LEARNING IN INTEGRATED OPTIMIZATION MODELS OF CLIMATE CHANGE AND ECONOMY

A Thesis
Presented to
The Academic Faculty

by

Soheil Shayegh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2014

LEARNING IN INTEGRATED OPTIMIZATION MODELS OF CLIMATE CHANGE AND ECONOMY

Approved by:

Dr. Valerie Thomas, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Hayriye Ayhan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Alexander Shapiro
School of School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Roshan Joseph Vengazhiyil
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Athanasios Nenes
School of Earth & Atmospheric Sciences
Georgia Institute of Technology

Date Approved: June 13, 2014

[To God]

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my adviser Dr. Valerie Thomas for providing endless support and encouragement throughout my PhD studies. This thesis would not have been accomplished without her constructive mentorship, guidance and critique. I would like to extend my deepest appreciation to my PhD committee, Dr. Hayriye Ayhan, Dr. Alexander Shapiro, and Dr. Roshan Joseph Vengazhiyil from the School of Industrial & Systems Engineering, as well as Dr. Athanasios Nenes from the School of Earth & Atmospheric Sciences at Georgia Institute of Technology for providing advice and intellectual insight into the research question.

I would like to thank all my professors at the Georgia Institute of Technology for what I learned from them especially Dr. Seymour Goodman from the School of International affairs, Dr. William Foster from Emory University, Dr. Juan Moreno-Cruz from the School of Economics, and Dr. Baabak Ashuri from the School of Building Construction for their support and consultation. I would especially like to thank all ISyE staff especially Ms. Pam Morrison and Mr. Mark Reese.

Studying a PhD degree at Georgia Tech was one of the highlights of my life and it would not be accomplished without having great support from my friends. I would like to thank all my friends especially Nathan Mayercsik, Mallory Soldner, Catharine and Dan Writh, Bahar Cavdar, Isil Alev, Alireza Khoshgoftar Monfared, Nassir Mokkaram, Nazanin Masoodzadehgan, Shafi and Adria Motiwalla, and Eric Tollefson.

Finally, I wish to thank my caring, loving, and supportive parents, Mehran and Masoud, and brothers, Nima and Shakib, and Jiji for their big hearts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Decision Variables:

Variable	Description	Unit
$a(t)$	Emission Control Rate	Percent/decade

Auxiliary Variables:

W	present value of social welfare (objective function)	utility units
$R(t)$	social time preference discount factor	percent/decade
$U(t)$	social utility	utility units
$u(t)$	utility per capita	utility
$C(t)$	consumption	$ trillion
$c(t)$	consumption per capita	$ million / capita
$Y(t)$	gross economic output (GDP)	$ trillion
$\Omega(t)$	climate change damage cost factor	percentage
$\Lambda(t)$	abatement cost factor	percentage
$Q(t)$	net economic output	$ trillion
$I(t)$	investment	$ trillion
$A(t)$	total factor productivity (TFP)	productivity unit
$A_g(t)$	TFP growth rate	percent/decade
$L(t)$	population (and labor)	millions of people
$L_g(t)$	population growth rate	percent/decade
$K(t)$	capital stock	$ trillion
$T_{at}(t)$	atmospheric temperature	° C above preindustrial
$T_{lo}(t)$	lower ocean temperature	° C above preindustrial
$F(t)$	total increase in radiative forcing since preindustrial	Watt/m^2
$M_{at}(t)$	mass of carbon in atmosphere	GTC
$M_{up}(t)$	mass of carbon in upper oceans	GTC
$M_{lo}(t)$	mass of carbon in lower oceans	GTC
$E(t)$	total carbon emission	GTC/year
$E_{ind}(t)$	industrial carbon emission	GTC/year
$E_{land}(t)$	carbon emission from land use	GTC/year
$\sigma(t)$	ratio of uncontrolled industrial emissions to output	MTC/$ 1000

$\sigma_g(t)$	growth rate of σ	percent/decade
$\pi(t)$	participation cost markup	Percentage
$\varphi(t)$	participation rate	Percentage
$\theta_1(t)$	abatement cost function coefficient	Dimensionless
$F_{ex}(t)$	exogenous forcing	Watt/m^2

Parameters:

t	time	decade
ρ	pure rate of social time preference	1/year
α	elasticity of marginal utility of consumption	dimensionless
β	elasticity of output with respect to capital	dimensionless
Ψ_1	damage coefficient on temperature	dimensionless
Ψ_2	damage coefficient on temperature squared	dimensionless
Ψ_3	exponent on damage	dimensionless
ξ_1	speed of adjustment	dimensionless
ξ_2	CO_2 doubling coefficient	dimensionless
ξ_3	coefficient of heat loss from atmosphere to oceans	dimensionless
ξ_4	coefficient of heat gain by deep oceans	dimensionless
A_{gd}	rate of decline in TFP growth rate	percent/year
L_A	asymptotic population	millions of people
θ_2	exponent of control cost function	dimensionless
ϕ_{11}	carbon cycle transition coefficients atmosphere to atmosphere	percent/decade
ϕ_{12}	carbon cycle transition coefficients atmosphere to biosphere	percent/decade
ϕ_{21}	carbon cycle transition coefficients biosphere to atmosphere	percent/decade
ϕ_{22}	carbon cycle transition coefficients biosphere to biosphere	percent/decade
ϕ_{23}	carbon cycle transition coefficients biosphere to deep oceans	percent/decade
ϕ_{32}	carbon cycle transition coefficients deep oceans to biosphere	percent/decade
ϕ_{33}	carbon cycle transition coefficients deep oceans to deep oceans	percent/decade
σ_{ga}	rate of increase in the growth rate of σ	percent/decade
σ_{gd}	rate of decrease in the growth rate of σ	percent/decade
P_b	cost of backstop 2005	1000$/tC
P_r	ratio initial to final backstop cost	dimensionless
P_d	initial cost decline backstop	percent/decade
ω_1	marginal retention rate of emissions in the atmosphere	dimensionless

ω_2	rate of transfer to the deep oceans	dimensionless

Initial Settings:

A_0	initial TFP	Productivity Unit
A_{g0}	initial TFP growth rate	Percent/decade
L_0	initial population	Millions of people
L_{g0}	initial population growth rate	Percent/decade
K_0	initial capital stock	$ trillion
Y_0	initial output	$ trillion
E_0	Initial carbon emissions from land use change	GTC per year
σ_0	initial σ	percent
σ_{g0}	initial growth rate of σ	percent/decade
F_{2000}	estimate of 2000 forcings of non-CO_2 GHG	Watt/m^2
F_{2100}	estimate of 2100 forcings of non-CO_2 GHG	Watt/m^2
t_{max}	terminal time	year
T_0^{at}	initial atmospheric temperature	° C
T_0^{lo}	initial temperature of deep oceans	° C
ΔR_f	estimated forcing of equilibrium CO_2 doubling	° C/(W/m^2)
ΔT	equilibrium temperature increase for CO_2 doubling	° C/(W/m^2)
M_0^{at}	quantity of carbon in atmosphere prior to industrialization	GTC
M_0^{lo}	quantity of carbon in upper ocean prior to industrialization	GTC
M_0^{up}	quantity of carbon in lower ocean prior to industrialization	GTC

SUMMARY

Integrated assessment models are powerful tools for providing insight into the interaction between the economy and climate change over a long time horizon. However, knowledge of climate parameters and their behavior under extreme circumstances of global warming is still an active area of research. In this thesis we incorporated the uncertainty in one of the key parameters of climate change, climate sensitivity, into an integrated assessment model and showed how this affects the choice of optimal policies and actions. We constructed a new, multi-step-ahead approximate dynamic programing (ADP) algorithm to study the effects of the stochastic nature of climate parameters. We considered the effect of stochastic extreme events in climate change (tipping points) with large economic loss. The risk of an extreme event drives tougher GHG reduction actions in the near term. On the other hand, the optimal policies in post-tipping point stages are similar to or below the deterministic optimal policies. Once the tipping point occurs, the ensuing optimal actions tend toward more moderate policies. Previous studies have shown the impacts of economic and climate shocks on the optimal abatement policies but did not address the correlation among uncertain parameters. With uncertain climate sensitivity, the risk of extreme events is linked to the variations in climate sensitivity distribution. We developed a novel Bayesian framework to endogenously interrelate the two stochastic parameters. The results in this case are clustered around the pre-tipping point optimal policies of the deterministic climate sensitivity model. Tougher actions are more frequent as there is more uncertainty in likelihood of extreme events in the near future. This affects the optimal policies in post-tipping point states as well, as they tend to

utilize more conservative actions. As we proceed in time toward the future, the (binary) status of the climate will be observed and the prior distribution of the climate sensitivity parameter will be updated. The cost and climate tradeoffs of new technologies are key to decisions in climate policy. Here we focus on electricity generation industry and contrast the extremes in electricity generation choices: making choices on new generation facilities based on cost only and in the absence of any climate policy, versus making choices based on climate impacts only regardless of the generation costs. Taking the expected drop in cost as experience grows into account when selecting the portfolio of generation, on a pure cost-minimization basis, renewable technologies displace coal and natural gas within two decades even when climate damage is not considered in the choice of technologies. This is the natural gas as a bridge fuel scenario, and technology advancement to bring down the cost of renewables requires some commitment to renewables generation in the near term. Adopting the objective of minimizing climate damage, essentially moving immediately to low greenhouse gas generation technologies, results in faster cost reduction of new technologies and may result in different technologies becoming dominant in global electricity generation. Thus today's choices for new electricity generation by individual countries and utilities have implications not only for their direct costs and the global climate, but also for the future costs and availability of emerging electricity generation options.

