The University of Chicago

An Adjoint Trajectory Model of the Perturbed Carbon Cycle

by

Dylan Gaeta

A thesis submitted to the advisory committee of:

Douglas MacAyeal
David Archer
Tiffany Shaw
Noboru Nakamura

in partial fulfillment for the degree of

Master of Science

in the

Department of the Geophysical Sciences

August 2018

Abstract

Here we present a new model framework for an adjoint trajectory model of the carbon cycle, which can be used to constrain carbon dioxide emission rates over a specific time interval given only a path constraint on the fraction of carbon dioxide that remains airborne over that interval. Earth's ocean and terrestrial biosphere act to quickly remove a large and variable portion of carbon dioxide emissions from the atmosphere, and thus measuring the atmospheric carbon dioxide concentration at any given time provides limited insight into the global emission flux that is responsible for driving the time-dependent behavior of the atmospheric concentration. Using the adjoint trajectory model presented here, we can, without having any information about the emission schedule, invoke the use of mathematical inverse methods to computationally search for an emission scenario that best reproduces a time series constraint of observed (or desired) atmospheric CO_2 concentrations. This adjoint trajectory model utilizes a simple yet powerful 3-box transfer model to partition CO_2 emissions between the atmosphere and ocean over time. The transfer functions that govern this 3-box model are also utilized to provide a physical constraint on the acceptable set of emission scenario solutions. This thesis describes the governing mathematical framework of the adjoint trajectory model and provides a general algorithm for determining the unknown emission scenario that optimally reproduces the observed (or desired) atmospheric carbon dioxide trend. We show an example of the model performance in its current stage of development, and discuss its limitations and key areas for future development.

Contents

Abstract i

Symbols iii

1 A Forward-Stepping Model of the Perturbed Carbon Cycle: 1
- 1.1 The DICE/BEAM carbon cycle model . . . 2
- 1.2 A 3-box transfer model of a conserved quantity . . . 3
- 1.3 Application of the 3-box model to the carbon cycle . . . 6
- 1.4 Discretization of the forward 3-box model . . . 8
- 1.5 Forward model driven by historical anthropogenic emissions . . . 10
- 1.6 Forward model driven by a future emission scenario . . . 13

2 An Adjoint Trajectory Model of the Perturbed Carbon Cycle: 15
- 2.1 Formulation of the cost function J . . . 16
- 2.2 The Euler-Lagrange equations of optimization . . . 19
- 2.3 Algorithm for minimizing the cost function J to find the forcing $\widetilde{\mathbf{F}}_i$ 20
- 2.4 An application of the adjoint trajectory model . . . 21
- 2.5 Future development of the adjoint trajectory model . . . 24

A Equilibrium carbonate chemistry and the α parameter 27
- A.1 Carbonate chemistry reactions and equilibrium constants . . . 28
- A.2 The DICE temperature model for computing T dependence of equilibrium constants . . . 29
- A.3 Parameterization of α . . . 31
- A.4 Calculating ocean pH . . . 32
- A.5 Calculation of upper ocean alkalinity . . . 33

Bibliography 35

Symbols

SYMBOL	SIGNIFICANCE	UNIT
Introduced in:	**Chapter 1**	
$[CO_2]$	total atmospheric carbon dioxide concentration	Gton C
ϕ_i^j	transfer coefficient from reservoir i to reservoir j	unitless
α	size/chemistry scaling factor between reservoirs 1 and 2	unitless
β	size/chemistry scaling factor between reservoirs 2 and 3	unitless
k_α	inverse of exchange timescale between reservoirs 1 and 2	yr^{-1}
k_β	inverse of exchange timescale between reservoirs 2 and 3	yr^{-1}
M_i	mass of carbon in reservoir i	Gton C
M_{at}	mass of carbon in the atmosphere	Gton C
M_{up}	mass of carbon in the upper ocean	Gton C
M_{lo}	mass of carbon in the lower ocean	Gton C
$\frac{dM_i}{dt}$	time rate of change of the mass of carbon in reservoir i	Gton C/yr
t_i	time at the discrete time step i	yr
Δt	size of the discrete time step	yr
N	number of discrete time steps in a time series	unitless
$E(t)$	carbon emission rate to atmosphere	Gton C/yr
$C(t_1, t_2)$	cumulative mass of carbon emissions between time 1 and time 2	Gton C
$\mathbf{M}_i$	3×1 mass vector at time t_i	Gton C

$\mathbf{K}_i$	3×3 transfer matrix at time t_i	yr^{-1}
$\mathbf{F}_i$	3×1 external forcing vector (i.e. emissions) at time t_i	Gton C/yr
$\frac{d\mathbf{M}_i}{dt}$	time rate of change of all 3 reservoirs of mass vector $\mathbf{M}$ at time t_i	Gton C/yr

Introduced in: **Chapter 2**

$\mathbf{D}_i$	3×1 matrix data constraint on $\mathbf{M}_i$ at the i^{th} discrete time step	Gton C
J	model-data cost function, i.e. sum of square "error" at each time step i	$(\text{Gton C})^2$
$\mathbf{X}^T$	transpose of matrix $\mathbf{X}$	unitless
$\mathbf{H}$	3×3 deficient identity matrix (reservoir 1 only)	unitless
$\widetilde{\mathbf{F}}$	guessed forcing term for the adjoint model	Gton C/yr
$\mathbf{F}_i^*$	final (converged) forcing term calculated by the adjoint model at time t_i	Gton C/yr
$\mathbf{P}_1$	3×1 matrix of Lagrange multipliers constraining the initial condition	unitless
$\mathbf{L}_i$	3×1 matrix of Lagrange multipliers constraining the model physics at time t_i	unitless

Introduced in: **Appendix A**

k_H	Henry's law constant	unitless
k_1	first dissociation constant of carbonic acid	mol/kg
k_2	second dissociation constant of carbonic acid	mol/kg
$[X]$	concentration of chemical species "X"	mol/kg
k_0	empirical function used to calculate k_H	mol/kg
pK_1	empirical function used to calculate k_1	unitless
pK_2	empirical function used to calculate k_2	unitless
T	temperature of the surface ocean, following atmospheric temperature anomaly	Kelvin
δ	atmosphere-to-upper ocean reservoir size ratio	unitless
DIC	dissolved inorganic carbon concentration	mol/kg
Alk	carbonate alkalinity	mol/kg

Chapter 1

A Forward-Stepping Model of the Perturbed Carbon Cycle:

Estimating atmospheric [CO_2] given a carbon emission schedule

As the anthropogenic carbon flux from the deep, slow pool of the carbon cycle to the atmosphere via fossil fuel burning remains active, it becomes increasingly important for researchers to monitor and predict how much of those emissions remain in the atmosphere. Knowing the fraction of anthropogenic CO_2 emissions that remain airborne over the next few centuries is a goal that is rapidly gaining attention from both scientists and public policy makers. Excess airborne CO_2 emissions directly suppress Earth's outgoing longwave radiation, inducing climate change and endangering life throughout the planet, and thus it is important that tools to monitor airborne CO_2 be developed continually.

Efforts to model the response of the carbon cycle to human-induced perturbations have resorted to increasingly complex models in order to accurately represent the dozens of processes that govern the carbon cycle. Doing so requires a tradeoff between either having a full representation of the processes that control the system, or having computational efficiency coupled with a lower-order qualitative understanding of the system.

In this thesis, we will turn our attention to a mathematically simple representation of Earth's carbon cycle in order to demonstrate that a simple three-box model

performs qualitatively similarly to higher complexity models, but at a fraction of the computational cost. Because of the high computational efficiency of the three-box model, we will show for the first time a demonstration of an adjoint trajectory model that is driven by a three-box model. Our development of the adjoint trajectory model presented in Chapter 2 provides the framework of a useful technique for solving the inverse problem in which we have an empirical constraint on the atmospheric CO_2 concentration, but no knowledge of the emission scenario that is responsible for driving the observed behavior.

The type of carbon cycle model presented here has potential applications in a variety of Earth system research subfields, including paleoclimate research and future projections of anthropogenic carbon forcing, and thus we will keep our formulation of the mathematical framework as general as possible.

1.1 The DICE/BEAM carbon cycle model

One of the simplest models of the carbon cycle to date is the Bolin and Eriksson Adjusted Model (BEAM) [9]. In this chapter we will describe a generalized version of BEAM, and in greater detail than is presented in the original publication for the sake of building a framework to study how conserved quantities are transferred between adjacent reservoirs given a transfer function. Following the Nordhaus DICE (Dynamic Integrated Climate-Economy) model [12], BEAM is formulated as three reservoirs of carbon, with exchange fluxes between adjacent reservoirs. The reservoirs are constructed to represent the atmosphere, the upper ocean, and the lower ocean, which yields some constraints on the physics of the model. The total mass of carbon in the system is conserved, except for mass that is added or removed through a perturbation term to the atmosphere reservoir (i.e. "emissions"). This external forcing can be thought of as being any flux of CO_2 into or out of the atmosphere that is not a direct exchange with dissolved inorganic carbon in the upper ocean. BEAM does not consider processes which remove carbon from the deep ocean (e.g. seafloor weathering), since they operate on time scales longer than which BEAM is intended to be used for ($> 10,000$ years).

The mass conservation constraint is satisfied by requiring that any flux out of a reservoir must be simultaneously counteracted by an equal flux into an adjacent reservoir. We will additionally require that the only net flux out of the atmosphere

reservoir and the deep ocean reservoir both go directly into the upper ocean reservoir; there is no terrestrial carbon sink from the atmosphere, and no geologic sink from the deep ocean. Thus, in a steadily positive carbon emission scenario, carbon will appear to accumulate in the deep ocean. Since BEAM was designed to emulate atmospheric concentrations of CO_2, we will focus our attention solely on its ability to accurately model just the atmosphere reservoir, and only use the behavior of the two ocean reservoirs as diagnostics of the model behavior.

It is important to note that size of the "upper ocean" and the "lower ocean" are defined somewhat arbitrarily. The two reservoirs are qualitatively designed to emulate Earth's modern ocean, such that the exchange timescale between the atmosphere and the upper ocean is faster than the exchange timescale between the upper ocean and the lower ocean. They are not intended to be representative of a physically significant depth. This construction allows BEAM to reproduce the long tail that is typically observed in model experiments of anthropogenic CO_2 ocean uptake. It is assumed that the exchange timescale between the atmosphere and upper ocean and between the upper and lower ocean does not change as carbon mass is added to or removed from any of the reservoirs; this timescale requirement is the only explicit assumption that will be made about atmosphere and ocean dynamics.

