not only for the projection of the future location of a given species but also for population projections. These projections can be used to proactively inform policy so as to avert problematic scenarios well before they occur. For example, given what is known about the habitual return of harp seals to the same whelping locations each year, the Canadian government could closely monitor and manage the seal hunt to prevent the extirpation of *Pagophilus groenlandicus* and thus preserve the Canadian seal hunt for years to come.

For others species, this modeling approach can be used to project whether the future range of a species may encroach on human populations, leading to more conflicts. This is particularly useful for species such as the Gray seal and Harbor Seal because they're already commonly found near population centers. For Hawaiian Monk Seals, this modeling approach can help to determine when the seals may be forced away from the small isolated islands they rely on and into conflict with humans inhabiting larger islands.

This study demonstrated that future sea ice projections can be used to understand population dynamics, but that there are many more inputs to this study that could be improved to arrive at better estimates. Better understanding of causal mechanisms is essential in mechanistic modeling, and better population estimates are important for future modeling efforts. Despite these limitations, this study shows that harp seals will likely be affected by declining sea ice under most warming scenarios, and serves as a good first attempt at constraining the prediction of harp seal fitness in a rapidly changing climate.

References

Antonelis, G.A., Baker, J.D., Johanos, T.C., Braun, RC, & Harting, AL. (2006). Hawaiian monk seal(Monachus schauinslandi): Status and conservation issues. *Atoll Research Bulletin, 543*, 75-101.

Burns, J.J. (2002). Harbor seal and spotted seal Phoca vitulina and P. largha. *Encyclopedia of Marine Mammals. Academic Press, San Diego, CA*, 552-560.

Dormann, Carsten F., Schymanski, Stanislaus J., Cabral, Juliano, Chuine, Isabelle, Graham, Catherine, Hartig, Florian, . . . Singer, Alexander. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography, 39*(12), 2119-2131. doi: 10.1111/j.1365-2699.2011.02659.x

Dublin, University College. (2006). The origin of numerical weather prediction. *Nuacht UCD.* Retrieved February 20, 2013, 2013, from http://www.ucd.ie/news/dec06/121506_weather_forecast.htm

Eaton, Brian, Gregory, Jonathan, Drach, Bob, Taylor, Karl, & Hankin, Steve. (2011). NetCDF Climate and Forecast (CF) Metadata Conventions.

Edwards, Paul N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*: MIT Press.

Elith, J., & Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677-697.

Friedlaender, A.S., Johnston, D.W., & Halpin, P.N. (2010). Effects of the North Atlantic Oscillation on sea ice breeding habitats of harp seals (Pagophilus groenlandicus) across the North Atlantic. *Progress in Oceanography, 86*(1), 261-266.

Hawkins, Ed, & Sutton, Rowan. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society, 90*(8), 1095-1107. doi: 10.1175/2009BAMS2607.1

Heithaus, M.R., Frid, A., Wirsing, A.J., & Worm, B. (2008). Predicting ecological consequences of marine top predator declines. *Trends in Ecology and Evolution, 23*(4), 202.

Johnston, DW, Friedlaender, AS, Torres, LG, & Lavigne, DM. (2005). Variation in sea ice cover on the east coast of Canada from 1969 to 2002: climate variability and implications for harp and hooded seals. *Climate Research, 29*(3), 209.

Johnston, David W., Bowers, Matthew T., Friedlaender, Ari S., & Lavigne, David M. (2012). The Effects of Climate Change on Harp Seals (<italic>Pagophilus groenlandicus</italic>). *PLoS ONE, 7*(1), e29158. doi: 10.1371/journal.pone.0029158

Kearney, Michael, & Porter, Warren. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters, 12*(4), 334-350. doi: 10.1111/j.1461-0248.2008.01277.x

Knutti, Reto, Masson, David, & Gettelman, Andrew. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, n/a-n/a. doi: 10.1002/grl.50256

Laidre, Kristin L., Stirling, Ian, Lowry, Lloyd F., Wiig, Øystein, Heide-Jørgensen, Mads Peter, & Ferguson, Steven H. (2008). QUANTIFYING THE SENSITIVITY OF ARCTIC MARINE MAMMALS TO CLIMATE-INDUCED HABITAT CHANGE. *Ecological Applications, 18*(sp2), S97-S125. doi: 10.1890/06-0546.1

Lowry, L.F., Laist, D.W., Gilmartin, W.G., & Antonelis, G.A. (2011). Recovery of the Hawaiian Monk Seal (Monachus schauinslandi): A Review of Conservation Efforts, 1972 to 2010, and Thoughts for the Future. *Aquatic Mammals, 37*(3), 397.

Lunneryd, S. G. (2001). Fish preference by the harbour seal (Phoca vitulina), with implications for the control of damage to fishing gear. *ICES Journal of Marine Science: Journal du Conseil, 58*(4), 824-829. doi: 10.1006/jmsc.2001.1073

Masson, D., & Knutti, R. (2011). Climate model genealogy. *Geophysical Research Letters, 38*(8), L08703. doi: 10.1029/2011GL046864

Moss, Richard H., Edmonds, Jae A., Hibbard, Kathy A., Manning, Martin R., Rose, Steven K., van Vuuren, Detlef P., . . . Wilbanks, Thomas J. (2010). The next generation of scenarios for climate change research and assessment. *Nature, 463*(7282), 747-756. doi: http://www.nature.com/nature/journal/v463/n7282/suppinfo/nature08823_S1.html

Nickisch, Curt. (2012). Seals' Comeback Spells Trouble For Mass. Coast. *All Things Considered*: WBUR.