CHAPTER 1

INTRODUCTION

Our understanding of natural systems, especially those with dynamics, is evolving over time. This is in addition to the fact that many of such systems possess inherited uncertainty in their behavior. Therefore, we often face at least two types of uncertainty: (1) parametric uncertainty which can be resolved over time via more observations and active learning, and (2) stochasticity that is an inherent characteristic of the system under study. In designing a mathematical model of a natural system, various types of uncertainties need to be taken into account [1], [2]. The process of acquiring knowledge and information about uncertainties is called "learning". In general, two types of learning are considered in different contexts of engineering modeling: passive and active. While in passive learning the decision making agent is simply the recipient of information and his decisions will not change the environment, in active learning the learner agent has the ability to act, to gather data, and to influence the world [3].

One of the natural systems that has received significant attention in recent years is the global climate system. Climate change dynamics combined with its socio-economic aspects present a complex challenge for policy making under uncertainty from environment and economy [4]. Integrated assessment models (IAMs) provide insight into the interaction between the economy and the climate. In general, these models are trying either to evaluate a specific policy and its long term effects on greenhouse gas (GHG) emissions and economic growth or to find the best policy among several options, that will provide the most social benefit (welfare maximization models) or the least cost of achieving particular goals (cost minimizing models). Stanton et al. have studied and classified over thirty IAMs in four key areas [5], [6]:

- Choice of model structure and the type of results produced

- Uncertainty in climate outcomes and the projection of future damages
- Equity across time and space
- Abatement costs and the endogeneity of technological change

Some IAMs are designed to solve the optimal path problem and give the best climate policy over a long but finite time horizon. However, there is uncertainty around estimation of some key parameters in any climate model, which makes the process of finding optimal policies stochastic.

Dynamic Integrated Model of Climate and Economy (DICE)

In this study we use the Dynamic Integrated Model of Climate and Economy (DICE 2007) as a reference framework for studying the impacts of uncertainty in some climate parameters [7]. It follows the standard Ramsey-Cass-Koopmans model structure to include GHG dynamics [8]. It models the world in a fairly transparent way by integrating economic inputs (capital, labor), climate change associated costs (abatement cost and damage costs), and policy options (carbon tax rate, saving rate) into a general macroeconomic model. It indicates the economic impact from global warming as a percentage of annual gross domestic product (GDP). It also assumes this percentage to be an exponential function of the global mean temperature. Due to its simple structure and interesting results, DICE 2007 has been widely considered to be a standard framework for modeling the climate economy. The model consists of two main modules: (1) the Standard Economic Growth Module, and (2) the Climate Change Module. The two main modules are developed separately and linked through the transient module, which includes the functions for climate damage and abatement cost. Appendix A provides a complete list of equations used in the original DICE 2007 model and shows the relationship between different modules.

At each time step t, an abatement action $a(t)$ defines what percentage of GHG emissions are being removed from the atmosphere. Taking any nonzero action is

nevertheless costly and will affect the total economic output (i.e. GDP) for the next decision epoch $t + 1$. On the other hand, doing nothing (taking the action zero) will let the GHG emissions rise and consequently increase the global atmospheric temperature. In the DICE 2007 model an increase in global atmospheric temperature is translated to a monetary loss through a damage function. The objective is to maximize social welfare over time, which is a function of total economic output in each time step. Therefore, the DICE 2007 model seeks an optimal policy that balances the total economic damage due to climate change and the costs of taking abatement measures.

As many researchers have pointed out, there are several assumptions in the model which may restrict its application and undermine its ability to prescribe the optimal climate policy in the long run. Therefore, there are a growing number of studies on how to modify different modules of the DICE model. We review these modifications in three categories: the climate module, the economic module, and the modeling assumptions.

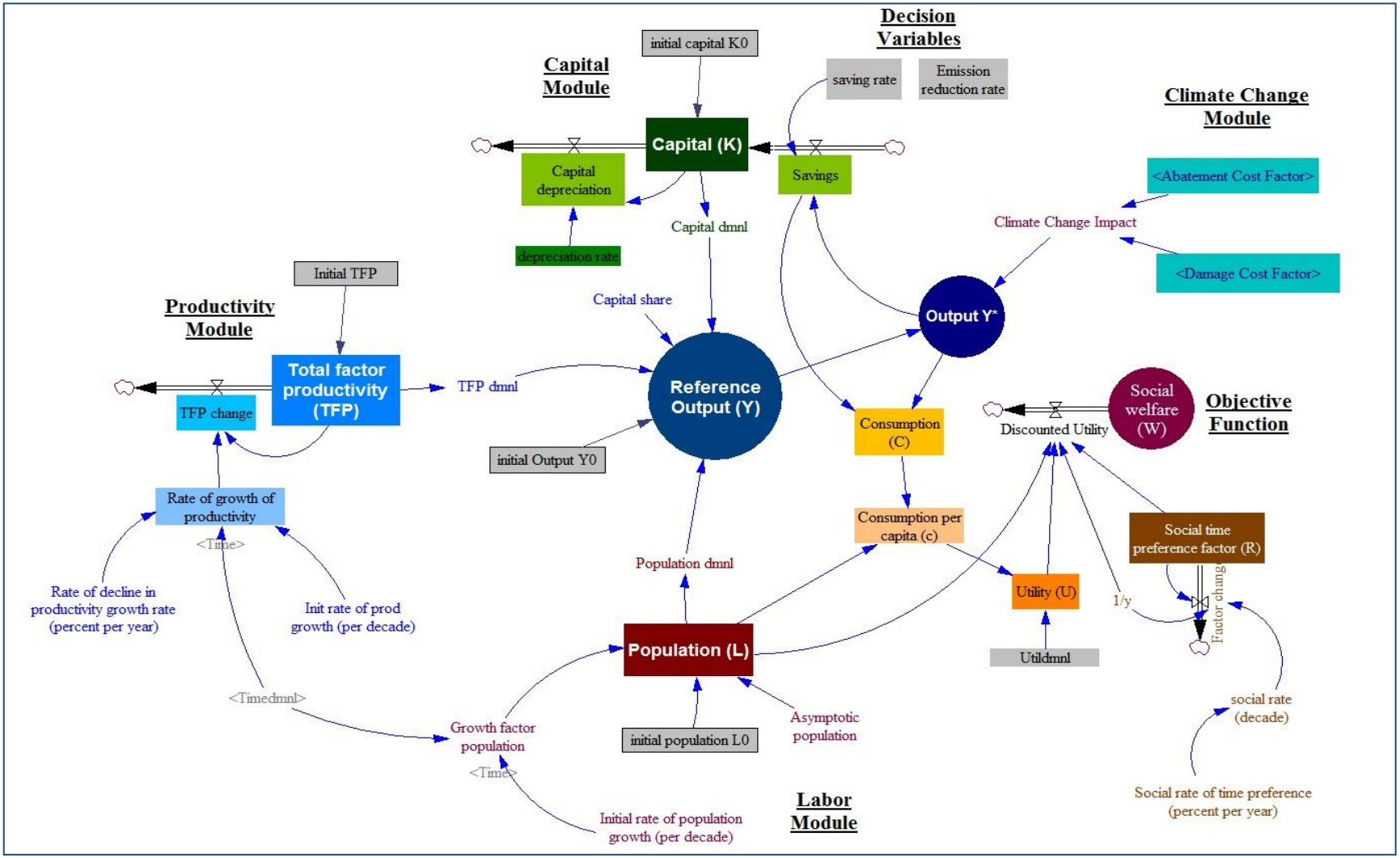


Figure 1: A general view of DICE 2007 model as a dynamic system with stock and flow variables

Modification of the Economic Module

Some studies have proposed modifications of the economic module in DICE. Introducing endogenous technological change into the DICE model has been the focus of several studies. Popp developed a model called ENTICE, which is a version of DICE with endogenous technological change [9]. He reported that induced innovation improves the objective function under optimal policy. Islam et al. modeled endogenous technical change in DICE to study the effect on abatement cost. Their ADICE model showed the need for policy intervention even with endogenous technical progress [10]. Kosugi et al. merged the DICE model with a life-cycle impact assessment (LCIA) model to find the share of global warming external costs in the 21st century [11]. De Burin et al. developed AD-DICE by introducing an adaptation module to the original DICE model. They concluded that adaptation reduces the potential costs of climate change in earlier periods while mitigation plays a bigger role in later periods [12]. Cai et al. have suggested a different form of the utility function in place of the standard constant relative risk aversion (CRRA) in the DICE model [13].