1.2 A 3-box transfer model of a conserved quantity

We now will define the rate of change of mass in each reservoir by considering all pathways through which mass could enter or leave a reservoir. We will assign each of these pathways a transfer coefficient ϕ_i^j, which denotes the *fraction* of mass M_i moving from reservoir i to reservoir j per year. Considering only transfer between adjacent reservoirs, we can scale the transfer coefficients by the mass of the reservoir from which the mass is leaving. Adding each pathway by which mass enters a reservoir and subtracting each pathway by which mass leaves a reservoir yields the net flux of the reservoir:

$$\frac{dM_i}{dt} = \left(\phi_{i-1}^i \cdot M_{i-1}\right) + \left(\phi_{i+1}^i \cdot M_{i+1}\right) - \left(1 - \phi_i^i\right) \cdot M_i \tag{1.1}$$

where $(1-\phi_i^i)\,M_i$ is the mass per year in reservoir i that gets transferred out of reservoir i; (ϕ_i^i is the fraction per year that stays in reservoir i). Due to our requirement that mass is conserved during reservoir exchanges, the transfer coefficients must not add together to transfer more or less than the total mass of the reservoir. Since the transfer coefficients each control a fraction of the mass of a given reservoir, the sum of the three fractions that the coefficients control must be equal to one. For example, if $\phi_i^1 = 0.2$, $\phi_i^2 = 0.5$, and $\phi_i^3 = 0.3$, then 20% of the mass in reservoir i is transferred to reservoir 1, 50% is transferred to reservoir 2, and 30% is transferred to reservoir 3. In this case, if $i = 2$, for example, then it means that 50% of the mass in reservoir 2 does not move. We can write this constraint on our transfer coefficients (for reservoirs $i = 1, 2, 3$) more succinctly as:

$$\phi_i^1 + \phi_i^2 + \phi_i^3 = 1 \tag{1.2}$$

Since carbon is only exchanged between adjacent reservoirs:

$$\phi_1^3 = \phi_3^1 = 0 \tag{1.3}$$

We can thus write our transfer coefficient constraints explicitly:

$$\phi_1^1 + \phi_1^2 = 1 \tag{1.4}$$

$$\phi_2^1 + \phi_2^2 + \phi_2^3 = 1 \tag{1.5}$$

$$\phi_3^2 + \phi_3^3 = 1 \tag{1.6}$$

Taking these constraints into account, we can then rewrite Equation 1.1 explicitly for each of the three reservoirs, giving us a full set of differential equations for the net flux (in the absence of an external mass forcing):

$$\frac{dM_1}{dt} = -\left(1-\phi_1^1\right) M_1 + \phi_2^1 M_2 \tag{1.7}$$

$$\frac{dM_2}{dt} = \phi_1^2 M_1 - \left(1-\phi_2^2\right) M_2 + \phi_3^2 M_3 \tag{1.8}$$

$$\frac{dM_3}{dt} = \phi_2^3 M_2 - \left(1-\phi_3^3\right) M_3 \tag{1.9}$$

We can then eliminate the self-exchange coefficients ϕ_i^i using the equivalences in Equations 1.4, 1.5, and 1.6:

$$\frac{dM_1}{dt} = -\phi_1^2 M_1 + \phi_2^1 M_2 \tag{1.10}$$

$$\frac{dM_2}{dt} = \phi_1^2 M_1 - \left(\phi_2^1 + \phi_2^3\right) M_2 + \phi_3^2 M_3 \tag{1.11}$$

$$\frac{dM_3}{dt} = \phi_2^3 M_2 - \phi_3^2 M_3 \tag{1.12}$$

We must note that this formulation of the transfer coefficients is only valid if the three reservoirs are identical in size and, in the context of the carbon cycle, contain carbon in the exact same chemical form. If our reservoirs represent the atmosphere and ocean, this is certainly not the case, and so we will modify Equations 1.10, 1.11, and 1.12 using two parameters, α and β, such that the mass of carbon in each reservoir is normalized by reservoir size and chemical activity. More generally, we can define the unitless α and β parameters as the ratio of the reverse exchange coefficient to the forward transfer coefficient, for the upper two reservoirs and lower two reservoirs, respectively.

$$\alpha = \frac{\phi_2^1}{\phi_1^2} \tag{1.13}$$

and

$$\beta = \frac{\phi_3^2}{\phi_2^3} \tag{1.14}$$

We can then rewrite our exchange coefficients ϕ_i^j using our reservoir size and chemistry normalizing parameters (α and β) to be physically meaningful exchange rate constants k_α and k_β, where k_α denotes exchange between the top two reservoirs, and k_β denotes exchange between the bottom two reservoirs, both with units of years^{-1}.

$$k_\alpha = \phi_1^2 = \frac{\phi_2^1}{\alpha} \tag{1.15}$$

$$k_\beta = \phi_2^3 = \frac{\phi_3^2}{\beta} \tag{1.16}$$

We can now rewrite Equations 1.10, 1.11, and 1.12 as an effective concentration gradient that decays with a characteristic timescale $\tau = 1/k$:

$$\frac{dM_1}{dt} = -k_\alpha \cdot (M_1 - \alpha \cdot M_2) \tag{1.17}$$
$$\frac{dM_2}{dt} = k_\alpha \cdot (M_1 - \alpha \cdot M_2) - k_\beta \cdot (M_2 - \beta \cdot M_3) \tag{1.18}$$
$$\frac{dM_3}{dt} = k_\beta \cdot (M_2 - \beta \cdot M_3) \tag{1.19}$$

Our constraints on the transfer coefficients are thus:

$$\phi_1^1 = 1 - k_\alpha \tag{1.20}$$
$$\phi_1^2 = k_\alpha \tag{1.21}$$
$$\phi_2^1 = \alpha \cdot k_\alpha \tag{1.22}$$
$$\phi_2^2 = 1 - \alpha \cdot k_\alpha - k_\beta \tag{1.23}$$
$$\phi_2^3 = k_\beta \tag{1.24}$$
$$\phi_3^2 = \beta \cdot k_\beta \tag{1.25}$$
$$\phi_3^3 = 1 - \beta \cdot k_\beta \tag{1.26}$$

Note that each of the three reservoirs satisfy the mass conservation requirement in Equations 1.4, 1.5, and 1.6.

1.3 Application of the 3-box model to the carbon cycle

In the context of the atmosphere, upper ocean, and lower ocean, we can add an emissions rate term $E(t)$ to the atmosphere reservoir (in units of mass per time) and write Equations 1.17, 1.18, and 1.19 using explicit subscripts:

$$\frac{dM_{at}}{dt} = -k_\alpha \cdot (M_{at} - \alpha \cdot M_{up}) + E(t) \tag{1.27}$$
$$\frac{dM_{up}}{dt} = k_\alpha \cdot (M_{at} - \alpha \cdot M_{up}) - k_\beta \cdot (M_{up} - \beta \cdot M_{lo}) \tag{1.28}$$
$$\frac{dM_{lo}}{dt} = k_\beta \cdot (M_{up} - \beta \cdot M_{lo}) \tag{1.29}$$

In this case, the parameter β is simply the unitless size ratio of the upper ocean to the lower ocean, and is a fixed parameter; (on a spherical planet, this is simply

the depth ratio of the two layers). Thus, the term $\beta \cdot M_{lo}$ is simply the mass of carbon that would be in the lower ocean reservoir if the lower ocean reservoir was the same size as the upper ocean reservoir.

The physical significance of the parameter α is a bit more complicated, and requires a careful treatment of equilibrium carbonate chemistry. This is done in Appendix A, but for now we will simply state that α is the unitless ratio of the mass of atmospheric carbon to the mass of upper ocean carbon that would be in equilibrium with one another at a given pH. It is calculated using empirically derived equations of chemical thermodynamic equilibrium.

The role of the α parameter in BEAM is to account for the diminishing ability of the ocean to take up excess CO_2 from the atmosphere over time. As more and more carbon dissolves in the ocean, the ocean carbonate system responds by reducing ocean pH and inhibiting further uptake of CO_2. Using a three-box model like the one described here, but without an active carbonate chemistry feedback will tend to underestimate the mass of carbon remaining in the atmosphere, since the ocean would continue to draw down atmospheric CO_2 at the same rate even after its buffering capacity has diminished.

In Chapter 2, we will treat α as a constant, but it is worth noting that it is computationally straightforward to update each α_i for $i \in \{1, ..., N\}$ using the carbon content and the pH of the upper ocean at the $i-1$ time step. All figures made in Chapter 1 use a forward model with a time-dependent α. The use of a time-dependent α in the adjoint trajectory model in Chapter 2 will require some further development, but would be an important area of improvement in future versions of the model. This modification has not been thoroughly considered yet, but we will mention here that it may not be as trivial as simply running the carbonate chemistry-including model backwards in time.

It is important to note that the sum of all three instantaneous rates of change at any given time t is simply equal to the external mass forcing rate $E(t)$ at time t, in units of mass per time.

$$\frac{dM_{total}}{dt} = \frac{d\left(M_1 + M_2 + M_3\right)}{dt} = \sum_{i=1}^{3} \frac{dM_i}{dt} = E(t) \tag{1.30}$$

In other words, the only source or sink of carbon mass in the system (i.e. the sum of the three reservoirs) is through the emission term $E(t)$ in the atmosphere

reservoir. If $E(t) > 0$ there is a net gain of carbon mass in the system at time t, and if $E(t) < 0$ there is a net loss of carbon mass from the system at time t. One could simply integrate $E(t)$ between any two times to calculate the cumulative mass of carbon $C(t_1, t_2)$ (in units of Gton C) entering/exiting the system over the time interval $[t_1, t_2]$.

$$C(t_1, t_2) = \int_{t_1}^{t_2} E(t)dt \tag{1.31}$$

The sources and sinks of carbon in this model can be interpreted to represent any process that releases carbon into the atmosphere as CO_2 or removes CO_2 molecules from the atmosphere, *except* the processes of dissolution into the ocean and outgassing from the ocean (i.e. except processes that are explicitly parameterized by the k_α coefficient). The emission term $E(t)$ can be thought of as the sum of:

- anthropogenic fossil fuel emissions
- anthropogenic land use emissions
- human-led carbon capture
- land biosphere carbon fluxes
- ocean biosphere carbon fluxes
- volcanic CO_2 outgassing

The model is not designed to be used on geologic timescales (>10,000 years), as there is no mechanism for the long-term removal or addition of carbon from geologic reservoirs, nor is there a weathering feedback in the 3-box model.