Pompa, Sandra, Ehrlich, Paul R., & Ceballos, Gerardo. (2011). Global distribution and conservation of marine mammals. *Proceedings of the National Academy of Sciences, 108*(33), 13600-13605. doi: 10.1073/pnas.1101525108

Ramirez-Villegas, Julian, & Jarvis, Andy. (2010). *Downscaling Global Circulation Model Outputs: The Delta Method Decision and Policy Analysis Working Paper No. 1*. International Center for Tropical Agriculture. Retrieved from http://www.ccafs-climate.org/downloads/docs/Downscaling-WP-01.pdf

Read, Andrew J., & Wade, Paul R. (2000). Status of Marine Mammals in the United States

Estado de los Mamíferos Marinos en los Estados Unidos. *Conservation Biology, 14*(4), 929-940. doi: 10.1046/j.1523-1739.2000.99107.x

Richardson, Mark I., Toigo, Anthony D., & Newman, Claire E. (2007). PlanetWRF: A general purpose, local to global numerical model for planetary atmospheric and climate dynamics. *Journal of Geophysical Research: Planets, 112*(E9), E09001. doi: 10.1029/2006JE002825

Smith, Richard J. (1997). Status of the Lacs des Loups Marins Harbour Seal, Phoca vitulina mellonae, in Canada. *Canadian Field Naturalist, 111*(2), 270 - 276.

Solomon, S, Qin, D, Manning, MR, Marquis, M, Averyt, K, Tignor, MH, . . . Chin, Z. (2007). Climate Change 2007: The Physical Science Basis (IPCC): Cambridge University Press.

Taylor, Karl E., Balaji, V., Hankin, Steve, Juckes, Martin, Lawrence, Bryan, & Pascoe, Stephen. (2011, 9 March 2011). *CMIP5 Data Reference Syntax (DRS) and Controlled Vocabularies*. Coupled Model Intercomparison Project.

Taylor, Karl E, Doutriaux, Charles, & Peterschmitt, J. (2004). *Climate Model Output Rewriter (CMOR)*: United States. Department of Energy.

Taylor, Karl E., Stouffer, Ronald J., & Meehl, Gerald A. (2011). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society, 93*(4), 485-498. doi: 10.1175/BAMS-D-11-00094.1

Tebaldi, Claudia, & Knutti, Reto. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365*(1857), 2053-2075. doi: 10.1098/rsta.2007.2076

Temte, Jonathan L. (1994). Photoperiod control of birth timing in the harbour seal (Phoca vitulina). *Journal of Zoology, 233*(3), 369-384. doi: 10.1111/j.1469-7998.1994.tb05271.x

van Vuuren, Detlef P, Edmonds, Jae, Kainuma, Mikiko, Riahi, Keywan, Thomson, Allison, Hibbard, Kathy, . . . Rose, StevenK. (2011). The representative concentration pathways: an overview. *Climatic Change, 109*(1-2), 5-31. doi: 10.1007/s10584-011-0148-z

Wiens, John A., Stralberg, Diana, Jongsomjit, Dennis, Howell, Christine A., & Snyder, Mark A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences, 106*(Supplement 2), 19729-19736. doi: 10.1073/pnas.0901639106

Williams, D. N., Drach, R., Ananthakrishnan, R., Foster, I. T., Fraser, D., Siebenlist, F., . . . Sim, A. (2009). The Earth System Grid: Enabling Access to Multimodel Climate Simulation Data. *Bulletin of the American Meteorological Society, 90*(2), 195-205. doi: 10.1175/2008BAMS2459.1

Williams, TM, Richter, B., Kendall, T., & Dunkin, R. (2011). Metabolic Demands of a Tropical Marine Carnivore, the Hawaiian Monk Seal (Monachus schauinslandi): Implications for Fisheries Competition. *Aquatic Mammals, 37*(3), 372.

Appendix I – Example Python Scripts and Climate Utility Commands

Code snippets use hanging indentation to indicate line wrapping as opposed to carriage returns because this distinction is important in programming. Additional support for these utilities is available in their corresponding manuals:

- NCO <http://nco.sourceforge.net/nco.html>
- CDO <http://www.mpimet.mpg.de/fileadmin/software/cdo/cdo.pdf>
- Python <http://docs.python.org/2/reference/>
- NetCDF CF <http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.6/cf-conventions-multi.html>

The code snippets below are just a sampling of the most important commands and scripts used to prepare the climate model ensembles.

A. Concatenating multiple NetCDF climate records

This step is necessary because some mode output obtained from the ESG is broken into multiple files by time period. It is helpful to have just one file per model per scenario when executing the subsequent commands, so the NetCDF Record Concatenator is used to merge the separate files together.

```
ncrcat -O sic_OImon_NorESM1-ME_rcp60_r1i1p1_200601-205012.nc
      sic_OImon_NorESM1-ME_rcp60_r1i1p1_205101-210112.nc
      sic_OImon_NorESM1-ME_rcp60_r1i1p1_200601-210112.nc
```

B. Averaging across hyperslabs to create climatologies

The NetCDF Record Averager is used in conjunction with hyperslabbing and Fortran indexing to extract a climatology along the time dimension of the input file. Record '902' corresponded to February 2081 and '1130' to February 2100, thus hyperslabbing with a duration of 12 skips all other months and averages the February SIC over a 20 year period. This process was repeated for each climate model, for each 20 year period, and for the month of March as well.

```
ncra -F -d time,902,1130,12
      sic_OImon_bcc-csm1-1_rcp26_r1i1p1_200601-230012.nc
      sic_OImon_bcc-csm1-1_rcp26_r1i1p1_feb2100.nc
```

