Predicting Past and Future Variations in Global Mean Surface Temperature with a Simple Model

Danya Levy

Advisor: Alexey Fedorov

Second Reader: Ronald Smith

May 1, 2019

A Senior Thesis presented to the faculty of the Department of Geology and Geophysics, Yale University, in partial fulfillment of the Bachelor's Degree.

ABSTRACT

General Circulation Models (GCMs) are critical to understanding our climate system but expensive to run and sometimes difficult to interpret. This study investigates the potential for a simple, physically based climate model, based on one differential equation, to perform some of the most important functions of GCMs: reproducing and predicting variations in global mean surface temperature (GMST). The model used in this project incorporates three forcings: equivalent carbon dioxide concentrations, shortwave reduction caused by volcanic eruptions, and effects of climate variability related to ENSO. These inputs are sufficient to reproduce observed GMST variations since 1880 with a root mean square error of only 0.08 C. In order to determine how to improve and understand this result, and investigate whether GCM data can be reproduced and predicted by the simple model, we ran a variety of experiments, altering both the equation and its inputs. The results indicated, for example, that in reproducing observed GMST, incorporating the Atlantic Multidecadal Oscillation did not significantly reduce error. For GCM data, in a majority of cases, the simple model reproduced GMST with a high degree of accuracy. Using combined Nino 3, 4 data with the warming trend removed provided the best results; so did selecting either atmospheric or equivalent carbon dioxide concentrations based on the GCM's aerosol interactivity. The model was less successful in predicting GCM temperature data into the future based on the historical period; adding another term to the model representing deep ocean temperatures helped to improve those predictions in many cases. Overall, the accuracy of these experiments demonstrates that while GCMs are extremely complex, when it comes to GMST, they essentially boil down to the same factors as the simple model and follow the basic physics of global warming. This has the potential to make climate modeling more accessible and easier to comprehend, with implications for policy and education.

TABLE OF CONTENTS

1. INTRODUCTION

a. Background on Simple Model

In their paper "The Extreme El Nino of 2015-2016 and the end of global warming hiatus," published in 2017, Alexey Fedorov and Shineng Hu used a simple model, based on one equation, to investigate recent patterns in global mean surface temperature (2017). Following the year 2000, rates of increase of global mean surface temperature (GMST) slowed; this has been dubbed the "global warming hiatus." Since the year 2014, however, rapid warming resumed. Hu and Fedorov note that "a number of physical mechanisms were proposed to explain the hiatus, including but not limited to eastern Pacific cooling, Walker Cell strengthening, enhanced ocean heat uptake, and changes in stratospheric water vapor and aerosols" (2017). In order to investigate the cause of these recent fluctuations, they built a simple, physically based model to reproduce GMST data during the 1880-2015 time period.

This model builds off of the connection between GMST and El Nino - Southern Oscillation, or ENSO, which defines the fluctuations in sea surface temperature in the equatorial Pacific. ENSO has a significant effect on interannual climate variability; generally, the "warmer or cooler than normal ocean temperatures can affect weather patterns around the world by influencing high and low pressure systems, winds, and precipitation" (Hu and Fedorov, 2017). Swings in ENSO affect vertical energy convergence in the Pacific, which in turn modulates heating rates of the tropical atmosphere and GMST. The model developed by Hu and Fedorov--a first-order differential equation--takes into account this connection, as well as two other crucial climate forcings: volcanic eruptions and carbon dioxide concentrations. While greenhouses gases other than carbon dioxide are also play an important role in global warming, it was initially assumed that their effect on GMST would be compensated by aerosols. The authors also noted that incorporating the Atlantic Multidecadal Oscillation (AMO)--a mode of natural variability quantified by sea surface temperatures in the North Atlantic--into the model might improve accuracy, as the model's errors "coincide with the negative and positive phases of the AMO" (Hu and Fedorov, 2017; NCAR).

While it simplifies the climate system to only three key forcings, this model reproduces GMST anomalies over the course of the 1880-2015 system very closely, with RMS error of approximately 0.08 C. Error decreases in the period after 1950, when observations became more accurate due to improved technology. The global warming hiatus and subsequent temperature rise are well captured as well, and isolating the forcings can indicate what caused these phenomena. Suppressing ENSO, and leaving carbon dioxide and volcanic aerosols, models a warming trend with GMST decreases due to volcanic activity; under these simulations, the global warming hiatus is not present. Running the model with volcanic forcing suppressed and ENSO activated, on the other hand, closely reproduces the hiatus and the return of global warming that followed. The authors conclude that volcanic activity did not contribute to the hiatus, but rather that "ENSO-related anomalous heating originating in the tropical Pacific act to modulate GMST on interannual to interdecadal time scales, shaping the hiatus and the subsequent temperature increase" (Hu and Fedorov, 2017). Training the model using the

1880-1995 time period and then predicting GMST variations through 2015 confirms these results.

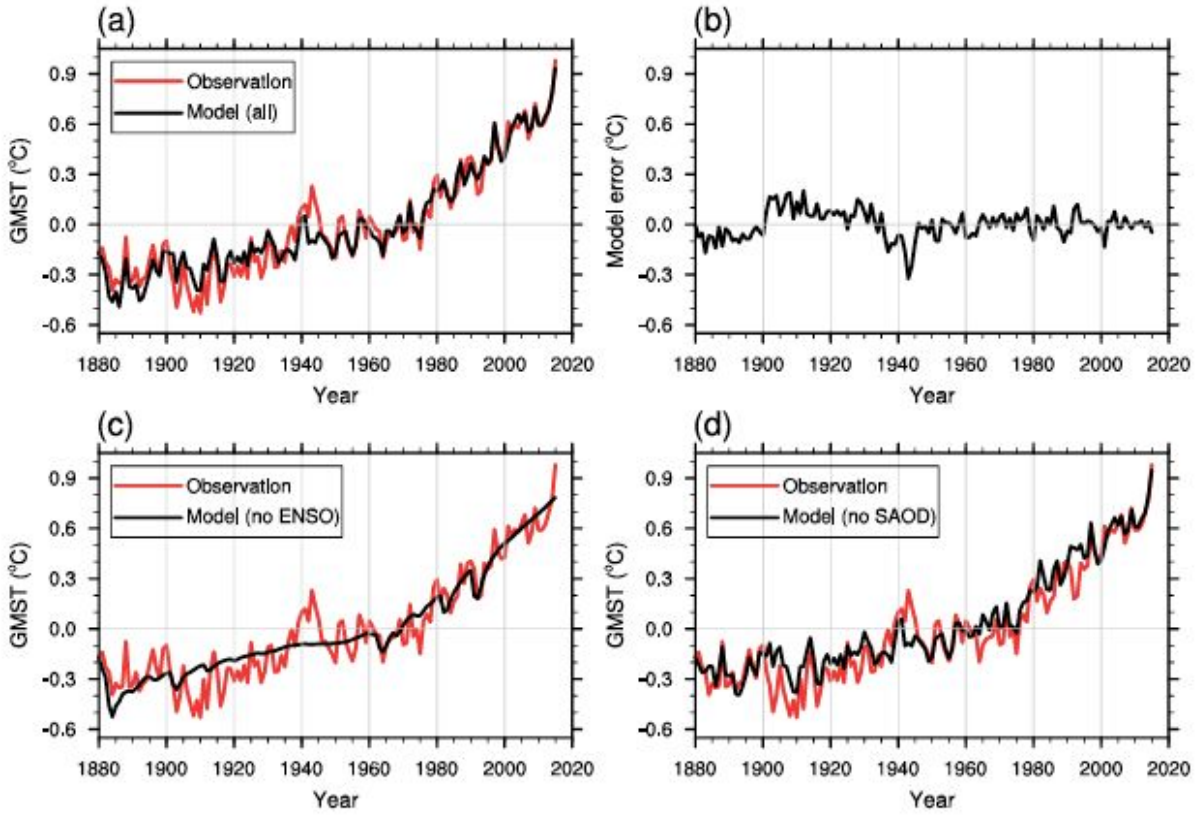


Figure 1: *Figure from Hu and Fedorov depicting the results of using the simple model and suppressing certain forcings (2017). Black indicates GMST calculated from the simple model; red indicates GMST from observations. a) All forcings used. b) Model error (differences between model and observations) for scenario a. c) Simple model with ENSO suppressed. d) Simple model with volcanic aerosol forcing suppressed.*

b. General Circulation Models and CMIP5

Climate models are the most important tools available to scientists to understand our ocean and atmosphere. General Circulation Models (GCMs) are the most comprehensive and complex type of climate model, providing numerical simulations on a three-dimensional grid with a typical horizontal resolution of 250-600 km and many layers in the atmosphere and ocean (IPCC Data Distribution Center). In order to provide simulations on this grid, the models solve a set of fundamental equations, including conservation of momentum, mass, energy and moisture, and equations of state. GCMs often take into account atmospheric chemistry and circulation, ocean

circulations, atmosphere-ocean interactions, the cryosphere and land surface, ecological changes, the carbon cycle, anthropogenic forcings, and other factors; the details can vary between models (Tel Aviv University). The IPCC suggests that "there continues to be very high confidence that models reproduce observed large-scale surface temperature patterns (pattern correlation of ~0.99), though systematic degrees are found in some regions" (Stocker, 2013, p. 743). In their assessment of coupled models, Reichler and Kim note that while current models are not yet ideal, considerable progress has been made since the earliest generation of climate models; they attribute these advancements to more sophisticated model parameterizations and to improved computational resources (2008). The authors also note that the multimodel mean, obtained by averaging the data produced by all simulations, usually provides more accurate results than any individual GCM.

The Coupled Model Intercomparison Project (CMIP5) provides a framework for scientists to systematize climate modeling experiments. The project provides a standard set of simulations so that researchers may evaluate model performance in both replicating the past and predicting the future. CMIP5's structure also allows scientists to identify and explain the differences in model predictions (Taylor et al., 2012). CMIP5 simulations are split into several groups: a "core" category, as well as "tiers" 1 and 2. (This is summarized in Figure 3). All groups complete the core simulations. The tier 1 simulations "examine specific aspects of climate model forcing, response, and processes" and the tier 2 experiments investigate those models in greater detail (Taylor et al., 2012). CMIP5 includes two focus areas: one near-term, or decadal, and one long-term, or over a century or more. This paper will mention only the long-term simulations. For these long-term experiments, the core experiments include Representative Concentration Pathways 4.5 and 8.5, and tier 1 and 2 simulations include RCPs 2.6 and 6.0, extensions of all RCPs to to the year 2300, and experiments quantifying aerosol forcing, among others.

These Representative Concentration Pathways are "based on selected scenarios from four modeling teams/models (NIES/AIM, IIASA/MESSAGE, PNNL/MiniCAM, and PBL/IMAGE)," known as Integrated Assessment Models (IAMs) (RCP Database; Meinshausen et al.). They complement and replace scenario-based projections from earlier stages of the Intercomparison

Project, and are used in the CMIP5 simulations (Meinshausen et al., 2011). Meinshausen et al. note that because the IAMs used to produce the concentrations differed from each other, it was necessary to compile historical concentrations and harmonize emissions "to common 2000-2005 emission levels," and then project to 2100; data past 2100 are known as ECPs (2011). The resulting atmospheric carbon dioxide and equivalent carbon dioxide for the RCPs are shown in Figure 2. Each pathway describes a different scenario. The RCP with the least warming, 2.6, includes GMST increases of 1.5 degrees C by 2100; the one with the most warming, 8.5, includes warming of 2.5 degrees by that time.

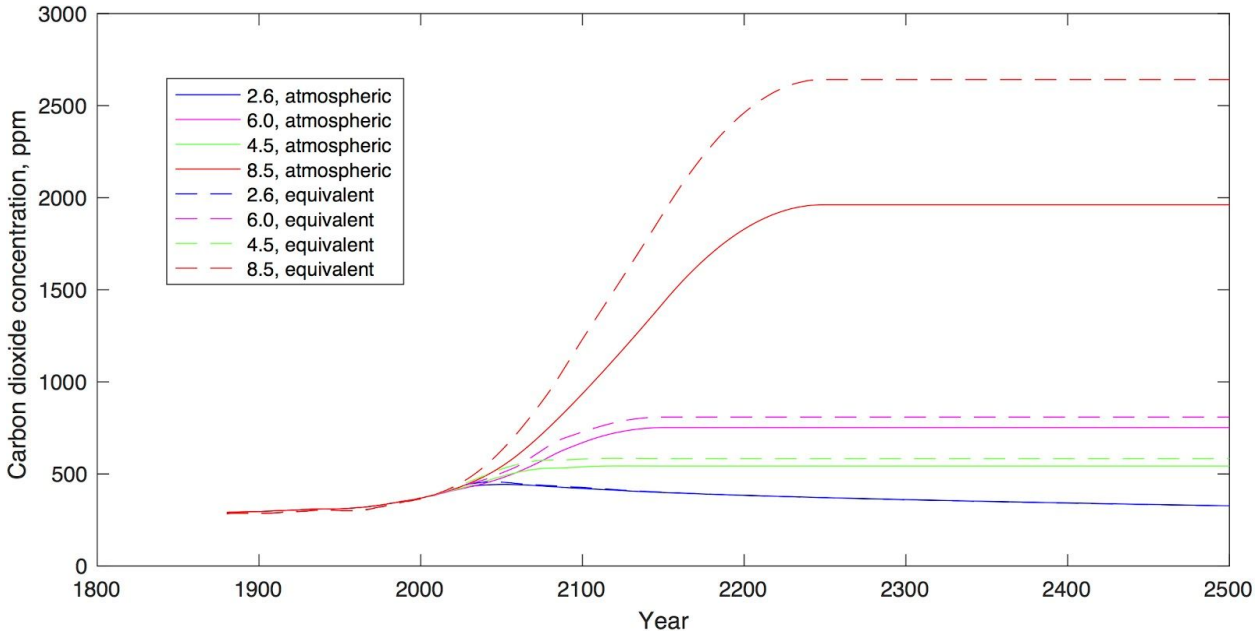


Figure 2: *Atmospheric carbon dioxide and equivalent carbon dioxide concentrations for the Representative Concentration Pathways, years 1880-2500.*

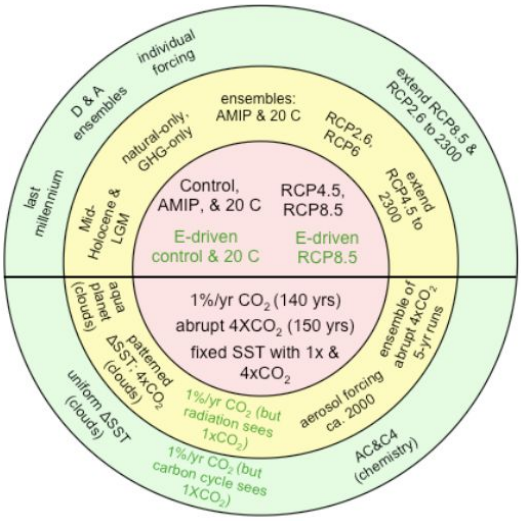


Figure 3*: Diagram summarizing the core, first and second tier experiments of the CMIP5 long-term simulations.*

c. Aerosols

Aerosols--solid or liquid particles suspended in the atmosphere--affect Earth's radiation balance by scattering and absorbing solar radiation. These particles can be both natural or anthropogenic, and are either emitted into the atmosphere directly or formed from gaseous precursors. The IPCC separates atmospheric aerosols into several categories: inorganic species, mineral species, organic species, black carbon and primary biological aerosol particles (Stocker, 2013, p. 595). The main natural source of aerosols is volcanic emissions, which primarily emit sulphate aerosols to the atmosphere. Other natural sources include terrestrial ecosystems and wind erosion. Anthropogenic sources span from combustion of fossil fuels and biofuels to agricultural practices (Stocker, 2013, p. 597).

The effects of aerosols on the atmosphere are still relatively poorly understood, and have a hand in some of the largest uncertainties in climate predictions (Stocker, 2013, p. 574). "Climate Change 2013: The Physical Science Basis," the first report in which the IPCC systematically assessed clouds and aerosols, notes that "our inability to better quantify non-greenhouse gas RFs [radiative forcings], and primarily those that result from aerosol-cloud interactions, underlie difficulties in constraining climate sensitivity from observations even if we had a perfect knowledge of the temperature record" (Stocker, 2013, p. 576). Overall, however, research suggests that anthropogenic aerosols have a net cooling effect, and have decreased the

amount of global warming that the atmosphere has experienced thus far (Stocker, 2013, p. 623). Because aerosols are often co-emitted with greenhouse gases, this complicates future efforts to fight global warming. For example, in their study on the impacts of removing anthropogenic aerosol emissions, Samset et al. indicate that removing anthropogenic aerosols would induce a global mean surface heating of 0.5 - 1.1 C (2018).

For experiments requiring greenhouse gas or aerosol data, the Integrated Assessment Model Consortium provides the "concentrations, emissions and time-evolving land use changes to be used in the simulations" (Van Vuuren et al., 2011). Aerosol data are included in the data provided by each Representative Concentration Pathway repository, along with data on land use, greenhouse gas concentrations, and other important factors; many of the most advanced climate models now take advantage of this dataset. For these simulations, the Community Atmosphere Model 3.5 generated the necessary gridded concentration data. Concentrations were provided for the historical period, RCP period (2005-2100) and the extension (2100-2300). Some of the GCMs involved, however, may not use the aerosol input from the RCPs.