Modification of the Climate Module

The DICE model, like many other IAMs, uses best estimates for economic and climate parameters. A study showed that social welfare parameters are more important than some climate parameters in assessing sustainable development [14]. However, one of the key parameters in the DICE model, which has received a great deal of attention in the literature, is climate sensitivity (the equilibrium increase in mean global surface temperature due to doubling of atmospheric CO_2 compared to the pre-industrial era). Some economists and climate scientists have argued that the best estimates fail to capture the small probability of climate catastrophes that could have a high irreversible impact on the economy [15]. However, some other studies showed that imposing an upper bound on future temperature change is justified and will not affect the optimal policy [16]. There

are several fat-tailed distribution functions introduced in the literature for the climate sensitivity parameter. Newbold and Daigneault used a Bayesian framework to summarize a large number of climate models to find the posterior distribution for climate sensitivity. They suggested that the posterior distribution function of the climate sensitivity parameter and the shape of the damage function play a crucial role in finding the best control policies [17]. Roe and Baker derived a simple analytical form for the distribution of the climate sensitivity parameter, which fits many published distribution estimates very well [18]. Other well-known fat-tailed distributions such as lognormal [19] and Cauchy [16] have also been considered in the literature. In studies of the impact of a fat-tailed distribution of climate sensitivity on the results of the DICE model, some researchers have reported that the uncertainty in the climate sensitivity parameter alone does not drastically change the optimal policy [19]. Other studies have found that under a given climate policy, using a fat-tailed distribution for the uncertainty in the damage function and climate sensitivity will result in substantially larger economic losses compared to the deterministic case [20].

However, most of the studies on applications of fat-tailed distribution in the DICE model are based on Monte Carlo analysis with sampling from a known distribution, rather than solving the stochastic dynamic problem and finding the optimal control. The result from simulation provides a range of possible outcomes of the model but fails to instruct on best policies to be implemented once a new observation of an unknown parameter is realized. One approach to solve the stochastic problem is to use different damage functions for different levels of the climate sensitivity parameter. In this case, at any given time in future, the climate status can switch from the current status to a catastrophic one with a low probability that positively depends on the surface temperature at that time period [13], [21]. Some studies, including the approach presented in this study, are based on approximate dynamic programming using value function approximation. However, the application of these studies have been limited either by

their choice of the time scale [22] or by their approximation complexity [23]. We introduce a novel but simple approximation technique based on the post-decision state framework presented in [24], [25].

Modeling Assumptions

Beside modifications to different modules of the DICE model, there are some critics who believe that it suffers from inherent deficiencies in its assumptions. Some researchers have argued that the damage function in DICE unrealistically understates the loss due to high atmospheric temperature increases and also have argued that that the mitigation is undervalued in DICE and therefore the combination of soft damage and cheap mitigation is generating a moderate optimal path, famous as the “policy ramp” [8]. Other researchers have questioned the use of large time steps in the model and suggested the use of continuous time instead of the original decadal steps [26].

CHAPTER 2

UNCERTAINTY IN INTEGRATED ASSESSMENT MODELING

As discussed in the previous chapter, integrated assessment models (IAMs) provide insight into the interaction between the economy and the climate [27]. IAM analysis should take into account the uncertainties in the climate system, as well as in the future of the economy. The nature of some of the uncertainties in the climate system suggests that consideration of uncertainties could substantially affect optimal climate policies [28]. Climate model uncertainties have been included in a number of integrated assessment models [29]. However, most previous attempts to include climate sensitivity uncertainty have used Monte Carlo analysis [30]; these studies provide a range of possible outcomes but fail to instruct on the best policies to be implemented once a new observation of an unknown parameter is realized [19], [20]. There have also been a few studies that have taken a dynamic programming approach to fully integrate uncertainties [22], [23]. In all of these models the uncertain parameters are assumed exogenous and independent. However, as in the case of climate and economic systems, uncertain parameters are often correlated and therefore demonstrate an inherent dependency. Here we develop an elegant approximation technique that allows for full integration of uncertainty within integrated assessment models in addition to a unique capability to update correlated uncertainties.

In the next section we introduce the baseline optimization model, its state variables, and the decision variable. Uncertainty in climate parameters of any integrated assessment model will propagate through the modeling time horizon, affecting both economic and climate modules and therefore, have an impact on the objective function.

Baseline Model

We develop and illustrate this technique through application to a well-known integrated assessment model, the Dynamic Integrated Model of Climate and Economy (DICE 2007) [7]. The model consists of two main modules: (1) the Standard Economic Growth Module, and (2) the Climate Change Module. The two main modules are separate and are linked through the transient module, which includes the functions for climate damage and abatement cost [31]. In this model, the state of the world at each time step t, denoted by S_t can be captured by six continuous variables: T_{at} is atmospheric temperature (degrees Celsius above preindustrial), T_{lo} is lower ocean temperature (degrees Celsius above preindustrial), M_{at} is atmospheric concentration of carbon (Gigatons of Carbon, GTC), M_{up} is concentration in biosphere and upper oceans (GTC), M_{lo} is concentration in deep oceans (GTC), and K is capital (trillion USD).

At each time step, an abatement action (control rate) a_t is taken which indicates the percentage reduction of GHG emissions in the next 10 years, compared to the uncontrolled level from the baseline economic emissions without consideration of climate impacts. The atmospheric temperature in the next state is defined by the temperature in the current state as well as the level of abatement taken in the current state. The atmospheric temperature, on the other hand, determines the economic impacts of climate change through an explicit damage function. The goal of the optimization model is to find the best level of abatement action a_t to maximize social welfare, taking into account both the costs and benefits of abatement. The utility U_t at each time step is defined as a constant-elasticity-of-substitution function of the flow of consumption per capita c and the level of population L_t in each state:

$$U_t = \frac{{c_t}^{1-\alpha}}{1-\alpha} \times L_t \quad (1)$$

where α denotes the elasticity of marginal utility of consumption. The consumption per capita c_t is the total production net of savings, climate mitigation, and climate damage. A

social planner decides on the level of carbon reduction at each stage and their decision will have both economic and climate impacts on future stages consequently. The objective function is the discounted accumulation of the social utility over a finite time horizon.

$$\max_{0\le a_t\le 1} \sum_{t=1}^{T} \gamma^t U_t \tag{2}$$

where γ is the social time preference discount factor and T indicates the terminal time horizon.

Introducing Uncertainty in the Integrated Assessment Modeling Framework

Uncertainty in Climate Sensitivity

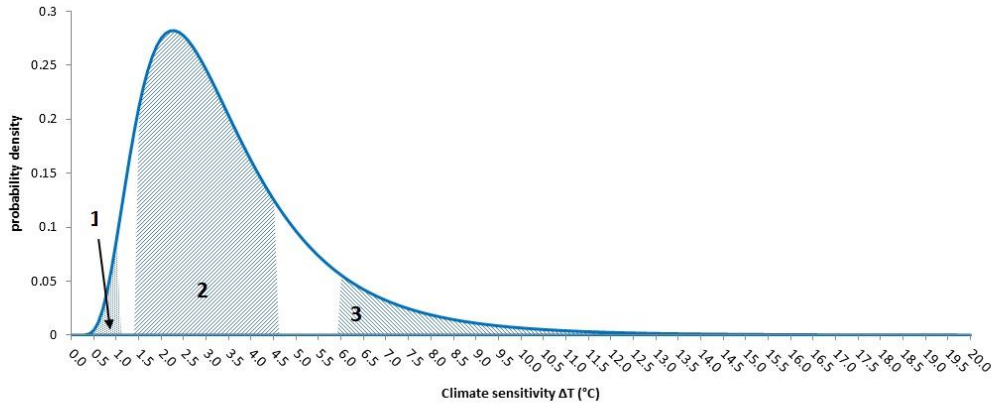
One of the key parameters in integrated assessment modeling which has received a great deal of attention in the literature is climate sensitivity, the equilibrium increase in mean global surface temperature due to doubling of atmospheric CO_2 compared to the pre-industrial era [32]–[34]. Climate sensitivity (ΔT) has been looked upon as the main source of uncertainty in many integrated assessment studies.