C. Creating the Anomaly File

The NetCDF Binary Operator is used in 'subtract' mode with 'overwrite' enabled to create an anomaly for the delta downscaling method. It simply subtracts across the record dimension (which at this point only contains 1 record – the month average created in the previous processing step). The anomaly file generated is the change in the variable (SIC) from the historical period.

```
ncbo -y sbt -O sic_OImon_CCSM4_rcp26_r1i1p1_mar2020.nc
      sic_OImon_CCSM4_historical_r1i1p1_mar2005.nc
      sic_OImon_CCSM4_rcp26_r1i1p1_mar2020_deltaAnom.nc
```

D. Regridding the Anomaly File

In order to overcome the unique gridding used by some of the modeling centers (GFDL uses a rotated pole (transverse) projection[8]), the anomalies must be re-projected so that they are in the global regular grid used by most other modeling centers. This is necessary so that GIS applications such as ArcMap can interpret the NetCDF file; it does introduce some error, however, because the existing points must be interpolated to create a regularly-space new grid. CDO is used for this because NCO is not capable of it, and the same resolution as was native to the original file (360x210) was specified for the new regridded file.

```
cdo remapnn,r360x210
      sic_OImon_GFDL-ESM2G_rcp26_r1i1p1_mar2020_deltaAnom.nc
      sic_OImon_GFDL-ESM2G_rcp26_r1i1p1_mar2020_deltaAnomRpj.nc
```

E. Spline Interpolating the Climate Model Centroids to Raster Surfaces with Python

This code iterates through each NetCDF file in the designated data directory and performs the thin-plate splining with barriers technique to create a smooth raster surface. Nested try-catch blocks are used in the code because there are different ways those climates modeling centers chose to name their latitude and longitude variables. The script creates a simple table of the climate model output first, and then projects this data onto a map using the standard projection. The points are then re-projected to a polar stereographic projection to match the

[8] For more information on this projection, see 'Appendix C' in Richardson et al. (2007).