The most recent IPCC Physical Science Basis identifies three categories of GCMs in relation to their usage of aerosol concentrations: fully interactive, semi-interactive and noninteractive. Interactivity implies "a physically based prognostic equation and at least a two-way coupling with another component, allowing climate feedbacks" (Stocker, 2013, p. 747). In the case of some GCMs--the iterations of GISS models in particular--aerosol interactivity may vary between different model runs. For models that do take into account aerosol effects, many "now include the basic features of the sulphur cycle" and several "are currently capable of simulating the mass, number, size distribution and mixing state of interacting multi-component and aerosol particles" (Stocker, 2013, p. 752). Despite these advancements, however, the IPCC identifies the usage of aerosol particles as a crucial source of uncertainty in CMIP5.

d. Equivalent CO_2

There are two primary methods of quantifying the climate perturbations that result from changing the concentrations of greenhouse gases or aerosols in the atmosphere. The first, radiative forcing, is measured in watts per square meter, and in the context of gases and aerosols

represents a difference in irradiance at the tropopause following a concentration change (Gohar and Shine, 2007). Gohar and Shine note that global mean surface temperature (GMST) is correlated to radiative forcing (RF) using the following equation:

$$\Delta T_s = \lambda RF$$

(Equation 1)

where ΔT_s represents GMST change, λ is climate sensitivity (to be discussed in depth later on in this paper), and RF is the radiative forcing (2007). Instead of measuring perturbations in terms of change in irradiance, equivalent CO_2 describes the concentration of CO_2 required to achieve the same radiative forcing. This metric is measured in terms of CO_2 concentrations because "it is currently believed to be the largest contributor to anthropogenic radiative forcing." While radiative forcing is a useful metric, the authors argue that equivalent CO_2 is easier for policymakers and the general public to understand.

The IPCC's Climate Change 2001 report provides the following simplified expression for calculating the radiative forcing due to carbon dioxide:

$$\Delta F = \alpha ln(\frac{C}{C_0})$$

(Equation 2)

where ΔF represents change in radiative forcing, C is the present carbon dioxide concentration, C_0 represents the pre-industrial carbon dioxide concentration of 278 ppm, and the constant α is equal to 5.35 W/m^2. The RCP Concentration Calculation & Data Group provides the same expression with its equivalent carbon dioxide data for the Representative Concentration Pathways, used in General Circulation Models (RCP Database). Other sources, however, may use slightly different values; for example, the Stern Review used $C_0 = 280$ ppm (Gohar and Shine, 2007). Using the Equation 2, Gohar and Shine derive the following expression to calculate equivalent carbon dioxide:

$$CO_2eq = C_0 e^{\frac{RF}{\alpha}}$$

(Equation 3)

This equation assumes that a certain value of radiative forcing due to any number of factors is known, and gives the carbon dioxide concentration required to achieve the same amount of radiative forcing.

The IPCC provides in its reports updated values for the radiative forcing caused by relevant gases and aerosols (Stocker, 2013). Using this information, different groups of forcings can be used to calculate equivalent carbon dioxide. For example, the Representative Concentration Pathway datasets provide two types of equivalent carbon dioxide values: one using all anthropogenic forcings, aerosols and tropospheric ozone, and one using only greenhouse gases controlled under the Kyoto Protocol (RCP data source). Generally, however, equivalent carbon dioxide takes into account forcing by greenhouse gases, tropospheric ozone and aerosols and does not represent natural forcings (Hare and Meinshausen, 2006).

e. Climate Sensitivity

Climate sensitivity describes the temperature change resulting from doubling the amount of carbon dioxide in the atmosphere (Stocker, 2013, p. 16). The two most important metrics for describing this value are equilibrium climate sensitivity and transient climate response. Equilibrium climate sensitivity (ECS) is defined as the change in GMST at equilibrium resulting from doubling the amount of carbon dioxide in the atmosphere relative to preindustrial levels. The IPCC puts the equilibrium climate sensitivity as likely to be in the range of 1.5 to 4.5 C (Stocker, 2013, p. 16). The transient climate response (TCR), on the other hand, represents the change in GMST following from carbon dioxide increases of 1% per year until concentrations are double those of preindustrial levels; the TCR is measured using the difference between the beginning of the time period studied and a 20-year period centered on the time in which carbon dioxide is doubled (Stocker, 2013, p. 817). The IPCC report "Climate Change 2013: The Physical Science Basis" provides the climate sensitivity values used in this project. For the models in CMIP5, TCR and ECS are shown to be linearly correlated (Stocker, 2013, p. 817)

The IPCC's climate sensitivity values are derived using the method of Gregory et al., who proposed performing a linear regression of the net downward heat flux over time against GMST (Stocker, 2013, p. 817). This regression is used to obtain the coefficient α, which is then applied to calculate the equilibrium climate sensitivity with the following equation:

$$\Delta T^{eqm} = F_{2x}/\alpha$$

(Equation 4)

Where F_{2x} is the positive downward imposed forcing due to a doubling of the carbon dioxide concentration in the atmosphere. This strategy is corroborated by other methods of obtaining climate sensitivities, and overcomes the challenge of determining the imposed forcing F in order to determine the coefficient α; it also allows for climate sensitivity to be determined from the Atmosphere-Ocean GCMs that are also used to make historical simulations and climate projections (Gregory et al., 2004). In the previous Physical Science Basis report, on the other hand, climate sensitivities had been determined using "slab models," or atmosphere GCMs coupled with mixed-layer oceans; such simulations equilibrate within only 10-20 years, making them cheaper to use (Stocker, 2013, p. 817). In those cases, after being run to equilibrium following instantaneous carbon dioxide concentration doubling, the slab models were used to determine the coefficient α (Gregory et al., 2004).

In their paper, Gregory et al. acknowledge that α may change over time. The authors note that if this change does occur, then in their linear regression of net downward heat flux against GMST, "the variation of the slope provides a means of diagnosing the dependence of the feedbacks on climate state" (Gregory et al., 2004). In investigating these changes, the metric of "effective climate sensitivity," or equilibrium climate sensitivity given that the feedbacks of a given point in time remain constant, is helpful. Some other studies corroborate this suggestion that climate sensitivity can change considerably over long periods of time. Senior and Mitchell, for example, ran a coupled ocean-atmosphere experiment over 800 years with doubled carbon dioxide concentrations; the effective climate sensitivity was found to increase by 40% over the course of the simulation (2000). If true, the time dependence of climate sensitivity poses challenges for simpler models, which will struggle to properly reproduce warming due to carbon dioxide over long periods of time. A later study by Williams et al., however, posited that "much of the apparent variation in effective climate sensitivity identified in previous studies is actually due to the comparatively fast forcing adjustment" (2008).

2. DATA AND METHODS

a. One-temperature Model and Observations

The version of the simple model developed by Hu and Fedorov and used in the first part of this project is represented by the following equation:

$$\frac{dTg}{dt} = -\frac{Tg}{\tau} + a * log(CO_2/CO_{2ref}) + b * T_{NINO} + c * SAOD + d$$

(Equation 5)

where the left-hand side of the equation represents the rate of change of annual mean GMST; if an initial value is known, the equation predicts and reproduces variations in GMST (2017). In the equation, Tg indicates GMST; CO_2 indicates atmospheric carbon dioxide concentrations; the natural logarithm is taken, as this function best represents the way in which CO_2 forces the atmosphere. T_{NINO} corresponds to atmospheric heating anomalies associated with ENSO with the long-term warming trend removed, and $SAOD$ represents shortwave scattering by stratospheric sulfate aerosols resulting from volcanic eruptions. The first term on the left-hand side describes linear damping with an e-folding time scale tau; in the original version of the simple model, tau was set to 2 years. The last part of the equation contains a constant term, which relates to the value of CO_{2ref}, here chosen to be 320 ppm. Doubling the CO_2 term in the equation will result in a temperature change of:

$$\Delta T = a * \tau * ln(2)$$

(Equation 6)

which represents a value for climate sensitivity, or the temperature change resulting from a doubling of carbon dioxide, as described previously.

The Earth observation data used in the historical period are the same as the data used by Hu and Fedorov (2017). GMST was taken from the Goddard Institute for Space Studies Surface Temperature Analysis with 1200 km smoothing (Goddard Institute for Space Studies Surface Temperature Analysis Team, 2017). ENSO variations were taken from the Extended Reconstruction Sea Surface Temperature v4 SST product; the T-NINO was calculated by "averaging SST anomalies within a large equatorial Pacific domain between 160 E - 90 W and 5

S - 5 N. This index incorporates both the Nino 3 and Nino 4 regions" (Hu and Fedorov, 2017). Mauna Loa in situ measurements downloaded from NOAA ESRL provided carbon dioxide data for years 1959-2015 (NOAA-ESRL). For years before that, data was taken from Law Dome DE09 and DE08-2 ice core reconstructions (CDIAC). For volcanic eruption data, a NASA GISS stratospheric aerosol optical depth data set is used (Sato et al., 1993).

The above datasets for atmospheric carbon dioxide concentrations, shortwave scattering by stratospheric sulfate aerosols, and atmospheric heating anomalies associated with ENSO for the 1850-2015 time period were used in the "simple model," Equation (1). The results were compared to the observational GMST data by calculating the Root Mean Square Error (RMSE). The coefficients found by Hu and Fedorov to minimize the RMSE for this time period were a = 1.76 C/yr, b = 0.122 1/yr, c = -1.47 C/yr, and d = 0.0134 C/yr; tau was set to 2 years, and the resulting RMSE was 0.08 C (Hu and Fedorov, 2017). In order to verify these coefficients and to test different values for tau, the damping timescale, we tested out the simple model for a range of values for each coefficient, and determined the combination that created the lowest error. To determine how a specific coefficient affected RMSE, we plotted the range of values of that coefficient against error; the error values used were from the combination of other coefficients that gave the lowest RMSE possible. We performed this test for multiple time periods: 1880-2015, 1880-1950, and 1950-2015, since the data varies in quality between these time periods. To test how pairs of coefficients changed the error, we did the same for two coefficients at a time, creating contour plots with a vertical RMSE axis. These contour plots were used to investigate how different variables constrained each other. We performed this operation for different combinations of a, tau and b.

In order to investigate whether adding the effects of the Atlantic Multidecadal Oscillation to the model would make the results more accurate, a new term was added to the model:

$$\frac{dTg}{dt} = -\frac{Tg}{\tau} + a * log(CO_2/CO_{2ref}) + b * T_{NINO} + c * SAOD + d + e * AMO$$

(Equation 7)

Where AMO represents the AMO index, which is based on the weighted temperature average of the North Atlantic, from 0 to 70 N (NOAA - ESRL). The provided datasets had monthly SST values; we calculated the annual mean to use in the simple model, with each year being

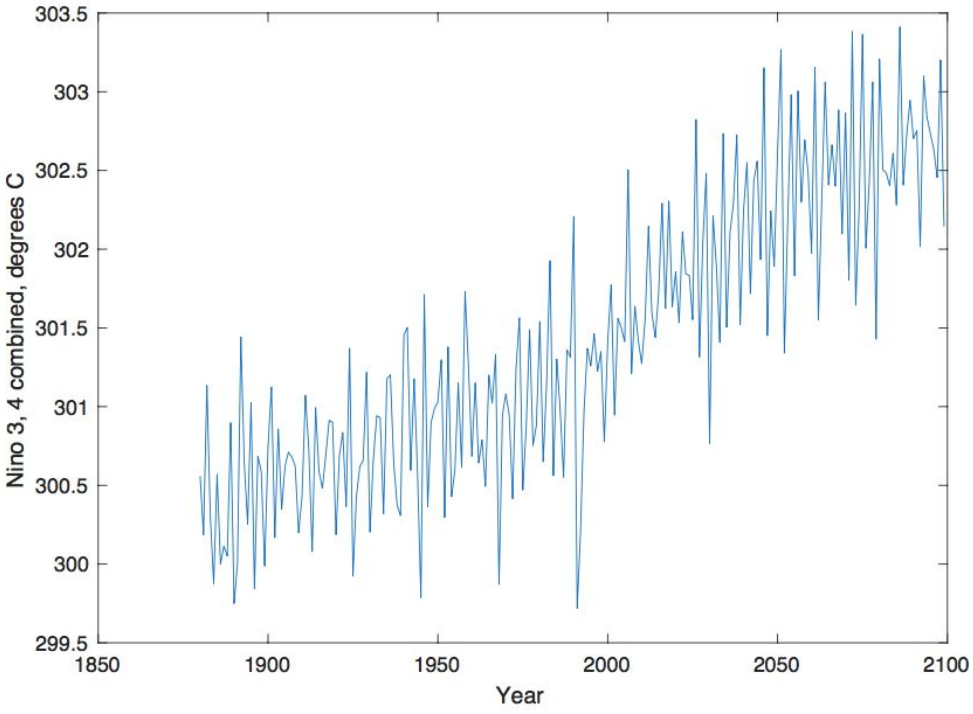
considered to be from July to June (i.e, the year 2015 is from July of 2015 to June of 2016). The optimal value for coefficient e was tested by running the simple model for a range of coefficients; the range was predicted by comparing the error found with a plot of the AMO index. We then calculated the lowest RMSE possible between the simple model and the observations resulting from incorporating the AMO term. We tried this with both smoothed and unsmoothed datasets from NOAA-ESRL.

b. Reproducing and Predicting GCM Data

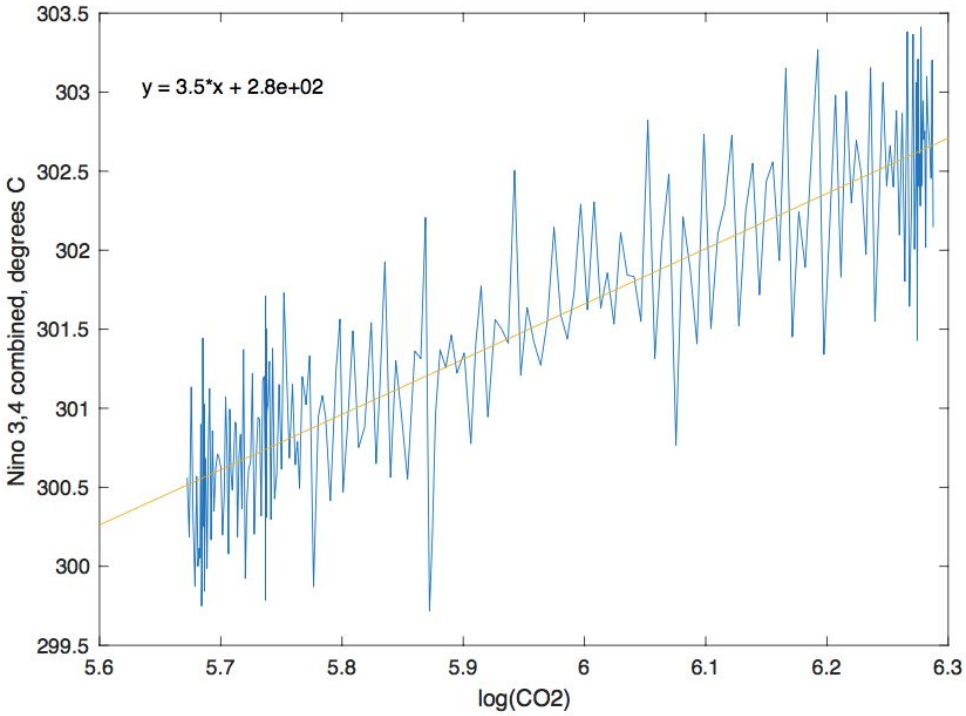
In the next phase of the project, the simple model (Equation 5) was used to reproduce data from General Circulation Models. For the first steps of the this phase, we used monthly GMST data from each model for Representative Concentration Pathway (RCP) 4.5. The data used was downloaded from World Data Center for Climate and spanned the years 1850-2100, combining the results of preindustrial forcings (the same for every RCP) with future forcing, past the year 2005. We also used El Nino SST data from the simulations to provide measurements for ENSO (GCM Data Source). El Nino data was provided for the Nino 3, 4, and combined regions. For each type of data, we calculated annual means, once again with each year being considered to be from July to June. Annual atmospheric carbon dioxide and equivalent carbon dioxide concentrations were downloaded from the RCP Database. Here, equivalent carbon dioxide concentrations were calculated using the method described in the Introduction and taking into account all anthropogenic forcings including greenhouse gases, aerosols and tropospheric ozone (RCP Database). Another equivalent carbon dioxide value is provided incorporating only forcings caused by gases controlled by the Kyoto Protocol; this value was not used. The volcanic aerosol measurements used depended on the particular GCM. The two primary datasets used were those of Sato et al. (1993), the same as previously used for running the simple model to reproduce observations, and Amman et al. (2003). We selected the appropriate dataset for each experiment depending on which was used by the original GCM (Driscoll et al.). While reproducing GCM data, we always used GMST output from RCP 4.5 and run r1i1p1.

For the carbon dioxide term, we used both equivalent and atmospheric carbon dioxide concentrations, in different runs of the model. Initially, Nino 3 data were used for the ENSO

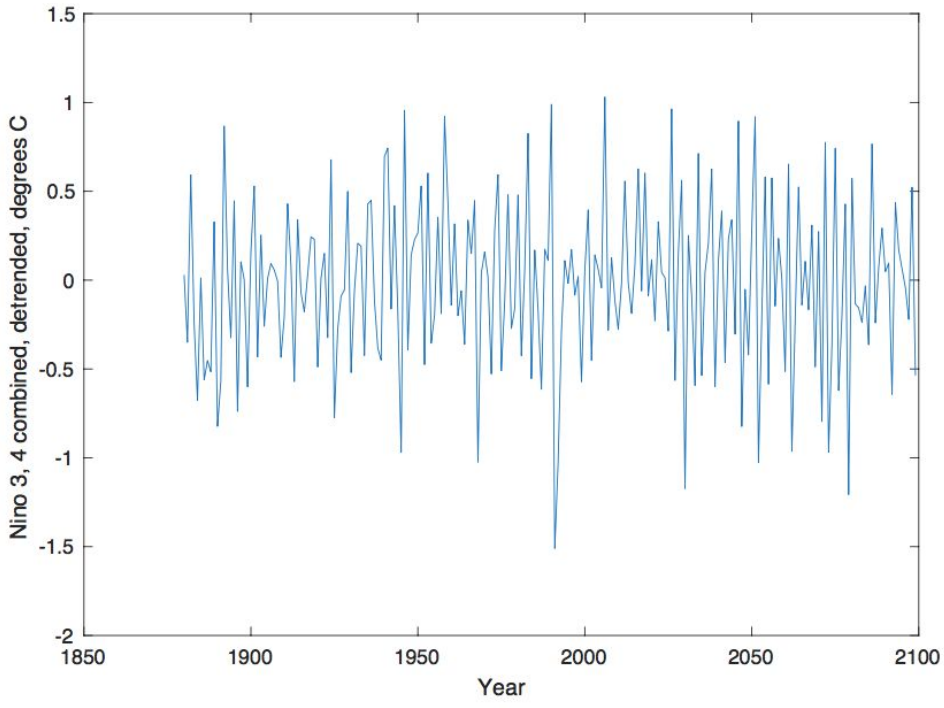
term of Equation 5; this is data for the 5N-5S, 150W-90W region of the Pacific (NCAR - Nino SST Indices). During this phase, the only calculation performed on the ENSO data was to determine annual averages. Later, we switched to using the combined Nino 3-4 data, which incorporates both the Nino 3 region and the Nino 4 region of 5N-5S, 160E-150W, and better encompasses the ENSO signal (NCAR - Nino SST Indices). Similarly, in the earlier phase, the Nino 3-4 data were used to calculate annual averages, and not altered further. During a later part of the project, the warming trend was removed from the Nino 3-4 before being used in the simple model. At first, this operation was performed using the Matlab "detrend" function. Later, we removed the warming trend by performing a linear regression of the ENSO data against the natural logarithm of carbon dioxide, determining the linear function that characterized the connection between the two, and subtracting that function from the ENSO SST data. The resulting detrended data were then used as the ENSO term in Equation 5. When equivalent carbon dioxide data was used for the CO_2 term, equivalent carbon dioxide was also used to remove the warming trend from the ENSO data. (Figure 4 depicts the process of removing the warming trend.)



a)



b)



c)

Figure 4: *An example of calculations performed on El Nino data, for the GISS-E2-H General Circulation Model, RCP 4.5. a) Nino 3, 4 combined annual means plotted over the 1880-2099 time period. b) Linear regression of*

logarithm of atmospheric carbon dioxide against annual means of Nino 3, 4 combined. c) Nino 3, 4 combined annual means with the warming trend removed using linear regression as seen in image b.