1.4 Discretization of the forward 3-box model

As we begin to think about solving our model numerically, it will be useful to reformulate our governing Equations 1.17, 1.18, and 1.19 in one succinct matrix form equation. Using bold font to denote that a variable is a matrix, we can write a mass vector $\mathbf{M}$:

$$\mathbf{M} = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} \tag{1.32}$$

a transfer matrix $\mathbf{K}$:

$$\mathbf{K} = \begin{bmatrix} -k_\alpha & k_\alpha \cdot \alpha & 0 \\ k_\alpha & -(k_\alpha \cdot \alpha) - k_\beta & k_\beta \cdot \beta \\ 0 & k_\beta & -k_\beta \cdot \beta \end{bmatrix} \tag{1.33}$$

and an external forcing vector $\mathbf{F}$, which includes only an atmospheric component:

$$\mathbf{F} = \begin{bmatrix} E(t) \\ 0 \\ 0 \end{bmatrix} \tag{1.34}$$

We can then write the governing differential equations in matrix form:

$$\frac{d\mathbf{M}}{dt} = \mathbf{K}\mathbf{M} + \mathbf{F} \tag{1.35}$$

For completeness, we can write Equation 1.35 using the explicit matrices from above:

$$\frac{d}{dt}\begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} -k_\alpha & k_\alpha \cdot \alpha & 0 \\ k_\alpha & -(k_\alpha \cdot \alpha) - k_\beta & k_\beta \cdot \beta \\ 0 & k_\beta & -k_\beta \cdot \beta \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} + \begin{bmatrix} E(t) \\ 0 \\ 0 \end{bmatrix} \tag{1.36}$$

which we can algebraically expand to confirm:

$$\begin{bmatrix} \frac{dM_1}{dt} \\ \frac{dM_2}{dt} \\ \frac{dM_3}{dt} \end{bmatrix} = \begin{bmatrix} -k_\alpha \cdot M_1 + k_\alpha \cdot \alpha \cdot M_2 + E(t) \\ k_\alpha \cdot M_1 - (k_\alpha \cdot \alpha - k_\beta) \cdot M_2 + k_\beta \cdot \beta \cdot M_3 \\ k_\beta \cdot M_2 - k_\beta \cdot \beta \cdot M_3 \end{bmatrix} \tag{1.37}$$

In order to numerically solve the model forward in time, we must map the continuous functions and variables in Equation 1.35 onto a discrete space. We will do so by mapping the continuous time interval $[t_0, t_f]$ onto the discrete time interval with N timesteps: $t_i \in [t_1, ..., t_N]$, using i to denote the discrete time step, and where each t_i is evenly separated by the discrete time step size Δt. We can then use the forward Euler method to write Equation 1.35 as a forward time-stepping finite difference model:

$$\frac{\mathbf{M}_{i+1} - \mathbf{M}_i}{\Delta t} = \mathbf{K}_i\mathbf{M}_i + \mathbf{F}_i \tag{1.38}$$

$$\mathbf{M}_{i+1} = \mathbf{M}_i + \Delta t \mathbf{K}_i \mathbf{M}_i + \Delta t \mathbf{F}_i \tag{1.39}$$

It is worth a statement drawing attention to the elegant simplicity of the three-box model described in Equation 1.39. Note that other computational methods could be used to numerically solve our differential equation (e.g. a leapfrog scheme), but we have chosen the forward Euler method because it simplifies the mathematics of the equations governing the adjoint trajectory model in Chapter 2. We find that this method is accurate as long as a sufficiently small time step $\Delta t < 1$ year is used.

The transfer matrix $\mathbf{K}_i$ can simply be written as $\mathbf{K}$ if α is treated as a constant, since α is the only time-dependent parameter in the transfer matrix. With a constant, known α, we also have a fully known $\mathbf{K}$ as long as we know k_α and k_β, both of which were "experimentally" determined via parameter optimization search constrained by higher-complexity model output. Glotter et. al. determined that parameter values of $k_\alpha = 0.2$ yr^{-1} and $k_\beta = 0.05$ yr^{-1} can be used in Equation 1.36 to accurately emulate higher complexity models forced by the same emission scenario [9]. These values correspond to a 5 year atmosphere-upper ocean concentration gradient erosion and a 20 year upper ocean-lower ocean concentration gradient erosion. The constraint that the former time scale be shorter than the latter is a constraint explicitly given to the parameter optimization search algorithm. We will use these parameter values in all calculations throughout this thesis.

1.5 Forward model driven by historical anthropogenic emissions

Before the model can be solved forward in time, we must first establish initial conditions for the differential Equations 1.27, 1.28, and 1.29. If we want to use the model to project forward from today, we need to initialize, or "spin up" the model from pre-industrial conditions until modern conditions, in order to account for the modern day disequilibrium of the system relative to pre-industrial conditions. We will begin by assuming that a 280 ppm CO_2 atmosphere is in equilibrium with the BEAM constructed upper ocean, and that the upper ocean is in equilibrium with

the lower ocean. This is calculated by requiring that:

$$\left(\frac{dM_{at}}{dt}\right)_{t=0} = 0 \tag{1.40}$$

$$\left(\frac{dM_{up}}{dt}\right)_{t=0} = 0 \tag{1.41}$$

$$\left(\frac{dM_{lo}}{dt}\right)_{t=0} = 0 \tag{1.42}$$

Using Equations 1.27, 1.28, and 1.29, we find that these conditions are met when:

$$M_{at} = \alpha \cdot M_{up} \tag{1.43}$$

$$M_{up} = \beta \cdot M_{lo} \tag{1.44}$$

Taking $M_{at} = 595.2$ Gton C (equivalent to ∼280 ppm), and $\alpha = 0.8369$ (see Appendix A for equations used to obtain this value of α), we calculate $M_{up} = 732.7$ Gton C. Taking $\beta = 1/50$ to be the ratio of the upper ocean to lower ocean depth, we can then calculate $M_{lo} = 36633.4$ Gton C.

We now have an initial $\mathbf{M}$ vector at $t = 0$, and so we just need the historical emission scenario $\mathbf{F}_i$ for $i \in \{1, ..., N\}$ before we can run the model forward. We will take the emission history from fossil fuel combustion in Figure 1.1 and use it to force the model out of equilibrium until the model reaches the modern value of CO_2 (∼400 ppm). The mass vector $\mathbf{M}$ at the time at which this occurs will be saved and used as the initial condition for the model when it is run forward in time to predict future atmospheric CO_2 concentrations using a hypothetical future emission schedule. The results of the fossil fuel forcing from a preindustrial equilibrium are shown in Figure 1.2.

Figure 1.2 demonstrates the remarkable ability of a simple three-box model to replicate historical observations of atmospheric CO_2 within a quite small time window (< 20 years). Further optimization of the exchange constants k_α and k_β, and the aspect ratio of the ocean reservoirs β could potentially yield an even closer replication of historical observations, but doing so will not be the focus of this thesis. It is not necessarily the case that the parameter values of k_α, k_β, and β used in this study are the optimal values for emulating atmospheric CO_2 concentrations with the three-box model described above, and one could potentially yield more

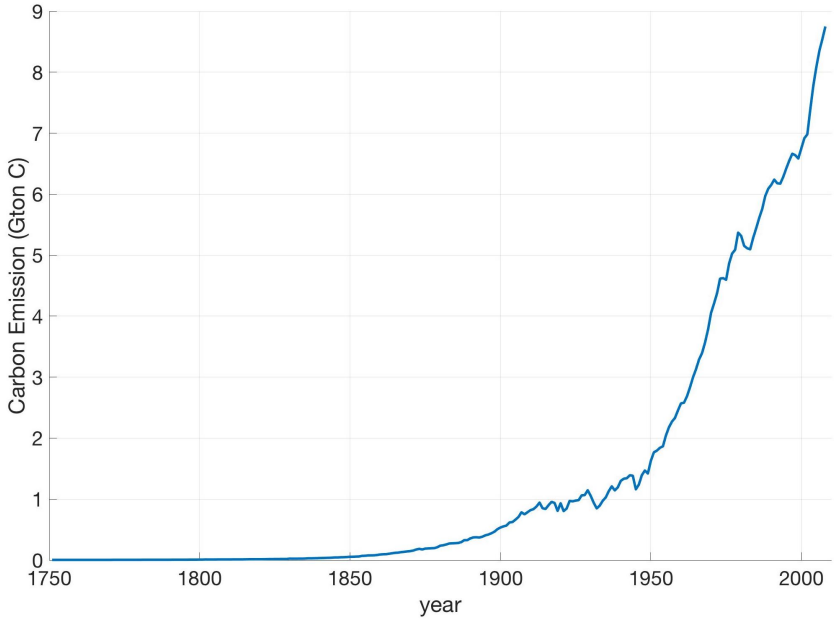

FIGURE 1.1: Carbon emissions from fossil fuel combustion; these data do not include the influence of terrestrial carbon uptake. Data from Boden et. al. [5]

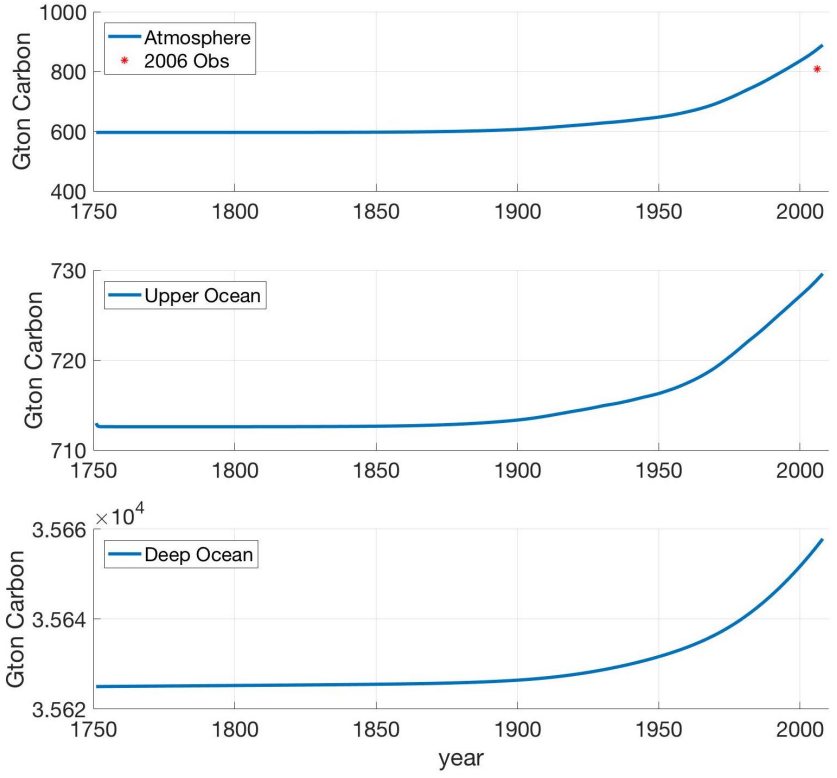

FIGURE 1.2: Model spin-up from an equilibrated pre-industrial atmosphere with 280 ppm CO_2. Calculated using $k_\alpha = 0.2$ yr^{-1}, $k_\beta = 0.05$ yr^{-1}, $\beta = 1/50$, a time-dependent α_i, and pre-industrial $pH = 8.29$.

accurate results with a more optimized triplet set $\{k_\alpha, k_\beta, \beta\}$, subject to the physical constraints that $k_\alpha > k_\beta$ and $\beta < 1$. It is also worth noting that the emission data used to drive the forward model out of pre-industrial equilibrium do not take into account the influence of a terrestrial land carbon sink, which would generally act to slow the growth of atmospheric CO_2 (though this effect would diminish with time) [3].