projection of the observation raster used in the downscaling procedure and the spline is run.

```
import os, sys, arcpy
arcpy.CheckOutExtension("Spatial")
arcpy.env.overwriteOutput = True

# Local variables:
dataDir = "E:\\Users\\hunter\\climate\\"
outDir = "C:\\Users\\Hunter\\Documents\\Masters
      Project\\netcdf\\splined\\"

# Set Geoprocessing environments
arcpy.env.outputCoordinateSystem =
      "PROJCS['Polar_Stereographic',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_
      1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Green
      wich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Stereog
      raphic_North_Pole'],PARAMETER['false_easting',0.0],PARAMETER['fal
      se_northing',0.0],PARAMETER['central_meridian',-
      45.0],PARAMETER['standard_parallel_1',70.0],UNIT['Meter',1.0]]"
arcpy.env.snapRaster = dataDir + "NSIDC-SIC_1988-2005_Mar.img"
arcpy.env.cellSize = ".25"
arcpy.env.outputCoordinateSystem =
      "PROJCS['Polar_Stereographic',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_
      1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Green
      wich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Stereog
      raphic_North_Pole'],PARAMETER['false_easting',0.0],PARAMETER['fal
      se_northing',0.0],PARAMETER['central_meridian',-
      45.0],PARAMETER['standard_parallel_1',70.0],UNIT['Meter',1.0]]"
arcpy.env.snapRaster = "NSIDC-SIC_1988-2005_Mar.img"
arcpy.env.extent = "DEFAULT"
arcpy.env.cellSize = "25000"
arcpy.env.geographicTransformations = ""
arcpy.env.mask = ""

#Processing Data Dir into List
climateFiles = os.listdir(dataDir)
chosenFiles = []
for filename in climateFiles:
    if "deltaAnom" in filename and "_feb" in filename:
        chosenFiles.append(filename)
for chosenFile in chosenFiles:
    arcpy.AddMessage("Processing: " + chosenFile)
```

```
    chosenFileParts = chosenFile.split("_")
    # Script arguments
    sic_OImon_CCSM4_rcp26_r1i1p1_mar2020_deltaAnom_nc = dataDir +
      chosenFile
    CCSM4 = chosenFileParts[2]

    # Local variables:
    sic_OImon_CCSM4_rcp26_r1i1p1 =
      sic_OImon_CCSM4_rcp26_r1i1p1_mar2020_deltaAnom_nc
    sic_lyr = sic_OImon_CCSM4_rcp26_r1i1p1

    sic_tmp_reproj_shp = sic_lyr
    continents_poly_rough_shp__2_ =
      "C:\\Users\\Hunter\\Documents\\Masters
      Project\\netcdf\\continents_poly_rough.shp"

    if chosenFileParts[0] == "":
        continue
    if "NorESM1-ME" not in chosenFileParts[2]:# or chosenFileParts[2]
      == "ENSEMBLE":
        continue
    ncVars = ""
    ncDims = ""
    if chosenFileParts[2] == "CSIRO-Mk3-6-0" or chosenFileParts[2] ==
      "FGOALS-s2" or chosenFileParts[2] == "GFDL-ESM2G" or
      chosenFileParts[2] == "GISS-E2-R" or chosenFileParts[2] ==
      "HadGEM2-ES" or chosenFileParts[2] == "MIROC-ESM" or
      chosenFileParts[2] == "MIROC-ESM-CHEM" or chosenFileParts[2] ==
      "MRI-CGCM3":
        ncVars = "sic"
        ncDims = "time;lat;lon"
    else:
        ncVars = "sic;lat;lon"
        ncDims = "time;j;i"

    # Process: Make NetCDF Feature Layer
    try:
      arcpy.MakeNetCDFTableView_md(sic_OImon_CCSM4_rcp26_r1i1p1_mar2020
      _deltaAnom_nc, ncVars, sic_OImon_CCSM4_rcp26_r1i1p1, ncDims, "",
      "BY_VALUE")
        arcpy.AddMessage("made " + ncVars + " | " + ncDims + " Table
      View")
    except:
```

```
        try:
      arcpy.MakeNetCDFTableView_md(sic_OImon_CCSM4_rcp26_r1i1p1_mar2020
      _deltaAnom_nc, "sic;lat;lon", sic_OImon_CCSM4_rcp26_r1i1p1,
      "time;rlat;rlon", "", "BY_VALUE")
            arcpy.AddWarning("made lat and lon Table View")
        except:
            arcpy.AddError(sys.exc_info()[:2])
            arcpy.AddWarning("CONTINUING ADDITIONAL ATTEMPTS")

    # Process: Make XY Event Layer
    arcpy.MakeXYEventLayer_management(sic_OImon_CCSM4_rcp26_r1i1p1,
      "lon", "lat", "sic_lyr_tmp",
      "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
      8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
      74532925199433]];-400 -400 1000000000;-100000 10000;-100000
      10000;8.98315284119522E-09;0.001;0.001;IsHighPrecision", "")


      #arcpy.mapping.AddLayer(arcpy.mapping.ListDataFrames('''C:\Users\
      Hunter\Documents\Masters
      Project\netcdf\Masking.mxd''')[0],"sic_lyr_tmp")
    arcpy.AddMessage("Made XY Event Layer")
      #arcpy.TableToTable_conversion("sic_lyr_tmp","C:/Users/Hunter/Doc
      uments/Masters Project/netcdf","debugging_tbl.dbf","#","""time
      "time" true true false 0 Date 0 8 ,First,#,sic_OImon_MIROC-
      ESM_rcp26_r1_Features,time,-1,-1;lat "lat" true true false 0
      Double 0 0 ,First,#,sic_OImon_MIROC-ESM_rcp26_r1_Features,lat,-
      1,-1;lon "lon" true true false 0 Double 0 0
      ,First,#,sic_OImon_MIROC-ESM_rcp26_r1_Features,lon,-1,-1;sic
      "sic" true true false 0 Float 0 0 ,First,#,sic_OImon_MIROC-
      ESM_rcp26_r1_Features,sic,-1,-1""","#")
    #arcpy.AddWarning("XY TABLE WRITTEN TO FILE - pre-projection")

    # Process: Project
    try:
        arcpy.AddMessage("Projecting - attempt 1")
        arcpy.Project_management("sic_lyr_tmp", "sic_lyr_prj",
      "PROJCS['Polar_Stereographic',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_
      1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Green
      wich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Stereog
      raphic_North_Pole'],PARAMETER['false_easting',0.0],PARAMETER['fal
      se_northing',0.0],PARAMETER['central_meridian',-
      45.0],PARAMETER['standard_parallel_1',70.0],UNIT['Meter',1.0]]",
```

```
    "",
    "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
    8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
    74532925199433]]")
  except:
      arcpy.AddError("original projection attempt failed - retrying
    import from NetCDF")
      arcpy.AddError(sys.exc_info()[:2])
      try:
          arcpy.Delete_management(sic_OImon_CCSM4_rcp26_r1i1p1)
          arcpy.Delete_management("sic_lyr_tmp")

    arcpy.MakeNetCDFTableView_md(sic_OImon_CCSM4_rcp26_r1i1p1_mar2020
    _deltaAnom_nc, "sic;lat;lon", sic_OImon_CCSM4_rcp26_r1i1p1,
    "time;rlat;rlon", "", "BY_VALUE")
          arcpy.AddWarning("made lat and lon Table View - attempt 2 -
    sic;lat;lon | time;rlat;rlon")

    arcpy.MakeXYEventLayer_management(sic_OImon_CCSM4_rcp26_r1i1p1,
    "lon", "lat", "sic_lyr_tmp",
    "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
    8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
    74532925199433]];-400 -400 1000000000;-100000 10000;-100000
    10000;8.98315284119522E-09;0.001;0.001;IsHighPrecision", "")
          arcpy.AddWarning("Made XY Event Layer")

          arcpy.Project_management("sic_lyr_tmp", "sic_lyr_prj",
    "PROJCS['Polar_Stereographic',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_
    1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Green
    wich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Stereog
    raphic_North_Pole'],PARAMETER['false_easting',0.0],PARAMETER['fal
    se_northing',0.0],PARAMETER['central_meridian',-
    45.0],PARAMETER['standard_parallel_1',70.0],UNIT['Meter',1.0]]",
    "",
    "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
    8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
    74532925199433]]")
      except:
          arcpy.AddError(sys.exc_info()[:2])
          try:
              arcpy.Delete_management(sic_OImon_CCSM4_rcp26_r1i1p1)
              arcpy.Delete_management("sic_lyr_tmp")
```

```
      arcpy.MakeNetCDFTableView_md(sic_OImon_CCSM4_rcp26_r1i1p1_mar2020
      _deltaAnom_nc, "sic;lat;lon", sic_OImon_CCSM4_rcp26_r1i1p1,
      "time;j;i", "", "BY_VALUE")
                arcpy.AddWarning("made lat and lon Table View - attempt
      3 - sic;lat;lon | time;j;i")

      arcpy.MakeXYEventLayer_management(sic_OImon_CCSM4_rcp26_r1i1p1,
      "lon", "lat", "sic_lyr_tmp",
      "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
      8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
      74532925199433]];-400 -400 1000000000;-100000 10000;-100000
      10000;8.98315284119522E-09;0.001;0.001;IsHighPrecision", "")
                arcpy.AddWarning("Made XY Event Layer")
                arcpy.Project_management("sic_lyr_tmp", "sic_lyr_prj",
      "PROJCS['Polar_Stereographic',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_
      1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Green
      wich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Stereog
      raphic_North_Pole'],PARAMETER['false_easting',0.0],PARAMETER['fal
      se_northing',0.0],PARAMETER['central_meridian',-
      45.0],PARAMETER['standard_parallel_1',70.0],UNIT['Meter',1.0]]",
      "",
      "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',637
      8137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.01
      74532925199433]]")
            except:
                arcpy.addWarning("NONE OF THE ATTEMPTED NETCDF IMPORT
      SETTINGS WORKED")
                arcpy.AddError(sys.exc_info()[:2])

# Process: Spline with Barriers
    newFolder = chosenFileParts[3] + "_" + chosenFileParts[4] + "_" +
      chosenFileParts[5] + "\\"
    arcpy.CreateFolder_management(outDir, newFolder)
    outFile = outDir + newFolder + chosenFileParts[2][0:13]
    arcpy.AddMessage("Creating: " + outFile)
    arcpy.gp.SplineWithBarriers_sa("sic_lyr_prj", "sic",
      continents_poly_rough_shp__2_, "25000", outFile, "0.2")
    arcpy.AddMessage("done with: " + chosenFile + "\n")
```