For each experiment, we used a method similar to the one used to verify the best coefficients of the simple model to reproduce observations, but with annual GMST from GCM output. We tested a range of values for each coefficient, trying out each combination and determining which producing the lowest RMSE between the output of the simple model and the GCM. To determine the starting ranges for a, b and c, we started with ranges several integers away from the optimal values found for the observations; to determine the starting range of d for each model, we performed a linear regression of the GCM's GMST output over time, and determined the temperature's rate of change. We gradually reduced the range of values being evaluated, and decreased the step of each range, in order to test more precise values for each coefficient, until we tested coefficient values to margins of 0.01. For the majority of the trials, we fixed tau, the damping coefficient, at 2, and tested values for a, b, c and d. We found the coefficients that yielded the lowest RMSE for a variety of scenarios for each model, including different combinations of equivalent carbon dioxide vs. atmospheric carbon dioxide concentrations, untrended ENSO vs. detrended using the Matlab "detrend" function vs. detrended using a linear regression, and Nino 3 vs. combined Nino 3-4. For the later stage of the project, we continued performing trials of both atmospheric and equivalent carbon dioxide for each model, but mostly used Nino 3-4 and removed the warming trend by performing a linear regression against the natural logarithm of carbon dioxide.

In terms of the time period of GCM data used, we performed three types of experiments. For the first, "historical" experiments, we used the historical time period of 1880-2005 to tune the simple model by determining the coefficients that best reproduced the GMST output, using the method outlined above. The starting year varied slightly based on the time period for which volcanic forcing data were available; for the GCMs relying on the Amman et al. data set, information was only available starting in 1890 (2003). Since the Sato et al. dataset indicates large eruptions in the 1880-1890 time period, we started GCM experiments relying on Amman et al. in 1890. The data also ended in 1998; initially we modeled GCMs using Amman et al. ending then, but as no large eruptions occurred in the 1998-2005 time period, we began to run the

simulations through the end of the historical period, without volcanic data for the last few years. For the "full period" experiments, we performed the same operation for the 1880-2099 or 1890-2099 time range. After performing the historical and full period experiments, we calculated the result for climate sensitivity from Equation 6 for each trial, and compared it to known values for Equilibrium and Transient Climate Sensitivity for each GCM from the IPCC (Stocker, 2013, p. 818). We then executed "predictive" experiments: using the coefficients found from the historical period for a specific GCM, we tested how the simple model could reproduce the GMST results past 2005. We used the resulting RMSE to determine the success of using the simple model predictively in that case. For these predictive runs, the coefficients used were those found in the historical experiment with the same inputted data (for example, equivalent CO_2 and Nino 3-4 combined detrended using linear regression).

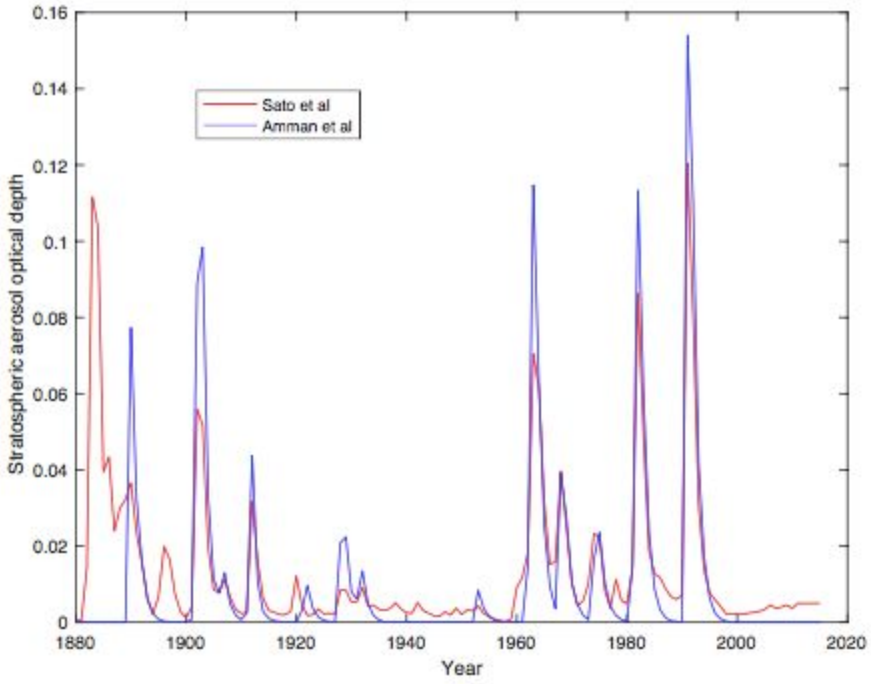


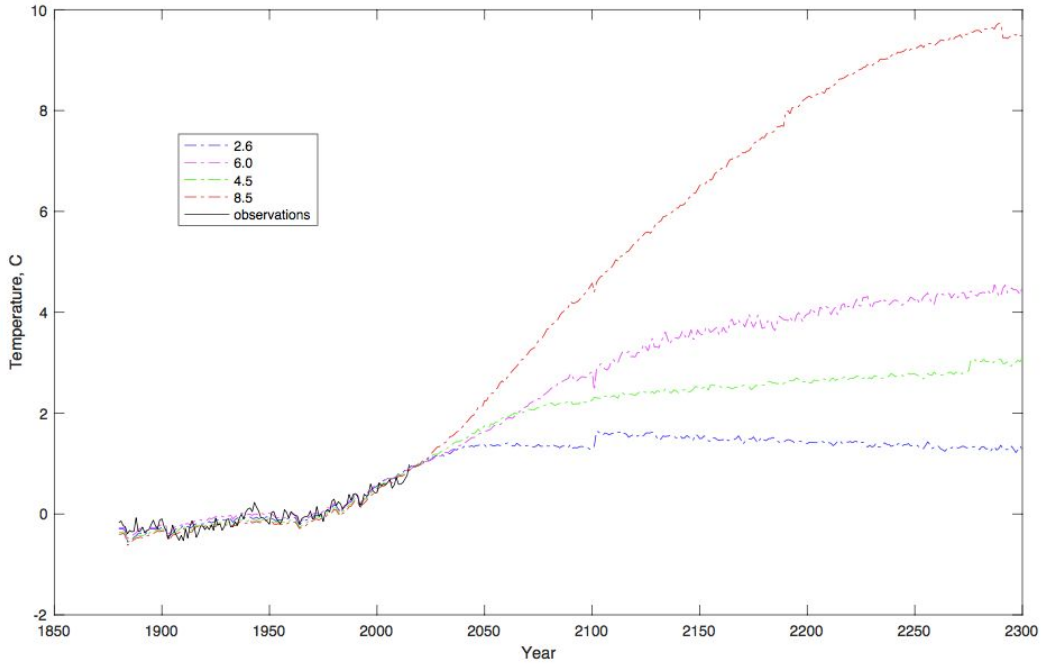
Figure 5: *Stratospheric aerosol optical depth (SAOD) due to volcanic eruptions from two data sets, Sato et al. (1993) and Ammann et al. (2003). Sato et al. covers the 1880-2005 time range, whereas Amman et al. only covers 1890-1998. The Sato et al. dataset demonstrates that SAOD was high during the 1880-1890 time period, indicating that simple model experiments using Amman et al. should start in 1890.*

c. GCM Multimodel Mean

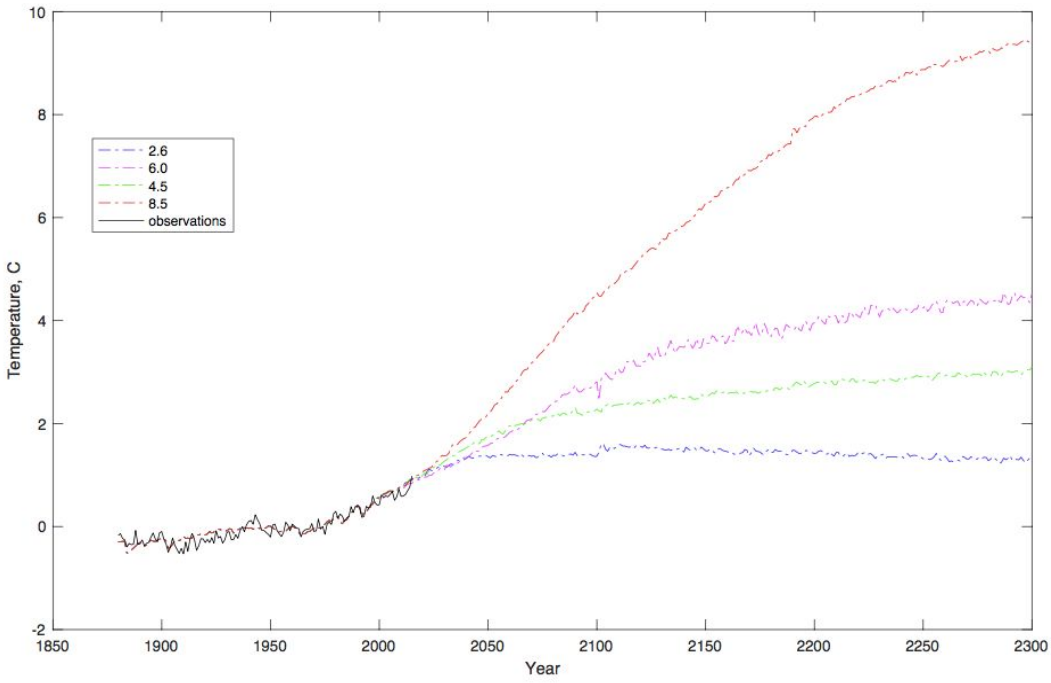
During a later stage of the project, we compared simple model output with the multimodel mean of the GCMs for each Representative Concentration Pathway. In this case, the simple model was forced using only atmospheric or equivalent carbon dioxide; SAOD and ENSO were set to zero. The multimodel means of the GCMs for each RCP was calculated using data from Climate Explorer. We calculated several versions of the multimodel mean for each RCP. The first, Mean A, was calculated using all of the GCMs available for each pathway. For Mean A, the result for each pathway was different in the historical period; this difference should not be present, since the historical forcing for each RCP is the same. This discrepancy was a result of each RCP corresponding to a different set of GCMs. After removing all GCMs from the set of models that did not have a simulation for each of the four RCPs, we calculated the multimodel mean again--Mean B--and the historical period discrepancy was removed. An additional discrepancy, however, was caused by the fact that not all of the simulations extended all the way to the year 2300; many stopped after 2100.

At first, for means A and B, we had calculated the multimodel mean using however many data points existed for each year for that particular RCP, and just used fewer data points after 2100. This method produced discrepancies in the data around the year 2100 for RCPs 2.6 and 6.0. Removing the GCMs that did not extend past 2100 from the multimodel mean--and thus calculating Mean C--eliminated the jump in temperature. Only a small group of GCMs, however, fit this criterion, and only one model, CCSM4, fit both that and the criterion of having simulations for each of the four RCPs. We also separated the GCMs by aerosol interactivity into two groups: either fully interactive with aerosols, or not/semi-interactive, as defined and determined by the IPCC (Stocker, 2013, p. 747). We calculated the multimodel mean, Mean D, of each pathway for both aerosol groups; similar to the case for Mean C, each group did not have enough models for the multimodel mean to be significant. As using only a few GCMs to calculate the multimodel mean renders it less likely to be significantly more accurate than any individual GCM, we ended up using Mean B to compare with the simple model output, while noting the cause of the discrepancies around the year 2100. The multimodel means (all shown in

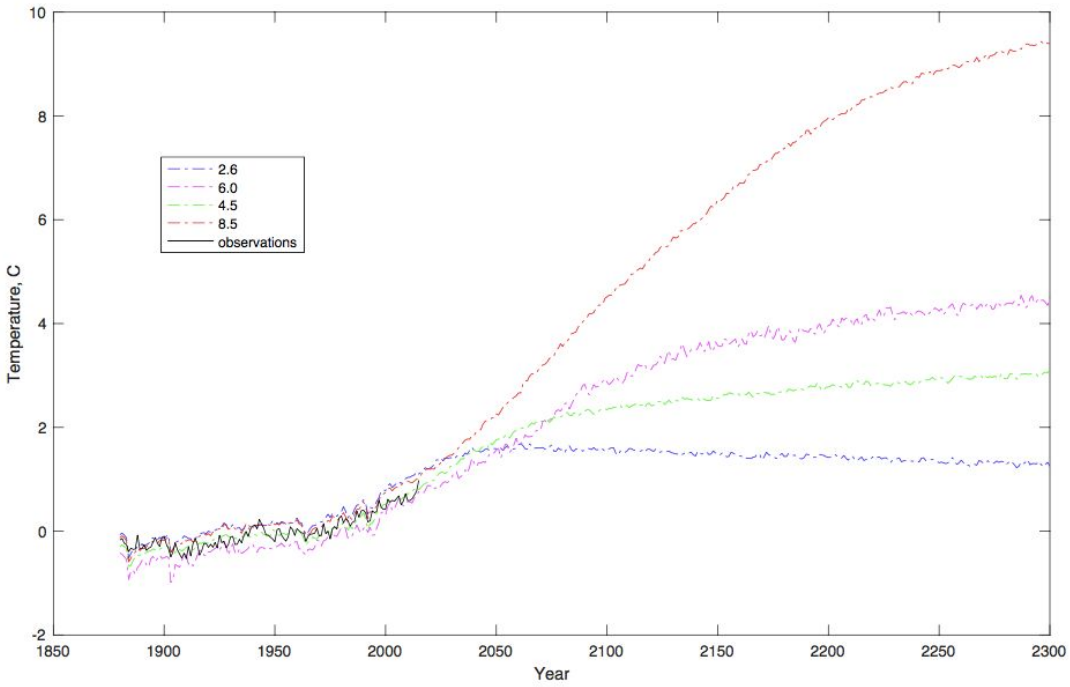
Figure 6) were all adjusted by the same amount until they approximately matched the historical period observations.



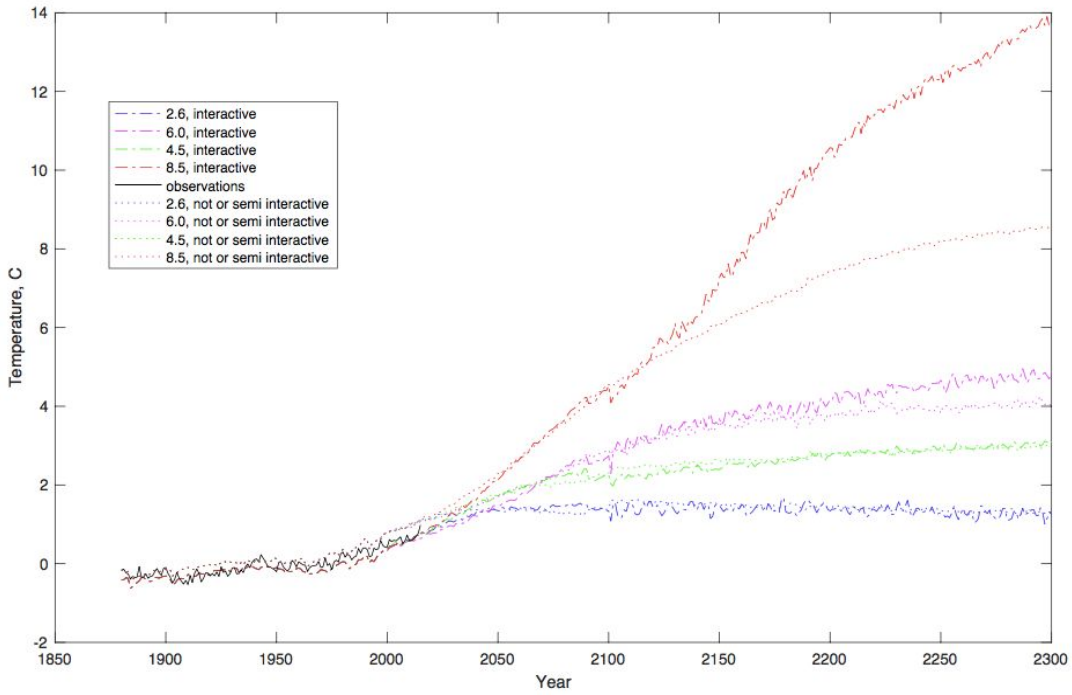
a)



b)



c)



d)

Figure 6*: a) Mean A, as originally calculated for each with all of the GCMs available. b) Mean B: the multimodel mean calculated only using GCMs available for all RCPs. c) Mean C: the multimodel mean calculated using only*

GCMs available for all RCPs that go through 2300. d) Mean D: the multimodel mean, split into two groups for each RCP - interactive with aerosols or not/semi interactive.

d. Two-temperature Model

In the final phase of this project, we replaced Equation 6 with a two-temperature model, in order to capture long-term GMST behavior and reconcile the simple model with the GCM multimodel mean. This new model consists of two equations, operating simultaneously:

$$\frac{dTg}{dt} = -\frac{Tg}{\tau} + a * log(CO_2/CO_{2ref}) + b * T_{NINO} + c * SAOD + d + (T_{deep} - T_g)/\tau_1$$

Equation 8a

$$\frac{dTd}{dt} = (T_g - T_d)/\tau_2$$

Equation 8b

Where T_d describes a new parameter, the temperature of the deeper ocean (~700 m depth), and τ_1 and τ_2 represent new damping timescales. These equations capture heat exchange with this deeper ocean layer, and are similar to the ones used by Held et al. to isolate the fast and slow components of global warming (2009). In their research, Held et al identify two primary components of the "physical climate system's response to changing radiative forcing": a fast component, with an e-folding time scale of less than 5 years, and a slow, "recalcitrant" component, which is "difficult to remove from the system by manipulating radiative forcing."