1.6 Forward model driven by a future emission scenario

Knowing that the model can be used to reasonably reproduce historical observations of atmospheric CO_2, it is also worth testing how BEAM performs relative to higher complexity carbon cycle models when forced with the same future carbon emission scenario. Here, we will give a demonstration of BEAM's ability to replicate the model output of two Earth System Models of Intermediate Complexity (EMICs), each driven by the same emission scenario given in Figure 1.3.

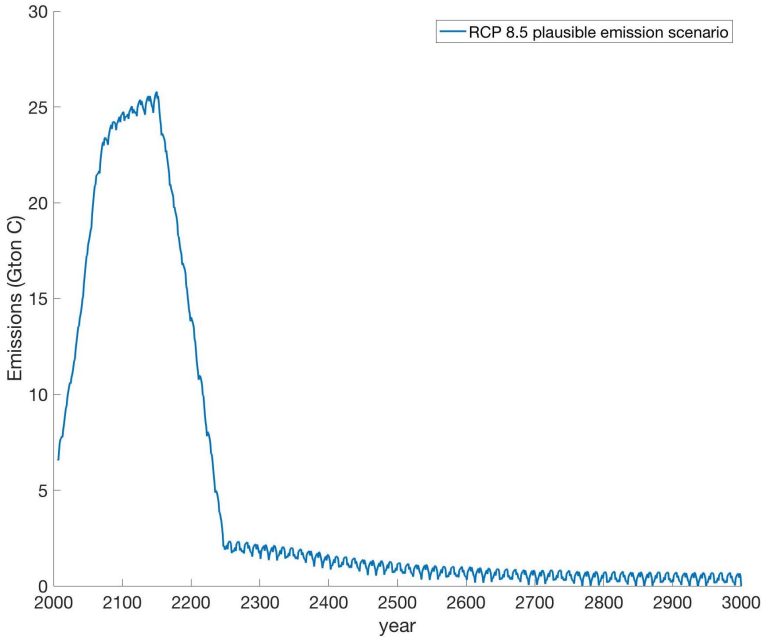

FIGURE 1.3: A plausible CO_2 emission scenario that could generate an 8.5 W/m^2 radiative forcing by 2100.

Given the atmospheric forcing term $\mathbf{F}_i$ in Figure 1.3, BEAM produces the following projection, plotted in Figure 1.4. The three-box model performs well within

the envelope of uncertainty surrounding the two EMIC projections, which is quite remarkable given the computational simplicity of the three-box model. Using a personal computer manufactured in 2016, the three-box model can be run 5000 years into the future on the timescale of a couple seconds, which makes the model quite a powerful tool for qualitative CO_2 emission research among physical scientists, economists, public policy researchers, etc.

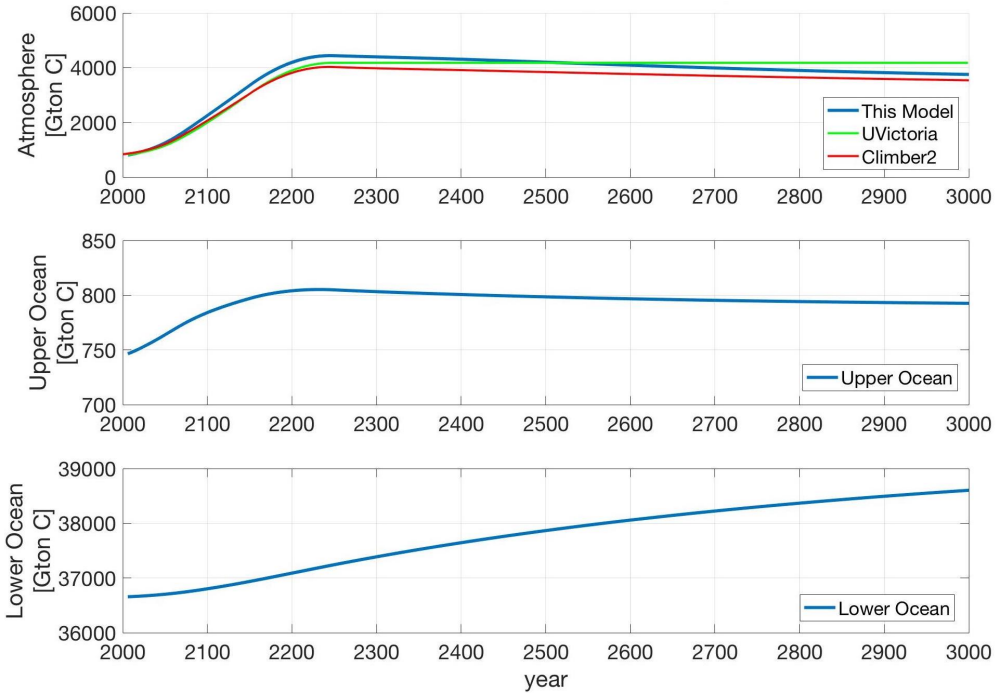

FIGURE 1.4: BEAM output given the emission scenario in Figure 1.3. Calculated using $k_\alpha = 0.2$ yr^{-1}, $k_\beta = 0.05$ yr^{-1}, $\beta = 1/50$, a time-dependent α_i, and non-equilibrium initial conditions from the red-starred year in Figure 1.2.

Chapter 2

An Adjoint Trajectory Model of the Perturbed Carbon Cycle:

Estimating a carbon emission scenario given atmospheric [CO_2] data

Chapter 1 provides a description of how a simple three-reservoir carbon cycle model can be ran forward in time to both reasonably replicate historical observations of atmospheric CO_2 given an empirically constrained emission history, and predict future atmospheric CO_2 concentrations given a projected emission scenario for the future. Both of these applications of the forward model require an *a priori* knowledge of the emission schedule $E(t)$.

However, one could imagine several scenarios in which time series data of atmospheric CO_2 concentrations are available, but the emission schedule that forces that time series is not known. Throughout the paleoclimate record, for example, we have data constraints on the atmospheric CO_2 concentration over several time intervals, but we have limited knowledge of the fluxes of CO_2 that drive the observed swings in the atmospheric concentration.

Similarly, one could imagine having a computationally projected constraint on the maximum concentration of CO_2 throughout a time series that would be allowable before damages from the resulting climate change become too costly. With this information, we would be interested to know what the optimum schedule of emissions within this time series would be, such that we could release as much CO_2

as possible (thereby maximizing economic growth) without exceeding our desired atmospheric concentration. [1]

It may also prove to be a useful tool for environmental regulation enforcement. After significant future development of the model, it would allow an enforcement official to determine the emission rate of CO_2 over some time interval without being given any information except a timeseries measurement of the airborne fraction remaining after ocean uptake. Enforcement officials would thus be able to determine whether an emission cap was violated without having to monitor each flux that contributes to the total emission. With the variety of potential applications of an adjoint trajectory model in mind, we will keep our formulation of the mathematical framework as general as possible (using 1's, 2's, and 3's to label the reservoirs) so that one could use this thesis as a framework for developing a similar model with different applications.

2.1 Formulation of the cost function J

With this type of problem in mind, we will invoke the use of mathematical inverse methods in order to perform an optimization of the forcing term $\mathbf{F}$ that will drive the model output $\mathbf{M}_i$ from Equation 1.39 to be as close as possible to an observational data constraint $\mathbf{D}_i$. This data constraint matrix $\mathbf{D}_i$ could be an empirical time series measurement, or it could be model output from a higher complexity model; either way, we will treat it as data that we trust to be more accurate than the simple model output $\mathbf{M}_i$.

We will start by defining a cost function J, which we will use as a metric of how closely our forward model $\mathbf{M}_i$ replicates the observational data $\mathbf{D}_i$ at each time step $i \in \{1, ..., N\}$.

$$J = \sum_{i=1}^{N} (\mathbf{M}_i - \mathbf{D}_i)^T (\mathbf{M}_i - \mathbf{D}_i) \tag{2.1}$$

where $\mathbf{M}_i = \left[M_i^{at}, M_i^{up}, M_i^{lo}\right]^T$ is the model output for the discrete time series from $i = \{1, ..., N\}$ and $\mathbf{D}_i$ is a data constraint on the same variable over the same discrete time series. However, due to our upper ocean and lower ocean reservoirs

[1] It is important to note that we are not advocating the use of this method in the allocation of carbon emissions, and it should not be used as a public policy tool or a marketing tool in carbon cap-and-trade decisions until much further development is made on the model.

being designed to not necessarily represent the real ocean, we generally do not have data constraints on the two ocean reservoirs. For example, with $\beta = 1/50$, a data constraint on the "upper ocean" would require a measurement of the carbon content of the globally-averaged upper 1/51st of the ocean; data such as these are unlikely to be available.

Given the general unavailability of D^{up} and D^{lo}, we must ensure that only the atmospheric components of the mass vector $\mathbf{M}$ contribute to the cost function J. To do this, we will define a deficient identity matrix $\mathbf{H}$ such that only the purely atmospheric reservoir components remain, and the rest get multiplied by zero.

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.2}$$

We can then rewrite the cost function using $\mathbf{H}$ so that the $(M_i^{up} - D_i^{up})^2$ and $\left(M_i^{lo} - D_i^{lo}\right)^2$ components are eliminated from the summation:

$$J = \sum_{i=1}^{N} \left(\mathbf{M}_i - \mathbf{D}_i\right)^T \mathbf{H}^T \mathbf{H} \left(\mathbf{M}_i - \mathbf{D}_i\right) \tag{2.3}$$

which we can rewrite more simply as:

$$J = \sum_{i=1}^{N} \left(M_i^{at} - D_i^{at}\right)^2 \tag{2.4}$$

Our objective at this point is to minimize J by finding $E(t)$ such that the forward model $\mathbf{M}_{i+1} = \mathbf{M}_i + \Delta t \mathbf{K} \mathbf{M}_i + \Delta t \mathbf{F}_i$ reproduces $\mathbf{D}_i$ as closely as possible at each $i \in \{1, ..., N\}$. Our optimization of the vector $\mathbf{F}_i$ will be performed on the N-dimensional vector space $\{1, ..., N\}$.

Before we begin to minimize J, we must first ensure that our cost function J satisfies two particular constraints. First, we will require that the initial conditions of our forward model are set by the observational data at the first discrete timestep $i = 1$. That is,

$$\mathbf{M}_1 = \mathbf{D}_1 \tag{2.5}$$

Second, we will require that the forward model satisfies the transfer function constraints set by $\mathbf{K}$ (i.e. the time-independent model physics), and also require that

emissions directly enter the atmospheric reservoir alone (via $\mathbf{H}$).

$$\mathbf{M}_{i+1} = \mathbf{M}_i + \Delta t \mathbf{K}\mathbf{M}_i + \Delta t \mathbf{H}\widetilde{\mathbf{F}}_i \tag{2.6}$$

Here we write $\widetilde{\mathbf{F}}_i$ using a tilde to indicate that our forcing vector is only a guess, and is not known *a priori*, as was the case in the forward model in Chapter 1. As this adjoint model is run forward, we will iteratively update $\widetilde{\mathbf{F}}_i$ using a mathematical method known as Lagrange multipliers in combination with a steepest descent down-gradient search method until our model error $(M_i^{at} - D_i^{at})$ at each $i \in \{1, ..., N\}$ is as small as possible. When our model error (i.e. J) is minimized at each i, the guessed forcing $\widetilde{\mathbf{F}}_i$ at that point can be saved and considered the optimized carbon forcing vector $\mathbf{F}_i^*$.