In its fifth assessment report, the Intergovernmental Panel on Climate Change has specified the range of the estimated values for climate sensitivity as the following [35]: *"Estimates of the Equilibrium Climate Sensitivity (ECS) based on multiple and partly independent lines of evidence from observed climate change indicate that there is high confidence that ECS is extremely unlikely to be less than* 1℃ *and medium confidence that the ECS is likely to be between* 1.5℃ *and* 4.5℃*and very unlikely greater than* 6℃*."*

In IPCC terminology "likely" observations have a probability of more than 66\%, "very unlikely" events are those with a probability less than 10%, and "extremely unlikely" have a probability less than 5%. This has led many researchers to assume a so-called fat-tailed probability distribution for climate sensitivity [16]–[20]. By definition, a

fat-tailed distribution is one whose upper tail declines to zero more slowly than exponentially [36].

The uncertainty in climate sensitivity is due to the lack of understanding of the physical processes, complexity of the relationships between the components of the climate system, and the chaotic nature of the system [18]. We use a truncated lognormal distribution for capturing uncertainty in the model which assigns zero probability for the values of climate sensitivity larger than 20℃ or less than 0℃. We assume that knowledge of climate sensitivity is represented by a probabilistic lognormal distribution with mean and standard deviation equal to 1.1 and 0.5 respectively. The confidence levels for different regions are shown and compared with IPCC recommended values in Figure 2.



Regions	Probabilities	IPCC 2013 characterization	lognormal distribution
1	$Pr(\Delta T < 1°C)$	less than 5%	2%
2	$Pr(1.5°C < \Delta T < 4.5°C)$	greater than 67%	67%
3	$Pr(\Delta T > 6°C)$	less than 10%	10%

Figure 2: Comparison of truncated lognormal distribution and IPCC characterization [35].

Using climate sensitivity as an uncertain parameter and considering other recursive relationships in the model strongly supports the case for viewing the DICE 2007 model as a Markov Decision Process (MDP) as shown in .

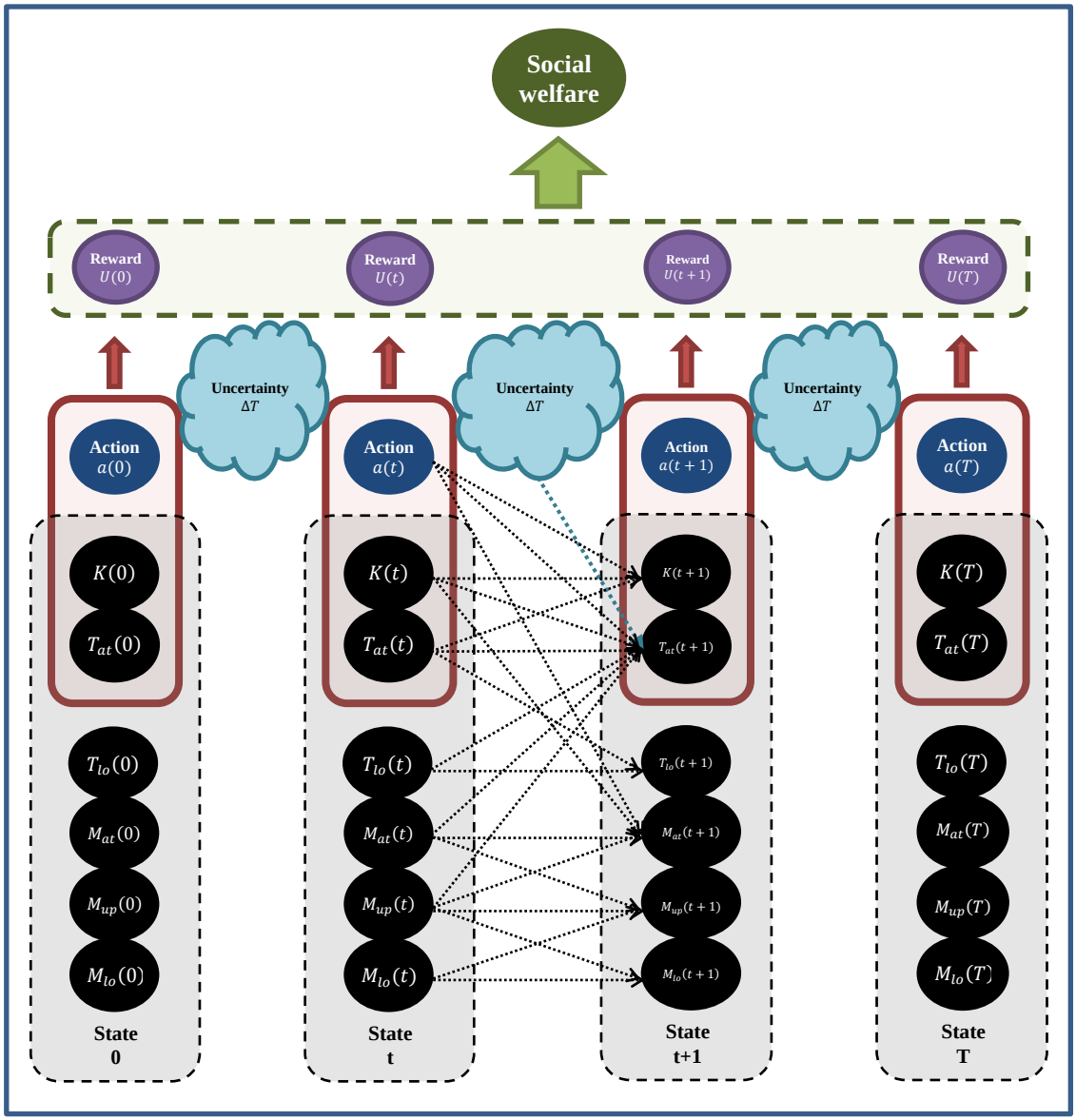


Figure 3: Markov Decision Process for DICE model with uncertain climate sensitivity.

Uncertainty in Extreme Events

Many researchers have discussed the possibility of irreversible outcomes from abrupt climate events in the near future; for a comprehensive review see [37]. The term tipping point has been defined as a state of the climate with strong feedback which triggers a sequence of irreversible catastrophic events, such as thermohaline circulation interruption, massive methane releases, or very rapid sea-level rise. Weitzman argues that the economic consequences of fat-tailed structural uncertainty in climate sensitivity

coupled with uncertain damages from high temperatures dominate the effects of discounting in climate change policy analysis [15]. In other words, such uncertainties call for more aggressive contingency plans for bad outcomes [38]. Since the nature of a tipping point is uncertain, we can only speculate about the probability of a tipping point happening in future. There is, however, agreement among climate scientists that the probability of having a tipping point is related to the mean global temperature [39]. The irreversibility of catastrophic outcomes from a tipping point indicates that the economic damage is asymmetric before and after a tipping point occurs.

The dynamic structure of the DICE 2007 model and interdependencies between its variables make it suitable for applying the Markov Decision Process (MDP) framework. An infinite horizon version of such model has been studied in the literature where the uncertainty is represented by a controlled jump process between two climate modes with different switching probabilities [21]. This model is shown in Figure 4 where $q_{01}(T)$ and $q_{10}(T)$ are transition probabilities between mode 0 (current situation) and mode 1 (climate threshold event), and ω_{01} and $d\omega_{01}$ are non-negative parameters. Modes in this model are influenced by temperature T and therefore the underlying assumption for this model is that global mean temperature increase will accelerate the transition from the current (normal) situation to a threshold event, *i.e.* Climate Extreme [40].