Similar to previous coefficient adjustments, we adjusted the coefficients and the starting T_d in order to reduce error as much as possible. This time, however, we first attempted to make the simple model GMST output as close as possible to the future multimodel mean predictions, and not the observations or any particular GCM. After tuning the two-temperature model to the multimodel mean and determining that it worked to bring the simple model output closer to the mean, we began to use it to reproduce individual GCM data. Similarly to the experiments with the original version of the model, for each GCM, we reproduced GMST output for the historical and full period, and then used the coefficients found from the historical period to make a prediction into the future. For each of those three experiments, the model was run once with atmospheric carbon dioxide, and once with equivalent. We initially tried to find the best results

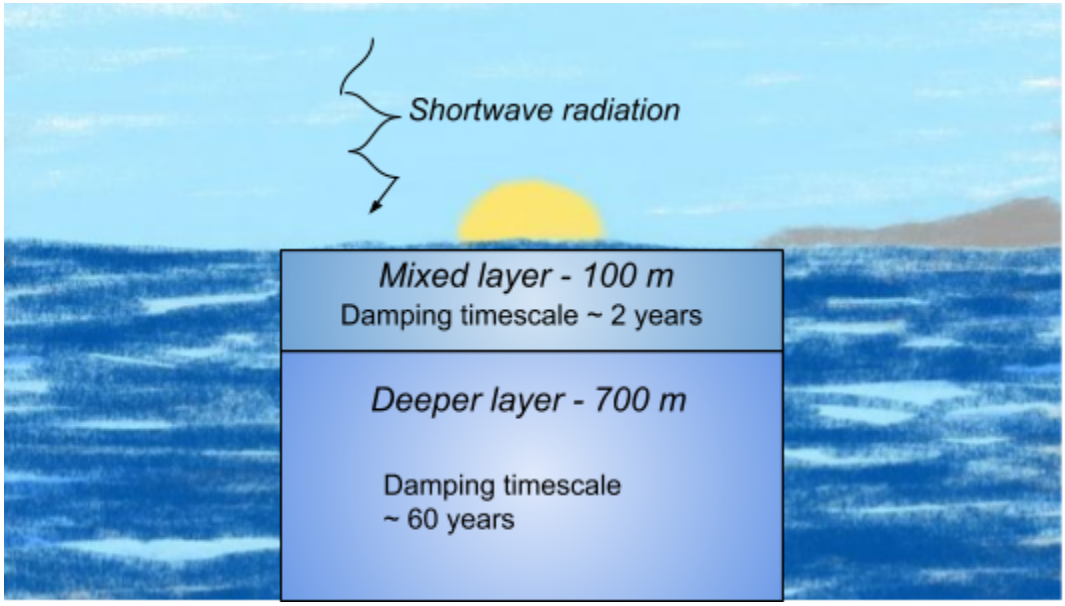


Figure 7: Description of the two-temperature model, which takes into account heat exchange of the atmosphere with two layers of the ocean: the upper, mixed layer, with a depth of 100 m, and a deeper layer, until 700 m.

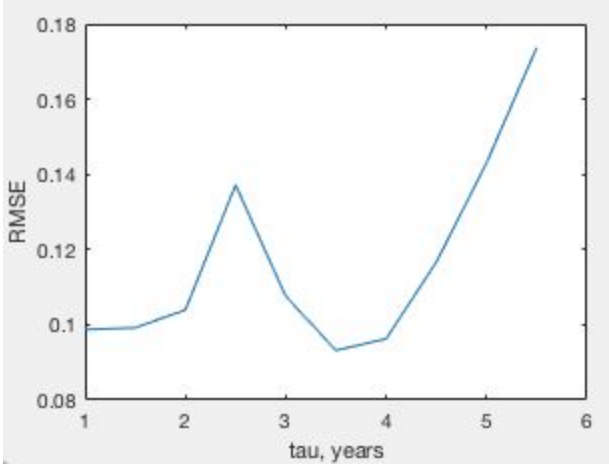
for all coefficients including both the damping coefficients and a, b, c and d, but the results did not converge neatly, and it was very time-intensive to run comprehensive enough experiments to test all the values simultaneously. As a result, we then fixed tau, tau1 and tau2 at the values found to reduce error the most for the observations (similarly to how we fixed tau = 2 years for the GCM experiments with the original simple model).

3. RESULTS

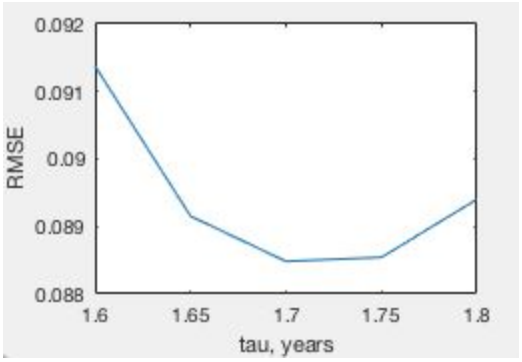
a. Reproducing Observations

In the first stage of the project, testing values of tau yielded two minima of error for the 1950-2015 time period. Decreasing the ranges of values tested and the intervals between those values eventually indicated minima at 1.70 and 3.65 years. For all of these tests, the values for c and d were fixed at those originally found. The coefficients a and b varied with intervals of 1.

a)



b)



c)

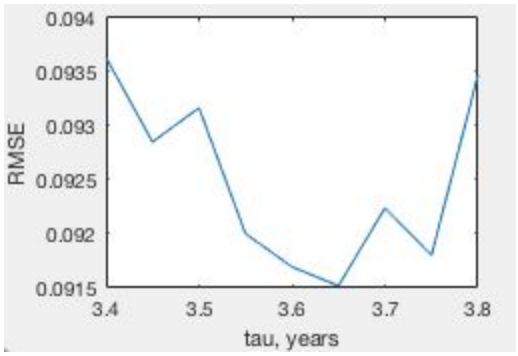
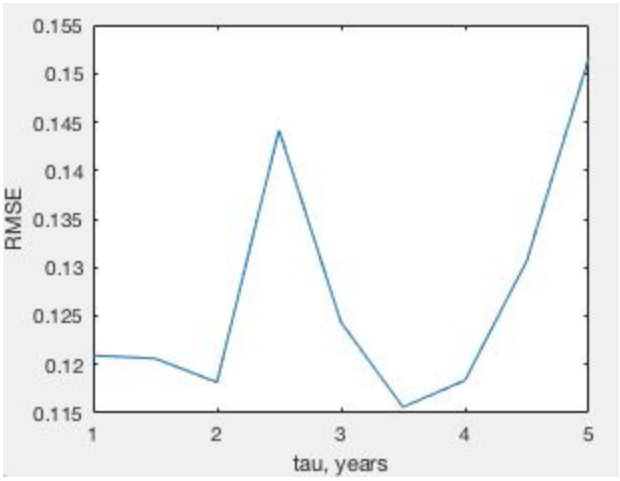
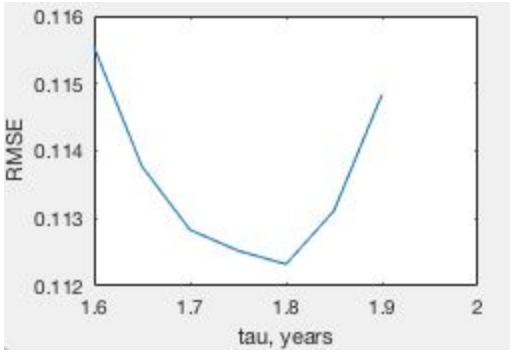


Figure 8: *a) RMSE plotted against tau for the 1950-2015 time period; a and b varied, and c and d were fixed. Two minima were observed for tau. b) After testing smaller intervals of values in the 1.6-1.8 range, a minimum was observed at 1.7. c) After testing smaller intervals of values in the 3.4-3.8 range, a minimum was observed at 3.65.*

Performing a similar operation for the 1880-2015 time period yielded comparable results. Two minima were observed for tau; further testing constrained the minima to 1.80 and 3.60 years.

a)



b) 

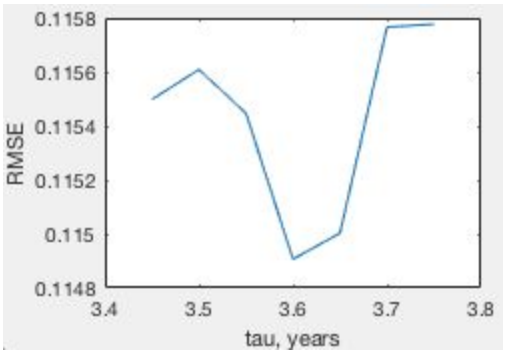
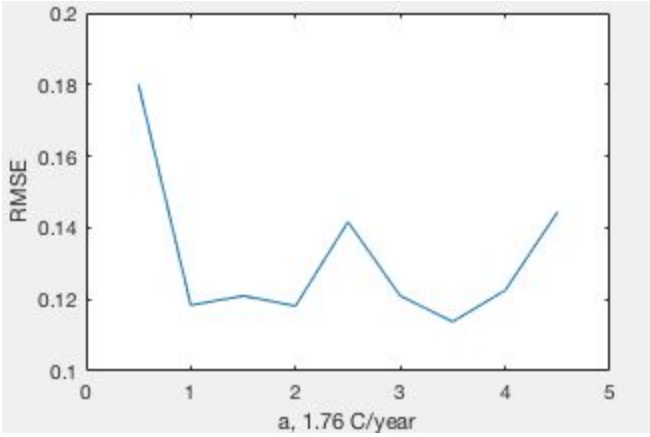
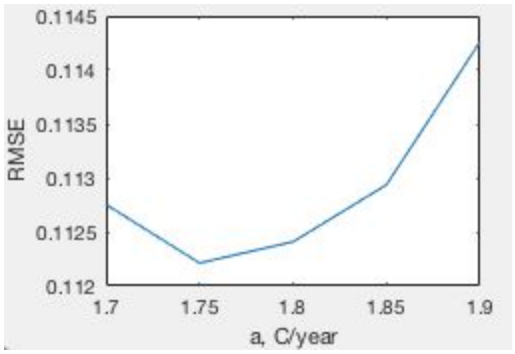
c) 

Figure 9*: a) RMSE plotted against tau for the 1880-2015 time period; a and b varied, and c and d were fixed. Two minima were observed for tau. b) After testing smaller intervals of values in the 1.6-1.9 range, a minimum was observed at 1.80. c) After testing smaller intervals of values in the 3.4-3.8 range, a minimum was observed at 3.60.*

Plotting errors resulting from varying the coefficient a yielded similar results for the 1880-2015 time period. Initially, when a large range was tested, two minima were observed; these minima were constrained to 1.75 and 3.5 C/year following further testing with more accurate intervals. For these tests, tau and b were varied with intervals of 1. The coefficients c and d were fixed.

a) 

b) 

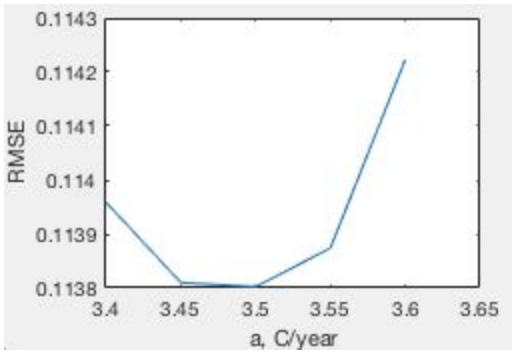
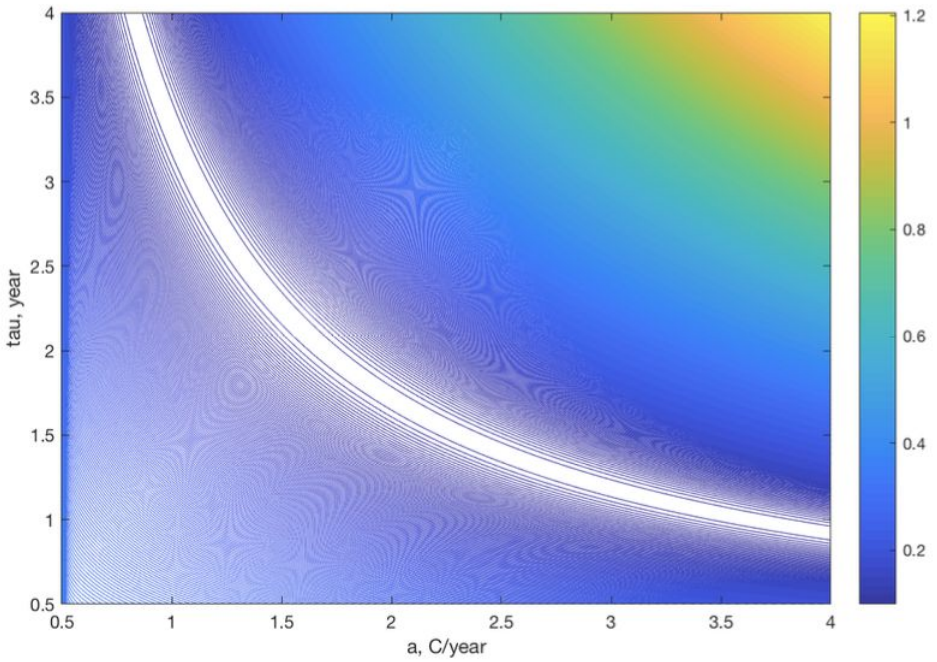
c) 

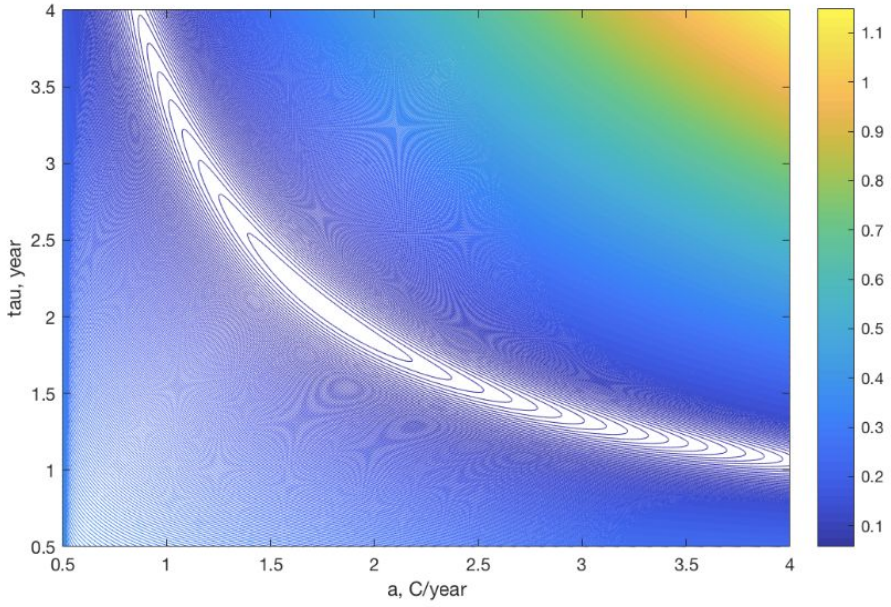
Figure 10*: a) RMSE plotted against a for the 1880-2015 time period; tau and b varied, and c and d were fixed. Two minima were observed for a. b) After testing smaller intervals of values in the 1.7-1.9 range, a minimum was observed at 1.75. c) After testing smaller intervals of values in the 3.4-3.65 range, a minimum was observed at 3.50.*

Testing smaller intervals, and using contour plots to display the results of testing two coefficients simultaneously, clarified that the values of the other coefficients affect the results for constraining a and tau. When tau and a were tested with all other coefficients equal to zero, the results formed a hyperbolic line, depicting a range of values for which error was lowest. When the values for b were not zero, however, a and tau were constrained. (These results are shown in Figure 11). Adding in non-zero c further constrained the values. The resulting range for which the error was lowest centered on approximately tau = 2.05 years and a = 1.75 C/year -- confirming the values initially found by Hu and Fedorov (2017). This experiment demonstrated that including carbon dioxide forcing can constrain tau*a; other forcings, however, must be included to further constrain tau and a. We also showed that there are not actually multiple values for tau that would significantly reduce error.

Contour plots of the other combinations of coefficients largely showed the expected values for minimizing error. The lowest error for b was found at approximately 0.1 1/year, and for a at approximately 1.7 C/year. (These results are shown in Figure 12.)