We can minimize the cost function J while only considering solutions which also satisfy Equations 2.5 and 2.6 by using the mathematical method of Lagrange multipliers. We will augment our cost function J using a Lagrange multiplier $\mathbf{P}_1$ to enforce the constraint from Equation 2.5 at the initial time step $i = 1$, and a vector-valued Lagrange multiplier $\mathbf{L}_i$ to enforce the constraint from Equation 2.6 on the interval $i \in \{1, ..., N-1\}$. When we differentiate J with respect to each of its variables (as we will do soon during our minimization algorithm), the Lagrange multipliers, by construction, can be used to search downgradient of J to find a more accurate $\mathbf{M}_i$, but only by considering solution curves that also meet the conditions of Equations 2.5 and 2.6. These Lagrange multipliers will yield a set of constraint equations that also must be satisfied (at each i) in order for the model output $\mathbf{M}_i$ to be considered optimized.

$$\begin{aligned} J = &\sum_{i=1}^{N} \left(M_i^{at} - D_i^{at}\right)^2 \\ &+ \mathbf{P}_1^T \cdot (\mathbf{M}_1 - \mathbf{D}_1) \\ &+ \sum_{i=1}^{N-1} \mathbf{L}_i^T \cdot \left(-\mathbf{M}_{i+1} + \mathbf{M}_i + \Delta t \mathbf{K}\mathbf{M}_i + \Delta t \mathbf{H}\widetilde{\mathbf{F}}_i\right) \end{aligned} \tag{2.7}$$

It is convenient to think of the Lagrange multipliers $\mathbf{P}_1$ and $\mathbf{L}_i$ as dummy variables that bookkeep the model error $(M_i^{at} - D_i^{at})$, and use it to drive the guessed forcing term $\widetilde{\mathbf{F}}_i$ towards an optimized value $\mathbf{F}_i^*$. As we will see in the next section, when our conditions of optimization are reached, neither of the Lagrange multiplier terms in Equation 2.7 will contribute to J; they will either be identically zero,

or multiplied by zero. Thus, once we have found $\mathbf{F}_i^*$, our cost function J can be effectively described by our original cost function in Equation 2.4.

2.2 The Euler-Lagrange equations of optimization

With the fully formulated cost function J in Equation 2.7, we then seek to minimize J with respect to $\widetilde{\mathbf{F}}_i$. We can achieve this via standard calculus procedures, i.e. taking the partial derivative of J with respect to each variable that it is a function of, setting those derivatives equal to zero, and then taking the resulting expressions as the conditions for optimization (also known as the Euler-Lagrange equations of optimization).

We can differentiate J with respect to our Lagrange multipliers $\mathbf{P}_1$ and $\mathbf{L}_i$, the model output $\mathbf{M}_i$, and the guessed forcing term $\widetilde{\mathbf{F}}_i$:

$$\frac{\partial J}{\partial \mathbf{P}_i} = \mathbf{H}\left(\mathbf{M}_1 - \mathbf{D}_1\right) \tag{2.8}$$

$$\frac{\partial J}{\partial \mathbf{L}_i} = -\mathbf{M}_{i+1} + \mathbf{M}_i + \Delta t \mathbf{K}\mathbf{M}_i + \Delta t \mathbf{H}\widetilde{\mathbf{F}}_i \tag{2.9}$$

$$\frac{\partial J}{\partial \mathbf{M}_i} = \begin{cases} 2\mathbf{H}\left(\mathbf{M}_1 - \mathbf{D}_1\right) + \mathbf{P}_1 + \mathbf{L}_1 + \mathbf{L}_1 \Delta t \mathbf{K}, & \text{i} = 1 \quad (2.10) \\ 2\mathbf{H}\left(\mathbf{M}_i - \mathbf{D}_i\right) - \mathbf{L}_{i-1} + \mathbf{L}_i + \mathbf{L}_i \Delta t \mathbf{K}, & \text{i} = 2, \ldots, \text{N-1} \quad (2.11) \\ 2\mathbf{H}\left(\mathbf{M}_N - \mathbf{D}_N\right) - \mathbf{L}_{N-1}, & \text{i} = \text{N} \quad (2.12) \end{cases}$$

$$\frac{\partial J}{\partial \widetilde{\mathbf{F}}_i} = \mathbf{L}_i \Delta t \mathbf{H} \tag{2.13}$$

It is worth noting that these equations are rewritten using the matrix identity $\mathbf{L}_i^T \mathbf{K} = \mathbf{K}^T \mathbf{L}_i$. Since $\mathbf{L}_i$ is used to drive the inverse model, the adjoint matrix $\mathbf{K}^T$ provides the name "adjoint trajectory model."

Equating Equations 2.8, 2.9, 2.10, 2.11, 2.12 and 2.13 to zero yields the Euler-Lagrange equations of optimization:

$$\mathbf{M}_1 = \mathbf{D}_1 \tag{2.14}$$

$$\mathbf{M}_{i+1} = \mathbf{M}_i + \Delta t \mathbf{K}\mathbf{M}_i + \Delta t \mathbf{H}\widetilde{\mathbf{F}}_i \tag{2.15}$$

$$\begin{cases} \mathbf{L}_{N-1} = 2\mathbf{H}\left(\mathbf{M}_N - \mathbf{D}_N\right), & \text{i = N} \quad (2.16) \\ \mathbf{L}_{i-1} = \mathbf{L}_i + \Delta t \mathbf{K}^T \mathbf{L}_i + 2\mathbf{H}\left(\mathbf{M}_i - \mathbf{D}_i\right), & \text{i = 2, ... , N-1} \quad (2.17) \\ \mathbf{P}_1 = -\mathbf{L}_1 - \Delta t \mathbf{K}^T \mathbf{L}_i - 2\mathbf{H}\left(\mathbf{M}_1 - \mathbf{D}_1\right), & \text{i = 1} \quad (2.18) \end{cases}$$

$$\mathbf{H}^T \mathbf{L}_i = 0 \quad (2.19)$$

Equation 2.19 is the condition under which $\frac{\partial J}{\partial \widetilde{\mathbf{F}}_i} = 0$, which is our ultimate goal. In other words, when $\mathbf{H}^T \mathbf{L}_i = L_i^{at} = 0$, the forcing term $\widetilde{\mathbf{F}}_i$ is optimized such that it produces $\mathbf{M}_i$ as close to $\mathbf{D}_i$ as possible for each $i \in \{1, ..., N\}$. Thus, our goal is to search down-gradient of $\frac{\partial J}{\partial \widetilde{\mathbf{F}}_i}$ to find a better $\widetilde{\mathbf{F}}_i$ to drive the model output $\mathbf{M}_i$ that yields $\mathbf{L}_i = 0$, as calculated by Equation 2.17. The model error $\mathbf{H}\left(\mathbf{M}_i - \mathbf{D}_i\right)$ is used to drive the adjoint model backwards in time via Equation 2.17.

We now have the full set of equations that are required to solve the adjoint model, and now we just need an algorithm for using them.

2.3 Algorithm for minimizing the cost function J to find the forcing $\widetilde{\mathbf{F}}_i$

Here we will use the subscript k to specify which iteration of the guess $\widetilde{\mathbf{F}}_i$ we are considering. The general algorithm for performing an iterative steepest descent down-gradient search method using Lagrange multipliers can be described as follows:

1. Guess an initial forcing $\left(\widetilde{\mathbf{F}}_i\right)_{k=1}$ for each $i \in \{1, ..., N\}$.
2. Use Equation 2.15 and $\left(\widetilde{\mathbf{F}}_i\right)_{k=1}$ to run the model forward in time, starting with the initial condition $\mathbf{M}_1 = \mathbf{D}_1$ (from Equation 2.14) in order to compute $\left(\mathbf{M}_i\right)_{k=1}$
3. Calculate the total error J_k using the data constraint $\mathbf{D}_i$ and the model output $\left(\mathbf{M}_i\right)_k$, via Equation 2.4
4. Use Equation 2.17 to solve backwards in time for the Lagrange multipliers $\left(\mathbf{L}_i\right)_k$ over $i \in \{N-1, ..., 2\}$, using Equation 2.16 as the initial condition.

(a) If $\mathbf{L}_i = 0$, then $\dfrac{\partial J}{\partial \widetilde{\mathbf{F}}_i} = 0$ and thus we have found $\mathbf{F}_i^*$, and so we are done.

5. Use Equation 2.13 and the $(\mathbf{L}_i)_k$ calculated in Step 4 to compute the new gradient of J. Use the sign of $\dfrac{\partial J}{\partial \widetilde{\mathbf{F}}_i}$ to determine whether to increase or decrease $\widetilde{\mathbf{F}}_i$ at each i. The magnitude of the adjustment will be computed in the next step.

6. Use an open-source optimization script to search for the magnitude by which $\widetilde{\mathbf{F}}_i$ should be adjusted in order to minimize $\mathbf{H}(\mathbf{M}_i - \mathbf{D}_i)$. For example, the MATLAB function *fminunc* is used in this edition of the model code to iteratively find the magnitude by which $\widetilde{\mathbf{F}}_i$ should be adjusted, using the adjustment magnitude from the previous iteration as a starting point in the search for a minimum.

7. At each i, increase or decrease $\left(\widetilde{\mathbf{F}}_i\right)_{k-1}$ by the amount determined in Step 6 to calculate the new guess $\left(\widetilde{\mathbf{F}}_i\right)_k$, then run the model forward using $\left(\widetilde{\mathbf{F}}_i\right)_k$ and Equation 2.15 to compute $(\mathbf{M}_i)_k$

8. Return to Step 3, and continue to follow this algorithm until Step 4a is reached.

2.4 An application of the adjoint trajectory model

Via the algorithm described above, Figure 2.1 shows an example of how this adjoint trajectory model performs, using empirically measured atmospheric CO_2 data from the Mauna Loa Observatory as the data constraint $\mathbf{D}_i$ [10]. The adjoint trajectory model does a remarkable job of fitting the small-scale structure in the empirical data, but notably fails at one of the near-endpoints of the time series (around the year 1958). This near-endpoint failure drives the entire solution curve of $\mathbf{F}_i^*$ in Figure 2.2 away from what we know to be the correct carbon emission history, as seen in the steadily rising curve in Figure 1.1.