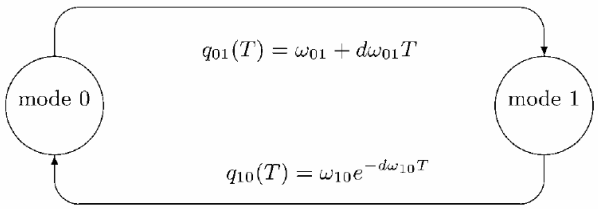


Figure 4: A two state jump process [21]

A more recent work used a Markov process with an absorbing state to represent the irreversibility of tipping point events [13]. We apply a similar idea for our finite horizon model. Climate change damage is explicitly modeled in DICE using a quadratic function of atmospheric temperature as a proxy for all climate related impacts:

$$D = \frac{1}{1 + \Psi_1 T_{at} + \Psi_2 {T_{at}}^2} \quad (3)$$

Here, Ψ_1 and Ψ_2 are constant parameters and D is the consumable portion of economic output after observing the temperature. As temperature T_{at} increases, it will cause more damage to the economy and D falls subsequently. To model the outcome of tipping points we use the following damage function with a stochastic factor as represented by Cai et al. [13]:

$$D = \frac{1 - J}{1 + \Psi_1 T_{at} + \Psi_2 {T_{at}}^2} \quad (4)$$

where J is a discrete Markov chain with non-decreasing values over time. For our analysis we take the benchmark case from Cai et al. and implant a new uncertain parameter to our stochastic model. In the benchmark case the two stage probability transition matrix for J is given by $\begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}$ where p is the probability of transforming from non-catastrophic to a catastrophic status and depends on the atmospheric temperature through this equation:

$$p = 1 - e^{-\upsilon \, max\{0,\ T_{at} - 1\}} \quad (5)$$

where υ is the hazard rate parameter with the initial value $\upsilon = 0.006$. We estimate this as an approximate fit to results of expert surveys [39], [41], [42]. The damage level $J = 0.025$ has been used for the DICE model with annual time steps. To modify for our model which uses a decadal time step we note that the probability of not having a tipping point in a decade is equal to the probability of not having a tipping point in a sequence of ten annual time steps: $1 - p^{decadal} = (1 - p^{annual})^{10}$. If we denote by υ' the decadal hazard rate parameter, we have $\upsilon' = 10\upsilon$. Finding the appropriate damage level for a decadal time step is more challenging. However, following the same logic we can approximate J to be about twenty-five percent. Once the extreme event happens even with a moderate annual $J = 0.025$, it stays in the new state for the rest of the modeling horizon and therefore the decadal damage would be the aggregate annual damage.

Stochastic Model

The first step toward incorporating uncertainty into the model is to try to model the transition from current state of the economy and climate to the next state using a Markov Decision Process (MDP). Fortunately, the finite horizon DICE 2007 can easily be modeled as a deterministic dynamic programming problem where one can find the optimal dynamic programming "policy" in each stage S_t by stepping backward through the time and finding the best action a_t by solving the Bellman equation in the deterministic case:

$$V_t(S_t) = \max_{a_t}\{U_t(S_t, a_t) + \gamma V_{t+1}(S_{t+1}|S_t, a_t)\} \quad (6)$$

where γ is the social time preference discount factor and V_{t+1} shows the value of future state S_{t+1}. By solving Equation (6) recursively, we can in theory obtain the optimal climate policy for any given time horizon. Therefore, we can define a dynamic programming policy as a function, or more precisely a set of tunable parameters of a function, which maps the information in each state of the model to an abatement decision. In other words, the MDP policy π can be defined as:

$$\pi = \{\hbar_t | a_t = \mathcal{M}(\hbar_t, S_t)\} \quad (7)$$

where $\mathcal{M}: \mathbb{S} \to \mathbb{A}$ is a mapping from the state space $\mathbb{S}$ to the action space $\mathbb{A}$ and $\hbar_t$ is a tunable parameter in $\mathcal{M}$, the mapping function. Given a state S_t and dynamic programming policy $\hbar_t$, the decision is determined by $\mathcal{M}(\hbar_t, S_t)$.

The solution to this dynamic programming problem cannot be found exactly, due to the continuity of the state space. The DICE 2007 model was originally solved as a nonlinear optimization problem with 60 unknown actions with a ten-year time interval. The approximation problem becomes harder once we introduce uncertainty to the model. There have been several attempts to address the issue of uncertainty in DICE 2007 and the closest to our work is the DSICE model where two stochastic shocks (economic and

climate) were introduced to the continuous time version of the original DICE 2007 model [23]. To have a meaningful comparison with the DICE 2007 model, we keep the original 60 time steps of the model and limit our study to climate shock only [7]. In the next chapter we introduce an approximation technique that can be applied to this problem.

As mentioned earlier, DICE 2007 has a continuous state space with six dimensions. Under uncertainty from the climate parameters, the general Bellman equation of Equation (6) can be rewritten as:

$$V_t(S_t) = \max_{a_t}\{U_t(S_t, a_t) + \gamma \mathbb{E}\big(V_{t+1}(S_{t+1}|S_t, a_t)\big)\} \tag{8}$$

where $\mathbb{E}(.)$ denotes the expectation that can be simplified if we use a discrete estimate of our continuous probability function:

$$\mathbb{E}\big(V_{t+1}(S_{t+1}|S_t, a_t)\big) = \sum_{s' \in \mathbb{S}} \mathbb{P}(s'|S_t, a_t) V_{t+1}(s') \tag{9}$$

where $\mathbb{P}(.)$ is the discrete probability function. Using approximate dynamic programming (ADP), we estimate the values of $V_{t+1}(s')$ by some parametric function $\bar{V}_{t+1}(s')$. We can broaden the use of the value function approximation by generating random paths and estimating the expected value of the next state as described thoroughly in [25]. In the core of this method lies the concept of a post-decision state variable S_t^a that is a transient state between the current state S_t and the next state S_{t+1}. This state is generated by implementing the chosen action a_t on the current state S_t but before realization of the random parameter ΔT_t. Figure shows the concept of the post-decision state in relation to other modeling variables.

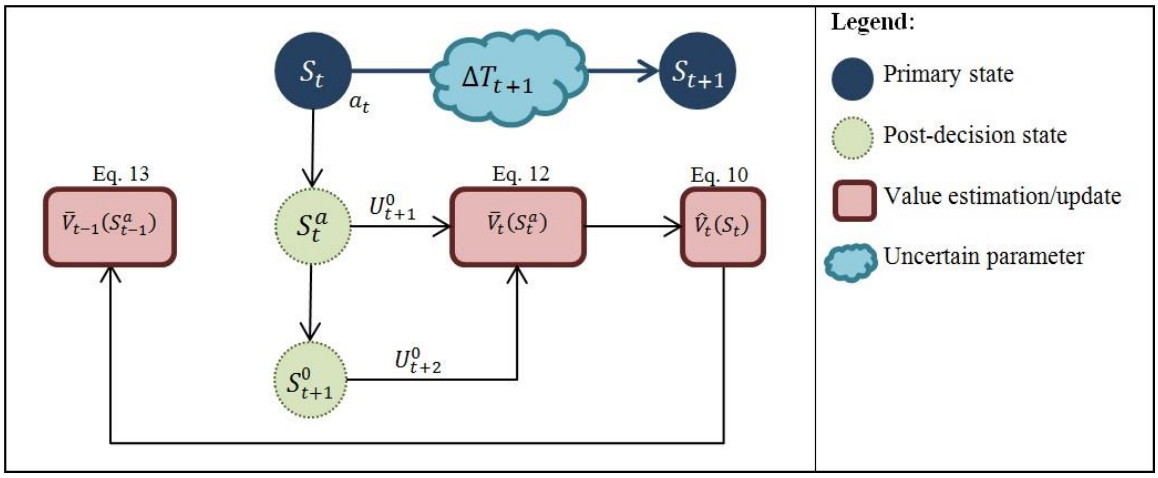


Figure 5: Calculating the optimal value and updating the post-decision state value function: the agent in state $\boldsymbol{S_t}$ takes action $\boldsymbol{a_t}$ without observing the realization of the random process $\boldsymbol{\Delta T_{t+1}}$ and ends up in the post-decision state $\boldsymbol{S^a_t}$. With the immediate reward $\boldsymbol{U^0_{t+1}}$, the agent then takes the predetermined null action 0 to end up in state $\boldsymbol{S^0_{t+1}}$ and from there taking the same null action again, observes the immediate reward $\boldsymbol{U^0_{t+2}}$, which through Equation (18) provides $\bar{\boldsymbol{V}}_{\boldsymbol{t}}(\boldsymbol{S^a_t})$, the approximate value of being in the post decision state $\boldsymbol{S^a_t}$. The value of being in state $\boldsymbol{S_t}$ is shown as $\hat{\boldsymbol{V}}_{\boldsymbol{t}}(\boldsymbol{S^a_t})$ and can be calculated from the Equation (16) using the approximation $\bar{\boldsymbol{V}}_{\boldsymbol{t}}(\boldsymbol{S^a_t})$. This value is used to update the coefficients of $\bar{\boldsymbol{V}}_{\boldsymbol{t-1}}(\boldsymbol{S^a_{t-1}})$, the approximate value of the post-decision state at time $t-1$, for the next iteration through Equation (19).