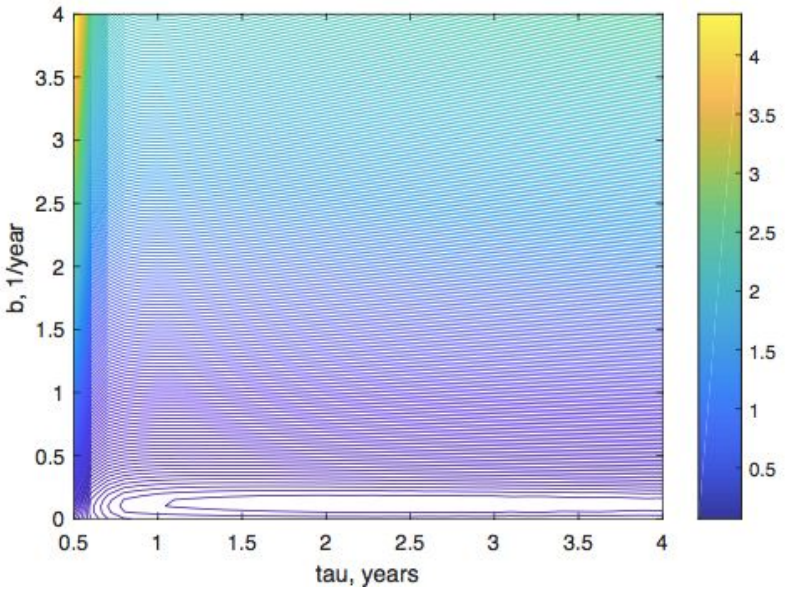


a)

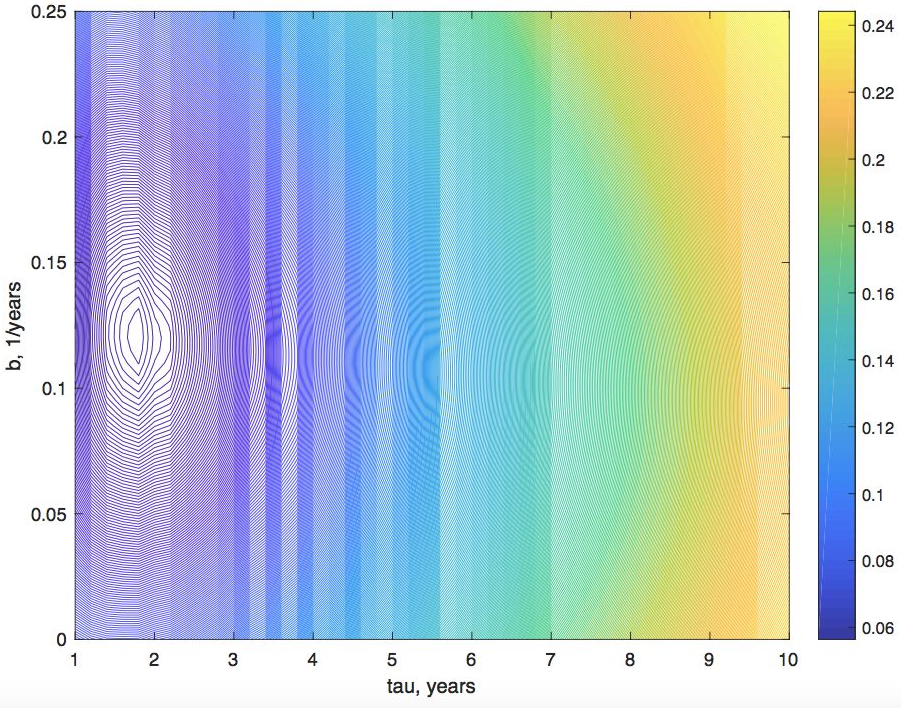


b)

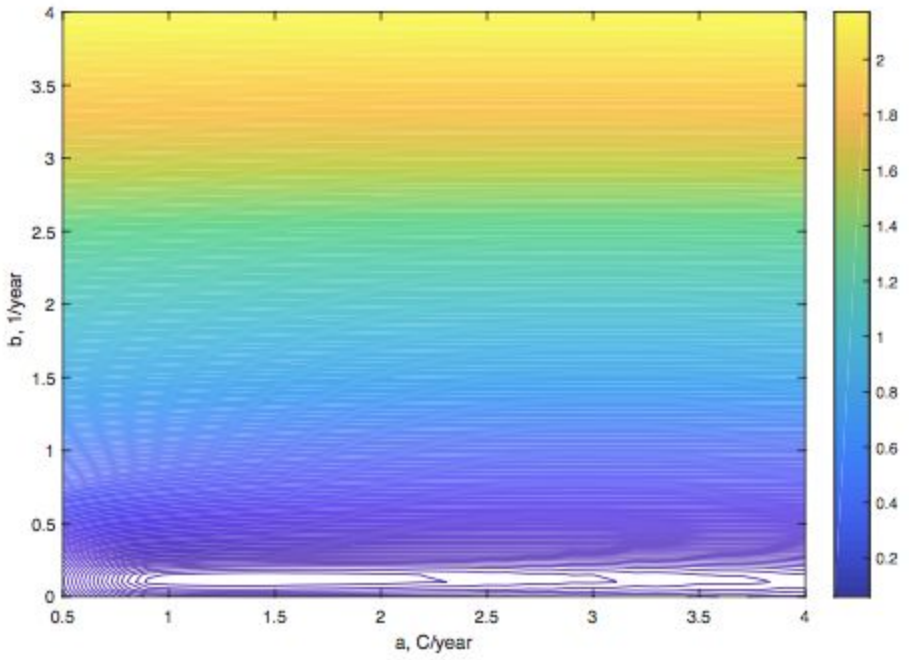
Figure 11*: a) Contour plot showing the RMSE (vertical axis) resulting from a range of values tested for a and tau, with c, b and d = 0. A range of optimal values are shown. b) Plot resulting from a range of values tested for a and tau when b is varied, and c is fixed at -1.5 C/year. As a result, the optimal tau and a are constrained.*


b, 1/year
tau, years
0
0.5
1
1.5
2
2.5
3
3.5
4

a)


b, 1/years
tau, years
0
0.05
0.1
0.15
0.2
0.25
1
2
3
4
5
6
7
8
9
10
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
0.22
0.24

b)



c)

Figure 12: *Examples of contour plots for other combinations of coefficients. a) Contour plot showing the RMSE (vertical axis) resulting from a range of values tested for b and tau. Here, a varies, and c = 0. b) Plot resulting from testing a and b, with c fixed at -1.5 C/year. c) Plot demonstrating error resulting from different values for a and b, with c = 0. Tau varies.*

Incorporating a term representing the AMO index reduced error slightly. Using the unsmoothed AMO dataset reduced RMSE by 0.006 C, with e = 0.14 1/year; using the smoothed dataset reduced RMSE by 0.008, with e = 0.076 1/year (see Equation 7). While this was a slight improvement in the simple model's accuracy, the difference was very small, especially compared to the effect the other three factors have on the error. (Removing ENSO from the equation, for example, increases error by 0.029 C.) As a result, we decided not to include AMO in future experiments.

b. Reproducing and Predicting GCM Data with the One-temperature Model

For the next stage of the project--reproducing GCM output--my first results involved using different types of ENSO data in the simple model. For a group of seven GCMs, comparing Nino 3 and combined Nino 3, 4 data indicated that the latter yields lower error for the simple model.

Here, we did not detrend the ENSO data in any way, and used atmospheric carbon dioxide concentrations; the simple model was used to reproduce historical period data. For all but one GCM, Nino 3, 4 combined yielded lower error. The average RMSE for Nino 3 was 0.089; the average for Nino 3, 4 was 0.081. The difference between the errors in each case also varied considerably among individual models. Because of these results, for later stages of the project, we chose to use combined Nino 3, 4 data only. (These results are shown in Figure 13.)

Results on detrending ENSO data were less definitive. For a group of six GCMs, comparing different ways of detrending the data--no detrending, detrending using the Matlab function, and detrending by removing the signal of $\log(CO_2)$--indicated that no single method was best in all or almost all cases. (In this group, we also reproduced historical period data. Figure 14 depicts the results.) We did observe, however, that removing the signal of carbon dioxide was especially helpful in reproducing full period data. This makes sense, given that the warming signal due to carbon dioxide in the ENSO data would have a larger effect beyond the historical period. As the larger goal of this project is to use the simple model predictively, we decided to remove the $\log(CO_2)$ signal from the ENSO data going forward.

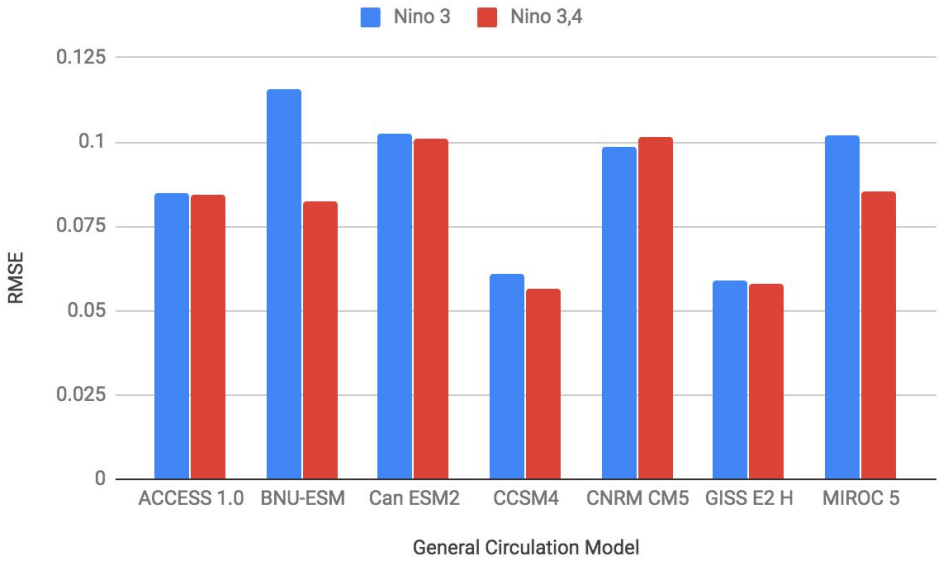


Figure 13: *Experiments in reproducing historical period GMST data for a group of General Circulation Models. For each GCM, one experiment was performed using Nino 3, and one using Nino 3, 4, which yielded the lowest RMSE.*

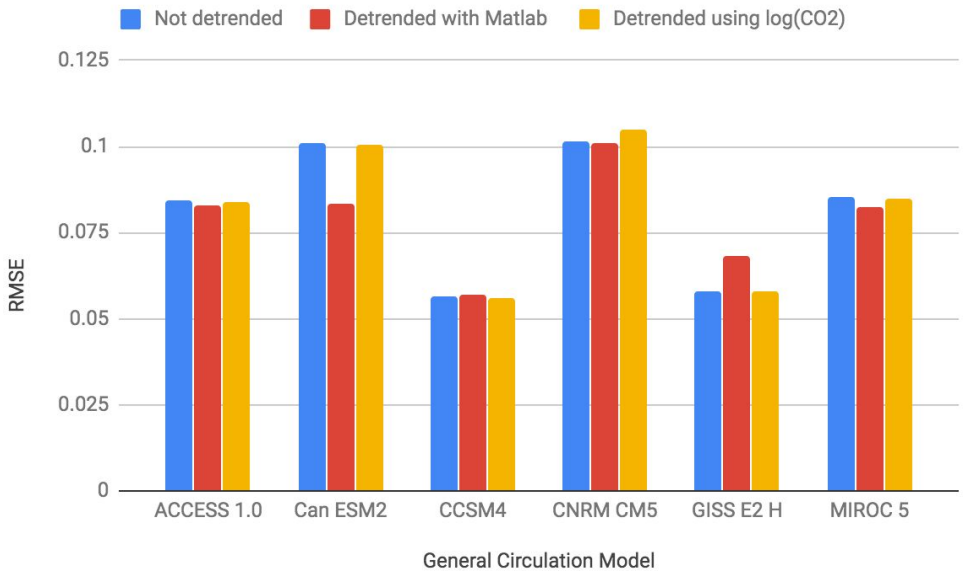


***Figure 14**: As in Figure 12, but for each GCM, one experiment was performed using original ENSO data, one using the Matlab detrend function, and one using log(CO_2).*

Trying out different types of carbon dioxide data initially did not yield clear results. For a group of seven GCMs, we reproduced historical period data with both atmospheric and equivalent carbon dioxide concentrations. For four GCMs, equivalent carbon dioxide reduced error more; for three GCMs, atmospheric carbon dioxide concentrations reduced error more, or did not change the results. When we performed this experiment for the same group of GCMs with the full period GMST data, the results were similar; equivalent carbon dioxide reduced error for the same four out of seven GCMs. Using the simple model predictively yielded slightly better results for equivalent carbon dioxide--five out of seven models had improved results, as opposed to two for which atmospheric concentrations reduced error the most. (For these and all experiments going forward, we used Nino 3, 4 combined data and detrended the data using the log(CO_2) signal.)

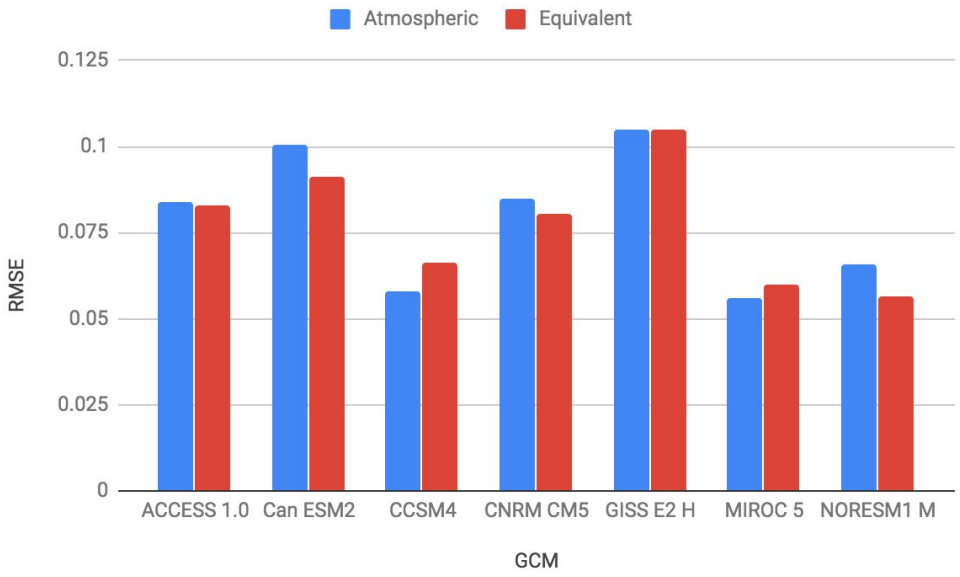


***Figure 15**: Experiments in reproducing historical period GMST data for a group of General Circulation Models. For each GCM, one experiment was performed using atmospheric carbon dioxide data, and one using equivalent carbon dioxide data. RCP 4.5 data was used.*

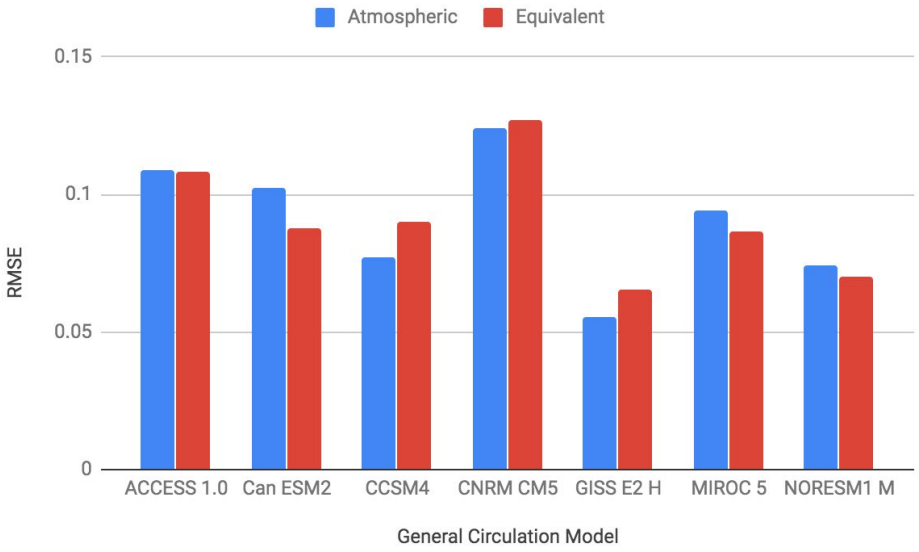


***Figure 16:** As in Figure 14, but for full period GMST data.*

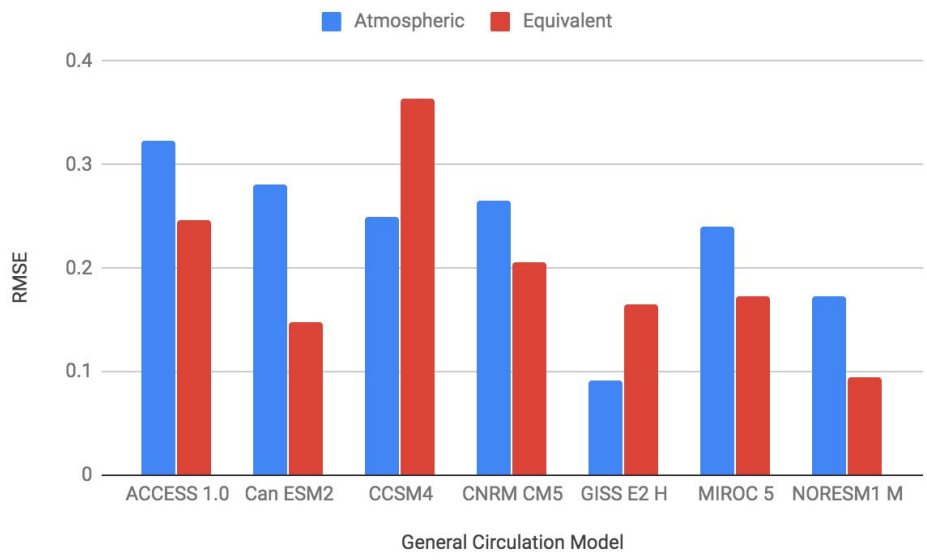


***Figure 17**: As in Figure 14, but for predictive GCM experiments: the model is tuned using historical period GMST, and then used to predict future period data.*

Noting the aerosol interactivity of each GCM explained why some models improved when equivalent carbon dioxide was used, and some did not. For those fully interactive with aerosols, as indicated in "Physical Science Basis 2013," atmospheric concentrations led to higher errors. For those that were either not interactive or semi interactive, atmospheric concentrations used in the simple model yielded lower errors. This was true in almost all cases for both historical period and full period runs, as well as when the model was used predictively. The only exception was CNRM CM5, which is not interactive with aerosols, when the simple model was used predictively; in this case, equivalent carbon dioxide reduced error the most. These results are summarized in Table 1.