Zooming in on the solution in Figure 2.2, we see in Figure 2.3 that the adjoint trajectory model does a notable job of reproducing the observed seasonal cycle in fossil fuel burning [15], with distinguished maxima occurring in the northern

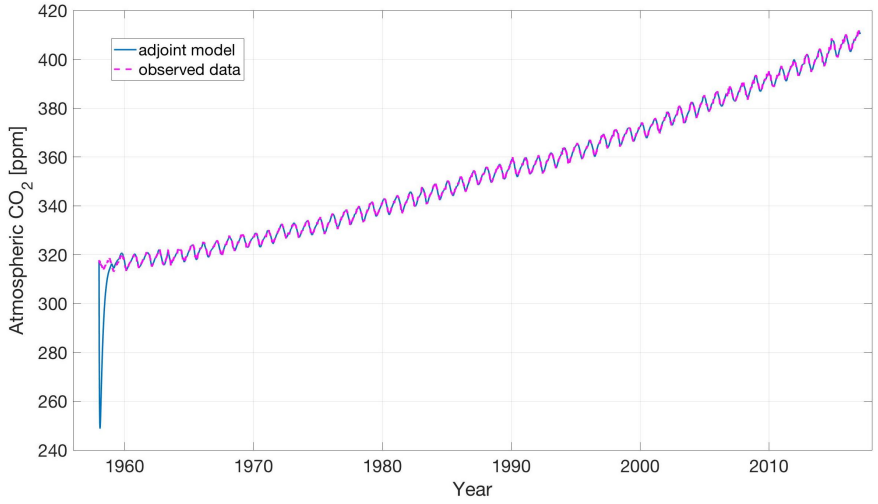

FIGURE 2.1: Atmospheric CO_2 data from Mauna Loa Observatory in Hawaii fitted using the adjoint trajectory model. Note that the sharp model-misfit near 1958 occurs a few data points after the initial condition of the time series, and does not violate the initial condition in Equation 2.14.

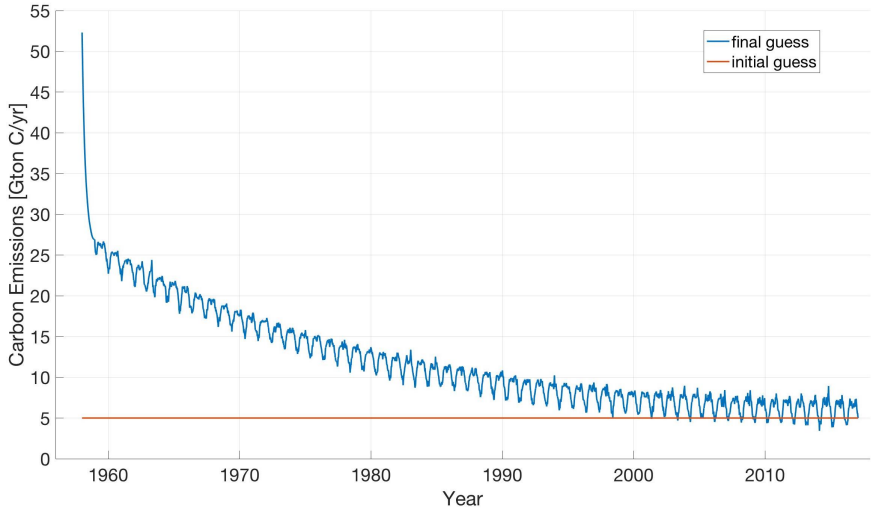

FIGURE 2.2: Guessed emission history $\mathbf{F}_i^*$ calculated by the adjoint trajectory model, and used to calculate the forward model fit in Figure 2.1

hemisphere winter. The model-calculated emission curve in Figure 2.3 successfully replicates the annual oscillatory pattern that is expected from energy consumption data [4], though it notably misses the local minimum that would be expected each year when terrestrial carbon uptake peaks [14].

The ability to extract small-scale structure in the guessed forcing term can be considered to be one of the larger successes of the model at this stage of development. The large-scale time series trend, however, is not yet able to be reliably computed

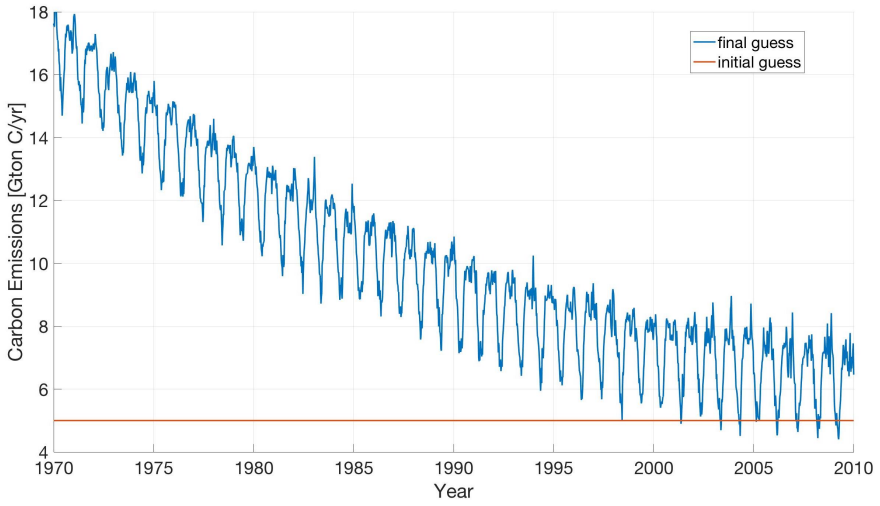

FIGURE 2.3: Zoom in on finer structure in Figure 2.2

using this model. Further development of the model should focus on the behavior of the model near the endpoints of the time series; the endpoint "tethering" behavior is something that was consistently observed using a variety of CO_2 time series data constraints.

Figure 2.4 demonstrates the iterative minimization of the total error over the time series, demonstrating the success of the model in its ability to continually search down-gradient for the optimized forcing term.

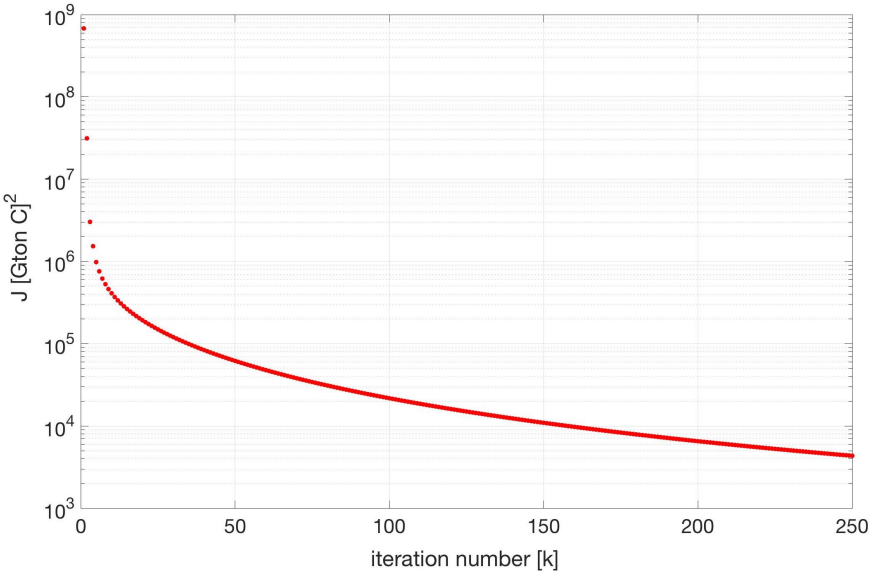

FIGURE 2.4: Iterative minimization of cost function J

2.5 Future development of the adjoint trajectory model

This thesis reports the first edition of a 3-box adjoint trajectory model of the carbon cycle, and we want to be careful to note that it does not represent the fully developed version of the model. There are several developments which could be made to improve the model such that it could be used to answer scientific research questions of greater complexity than the proof-of-concept questions addressed in this thesis. A few of those improvement areas will be addressed in this section.

Looking at Figure 2.1, there is an area of improvement which stands out immediately. The model output $\mathbf{M}_i$ in blue fails to fit anywhere near the data constraint $\mathbf{D}_i$ in magenta for a few data points near the beginning of the time series. The explanation for this is most likely that a self-inconsistent initial condition is used in the forward model. Since the example adjoint model run in Chapter 2 begins in the year 1958, the initial condition of the forward model ($\mathbf{M}_1$) should be taken to be the mass of each reservoir in the year of Figure 1.2 where M_{at} is equal to the 1958 value of atmospheric carbon dioxide concentration. At this time step, one would then need to determine how much the pH of the ocean had decreased since the pre-industrial $pH = 8.29$ and then use that to determine an initial value of the pH-dependent parameter α.

It is most likely our lack of self-consistent initial conditions in the adjoint model that cause the model-data misfit to spike near the beginning of the time series in Figure 2.1. The forward BEAM in Chapter 1 has quite sensitive numerical stability criterion, such that an incorrect value of α, M_1^{at} or M_1^{up} cause a large effective concentration gradient to be calculated at the first time step (via Equations 1.27 and 1.28), thus causing a large transfer of mass between the two reservoirs to occur over the next few time steps. There is evidence of this instability manifesting in Figure 2.1.

In order to have a fully self-consistent set of initial conditions, we would need to spin-up the model from pre-industrial equilibrium to determine the initial condition for the model $\mathbf{M}_1$ and for the reservoir normalization parameter α. The adjoint trajectory model in Chapter 2 uses a constant value of $\alpha = 1$, which is unlikely to be the value of α that would be calculated in the model spin-up. Thus,

the effective concentration gradient between the atmosphere and upper ocean is miscalculated, causing mass to rapidly flow.

If the adjoint trajectory model is unable to fit the model output $\mathbf{M}_i$ to the data constraint $\mathbf{D}_i$, it is virtually impossible for the cost function minimization algorithm to accurately compute a forcing $\widetilde{\mathbf{F}}$ at each i that it was unable to fit precisely. Thus, there is likely much to be said about the ability of the adjoint trajectory model to replicate an emission scenario in a physically reasonable manner, but since the error in Figure 2.2 relative to the known emission scenario appears to be almost entirely driven by the model misfit near time zero in Figure 2.1, there is little that we can conclude about how to enhance the physical reasonability of the calculated forcing term.

Another area of future exploration would be to methodically inquire whether the initial guess for the forcing matters. Does the model equilibrate to a different solution $\mathbf{F}_i^*$ if given a different initial guess $\widetilde{\mathbf{F}}_i$ (i.e. a different constant value guess, or a different shape curve), or does the optimization algorithm in Section 2.3 always converge to an absolute maximum solution $\mathbf{F}_i^*$? Answering this question should be relatively straightforward to implement by simply trying multiple initial guesses for different data constraints and analyzing how the initial guess affects the optimized forcing term.