F. Creating the ensemble with Python

This code takes the splined anomaly rasters from each climate model per time period and scenario and spatially averages their values to create an ensemble raster.

```
import os, arcpy
arcpy.CheckOutExtension("Spatial")
arcpy.env.overwriteOutput = True
from arcpy import env
from arcpy.sa import *

#PREP FILE AND DIR LIST
baseDirectory = 'C:\\\\Users\\\\Hunter\\\\Documents\\\\Masters
      Project\\\\netcdf\\\\splined\\\\'
anomalyRasters = ['CCSM4','CSIRO-Mk3-6-0','FGOALS-s2','FIO-ESM','GFDL-
      ESM2G','GISS-E2-R','HadGEM2-ES','IPSL-CM5A-LR','MIROC-ESM-
      CHE','MIROC-ESM','MRI-CGCM3','NorESM1-M','NorESM1-ME']

workspaceList = os.listdir('C:\\Users\\Hunter\\Documents\\Masters
      Project\\netcdf\\splined')

def isDirectory(i):
    if i.find(".") == -1 and i.find("info") == -1:
        return True
    else:
        return False

workspaceList = filter(isDirectory,workspaceList)
print workspaceList
workspaceList = ['rcp26_r1i1p1_feb2060']

#RUN MAP ALGEBRA ITERATIVELY
for workspace in workspaceList:
    rasterList = []
    for raster in anomalyRasters:
        rasterList.append('Con(IsNull(arcpy.Raster("' + baseDirectory +
      workspace + '\\'+ raster + '")),0,arcpy.Raster("' + baseDirectory
      + workspace + '\\'+ raster + '")) + ')
    rasterListStr = ''.join(rasterList)
    print rasterListStr
    #outraster = ( arcpy.Raster("C:\\Users\\Hunter\\Documents\\Masters
      Project\\netcdf\\splined\\rcp26_r1i1p1_feb2020\\FGOALS-s2") +
```

```
    arcpy.Raster("C:\\Users\\Hunter\\Documents\\Masters
    Project\\netcdf\\splined\\rcp26_r1i1p1_feb2020\\CCSM4") ) / 2
  #outraster.save(baseDirectory + workspace + '\\ensemble_anom')
  outraster = eval( rasterListStr[:-3] ) / 13
  outraster.save(baseDirectory + workspace + '\\ensemble_anom')
  print "Finished preparing ensemble_anom for: " + workspace
```

Appendix II – A Primer on NetCDF and Multidimensional data in ArcMap

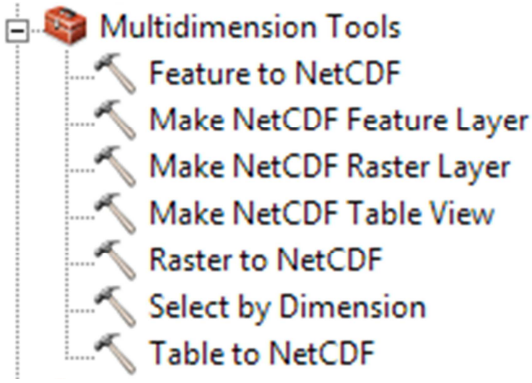
This guide was prepared during the summer of 2012 while at my research internship at Columbia University. I worked with a small group that produces an educational climate model (EdGCM) from a simplified version of the NASA GISS Model II climate model. The intent of the guide is to inform users of EdGCM how to understand and display the NetCDF output from EdGCM in ArcMap GIS software.

Below is an overview of the process for importing EdGCM output (in the NetCDF format) into ESRI's ArcMap 10 Geospatial Analysis software. If basic visualization of the EdGCM outputs is desired, we recommend use of EVA or Panoply instead. The utility of an advanced geospatial analysis package such as ArcMap is advantageous, however, if the outputs are being used for further analysis. Performing network analyses for hydrological modeling or performing impact studies requiring use of many other geospatial layers such as land use / land cover satellite imagery, for example, may require the additional capabilities of a geospatial analysis platform.

While ESRI's ArcGIS package is the chosen tool for this tutorial because it is widely used, many other geospatial tools exist which can read EdGCM's NetCDF file output. This guide assumes modest familiarity with both EdGCM and ArcMap software.