GCM	Aerosol interactivity	RMSE (equivalent) - RMSE (atmospheric)		
		Historical period	Full Period	Predictive
ACCESS 1.0	Full	-0.0014	-0.0007	-0.0766
Can ESM2	Full	-0.0095	-0.0143	-0.133
CCSM4	Semi	0.0004	0.0126	0.1138
CNRM CM5	No	0.0002	0.0027	-0.0609
GISS E2 H	Semi	0.0086	0.0101	0.0741
MIROC 5	Full	-0.0048	-0.0078	-0.0661
NORESM1 M	Full	-0.0063	-0.0044	-0.0783

Table 1: *Comparing errors for equivalent vs. atmospheric carbon dioxide in historical period, full period and predictive runs of the simple model. Negative results in the error column indicate that for that GCM, equivalent carbon dioxide yielded better results for the simple model; this correlates to full aerosol interactivity in almost all cases.*

Overall, the simple model reproduced GCM historical period GMST well. For many GCMs, the model's GMST output is nearly indistinguishable from the simple model output, especially for the full period and historical period runs. **These findings present one of the central results of this project.** Table 2 summarizes these results for each model. Figure 18 depicts the visual results for two GCMs.

Calculating a*tau*log(2) for each experiment did not provide results that pointed definitively to the value being equal to either equilibrium or transient climate sensitivity. For historical period runs, a*tau*log(2) was closer to transient sensitivity in 8/14 cases; for the full period runs, the value was closer to transient sensitivity in 4/14 cases. The average difference between a*tau*log(2) and transient sensitivity for the historical period was 0.77 C; for equilibrium sensitivity it was 0.88 C. For full period runs, the average difference between a*tau*log(2) and transient sensitivity was 0.84 C; for equilibrium sensitivity it was 0.54 C. For certain experiments, the difference was zero; for others, it was as high as 2.09. In general, that the longer time period would provide more accurate results for equilibrium sensitivity makes sense. The lack of definitive results for a*tau*log(2), however, is puzzling. For the historical period, analyzing the difference between atmospheric and equivalent carbon dioxide experiments

did not provide clear results, either. The type of carbon dioxide that provided the lowest model error for reproducing GCM data did not necessarily provide the closest match of a*tau*log(2) to either transient or equilibrium sensitivity; the same goes for the full period experiments. For the longer period, however, there was an interesting pattern: using atmospheric carbon dioxide yielded lower error for the equilibrium sensitivity for all GCMs but one, and using equivalent carbon dioxide yielded lower error for the transient sensitivity in all cases. We were not sure what might explain this, and noted that because the other climate sensitivity results were not clear, this result may also be somewhat random.

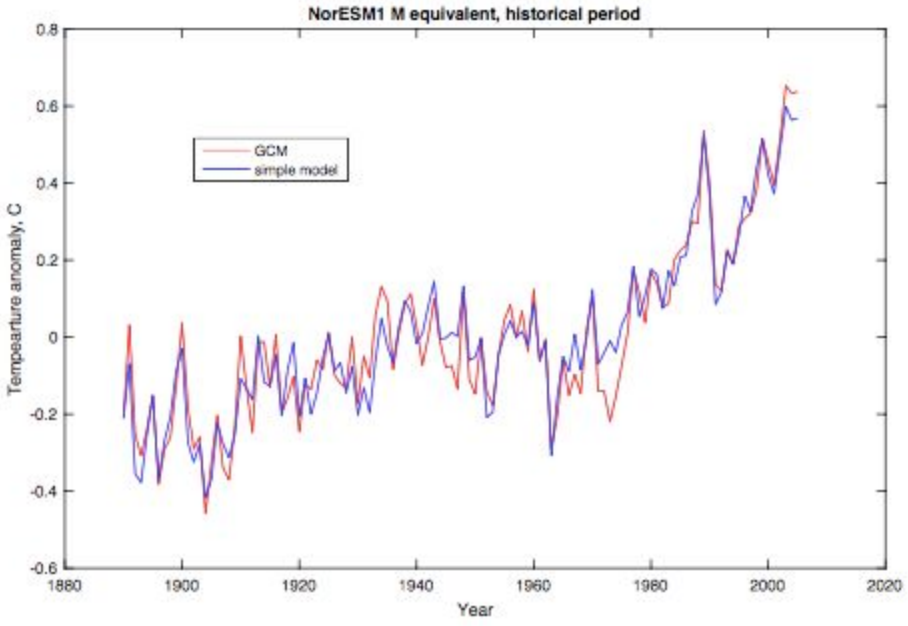
		RMSE of best experiment	
GCM	**CO_2 used**	**Historical period**	**Full Period**
ACCESS 1.0	Equivalent	0.0827	0.108
Can ESM2	Equivalent	0.0912	0.0879
CCSM4	Atmospheric	0.0559	0.0772
CNRM CM5	Atmospheric	0.1048	0.1244
GISS E2 H	Atmospheric	0.0578	0.0554
MIROC 5	Equivalent	0.0802	0.0865
NORESM1 M	Equivalent	0.0597	0.0699

Table 2: *Summary of results for reproducing both historical and full period GMST data. For each GCM, the best result found is shown, and the CO_2 data used in that run is indicated. For all experiments in this table, combined Nino 3, 4 data was used, and ENSO was detrended by removing the signal of log (CO_2).*

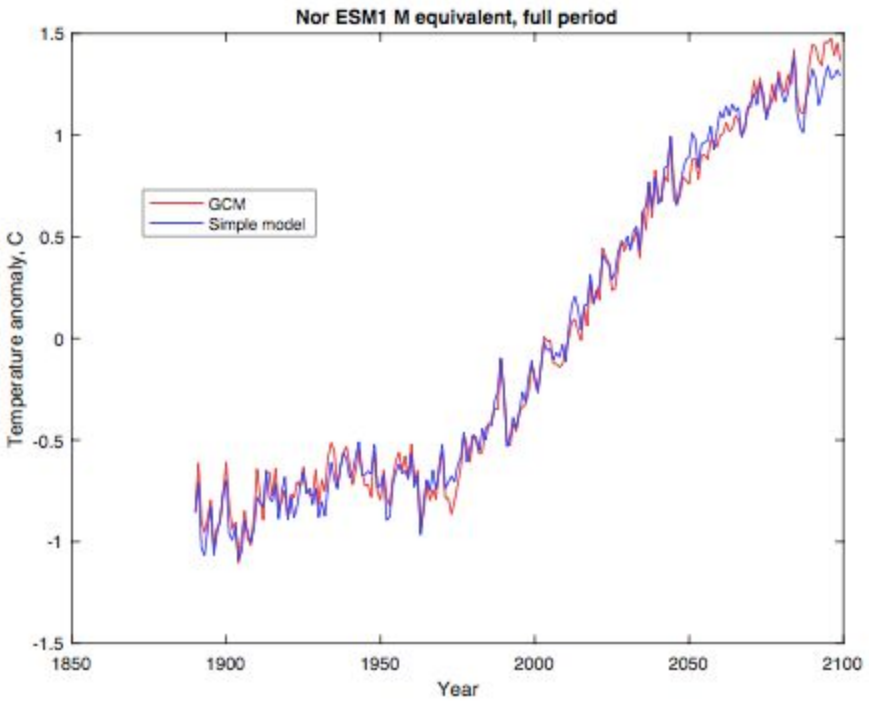
While tuning the model using the full period yielded good results, the error for using the model predictively (using coefficients from the historical period) was considerably higher; in most cases, the simple model and the GCM data diverge in the second half of the 21st century. More specifically, as time goes on, GCM temperature rises at a rate that the simple model does not reproduce. (GISS E2 H was a notable exception.) Table 3 describes the results for these predictive experiments. Figure 18 provides a visual example of the historical period, full period and predictive results for two GCMs, Nor ESM1 M and CNRM CM5.

GCM	CO_2 used	RMSE
ACCESS 1.0	Equivalent	0.2459
Can ESM2	Equivalent	0.1486
CCSM4	Atmospheric	0.2503
CNRM CM5	Equivalent	0.205
GISS E2 H	Atmospheric	0.0914
MIROC 5	Equivalent	0.1734
NORESM1 M	Equivalent	0.0943

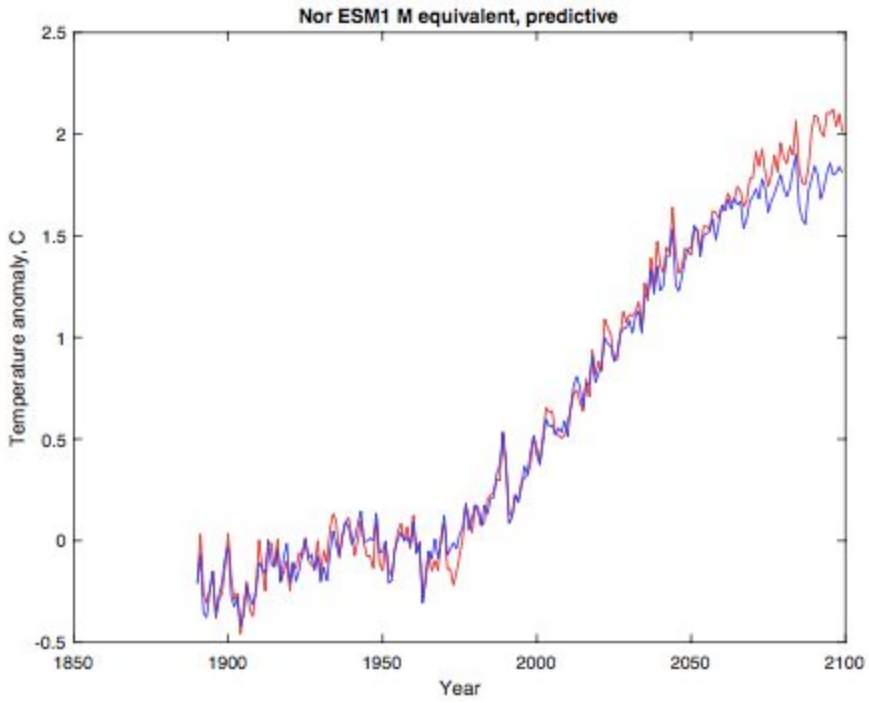
Table 3: *As in Table 2, but for predictive experiments: the simple model was tuned using historical period GCM data, and the resulting coefficients were used to predict data for the future period.*



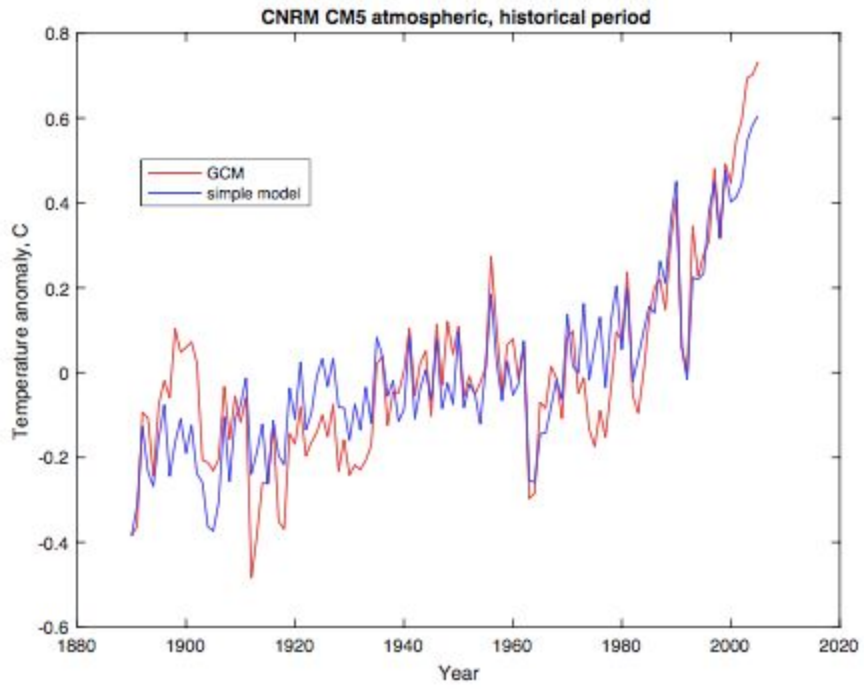
a)


Nor ESM1 M equivalent, full period
GCM
Simple model
Temperature anomaly, C
Year
1.5
1
0.5
0
-0.5
-1
-1.5
1850
1900
1950
2000
2050
2100

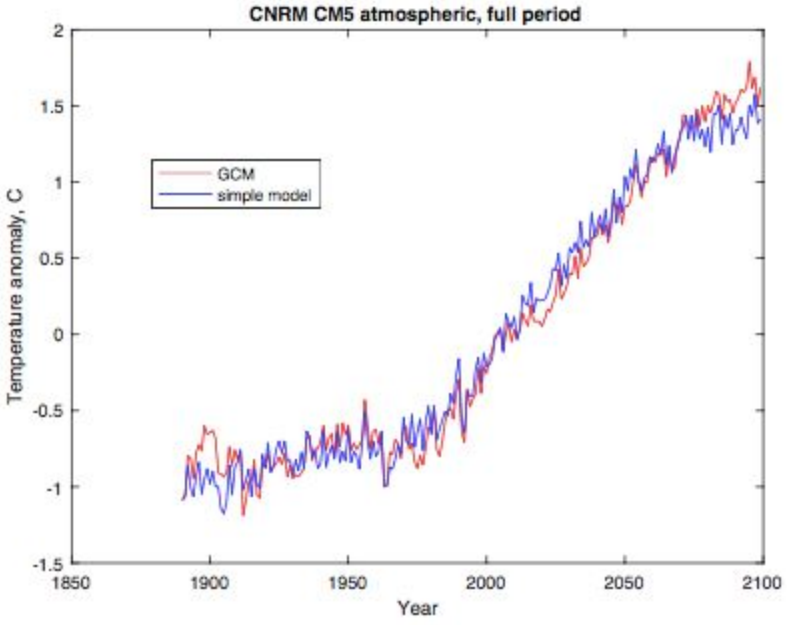
b)


Nor ESM1 M equivalent, predictive
Temperature anomaly, C
Year
2.5
2
1.5
1
0.5
0
-0.5
1850
1900
1950
2000
2050
2100

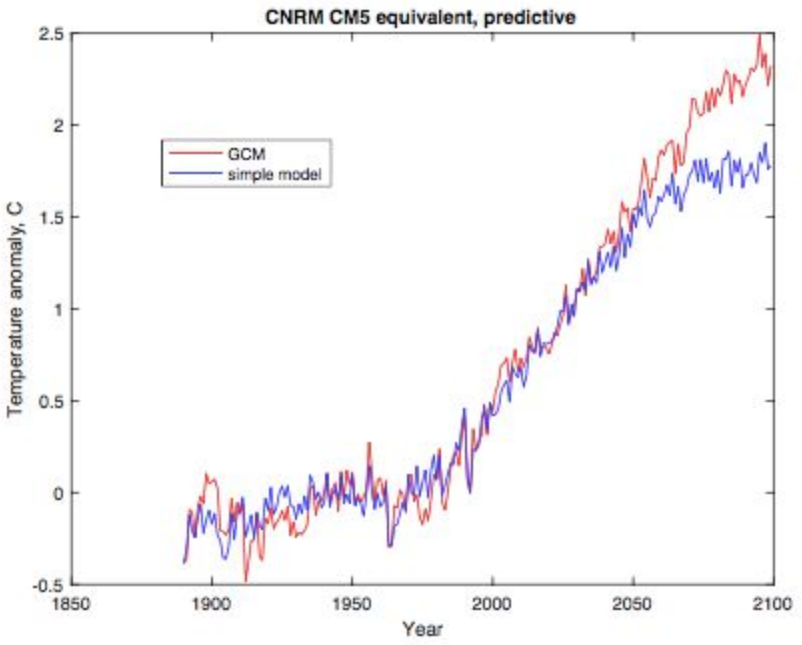
c)


CNRM CM5 atmospheric, historical period
GCM
simple model
Temperature anomaly, C
Year
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1880
1900
1920
1940
1960
1980
2000
2020

d)


CNRM CM5 atmospheric, full period
GCM
simple model
Temperature anomaly, C
Year
2
1.5
1
0.5
0
-0.5
-1
-1.5
1850
1900
1950
2000
2050
2100

e)



f)

Figure 18*: GCM data and simple model for six experiments based on the NorESM1 M GCM and CNRM CM5 models. For all three NORESM1 M experiments, equivalent carbon dioxide is used, which provided the best results. For the historical and full period CNRM CM5 experiments shown, atmospheric carbon dioxide was used. For the predictive CNRM CM5 experiment, equivalent carbon dioxide was used. a) The simple model tuned using the historical period for NORESM1 M. b) The simple model tuned using the entire period, 1890-2099, for NORESM1 M. b) The simple model tuned using the historical period, and then used to predict temperature anomalies through 2099 for NORESM1 M. The simple model and the GCM data diverge during the later part of the 21st century. d) The simple model tuned using the historical period for CNRM CM5. e) The simple model tuned using the entire period, 1890-2099, for CNRM CM5. f) The simple model tuned using the historical period, and then used to predict temperature anomalies through 2099 for CNRM CM5. As for NORESM1 M and most other GCMs used, the simple model and the GCM data diverge during the later part of the 21st century.*

c. Comparing with the Multimodel Mean

The predictive results detailed previously motivated us to compare the simple model with the multimodel mean. This comparison yielded similar results to the predictive experiments. Using the coefficients found to best reproduce the observations in simple model (without ENSO or volcanic eruptions) showed that the multimodel mean's temperatures rose above the simple model, especially after 2100.