Estimating the value function in the post-decision state S_t^a provides a significant computational advantage over the strategy adopted in Equation (9) by eliminating the need to calculate the expected value of the next state S_{t+1}. The new equation for calculating the approximate value of state S_t can be expressed as:

$$\hat{V}_t(S_t^a) = \max_{a_t}\{U_t(S_t, a_t) + \gamma \bar{V}_t(S_t^a)\} \tag{10}$$

where $\bar{V}_t(S_t^a)$ is the optimal value of state S_t based on the value approximation of the post-decision state S_t^a. The general ADP algorithm for value iteration is presented in the next chapter.

There are several well-known methods for approximating the value function; for an exhaustive survey see [43]. One widely used method uses a parametric function (basis function) of the state variables to construct the approximate value function. Here we introduce an elegant method to approximate the value function. We draw on the idea behind the deterministic rolling horizon or receding heuristics, to look a finite number of

steps into the future and solve a smaller problem than the original one [44]. We use a linear combination of utility functions from two steps ahead under the deterministic assumption to approximate the value of the current state and solve the stochastic problem iteratively. The detailed discussion of this method is presented in the next chapter. Here we discuss its application in the context of our stochastic problem. In each state S_t and under a candidate action a_t and the predetermined level of climate sensitivity, we first construct the post-decision state S_t^a. To approximate the value function of this transient state we apply a null action $a = 0$ to obtain S_{t+1}^0, the next-post-decision state. If we take a one-step-ahead approximation we can use the immediate reward from taking the null action $a = 0$ at post-decision state S_t^a and find the approximate value function as $\bar{V}_t(S_t^a) = U_{t+1}(S_t^a, 0)$. A more elaborate method is to use a tunable coefficient $\hbar_t$ for the approximation and to update it after each iteration:

$$\bar{V}_t(S_t^a) = \hbar_t U_{t+1}(S_t^a, 0) \tag{11}$$

However, taking only one step forward in this strategy might not give an accurate approximation for the future states. Looking forward into the future and taking into account the utilities of the two next states ahead will capture the tradeoff between a myopic policy to maximize the value of the current state only and a lookahead policy which maximizes the value of current and future states as a single function.

$$\bar{V}_t(S_t^a) = \hbar_t U_{t+2}(S_{t+1}^a, 0) \tag{12}$$

This technique provides a fast and robust solution for both deterministic and stochastic cases. The rate of learning (convergence) in this method depends on the updating scheme. We can employ a simple stochastic gradient algorithm [45] to update this coefficient similar to the one shown in Equation (13).

$$\hbar_t^n = \hbar_t^{n-1} - \alpha_{n-1}\left(\bar{V}^n - \hat{V}^n\right) \tag{13}$$

Application to stochastic modeling of climate tipping points

We incorporate the tipping point event as a new stochastic parameter in the approximate dynamic programming model, using the tipping point formulation and the two-step-ahead algorithm developed in the previous section. To explore the implications of a stochastic tipping point, in this section we assume a deterministic value for climate sensitivity ($\Delta T = 3$℃). Figure 6 shows the resulting tipping point probability over time. The histogram shows the distribution of first occurrence of tipping points after 1000 runs of simulation.

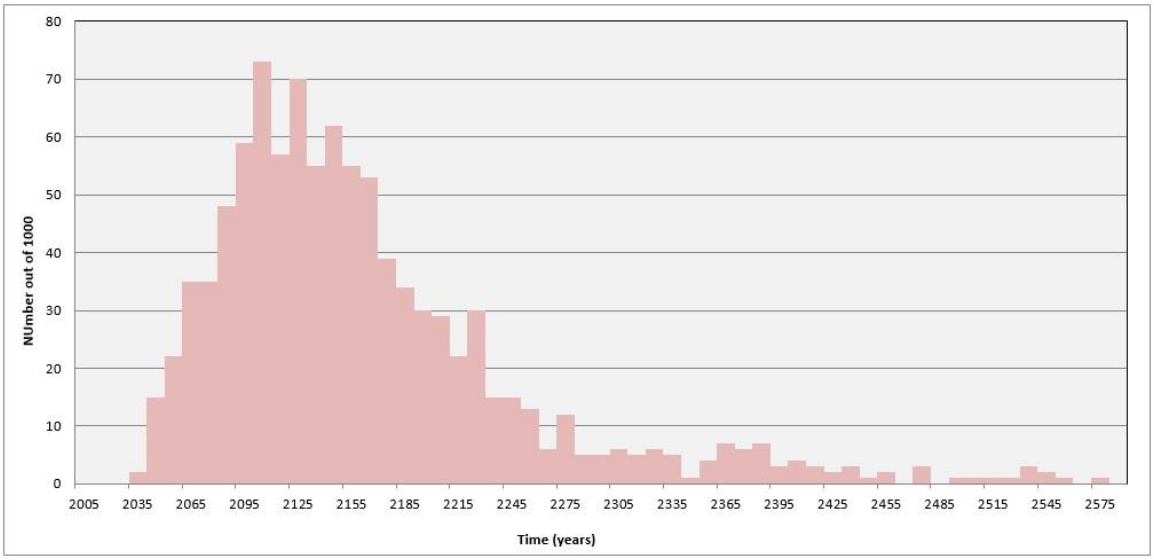


Figure 6: The experimental results from 1000 runs of simulation demonstrate that the number of runs with the first extreme event increases as temperature increases and falls after a peak around year 2125.

As shown in the figure, the earlier states have the most likelihood of having a tipping point while in the later states this possibility decreases dramatically. There are two reasons for this fast convergence; first, the tipping point uncertainty, as defined by its transition matrix, has a binary characteristic. It switches from pre-tipping point state to post-tipping point state and therefore there are limited fluctuations around the optimal

policy (the coefficient of the basis function). Second, p_t is an absorbing Markov chain: as the iterations grow larger, more incidents of hitting the absorbing state (tipping event) happen. Every time the simulation hits the absorbing state it continues to remain in the state of post-tipping for the rest of the modeling time horizon.

Figure 7 and Figure 8 show the integrated assessment model results with a tipping point at $t = 10$ (100 years). Compared to the deterministic model with no tipping points, the risk of tipping point events stimulates a higher reduction in carbon emission. However, after a tipping point happens, the optimal reduction drops dramatically even below the level of optimal reduction rate in the deterministic model. The reason is that once the tipping point hits the climate system, the global economy shrinks with an unprecedented scale (twenty five percent of global economic output) and stays in that state afterward; also the pre-tipping point reductions took into account the risk of both direct climate change and the tipping point.

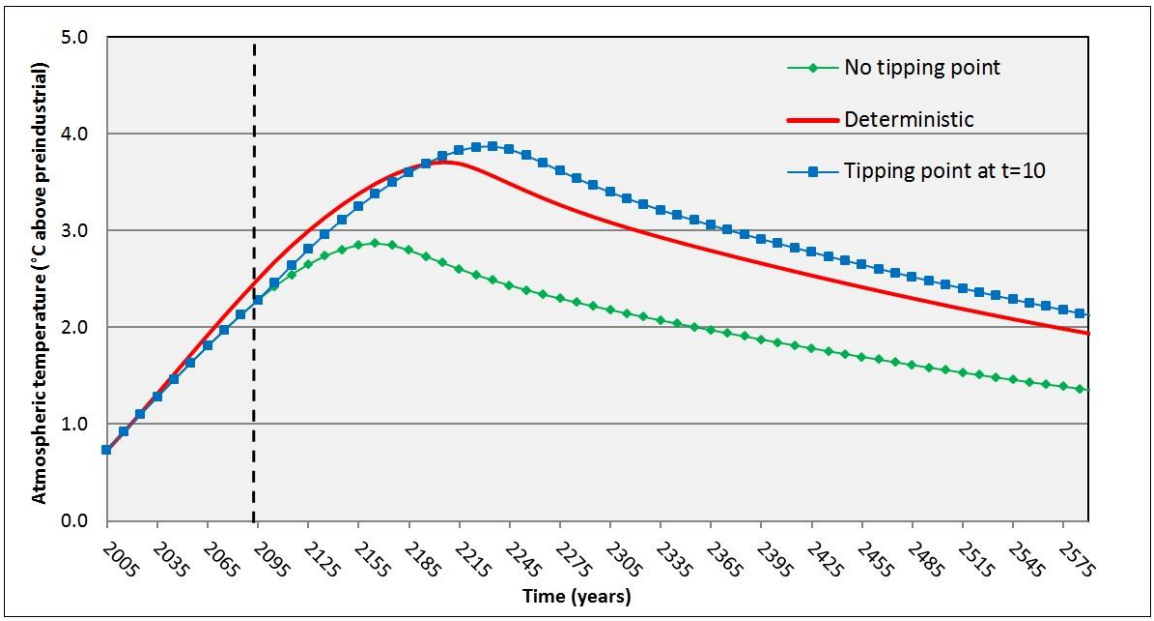


Figure 7: Taking steep reduction in GHG emissions will induce a small probability of having extreme events and controls the global mean temperature increase below **3**℃ (lower graph). The optimal abatement in the deterministic case keeps the temperature around $\mathbf{3.7}$℃ (middle line). In the case that the extreme event happens at $\boldsymbol{t = 10}$ (upper graph), the temperature first follows the "no-tipping" path (lower graph) and after the extreme event happens follows a trajectory similar to the one in optimal deterministic case with a shift due to lower post-extreme abatement actions.