The novelty of BEAM is its ability to explicitly account for the diminishing rate of ocean carbon uptake in a steadily positive carbon dioxide emission scenario [9]. With this chemical constraint in mind, it is important that future development of the adjoint trajectory model also addresses the carbon cycle constraints imposed by carbonate chemistry by making the α parameter time-dependent in the inverse model. It is possible that the variation in emission rates between guess iterations is small enough such that a time-dependent $\mathbf{K}_i$ can be calculated via the forward BEAM model and used as a constant in the inverse model for every k^{th} iteration of the optimization sequence. If they are not small enough, one would have to develop a different method for calculating the pH dependence of α in the inverse part of the model.

Considering the potential paleoclimate applications of the model, one might not be interested in obtaining a single, absolutely optimized guessed forcing, but rather a set of possible forcings that *could* drive the observed behavior in a set of empirical measurements (e.g. the Vostok ice core record). If a paleoclimate researcher could

be presented with a set of possible emission scenarios, they could combine that set with other empirical geochemical evidence to eliminate certain scenarios, and, with a careful understanding of the assumptions made by the adjoint trajectory model, try to constrain the range of plausible emission scenarios over periods in Earth's history. If this could be done successfully, it would allow one to then explore the plausible mechanisms that controlled Earth's climate in the distant past. An advanced version of this adjoint trajectory model would thus include the ability to explore local minimums of the optimization space of $\widetilde{\mathbf{F}}_i$, rather than the absolute minimum in the present state of the model.

The adjoint trajectory model presented here is still quite far from that idealized paleoclimate application, but the process of adding additional constraints to the model cost function J and then minimizing the cost function using inverse methods is a broadly applicable technique that could be continuously augmented. There is no technical limit to the number of constraints we can place on our system; our only practical limitation would be that the assumptions we make to arrive at our constraints may be contradictory such that the simultaneous minimization along each optimization constraint yields no solution. A solution to an overdetermined system of this type could be found by simply relaxing the constraints that require the most unrealistic assumptions about the physical system.

Appendix A

Equilibrium carbonate chemistry and the α parameter

Much of what is presented here can be found in the original BEAM paper [9] and references therein, but we present the following exposition for completeness and clarification of the original paper, plus some corrections. Note that Glotter et. al. [9] separate our parameter α into two parameters, A and B, such that $\alpha = A \cdot B$.

The parameter α was introduced in Equation 1.13 in order to account for the possibility that reservoirs 1 and 2 are different sizes and/or contain different chemical forms of carbon. In the context of the atmosphere and the upper ocean, both are the case. Glotter et. al. [9] first introduced a parameterization of α in order to account for the fact that the ocean's ability to uptake atmospheric CO_2 decreases as the concentration of dissolved inorganic carbon increases.

A.1 Carbonate chemistry reactions and equilibrium constants

We will start by defining the chemical equilibrium reactions that describe the process of atmospheric CO_2 dissolution in the ocean.

$$CO_{2(g)} \overset{k_H}{\rightleftharpoons} CO_{2(aq)} \tag{A.1}$$

$$CO_{2(aq)} + H_2O \overset{\text{fast}}{\rightleftharpoons} H_2CO_{3(aq)} \tag{A.2}$$

$$H_2\,CO_{3(aq)} \overset{k_1}{\rightleftharpoons} H^+ + HCO_3^- \tag{A.3}$$

$$HCO_3^- \overset{k_2}{\rightleftharpoons} H^+ + CO_3^{-2} \tag{A.4}$$

The aqueous subscript (aq) is omitted from any charged chemical species, as it is assumed that charged species are only stable in the aqueous phase.

The fast reaction in Equation A.2 means we can assume that the concentrations:

$$\left[CO_{2(aq)}\right] = \left[H_2CO_{3(aq)}\right] \tag{A.5}$$

We can then write the equilibrium constants k_H, k_1, and k_2 as:

$$k_H = \frac{\left[CO_{2(g)}\right]}{\left[CO_{2(aq)}\right]} \tag{A.6}$$

$$k_1 = \frac{\left[H^+\right]\left[HCO_3^-\right]}{\left[H_2CO_{3(aq)}\right]} \tag{A.7}$$

$$k_2 = \frac{\left[H^+\right]\left[CO_3^{-2}\right]}{\left[HCO_3^-\right]} \tag{A.8}$$

Here, k_H is the Henry's Law constant and is unitless, and k_1 and k_2 are the first and second dissociation constants of carbonic acid, and have units mol/kg. All three equilibrium constants are slightly temperature dependent. One could choose to omit their temperature dependence without significantly changing the model output of atmospheric CO_2, but their explicit dependence will be presented here for completeness. If one were looking to use a constant value of k_H, k_1, and k_2, they could be calculated using $T = 285$ K.

$$k_H(T) = \frac{1}{k_0} \cdot \frac{1\text{ L}}{1.027\text{ kg}} \cdot \frac{55.57\text{ mol}}{1\text{ L}} \tag{A.9}$$

where k_0 is an explicit, empirically derived function of temperature T:

$$k_0(T) = \exp\left[\frac{9345.17}{T} - 60.2409 + 23.3585 \cdot \ln\left(\frac{T}{100}\right) + S \cdot \left(0.023517 - 0.00023656 \cdot T + 0.0047036 \cdot \left(\frac{T}{100}\right)^2\right)\right] \tag{A.10}$$

$$k_1(T) = 10^{-pK_1} \tag{A.11}$$

$$k_2(T) = 10^{-pK_2} \tag{A.12}$$

where pK_1 and pK_2 are unitless, empirically derived functions of temperature T:

$$pK_1(T) = -13.721 + 0.031334 \cdot T + \frac{3235.76}{T} + 1.3 \cdot 10^{-5} \cdot S \cdot T - 0.1031 \cdot \sqrt{S} \tag{A.13}$$

$$pK_2(T) = 5371.96 + 1.761221 \cdot T + 0.22913 \cdot S + 18.3802 \cdot \log_{10}(S) - \frac{128375.28}{T} - 2194.30 \cdot \log_{10}(T) - 8.0944 \cdot 10^{-4} \cdot S \cdot T - 5617.11 \cdot \frac{\log_{10}(S)}{T} + 2.136 \cdot \frac{S}{T} \tag{A.14}$$

Here, S is the salinity:

$$S \sim 35 \frac{\text{g}}{\text{kg}} \text{ seawater}$$

A.2 The DICE temperature model for computing T dependence of equilibrium constants

For each of the temperature-dependent equations in section A.1, the water temperature T is in units of Kelvin, and is assumed to follow the global atmospheric temperature anomaly [2]. Assuming a globally averaged pre-industrial ocean temperature of 283 K, we can write T as it appears in Equations A.9 – A.14 as:

$$T_i = 283 \text{ K} + \left(T_i^{at} - T_1^{at}\right) \tag{A.15}$$

The global atmospheric temperature anomaly $(T_i^{at} - T_1^{at})$ (over the time interval between the discrete time step i and the initial condition) associated with a change in the atmospheric carbon content can be reasonably well described by the 2007 DICE temperature model [13]. This particular temperature model is used because

of its computational simplicity, but other temperature models could be substituted to calculate the temperature dependence of k_H, k_1, and k_2. The DICE temperature model is written:

$$T_i^{at} = T_{i-1}^{at} + \mu_{at} \cdot \left[\Lambda \cdot \left(T_i^{eq} - T_{i-1}^{at}\right) - \gamma \cdot \left(T_{i-1}^{at} - T_{i-1}^{lo}\right)\right] \tag{A.16}$$

$$T_i^{lo} = T_{i-1}^{lo} + \mu_{lo} \cdot \left[\gamma \cdot \left(T_{i-1}^{at} - T_{i-1}^{lo}\right)\right] \tag{A.17}$$

where T_i^{eq} is the equilibrium atmospheric temperature that the Earth would reach if the radiative forcing at time step i was allowed to relax to zero via heating the atmosphere to increase Earth's outgoing longwave radiation, with no further radiative forcing after time step i:

$$T_i^{eq} = \frac{F_i}{\Lambda} \tag{A.18}$$

where F_i is the radiative forcing away from equilibrium due to CO_2 (in W/m^2):

$$F_i = \eta \cdot \log_2 \left(\frac{M_i^{at}}{M_{PI}^{at}}\right) \tag{A.19}$$

where M_{PI}^{at} is the preindustrial mass of atmospheric CO_2, 596.4 Gton C ($\sim$280 ppm) and η is the radiative forcing that results from a doubling in the CO_2 concentration (assumed to be constant). Glotter et. al. [9] use α to denote this parameter, but we instead use η here since α is already taken.

$$\eta = 3.8 \frac{\mathrm{W}}{\mathrm{m}^2 \cdot 2\mathrm{xCO}_2}$$

and where Λ is Earth's climate sensitivity:

$$\Lambda = 1.3 \frac{\mathrm{W}}{\mathrm{m}^2 \cdot \mathrm{K}}$$

The parameter γ sets the rate of heat anomaly transfer between the atmosphere and lower ocean:

$$\gamma = 0.3 \frac{\mathrm{W}}{\mathrm{m}^2 \cdot \mathrm{K}}$$

μ_{at} and μ_{lo} are defined as the inverse characteristic timescales of the temperature anomaly erosion:

$$\mu_{at} = 0.059 \frac{\mathrm{m}^2}{\mathrm{W}}$$
$$\mu_{lo} = 0.018 \frac{\mathrm{m}^2}{\mathrm{W}}$$

Note that T^{lo} in only used to drive the temperature evolution of the atmosphere, and is not meant to parameterize the true temperature of the "lower" ocean.

A.3 Parameterization of α

Now that we are able to explicitly calculate our equilibrium constants k_H, k_1, and k_2 even as the temperature changes due to atmospheric CO_2 fluctuations, we now want to turn our attention back to Equations A.6, A.7, and A.8 in order to produce the final parameterization of α.