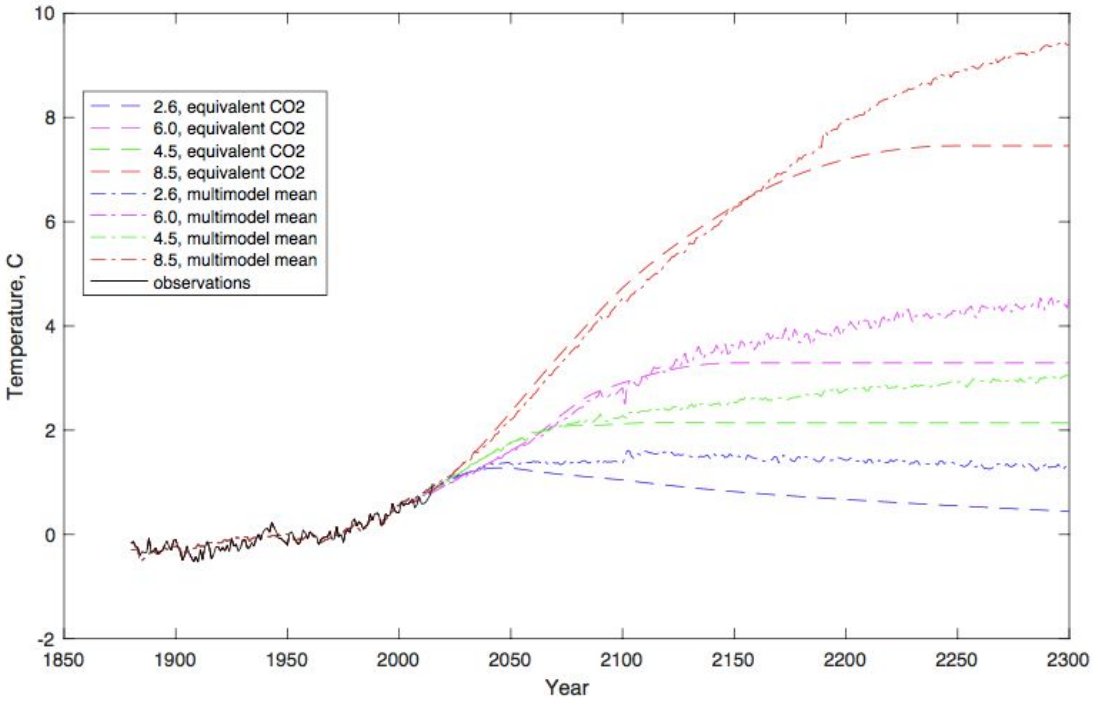


Figure 19*: Multimodel means of GCMs for each Representative Concentration Pathway plotted against temperatures produced by the original version of the simple model. Here, Mean B is used (dash-dot lines; see Methods section). For the simple model (dashed lines), the ENSO and volcanic terms are ignored, and equivalent carbon dioxide concentrations are used. The coefficients found best to best reproduce the historical period observations are applied here: tau = 2 years, a = 1.76 C/year, b = 0.122 1/year, c = -1.47 C/year, d = 0.0134 C/year. Here, the multimodel means show higher temperatures than the simple model, starting in the later part of the 21st century.*

These results indicated that the simple model was not capturing something critical to long-term GMST patterns. We hypothesized that we needed to add something to the model so that it could take into account the climate system's inertia and, more specifically, the heat absorbed and released by the deeper ocean. This led us to Equations 8a and 8b. Using this new version of the simple model with the coefficients tau = 4 years, tau1 = 4 years, tau2 = 56 years, a = 1.21 C/year and d = 0.03 C/year, the simple model with equivalent carbon dioxide closely resembled the multimodel mean. This is shown in Figure 20. (The ENSO and SAOD terms were not used.) Adjusting a to 1.36 C/year made the simple model with atmospheric carbon dioxide more closely resemble the multimodel mean. In both cases, the simple model least resembled the multimodel mean for RCP 2.6; the difference may be exacerbated by the jump observed at the

year 2100, due to the fact that not all models extend past that year. The jump is eliminated in Mean C, which was unfortunately less useful for comparison, for reasons outlined in the Methods section. **Figure 20 represents another one of the central findings of this study.**

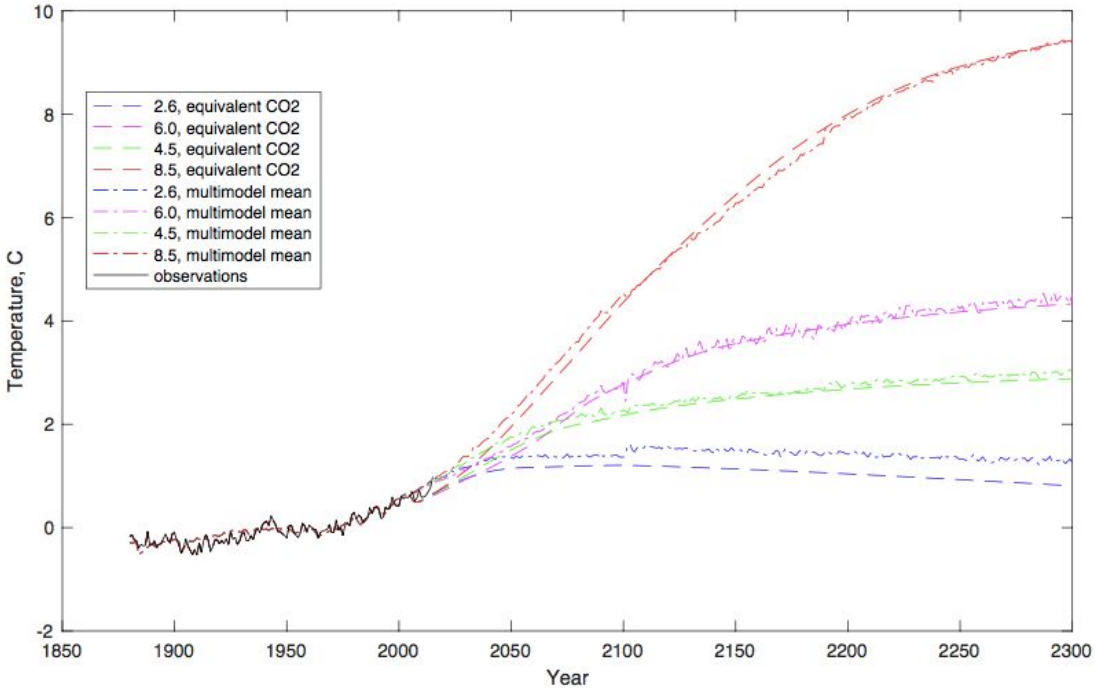


Figure 20: *As in Figure 18, but with the two-temperature version of the simple model. The following coefficients were used: tau = 4 years, tau1 = 4 years, tau2 = 56 years, a = 1.21 C/year, d = 0.03 C/year. This figure represents one of the central findings of this study: that the two-box model, taking into account the basic physics of global warming, can closely reproduce GMST projections as represented by the GCM multimodel mean.*

For the original version of the simple model, a particular GCM's aerosol interactivity correlated to the type of carbon dioxide data that minimized model error. As a result, we hypothesized that comparing the results of this new version of the simple model to Multimodel Mean D, for which the GCMs were separated by aerosol interactivity, might be useful. The results, however (shown in Figure 21) were inconclusive. When the coefficients that worked to match Mean B were used in the simple model, the mean for the GCMs interactive with aerosols did not neatly match the simple model with equivalent carbon dioxide. This might be due to the nature of Mean D; as outlined in the Methods section, there were not enough GCMs in each group to generate a meaningful multimodel mean.

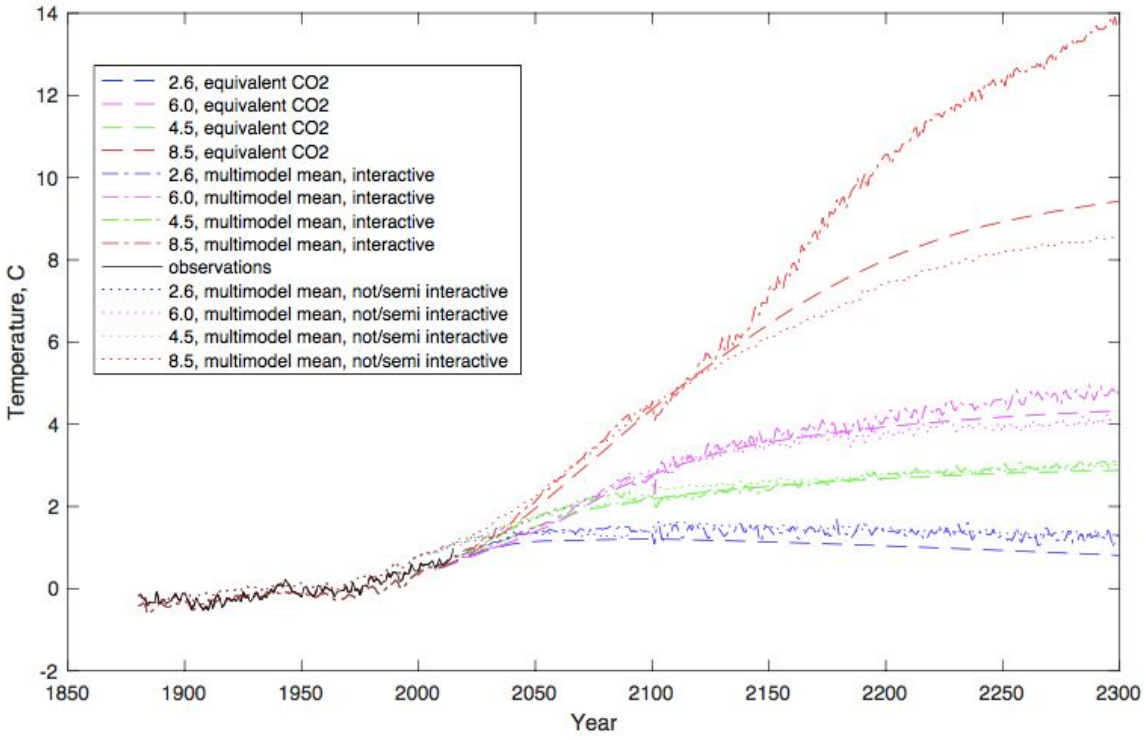


Figure 21*: As in Figure 20, but with Multimodel Mean D instead of Mean B.*

d. Reproducing and Predicting GCM Data with the Two-temperature Model

Applying the new version of the simple model to individual GCMs yielded mixed results. For the historical period experiments, using the two-temperature model increased model error in 9/14 cases. This was expected, as the model was already functioning well without taking into account deep ocean temperatures, and deep ocean temperatures should have a small effect during this period. This matches the presumption of Held et al. that for early enough times, only the fast component of global warming is critical (2009). In addition, the differences between the error for the one- and two-temperature models were very small; the average magnitude of the difference in RMSE was 0.005. For the experiments in reproducing full period data, the new version of the simple model increased model error in 6/14 cases. While deep ocean temperatures should affect warming more in the full period, we only made predictions through 2099. As Figures 18 and 19 show, the warming that the two-temperature model helps to replicate is most important in the years that follow. Due to this, and the fact that in many cases the one-temperature model was

already working well, the results from the two-temperature model made sense. Tables 4 and 5 show the best results for these experiments for each GCM.

GCM	**CO_2 Used**	**RMSE**	**RMSE (One temperature model) - RMSE (Two temperature model)**
GISS E2 H	Atmospheric	0.0494	0.0084
GISS E2 H	Equivalent	0.0525	0.0139
MIROC 5	Atmospheric	0.0885	-0.0035
MIROC 5	Equivalent	0.0834	-0.0032
Can ESM2	Atmospheric	0.1089	-0.0082
Can ESM2	Equivalent	0.1	-0.0088
ACCESS 1.0	Atmospheric	0.0869	-0.0028
ACCESS 1.0	Equivalent	0.0837	-0.001
CNRM CM5	Atmospheric	0.1006	0.0042
CNRM CM5	Equivalent	0.1001	0.0049
CCSM4	Atmospheric	0.0599	-0.004
CCSM4	Equivalent	0.0554	0.0009
NORESM1 M	Atmospheric	0.0694	-0.0034
NORESM1 M	Equivalent	0.0633	-0.0036

Table 4: *The results from reproducing historical period data with the two-temperature model. The type of carbon dioxide data used in each experiment is indicated. For all experiments in this table, combined Nino 3, 4 data was used, and ENSO was detrended by removing the signal of log (CO_2). The fourth column indicates the difference between the RMSE resulting from the one- and two-temperature models.*

GCM	CO_2 Used	RMSE	RMSE (One temperature model) - RMSE (Two temperature model)
GISS E2 H	Atmospheric	0.0785	-0.0231
GISS E2 H	Equivalent	0.0811	-0.0156
MIROC 5	Atmospheric	0.0819	0.0124
MIROC 5	Equivalent	0.0744	0.0121
Can ESM2	Atmospheric	0.1129	-0.0107
Can ESM2	Equivalent	0.1044	-0.0165
ACCESS 1.0	Atmospheric	0.0825	0.0262
ACCESS 1.0	Equivalent	0.0785	0.0295
CNRM CM5	Atmospheric	0.088	0.0364
CNRM CM5	Equivalent	0.086	0.0411
CCSM4	Atmospheric	0.1163	-0.0391
CCSM4	Equivalent	0.1201	-0.0303
NORESM1 M	Atmospheric	0.069	0.0053
NORESM1 M	Equivalent	0.0641	0.0058

Table 5: *As in Table 4, but for full period GCM data.*

As expected, the predictive experiments improved the most when deep ocean temperatures were incorporated into the model: RMSE increased for only 5/14 experiments. **The predictive improvements of the two-temperature model represents a central finding of this study.** The differences in error were more significant for the predictive runs than the full period or historical period experiments. In cases where the two-temperature model decreased error, the average difference was 0.1 C. For GISS E2 H, the original model had already succeeded in predicting future period temperatures with a low error, and the two-temperature model did not improve predictions. While NORESM1 M had one bad result from the two-temperature model, the difference from the one-temperature model was very small. All other models improved from the two-temperature model with the exception of CCSM4. Table 6 summarizes these results, and Figure 22 visually shows the model comparison for two GCMs.

GCM	CO_2 Used	RMSE	RMSE (One temperature model) - RMSE (Two temperature model)
GISS E2 H	Atmospheric	0.1647	-0.0733
GISS E2 H	Equivalent	0.2817	-0.1162
MIROC 5	Atmospheric	0.1353	0.1042
MIROC 5	Equivalent	0.0753	0.0981
Can ESM2	Atmospheric	0.151	0.1306
Can ESM2	Equivalent	0.122	0.0266
ACCESS 1.0	Atmospheric	0.2325	0.09
ACCESS 1.0	Equivalent	0.143	0.1029
CNRM CM5	Atmospheric	0.0988	0.1671
CNRM CM5	Equivalent	0.1109	0.0941
CCSM4	Atmospheric	0.4225	-0.1722
CCSM4	Equivalent	0.5275	-0.1634
NORESM1 M	Atmospheric	0.0714	0.1012
NORESM1 M	Equivalent	0.102	-0.0077

Table 6: *As in Table 4, but for predictive GCM experiments.*

The aerosol interactivity matched the type of carbon dioxide data that provided the best result in all but four cases for 42 experiments. The exceptions were: CNRM CM5 full and historical period experiments; CCSM4 historical period experiments; and NORESM1 M predictive experiments. In all other cases, equivalent carbon dioxide data reduced error the most for fully interactive GCMs, and atmospheric carbon dioxide data reduced error the most for semi-interactive or not interactive GCMs. While this was not as consistent as the one-temperature model experiments, for which there was a match in all but one case, it is still a strong correlation. In addition, for all but the last case, the difference in RMSE between the atmospheric and equivalent carbon dioxide results was under 0.01, indicating that those differences were not very significant.