Therefore the amount of emissions and the consequent abatement level drops down, lower than the deterministic case. In a model with more than one tipping point, the result could be different; for example if the occurrence of one tipping point increases the probability of another tipping point, the optimal reductions might continue to increase. Further detailed research on extended tipping point models could explore this issue. As shown in Figure 8, in the case of a tipping point event at time step $t = 10$, the action falls from a pre-tipping point level of 47% in the previous time epoch to 37% in the post-tipping point epoch and continues to stay below the level in the deterministic model until the emission reduction reaches 100%. The impact of different abatement strategies on the global mean temperature is demonstrated in Figure 7 where pre-tipping strategies with higher abatement rates result in lower increase in atmosphere temperature, even lower than the optimal level under deterministic conditions. However, the temperature rises more after the occurrence of the tipping point event due to the sharp decline in abatement.

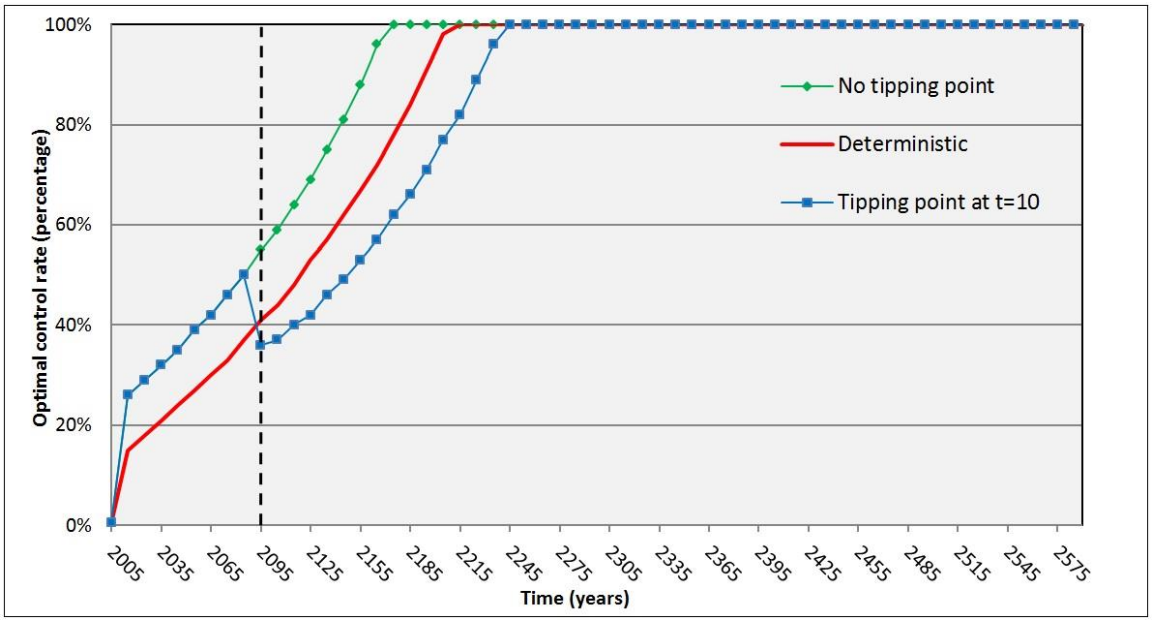


Figure 8: To avoid the risk of tipping point events, higher abatement rate is induced (upper graph) which is significantly above the optimal level of abatement in the deterministic case (middle line). In the case where the extreme event happens at $t=10$ (lower graph), the abatement falls below the level of the deterministic case and follows a trajectory parallel to it.

Figure 9 shows the optimal greenhouse gas reduction results for the first 30 decades. The shaded area demonstrates the possible pathways from pre-tipping point optimal policies to post-tipping point optimal policies (lower graph). In this 1000-run simulation, note that the first tipping point events occur about 30 years into the simulation ($t = 3$).

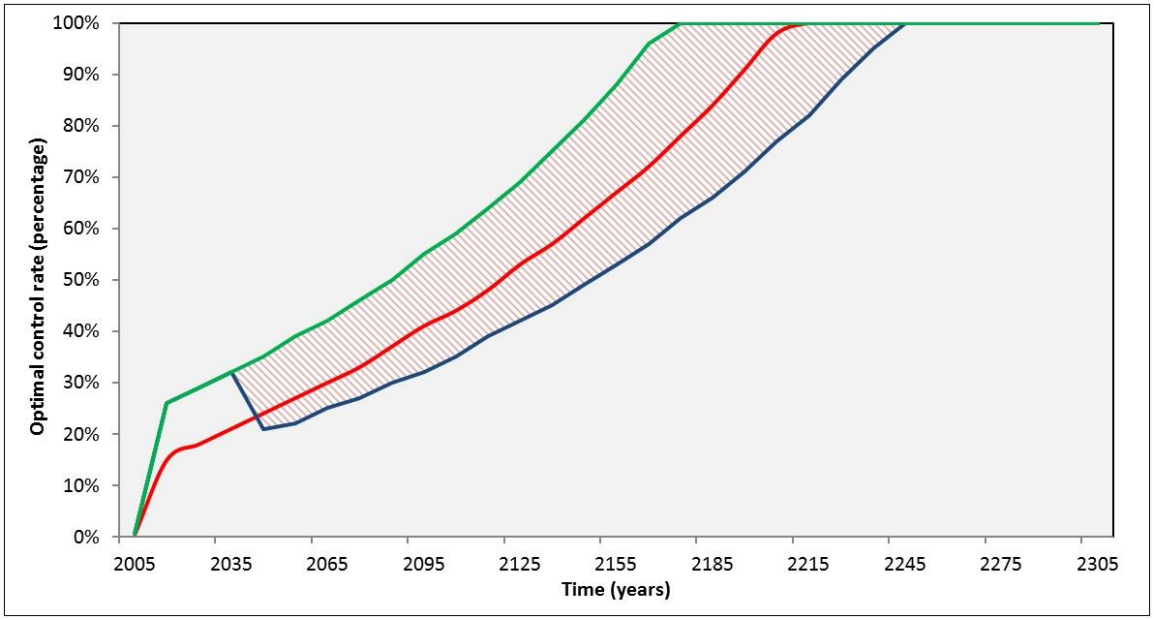


Figure 9: Optimal policies in pre and post tipping point states. The optimal actions form a trajectory (upper envelope) above the level of abatement in the deterministic case (middle graph). If a tipping point happens the abatement rate falls to a level (lower envelope) below the deterministic case.

CHAPTER 3

A MULTISTEP LOOKAHEAD ALGORITHM FOR APPROXIMATE DYNAMIC PROGRAMING

In this chapter we study the problem of decision making under uncertainty in finite horizon and with continuous state space. We develop a method for value function approximation in approximate dynamic programming that combines offline calculation with online rolling horizon methods. This method consists of using an H-step-ahead approximation for estimating the value function and finding the optimal action online, and a value iteration algorithm to update the parameters of the approximated value function offline. Conditions on the step size that guarantee the convergence of the value function approximation are derived. We apply the approach to an integrated model of climate and the world economy. We analyze the impact of discount factor on the choice of approximation and provide insight into the robustness of approximation.