ArcMap Multidimensional Tools

ArcMap's tools for handling NetCDF files are located in the 'Multidimension Tools' toolbox. This toolbox includes 3 import tools (Make NetCDF…), 3 export tools (… to NetCDF), and one selection tool (Select by Dimension).



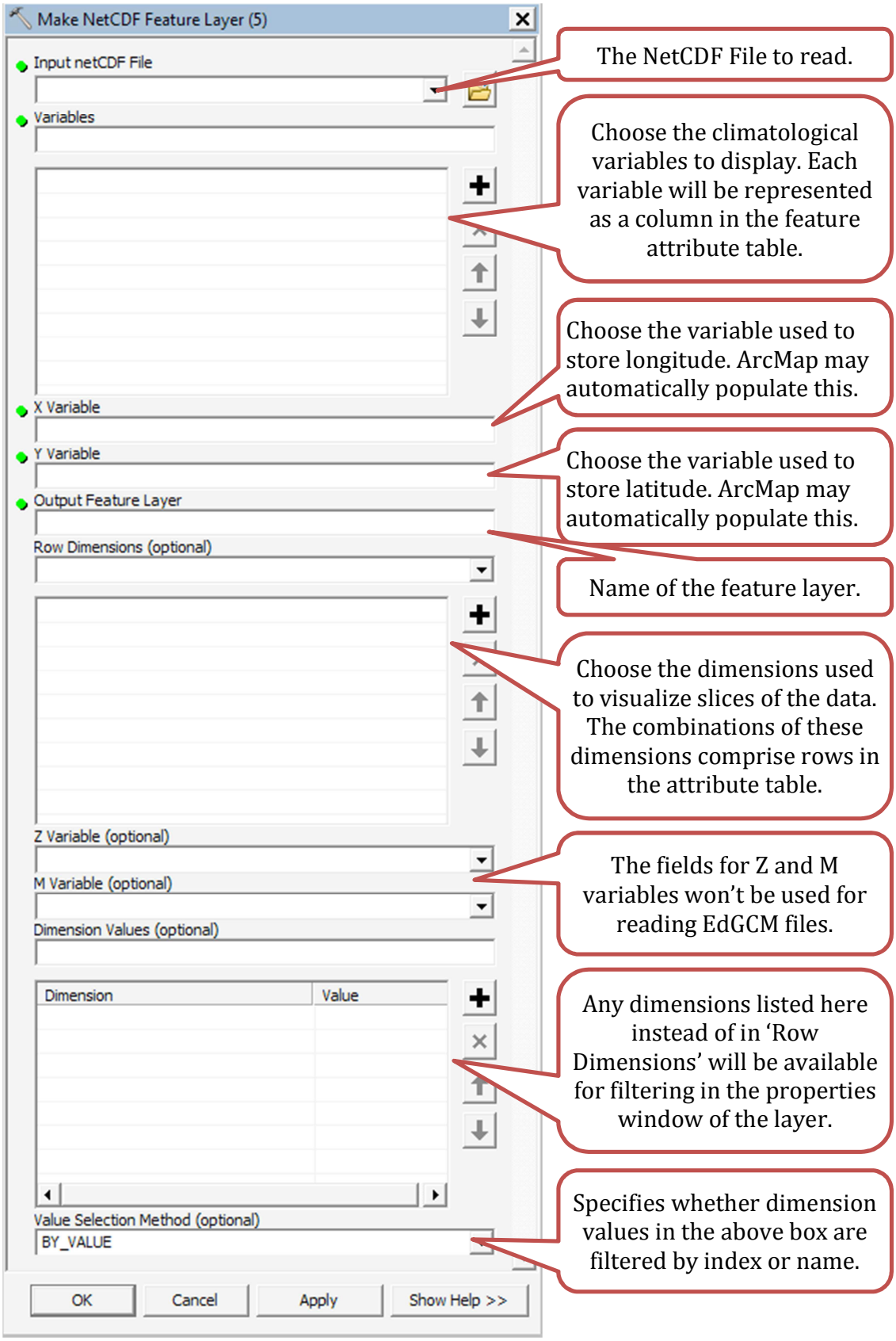
The import tools allow for the conversion of NetCDF array data to three of the fundamental data types that ArcMap can handle: vectors (features), rasters, and tables. Importing to a Feature Layer is the best option for EdGCM data because it most accurately represents the output from the model (NetCDF data is prepared as an array of discrete data points per time step rather than continuous values). For this reason, this method will be the only one covered in this guide.

The export tools will take a raster, feature, or table and output a NetCDF file based on the variables and dimensions chosen in the dialog box. Since ArcMap outputs files in CF (1.0) format, the tool will constrain certain choices in the dialog boxes to adhere to this standard. Use of this feature, while not covered in detail here, enables the export of EdGCM outputs that have been processed by ArcMap. If exported correctly, these files are viewable in any NetCDF reader.

The selection tool 'Select by Dimension' allows for programmatic control of the display of dimensional data. It allows users to select a specific value for a chosen dimension (ex. selecting to display "June" out of all available values for "Month").