The goal of the parameter α is to scale the mass of the atmosphere reservoir M_{at} to the mass of the upper ocean reservoir M_{up} by size and by chemical activity. Thus, we are looking for the ratio of total atmospheric carbon to the total dissolved inorganic carbon (DIC), scaled by the difference in size between the two reservoirs. We will call this size scaling factor δ, and assign it a value soon. We can now write our first edition of the parameter α:

$$\alpha = \frac{\left[CO_{2(g)}\right]}{[DIC]} \cdot \delta \tag{A.20}$$

Following Zeebe and Wolf-Gladrow [19], the total DIC concentration can be written:

$$[DIC] = \sum_{aq} CO_2 = \left[H_2CO_{3(aq)}\right] + \left[HCO_3^-\right] + \left[CO_3^{-2}\right] \tag{A.21}$$

which we can rewrite using the equilibrium constant expressions A.7 and A.8:

$$[DIC] = \left[H_2CO_{3(aq)}\right] \cdot \left(1 + \frac{k_1}{[H^+]} + \frac{k_1 k_2}{[H^+]^2}\right) \tag{A.22}$$

We can then substitute our expression for $[DIC]$ into our first edition of the parameter α:

$$\alpha = \frac{\left[CO_{2(g)}\right]}{\left[H_2CO_{3(aq)}\right] \cdot \left(1 + \frac{k_1}{[H^+]} + \frac{k_1 k_2}{[H^+]^2}\right)} \cdot \delta \tag{A.23}$$

Using the Henry's Law constant expression A.6 and the equivalence due to fast equilibration in Equation A.5, we can finally write our final parameterization of α (as a function of temperature and pH), as:

$$\boxed{\alpha = \frac{k_H \cdot \delta}{\left(1 + \frac{k_1}{[H^+]} + \frac{k_1 k_2}{[H^+]^2}\right)}} \tag{A.24}$$

In this expression, the unitless size ratio parameter δ can be defined:

$$\delta = \frac{N_{at}}{N_{oc}} \cdot \left(\frac{1}{\beta} + 1\right) \tag{A.25}$$

where $N_{at} = 1.77 \times 10^{20}$ mol is the total number of moles of air in the atmosphere and $N_{oc} = 7.8 \times 10^{22}$ mol is the total number of moles of water in the ocean (references within [9]), and β is the size ratio of the upper and lower ocean reservoirs, as defined in Chapter 1. Multiplication by $\left(\frac{1}{\beta} + 1\right)$ is equivalent to dividing the number of moles of water N_{oc} between the two ocean reservoirs, and taking just the fraction that exist in the upper ocean reservoir. For example, if $\beta = 1/50$, then $N_{oc}/(50+1)$ is the number of moles of water in the upper 1/51st of the ocean.

Note that we now have all the information we need to compute α except for the H^+ concentration at each time step, which we will discuss in the next section.

A.4 Calculating ocean pH

We lastly need a method for calculating $[H^+]$. To do so, we will introduce the concept of carbonate alkalinity, the definition of which can be simplified to represent the charge-neutralizing capability of carbonate species in an aqueous solution. In this application, we can ignore the aqueous charged species that generally contribute to the total alkalinity, and consider only the single-charge neutralizing

bicarbonate ion and the two-charge neutralizing carbonate ion to approximate:

$$Alk \sim \left[HCO_3^-\right] + 2\left[CO_3^{-2}\right] \tag{A.26}$$

Using Equations A.7 and A.8 we can rewrite the expression for alkalinity as:

$$Alk = \left(\frac{k_1}{[H^+]} + \frac{2k_1k_2}{[H^+]^2}\right) \cdot \left[H_2CO_{3(aq)}\right] \tag{A.27}$$

Following Archer et. al. [3], and Zeebe and Wolf-Gladrow [19], we can reasonably assume constant alkalinity on the timescale of $< 10,000$ years. See Archer et. al. [1] for more detail on the counteracting chemical feedbacks that would make this assumption break down.

We can define a new ratio:

$$\frac{Alk}{[DIC]} = \frac{\left[HCO_3^-\right] + 2\left[CO_3^{-2}\right]}{\left[H_2CO_{3(aq)}\right] + \left[HCO_3^-\right] + \left[CO_3^{-2}\right]} \tag{A.28}$$

which, considering the equilibrium relationships in Equations A.7 and A.8, gives us:

$$\begin{aligned}\frac{Alk}{[DIC]} &= \frac{\left(\frac{k_1}{[H^+]} + \frac{2k_1k_2}{[H^+]^2}\right) \cdot \left[H_2CO_{3(aq)}\right]}{\left[H_2CO_{3(aq)}\right] \cdot \left(1 + \frac{k_1}{[H^+]} + \frac{k_1k_2}{[H^+]^2}\right)} \\ &= \frac{\left(\frac{k_1}{[H^+]} + \frac{2k_1k_2}{[H^+]^2}\right)}{\left(1 + \frac{k_1}{[H^+]} + \frac{k_1k_2}{[H^+]^2}\right)}\end{aligned} \tag{A.29}$$

Since $[DIC]$ and Alk are both given by the same set of variables, we have now arrived at an expression that is a function of $[DIC]$ and $[H^+]$ only (given k_1 and k_2 at the same fixed temperature).

A.5 Calculation of upper ocean alkalinity

We lastly need a value for Alk, which can be derived in units of Gton C in the upper ocean reservoir by taking $[DIC] = M_{up}$ (and converting to standardized units). In order to calculate M_{up}, we need an M_{at} and a corresponding α, and then we can use Equation 1.43 to determine the M_{up} that would be in equilibrium

with the atmosphere M_{at}. Since we have assumed that our alkalinity is constant on the timescale that this model is designed for, we can use pre-industrial conditions to calculate the alkalinity. Assuming a pre-industrial ocean pH value of 8.29, we can calculate $[H^+]$ in units of mol/kg:

$$\left[H^+\right] = -\log_{10}\left(pH\right) \cdot \left(\frac{1 \text{ L}}{1.027 \text{ kg}}\right) \tag{A.30}$$

which we can then use to calculate α using Equation A.24. Knowing $M_{at}^{PI} \sim 596.4$ Gton C ($\sim$ 280 ppm CO_2) and α, we then calculate $M_{up} = 732.7$ Gton C using Equation 1.43. The alkalinity, which will be taken as a constant in the model, can then be calculated using Equation A.29:

$$Alk = M_{up} \frac{\frac{k_1}{[H^+]} + \frac{2k_1k_2}{[H^+]^2}}{\left(1 + \frac{k_1}{[H^+]} + \frac{k_1k_2}{[H^+]^2}\right)} = 790.3 \text{ Gton C}$$

This calculation serves as a correction to a mistake made in the assumed alkalinity of Glotter et. al. [9]. At zero temperature anomaly, the equilibrium coefficients used in this calculation are $k_1 = 7.98 \times 10^{-7}$ mol/kg and $k_2 = 4.63 \times 10^{-10}$ mol/kg, and the acidity $[H^+] = 4.99 \times 10^{-9}$ mol/kg (equivalent to pH= 8.29, assuming a seawater density of 1.027 kg/L).

Finally, with all of our constants known and temperature-dependent terms parameterized, we can calculate the $[H^+]_i$ at each time step by solving the quadratic:

$$\frac{Alk}{(M_{up})_{i-1}} = \frac{\frac{k_1}{[H^+]_i} + \frac{2k_1k_2}{[H^+]_i^2}}{\left(1 + \frac{k_1}{[H^+]_i} + \frac{k_1k_2}{[H^+]_i^2}\right)} \tag{A.31}$$

and then use that $[H^+]_i$ to determine α_i via Equation A.24, which allows us to fully know the transfer matrix $\mathbf{K}_i$ in Equation 1.33 that is used to drive the model forward.

Bibliography

[1] Archer D., H. Kheshgi, E. Maier-Reimer. (1998). Dynamics of Fossil Fuel CO_2 Neutralization by Marine $CaCO_3$. *Global Biogeochemical Cycles*, 12(2): 259–276.

[2] Archer, D., P. Martin, B. Buffett, V. Brovkin, S. Rahmstorf, A. Ganopolski. (2004). The Importance of Ocean Temperature to Global Biogeochemistry. *Earth and Planetary Science Letters*, 222(2): 333–348

[3] Archer D., M. Eby, V. Brovkin, A. Ridgwell, L. Cao, U. Mikolajewicz, K. Caldeira, K. Matsumoto, G. Munhoven, A. Montenegro, K. Tokos. (2009). Atmospheric Lifetime of Fossil Fuel Carbon Dioxide. *Annual Review of Earth and Planetary Sciences*, 37: 117–134.

[4] Blasing, T.J., C.T. Broniak, G. Marland. (2005). The Annual Cycle of Fossil-Fuel Carbon Dioxide Emissions in the United States. *Tellus*, 57(B): 107-115.

[5] Boden, T.A., G. Marland, R.J. Andres. (2010). Global, Regional, and National Fossil-Fuel CO_2 Emissions. *Carbon Dioxide Information Analysis Center*, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge.

[6] Dutkiewicz, S., M.J. Follows, P. Heimbach, J. Marshall. (2006). Controls on Ocean Productivity and Air-Sea Carbon Flux: An Adjoint Model Sensitivity Study. *Geophysical Research Letters*, 33, L02603.

[7] Errico, R.M. (1997). What is an Adjoint Model? *Bulletin of the American Meteorological Society*, 78(11): 2577–2592.

[8] Giering, R., T. Kaminski. (1998). Recipes for Adjoint Code Construction. *ACM Transactions on Mathematical Software*, 24(4): 437–474.

[9] Glotter, M.J., R.T. Pierrehumbert, J.W. Elliott, N.J. Matteson, E.J. Moyer. (2014). A Simple Carbon Cycle Representation for Economic and Policy Analyses. *Climatic Change*, 126: 319–335.

[10] Keeling, C.D., S.C. Piper, R.B. Bacastow, M. Wahlen, T.P. Whorf, M. Heimann, and H.A. Meijer. (2001). Exchanges of Atmospheric CO_2 and $^{13}CO_2$ with the Terrestrial Biosphere and Oceans from 1978 to 2000. *I. Global aspects, SIO Reference Series*, No. 01-06, Scripps Institution of Oceanography, San Diego, 88 pages.

[11] MacAyeal, D.R., J. Firestone, E. Waddington. (1991). Paleothermometry by Control Methods. *Journal of Glaciology*, 37(127): 326–338.

[12] Nordhaus, W.D. (1992). The "DICE" Model: Background and Structure Of a Dynamic Integrated Climate-Economy Model of the Economics of Global Warming. *Cowles Foundation Discussion Paper No. 1009*, Cowles Foundation for Research in Economics, Yale University.

[13] Nordhaus, W.D. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies.* Yale University Press.

[14] Randerson, J.T., M.V. Thompson, T.J. Conway, I.Y. Fung, C.B. Fiel. (1997). The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Global Biogeochemical Cycles*, 11(4): 535–560.

[15] Rotty, R.M. (1987). Estimates of Seasonal Variation in Fossil Fuel CO_2 Emissions. *Tellus*, 39B: 184–202.

[16] Sandu, A., D.N. Daescu, G.R. Carmichael, T. Chai. (2004). Adjoint Sensitivity Analysis of Regional Air Quality Models. *Journal of Computational Physics*, 204(2005): 222–252.

[17] Wunsch, C. (1988). Transient Tracers as a Problem in Control Theory. *Journal of Geophysical Research*, 93(C7): 8099–8110.

[18] Yumimoto, K., I. Uno, N. Sugimoto, A. Shimizu, Z. Liu, D. M. Winker. (2008). Adjoint Inversion Modeling of Asian Dust Emission using LIDAR Observations. *Atmospheric Chemistry and Physics*, 8: 2869–2884.

[19] Zeebe R., D. Wolf-Gladrow. (2001). *CO_2 in Seawater: Equilibrium, Kinetics, Isotopes*. Elsevier Oceanography Series, 65.