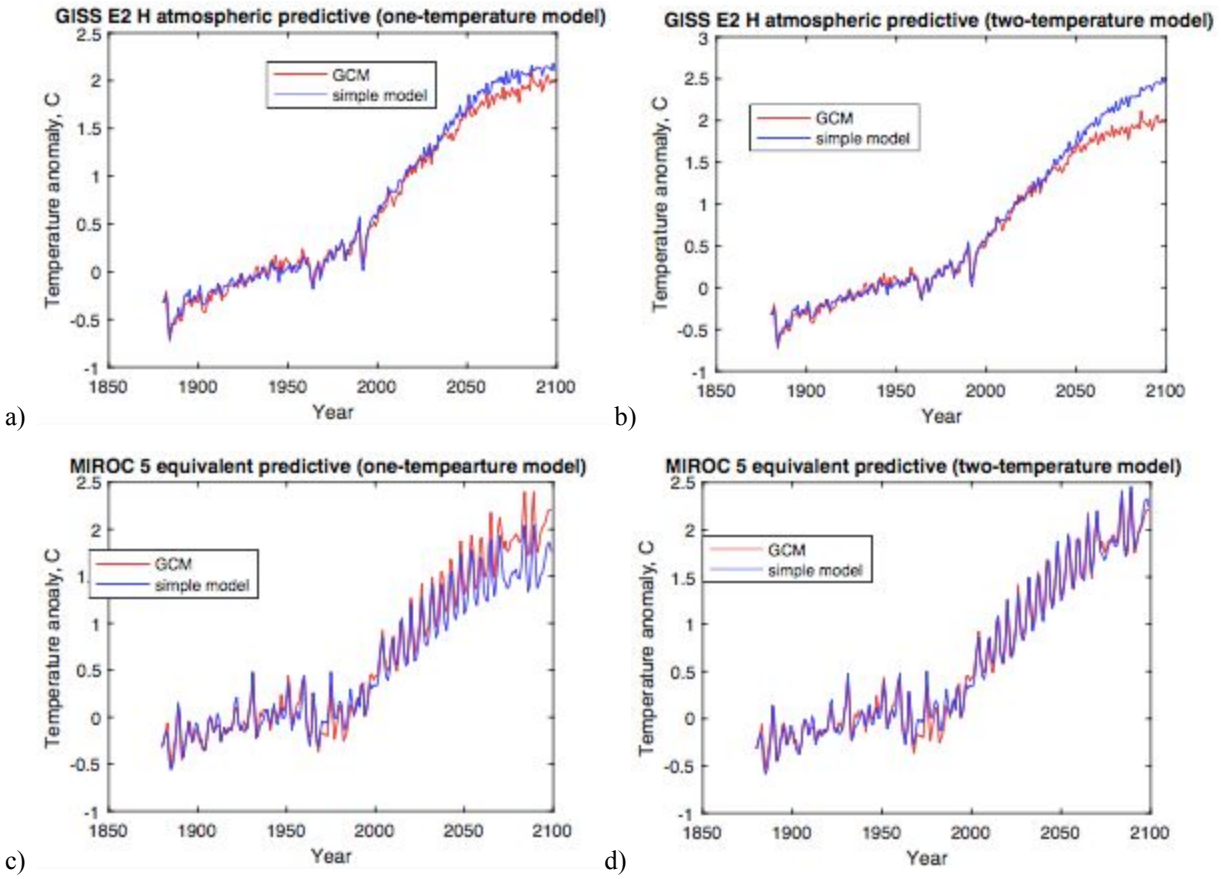


Figure 22*: Predictive experiments with both the one-temperature and two-temperature versions of the simple model for two GCMs . The experiment shown for each is the one with the type of carbon dioxide data that yielded the lowest error. a) GISS E2 H predictive experiment with the one-temperature model and atmospheric carbon dioxide. b) GISS E2 H predictive experiment with the two-temperature model and atmospheric carbon dioxide. c) MIROC 5 predictive experiment with the one-temperature model and equivalent carbon dioxide. d) MIROC 5 predictive experiment with the two-temperature model and equivalent carbon dioxide.*

DISCUSSION

Overall, the simple model reproduces historical and full period data remarkably well for both the observations and the GCMs. While error varied among GCMs, in most cases, the simple model captures the most important features of GMST. This demonstrates the core finding of this project: while GCMs are much more complex, they follow the same basic physics of global warming as the simple model. While errors for using the simple model predictively were higher,

incorporating deep ocean temperatures helped to capture warming inertia and reproduce the warming that continues after emissions flatten.

Using combined Nino 3, 4 data instead of Nino 3 lowered the simple model RMSE. This result was expected, as Nino 3, 4 characterizes a larger area of the Pacific and helps to fully capture El Nino oscillations. Detrending the ENSO data by removing the warming signal from carbon dioxide also helped to calibrate the simple model and reduce RMSE by isolating the signals of El Nino and warming from greenhouse gases. The assumption that using only atmospheric carbon dioxide concentrations, as the forcing from other greenhouse gases and aerosols would cancel each other out, did prove true, as the simple model with atmospheric concentrations reproduced GMST well in most cases. For the GCMs that are fully interactive with aerosols, however, the simple model was also improved by using equivalent carbon dioxide concentrations in a majority of experiments. This makes sense, as the forcing from aerosols would be stronger in those cases, and capturing that using equivalent concentrations would make the simple model more accurate. Including AMO in the model did not significantly improve accuracy; this underlines that GMST trends are largely dependent on the most basic physics of global warming, and that while forcings other than the original three included factors are important, they are not critical to modeling GMST behavior.

Comparing the simple model results to the multimodel mean was useful, but had limitations. While the multimodel mean is somewhat successful in eliminating the biases of individual GCMs and thus useful for comparison the simple model, a limited amount of GCM data was available for all four Representative Concentration Pathways. The differences between the multimodel means calculated in the Methods section demonstrates the extent to which the multimodel mean we used for comparison is not definitive. Despite its limitations, using the mean for comparison to the simple model made the case for using the two-temperature equation in order to capture the slow component of global warming. This method helped the simple model output correlate closely to the multimodel mean, and also ended up reducing predictive errors significantly for many GCMs. The predictive success of the two-temperature model, and its ability to closely replicate the GCM multimodel mean, represent some of the most important findings of this paper.

Fixing the damping coefficients at the values that best helped the two-temperature model to reproduce the multimodel mean was successful in most cases for reproducing individual GCM data. For the historical and full period experiments, the two-temperature model did not greatly alter results, and in about half of cases made them slightly worse. This correlates with the assumption that the slow component of global warming captured by the addition of the deep ocean temperature term is not very important in the historical period, and is most important for predictions, and especially for the later part of the full period. For a majority of experiments, the simple model was able to successfully make predictions with either the one-temperature or, as in most cases, the two-temperature model. This result demonstrates that even GCM predictions boil down to the basic factors included in the simple model, and that when only GMST is required, it might be possible to make predictions without GCMs. Making predictions with the simple model past 2100 for GCMs for which GMST output is available through 2300 might help to further test the predictive power of the two-temperature model, as the warming captured by the inclusion of deep ocean temperatures is particularly important after 2100. The results of this project could also be used in the future to investigate what causes GCMs to exhibit different GMST trends.

The model also has implications for policy and education. The focus on only three factors--ENSO, carbon dioxide, and volcanic aerosols--and the simplicity of the output (only GMST) make it a useful tool for educating people about the most important forcings that affect our planet's temperatures. To that end, we are developing a simple tool that will provide easy access to the model online. The tool consists of a screen that depicts the simple model output and observed GMST on the same graph. Users can adjust sliders to increase or decrease the strength of the three forcings, and see how turning them on and off changes GMST. We are planning to add a tutorial, background information, and a challenge for users: to adjust the forcings until they can determine what caused the "global warming hiatus" in the beginning of the 21st century. This last point is particularly important, as the hiatus was commonly cited as evidence that global warming is not real (Cohn, 2013). In adjusting the model's forcings, users can see for themselves that the hiatus was just an aspect of normal variability, likely caused by ENSO. In a political climate in which climate denial is well funded and climate education--and science education in

general--is not, tools like this one, which requires no mathematical or scientific background, can prove helpful in educating the public.

The simple model's success can also have implications for climate policy. This might be especially true at the international level, where the success of efforts like the Paris Agreement is measured in degrees of warming. The Agreement requires that individual countries present Nationally Determined Contributions (NDCs), their plans for reducing emissions, with a global target of reducing warming to 1.5-2 degrees C above pre-industrial levels (UNFCCC - Nationally Determined Contributions). Current NDCs, however, are on track to result in 3.3 degrees of

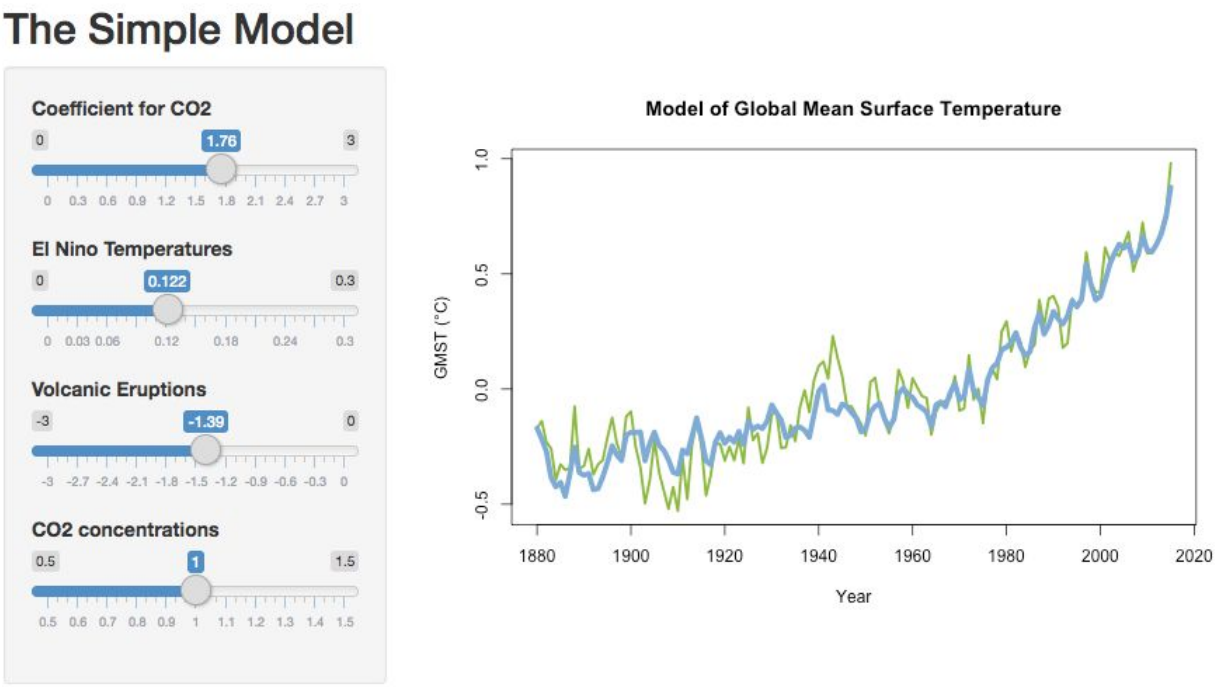


Figure 23: *A screenshot of the educational tool that we are developing. The panel on the right displays the simple model in blue, and GMST observations in green. The panel on the left allows the user to adjust the strength of coefficients a (carbon dioxide forcing), b (ENSO), c (volcanic eruptions) and an additional switch that changes the carbon dioxide concentrations proportionally throughout the model. This is an early draft, and changes will be made in the future.*

warming above pre-industrial levels--and that's only if the NDCs are actually achieved, which is not guaranteed (Climate Action Tracker). There is still opportunity for improvement, however: the Agreement requires that countries submit more ambitious NDCs over time (Rogelj et al.).

This past December, I was able to travel to the UN Climate Conference in Katowice, Poland, also known as COP24, where countries negotiated the rulebook for implementing the Paris Agreement. My experience there affirmed the importance of widely available information on climate modeling to the success of the Paris Agreement. At the conference, I observed how non-party stakeholders--these can include local governments, international organizations, non-profits, activists, academic institutions, businesses and others--are not only invited to COPs, but provided opportunities for input to negotiations (UNFCCC - Overview). The Paris Agreement is not legally binding, and in the vast majority of cases, governments will not act on their own. Pressure from these non-party stakeholders--both those participating in the negotiations and those organizing in their home countries--is critical in convincing their leaders and the international community to continuously improve their NDCs. In order to wield influence, it is often vital that these stakeholders can speak the language of climate science data used by member states and international organizations. Often, however, they do not have access to the same information and education used by national governments. And more specifically, knowing the effects that current NDCs will have on the planet, and why, can be helpful in pressing leaders to ratchet up ambition.

The educational tool we are developing and the model itself, which can be run on any laptop, can help people understand the basics of what affects global temperatures, and, as a result, why ambitious NDCs are important. Adjusting GMST and observing the warming that has occurred since preindustrial times can help users comprehend how close we are to 1.5 and 2 degrees of warming, and thus, why drastic climate action is necessary. The simple model also allows users to examine how much warming specific changes in greenhouse gas emissions would cause, which can help to analyze the NDCs of specific countries and the international community as a whole. Overall, the more ways there are to make climate science information accessible, the better, and we hope that the simple model outlined in this paper can help, even if in small ways.

5. ACKNOWLEDGEMENTS

I would like to thank Professor Alexey Fedorov for his mentorship, support and patience throughout this project, Shineng Hu for guiding me throughout, Wei Liu for helping me with GCM data, Professor Ronald Smith for his time in being a second reader, the Statlab consultants for advising me through multiple coding crises, and Tom Coile and Ivy Fan for helping me develop the educational tool. I greatly appreciate the financial support for this project provided by the Karen L. Von Damm '77 Undergraduate Research Fellowships in Geology & Geophysics, the Hopper College Mellon Grant, and Energy Studies. I would also like to acknowledge my sophomore advisor, Frank Robinson, and the Directors of Undergraduate Studies throughout my time in the Geology & Geophysics Department: David Bercovici, Mary-Louise Timmermans, and Mark Brandon. Lastly, I would like to express my gratitude to all of the climate scientists who created the General Circulation Models I used in this project.

6. REFERENCES CITED

Ammann, C.M., Meehl, G.A., Washington, W.M. and Zender, C.S.: 2003. A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophysical Research Letters*, *30*(12).

Climate Action Tracker, *Temperatures,* accessed April 20 2019, https://climateactiontracker.org/global/temperatures/.

Cohn, N.: 2013, *Where Did the Heat Go?,* The New Republic, accessed April 20 2019, https://newrepublic.com/article/113533/global-warming-hiatus-where-did-heat-go

Driscoll, S., Bozzo, A., Gray, L.J., Robock, A. and Stenchikov, G.: 2012. Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, *117*(D17).

Earth System Research Laboratory Global Greenhouse Gas Network Team, Trends in Atmospheric Carbon Dioxide, NOAA-ESRL. [Available at https://www.esrl.noaa.gov/gmd/ccgg/trends/.]

Etheridge, D.M., Steele, L.P., Langenfelds, R.L., Francey, R.J., Barnola, J.-M., and Morgan, V.I.: 1998. Historical CO_2 record derived from a spline fit (20 year cutoff) of the Law Dome DE08 and DE08-2 ice cores. [Available at https://cdiac.ess-dive.lbl.gov/ftp/trends/CO2/lawdome.smoothed.yr20.]

German Climate Computing Center, CMIP5, World Data Center for Climate. [Available at https://cera-www.dkrz.de/WDCC/ui/cerasearch/.]

Goddard Institute for Space Studies Surface Temperature Analysis Team (2017), GISS Surface Temperature Analysis (GISTEMP), NASA Goddard Institute for Space Studies. [Available at https://data.giss.nasa.gov/gistemp/.]

Gohar, L.K. and Shine, K.P.: 2007. Equivalent CO_2 and its use in understanding the climate effects of increased greenhouse gas concentrations. *Weather*, *62*(11), pp.307-311.

Gregory, J.M., Ingram, W.J., Palmer, M.A., Jones, G.S., Stott, P.A., Thorpe, R.B., Lowe, J.A., Johns, T.C. and Williams, K.D.: 2004. A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3).

Hare, B. and Meinshausen, M.: 2006. How much warming are we committed to and how much can be avoided?. *Climatic Change*, *75*(1-2), pp.111-149.

Held, I.M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. and Vallis, G.K.: 2010. Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), pp.2418-2427.

Hu, S. and Fedorov, A.V., 2017: The extreme El Niño of 2015–2016 and the end of global warming hiatus. *Geophysical Research Letters,* 44(8), pp.3816-3824.

IPCC, date unknown, *What is a GCM?,* accessed 23 March 2019, https://www.ipcc-data.org/guidelines/pages/gcm_guide.html.

Meinshausen, M., Smith, S.J., Calvin, K., Daniel, J.S., Kainuma, M.L.T., Lamarque, J.F., Matsumoto, K., Montzka, S.A., Raper, S.C.B., Riahi, K. and Thomson, A.G.J.M.V.: 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change*, *109*(1-2), p.213.

NOAA-ESRL, date unknown, *El Nino Southern Oscillation (ENSO),* accessed 23 March 2019, https://www.esrl.noaa.gov/psd/enso/.

NOAA Earth System Research Laboratory Team (2019), Atlantic Multidecadal Oscillation Index, OAA Earth System Research Laboratory. [Available at https://www.esrl.noaa.gov/psd/data/timeseries/AMO/.]

Reichler, T. and Kim, J.: 2008. How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, 89(3), pp.303-312.

Rogelj, J., Den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K. and Meinshausen, M.L 2016: Paris Agreement climate proposals need a boost to keep warming well below 2 C. *Nature*, *534*(7609), p.631.

Samset, B.H., Sand, M., Smith, C.J., Bauer, S.E., Forster, P.M., Fuglestvedt, J.S., Osprey, S. and Schleussner, C.F.: 2018. Climate impacts from a removal of anthropogenic aerosol emissions. *Geophysical Research Letters,* 45(2), pp.1020-1029.

Sato, M., Hansen, J.E., McCormick, M.P. and Pollack, J.B.: 1993. Stratospheric aerosol optical depths, 1850–1990. *Journal of Geophysical Research: Atmospheres*, *98*(D12), pp.22987-22994.

Senior, C.A. and Mitchell, J.F.: 2000. The time‐dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), pp.2685-2688.

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M.: 2013. Climate change 2013: The Physical Science Basis.

Taylor, K.E., Stouffer, R.J. and Meehl, G.A.: 2012. An overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society,* 93(4), pp.485-498.

Tel Aviv University, date unknown, *General Circulation Models,* accessed 23 March 2019, https://www.tau.ac.il/~colin/courses/CChange/CC5.pdf

Trenberth, K., *Atlantic Multidecadal Oscillation,* NCAR, accessed 25 Mar 2019, https://climatedataguide.ucar.edu/climate-data/atlantic-multi-decadal-oscillation-amo.

Trenberth, K., *Nino SST Indices,* NCAR, accessed 25 March 2019, https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni

UNFCCC, *Nationally Determined Contributions,* accessed April 20 2019, https://unfccc.int/process/the-paris-agreement/nationally-determined-contributions/ndc-registry.

UNFCCC, *Overview,* Accessed April 20 2019, https://unfccc.int/process/parties-non-party-stakeholders/non-party-stakeholders/overview

van Oldenborgh, G. J., Monthly CMIP5 Scenario Runs, Climate Explorer. [Available at https://climexp.knmi.nl/getindices.cgi?WMO=CMIP5/Tglobal/global_tas_Amon_modmean_rcp45_000&TATION=CMIP5_RCP45_Tglobal&TYPE=i&id=someone@somewhere.]

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.F. and Masui, T.: 2011. The representative Concentration Pathways: an Overview. *Climatic change*, *109*(1-2), p.5.

van Vuuren, D., Calvin, K., Masui, T. and Riahi, K.: 2006-2007, RCP Database. [Available at https://tntcat.iiasa.ac.at/RcpDb/dsd?Action=htmlpage&page=about.]

Williams, K.D., Ingram, W.J. and Gregory, J.M.: 2008. Time variation of effective climate sensitivity in GCMs. *Journal of Climate*, *21*(19), pp.5076-5090.