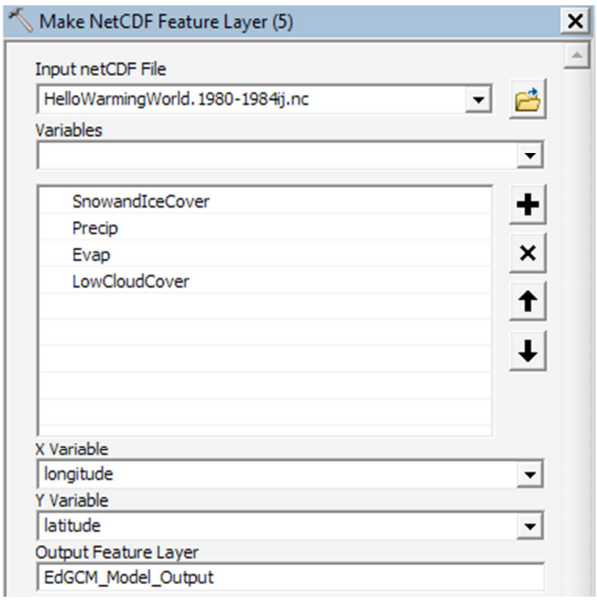
Making a NetCDF Feature Layer

The most useful NetCDF tool for analysis of EdGCM output in ArcMap's 'Multidimension Tools' toolbox is 'Make NetCDF Feature Layer'. This tool reads NetCDF files to make a point feature set. Each point is spatially located using EdGCM's geographic coordinates, and contains values for all exported variables. Below is an overview of the form used to set tool parameters:


Make NetCDF Feature Layer (5)
Input netCDF File
Variables
X Variable
Y Variable
Output Feature Layer
Row Dimensions (optional)
Z Variable (optional)
M Variable (optional)
Dimension Values (optional)
Dimension
Value
Value Selection Method (optional)
BY_VALUE
OK
Cancel
Apply
Show Help >>
The NetCDF File to read.
Choose the climatological variables to display. Each variable will be represented as a column in the feature attribute table.
Choose the variable used to store longitude. ArcMap may automatically populate this.
Choose the variable used to store latitude. ArcMap may automatically populate this.
Name of the feature layer.
Choose the dimensions used to visualize slices of the data. The combinations of these dimensions comprise rows in the attribute table.
The fields for Z and M variables won't be used for reading EdGCM files.
Any dimensions listed here instead of in 'Row Dimensions' will be available for filtering in the properties window of the layer.
Specifies whether dimension values in the above box are filtered by index or name.

Choosing the Climatological Variables for Import

The first section of the tool allows you to specify the input file, which then populates the remaining dropdown lists in the form with appropriate value choices. In this example, a NetCDF file output with 4 variables was chosen, and the 4 variables were all selected to be included when the file is imported.



The other values listed in the 'Variables' drop-down list (latitude, longitude, and months), are "coordinate variables" which are used to name the dimensions rather than having to refer to them by index number. Not all (non-coordinate) variables must be included, but each variable included will be placed in its own column in the attribute table of the resulting point feature set in ArcMap, so it is best to include all available variables at this point and separate them later.

The 'X Variable' and 'Y Variable' fields correspond to longitude and latitude, and should be automatically populated by ArcMap due to its compliance with the CF convention. If not, simply select the appropriate values.

Since it is not common to specify the datum (a critical attribute in geospatial analysis) in a NetCDF file, ArcMap assumes the data is in the Geographic Coordinate System based on the World Geodetic System 84 Datum (GCS_WGS_1984), which is correct for EdGCM output.

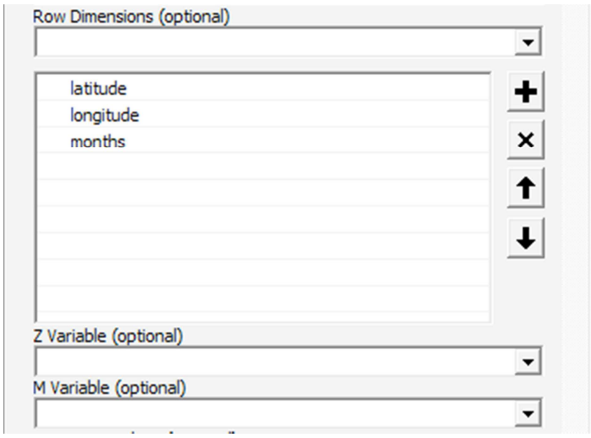
The Output Feature Layer should be set to the desired name of the layer.

Setting the Dimensions for Importing Climatological Variables

The next section concerns setting the dimensions that ArcMap should use to import and display the chosen variables. This section must be set using all available dimensions for the variables chosen (those listed in the header of the NetCDF file):

```
float SnowandIceCover(months, latitude, longitude) ;
float Precip(months, latitude, longitude) ;
float Evap(months, latitude, longitude) ;
float LowCloudCover(months, latitude, longitude) ;
```

See Appendix I for the full NetCDF header of the example file processed here.



By choosing these dimensions, ArcMap correctly transforms the multidimensional data into a flat table by performing a Cartesian product operation. Since the dimension sizes of longitude, latitude, and month are 36, 24, and 12, and since there is a data value at the intersection of each dimension, ArcMap will generate 36 * 24 * 12 = 10,368 rows of geo-referenced data for every possible dimension combination in the feature attribute table.

Note: this example included only 12 calendar months from EdGCM, but ArcMap can read output of seasonal or annual averages as well – all in the same file.

FID	Shape	latitude	longitude	months	Precip	SnowandIceCover	LowCloudCover	Evap
0	Point	-90	-175	January	0.354954	100	32.147652	0.549215
1	Point	-90	-165	January	0.354954	100	32.147652	0.549215
2	Point	-90	-155	January	0.354954	100	32.147652	0.549215
3	Point	-90	-145	January	0.354954	100	32.147652	0.549215
4	Point	-90	-135	January	0.354954	100	32.147652	0.549215
5	Point	-90	-125	January	0.354954	100	32.147652	0.549215
6	Point	-90	-115	January	0.354954	100	32.147652	0.549215
7	Point	-90	-105	January	0.354954	100	32.147652	0.549215
8	Point	-90	-95	January	0.354954	100	32.147652	0.549215
9	Point	-90	-85	January	0.354954	100	32.147652	0.549215
10	Point	-90	-75	January	0.354954	100	32.147652	0.549215
11	Point	-90	-65	January	0.354954	100	32.147652	0.549215
12	Point	-90	-55	January	0.354954	100	32.147652	0.549215
13	Point	-90	-45	January	0.354954	100	32.147652	0.549215
14	Point	-90	-35	January	0.354954	100	32.147652	0.549215
15	Point	-90	-25	January	0.354954	100	32.147652	0.549215

0 (0 out of 10368 Selected)

Dimensions can be entered in any order; ArcMap will still correctly compute the Cartesian product based on the order defined in the NetCDF file. The order specified in the input tool only alters column order in the attribute table. However, take care not to omit any dimension specified in the variable definition in the header of the NetCDF file, as ArcMap will still read this file, incorrectly, without warning. For example, if the 'months' dimension is omitted, only data for the first value of this dimension will be taken (January, in this case). Evidence of this is apparent in the resultant attribute table, which only will contain 36 * 24 * 1 = 864 rows.

Filtering by Dimension Value

The final section of the import form allows for the specification of 'Dimension Values' that can act as a filter so that only certain values in particular dimensions are shown. EdGCM output uses a 2-dimensional character array to label months with their names, but ArcMap cannot read these values properly. Thus, instead of

using this functionality, it is recommended that filtering be done after the data is imported using the 'Select by Attribute' tool.

Working with EdGCM data in ArcMap

Once the data is imported as point features, they will be displayed in ArcMap as seen below. If no dimension filters were set, all data points for each time step will be rendered on top of each other at each cell centroid so that only the first time step (at index 0) will be visible (ESRI Light Gray Basemap added for context):

At this point, there are many ways to manipulate the data in ArcMap depending on the type of analysis to be performed. The method discussed above for importing EdGCM data prepares the feature layer to be flexible so that further changes to symbology or calculations can be readily performed. For example, selection of a particular month's data is as simple as modifying the 'Definition Query' in 'Layer Properties' using basic SQL (ex. "months" = 'June'). If a new shapefile per time step is desired, using the 'Select by Attribute' function iteratively with 'Copy Features' would output layers by time step.

Appendix III: Example NetCDF Output

EdGCM data prepared in the NetCDF format is readable by any NetCDF viewer or libraries. These files contain a header that defines the dimensions of the data, variables, and attributes of those variables. EdGCM's NetCDF output contains the following descriptive header components (comments added) produced using the 'ncdump' command of NetCDF Operators:

```
netcdf FileName {
dimensions:
        longitude = 36 ;        // 10° * 36 = 360° global longitudinal coverage
        latitude = 24 ;         // 22 * 8° boxes + 2 * 2° boxes = 180° coverage
        months = 12 ;           // 12 in this case pertaining to calendar months
        bounds = 2 ;            // 2 latitudinal bounds (where cell edges are)
        cbounds = 10 ;          // used to create a 2D array for string support
                                // [http://tinyurl.com/6na886] for more info
variables:
// 4 standard variables always present in output.
// These "coordinate variables" are named after a dimension, and should be
// selected as dimensions when importing EdGCM NetCDF data into ArcMap.
        float longitude(longitude) ;
                longitude:long_name = "Longitude" ;
                longitude:units = "degrees_east" ;
                longitude:topology = "circular" ;
                longitude:modulo = 360.f ;
        float latitude(latitude) ;
                latitude:long_name = "Latitude" ;
                latitude:units = "degrees_north" ;
                latitude:bounds = "bounds_lat" ;
        float bounds_lat(latitude, bounds) ;
        char months(months, cbounds) ;
                months:long_name = "month or seasonal phase" ;
// The following variables will change based on user selection in EdGCM;
// these are the variables that can be plotted spatially in ArcMap.
        float SnowandIceCover(months, latitude, longitude) ;
                SnowandIceCover:long_name = "Snow and ice coverage" ;
                SnowandIceCover:units = "percent" ;
        float Precip(months, latitude, longitude) ;
                Precip:long_name = "Precipitation" ;
                Precip:units = "mm/day" ;
        float Evap(months, latitude, longitude) ;
                Evap:long_name = "Evaporation" ;
                Evap:units = "mm/day" ;
        float LowCloudCover(months, latitude, longitude) ;
                LowCloudCover:long_name = "Low level cloud cover" ;
                LowCloudCover:units = "percent" ;
```

```
// global attributes:
                :Conventions = "GISS-02" ;   //future output may use "CF"
                :title = "runID" ;
                :institution = "Goddard Institute for Space Sciences" ;
                :source = "GISS Model II" ;
                :run_number = "runID" ;
                :kMDItemAuthors = "Your Name" ;
                :kMDItemComment = "Maps" ;
                :kMDItemContactKeywords = "edgcm.columbia.edu" ;
                :kMDItemContentType = "application/x-netcdf" ;
                :kMDItemCreator = "EdGCM 3.2 (926)" ;
                :kMDItemDescription = "runDescription" ;
                :kMDItemHeadline = "runHeadline" ;
                :kMDItemKind = "Maps for years dateRange" ;
                :kMDItemNumberOfPages = "3" ;
                :kMDItemOrganizations = "edgcm.columbia.edu" ;
                :kMDItemProjects = "projectID" ;
                :kMDItemTitle = " runID" ;
                :kMDItemVersion = "Model II 1.0.7" ;
                :kMDItemWhereFroms = "computerName, operatingSystem" ;
```

Additional information on the NetCDF Format:
http://www.unidata.ucar.edu/software/netcdf/
Additional information on the CF Convention:
http://cf-pcmdi.llnl.gov/
Additional support for ArcMap Analyses:
http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html