Introduction

Reinforcement learning (RL) is one of the main branches of artificial intelligence, concerned with learning the optimal decision in a dynamic environment [46]. This can be done by finding the optimal expected value of each state and recursively finding the best action that maximizes (minimizes) the sum of the immediate reward (cost) of taking that action and the expected value of the future states. In the stochastic domain the value function of a given state is the expected value of that state under a certain policy (i.e. the cost-to-go function, which evaluates the expected future cost to be incurred, as a function of the current state). When facing a large state and action space, using traditional dynamic programming techniques is impractical and inefficient [47]. In this case, the most common approach is to estimate the value function with a direct approximator [48].

In offline methods, the policy (or parameters of the value function) is pre-calculated and stored to find the optimal action at any specific state [49]. On the other hand, in online methods the optimal action is calculated at the decision time [50]. Value iteration and policy iteration are two of the most common offline methods for finding optimal value function or optimal policy in finite state and action problems while rolling horizon is commonly used as an online strategy [51]. By optimal policy, here we mean a mapping from the state space to the action space [52]. Therefore a policy in approximate dynamic programing setting can be looked upon as the set of tunable parameters of the value function approximation.

The problem of finding the optimal policy or value function is more challenging in the continuous state and action spaces [53]. In this case, approximations should be made not only for calculating the value function but also for stage-state representations in the case of finite models. In this paper we focus our attention to finite-horizon discrete-time problems. We combine an online rolling horizon technique and an offline value iteration method and develop a new algorithm for approximating the value function in finite horizon problems. This method consists of using an H-step-ahead approximation for estimating the value function and finding the optimal action online and a value iteration algorithm to update the parameters of the approximated value function offline.

Approximate Dynamic Programing Framework

Consider a finite-horizon dynamic programing problem with continuous state space S, continuous action space A, and discount factor $\gamma \in (0, 1)$. To study the behavior and properties of this problem a Markov decision process framework can be developed. Given the state of the system S_t at time t, taking an action a_t under realization of the uncertain parameter W_t will transform the system to a new state S_{t+1}, and an immediate reward U_t will be utilized. The objective is to maximize the cumulative discounted rewards over the modeling horizon T as shown in Equation (14):

$$\max_{\pi} \mathbb{E}\left\{\sum_{t=0}^{T} \gamma^t U_t^{\pi}(S_t, a_t^{\pi}(S_t))\right\} \quad (14)$$

where π represents the policy for choosing optimal actions. The value of each state is calculated using the Bellman equation:

$$V(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \mathbb{E}_{W_t}\{V_{t+1}(S_{t+1})|S_t, a_t\}\right) \quad (15)$$

Deploying the Value Iteration (VI) algorithm, the value of each state $V(S_t)$ is calculated recursively as the maximum sum of the immediate reward and the expectation of the discounted optimal value of the next state as shown in Equation (16). In the idealized case of countable state space with finite actions, the value of each state converges to its optimal value (Banach Fixed-Point Theorem) [54]. Equation (16) shows the so-called Bellman operator for calculating the optimal value at the nth iteration.

$$V^n(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \mathbb{E}_{W_t}\{V_{t+1}^{n-1}(S_{t+1})|S_t, a_t\}\right) \quad (16)$$

where $V^n(S_t)$ represents the optimal value of the state S_t after n iterations. The problem occurs when the state space is continuous or is large with multiple dimensions as is the case in climate change modeling. Discretizing the space state and probability distributions of uncertain parameters is computationally expensive if not impossible. Therefore, we adopt a class of techniques known as approximate (adaptive) dynamic programming (ADP) to deal with problems with continuous (or large) state, action, and probability distributions [24]. The main idea here is to move forward in time and calculate the value of the current state by estimating the value of future states.

$$\hat{V}_t(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}(S_{t+1})\right) \quad (17)$$

where $\hat{V}_t(S_t)$ is the optimal value of state S_t based on the value approximation of the next state S_{t+1}. The general ADP algorithm for value iteration is presented in Table 1. After initialization it has three main steps: generating random paths, stepping forward in time and looping over the modeling time horizon by approximating the value function, and finally updating the approximation parameters. The main issue of the value iteration

algorithm is its updating scheme. On the one hand we would like to have a fast converging algorithm (exploitation) to save time and computational resources, and on the other hand a pure exploiting strategy may not be stable or optimal. A residual approach is one of the most common approaches developed to estimate the value of $\bar{V}_{t+1}$ which can be applied to either Value or Policy iteration algorithms [43]. This technique is based on calculating the difference between the true value of the state $\hat{V}_t(S_t)$ and its estimation $\bar{V}_t(S_t)$ at each iteration and trying to minimize this difference (temporal difference or TD [55]) by improving the optimal policy. Since $\bar{V}_t(S_t)$ is the estimate of the true value of the state, TD is an approximation of the Bellman error. If the estimated value $\bar{V}_t(S_t)$ is an explicit parametric function with parameter θ_t, we can use a direct gradient descent algorithm to update its value in each iteration [56]. Let F be a functional representation of the TD, for example

$$F(\bar{V}_t, \hat{V}_t) = \frac{1}{2}\left(\bar{V}_t(S_t) - \hat{V}_t(S_t)\right)^2 \tag{18}$$

We can update parameter θ of the estimated value function $\bar{V}_t$ after n iterations using a stepsize α_n

$$\theta_t^{n+1} = \theta_t^n - \alpha_n \nabla_\theta F(\bar{V}_t, \hat{V}_t) = \theta_t^n - \alpha_n(\bar{V}_t^n - \hat{V}_t^{n+1})\frac{\partial \bar{V}_t^n}{\partial \theta_t^n} \tag{19}$$

There are two challenges associated with this technique. First, we need to find a good approximation of the value function $\bar{V}_t^n$. Second, the stepsize α_n should be defined so to provide a balance between exploration and exploitation in the state space. If α_n is chosen close to zero ($\bar{V}_t^{n+1} = \bar{V}_t^n$), no learning is happening and the algorithm will return the current policy as the optimal one. We address these two challenges in the next two sections, respectively.

Table 1: ADP Value Iteration Algorithm

Step 0: Initialize parameters for value function approximation

Step 1: Generate sample paths from the distribution function of the stochastic parameter

Step 2: Stepping forward through time, for each time epoch t calculate

$$\hat{V}_t^n(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$
$$a_t^*(S_t) = \underset{a_t}{argmax} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$

Step 3: Update the value function approximation parameters

$$\bar{V}_t^n(S_t) = \bar{V}_{t-1}^{n-1}(S_t) - \alpha_{n-1} \left(\bar{V}_{t-1}^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)$$

Step 4: Repeat **Step 1** through **Step 3** for n (large number of runs) times

H-step-ahead Value Function Approximation

Rolling horizon algorithms are powerful techniques for approximating optimal policies in deterministic problems with low dimensional state and action space. The idea is that a greedy agent solves a dynamic programming problem by maximizing the sum of the immediate reward and the estimated value of the next state $U_t(S_t, a_t) + \gamma \bar{V}_{t+1}(S_{t+1})$, where $\bar{V}(.)$ is an estimation function [50]. Such a greedy algorithm can be thought of as a one-step Lookahead local search technique. A more extensive technique will perform a search of depth H and will return the optimal policies for H steps ahead [44]. As a result, instead of solving the whole optimization problem which spans a large time interval (or infinity), from starting at $t = 0$ to the terminal period T, we can solve a smaller problem for H time periods ($H \ll T$) starting from $t = 0$ and then iteratively stepping forward in time, and solving another optimization problem for the next H time periods (from $t = 1$ to $t = H + 1$). The general form of the rolling horizon (receding) procedure at any time epoch t is presented below

$$\hat{V}_t(S_t) = \max_{a_t, \dots, a_{t+H}} \sum_{t'=t}^{t+H-1} \gamma^{t'-t} U_{t'}(S_{t'}, a_{t'}) + \gamma^H \bar{V}_{t+H}(S_{t+H}) \quad (20)$$

where $\hat{V}_t(S_t)$ is the value of being in state S_t as before. We can limit the domain of search for the optimal policies through what are known as constrained local search (CLS)

algorithms [50]. For instance, we may consider only a subset of all feasible policies with certain structure such as increasing or decreasing actions or, it may be sufficient to sample only a small set of neighboring states in each time step and calculate the optimal policy using this small set. Although this technique reduces the complexity of the original problems especially in models with a long time horizon, the value of terminal states $(\bar{V}_H, \bar{V}_{H+1}, \ldots, \bar{V}_T)$ in each iteration of the algorithm still need to be estimated using value function approximation techniques. One way to overcome this problem is by assuming a null value for the terminal states and calculating the average value of intermediate states through simulation. This technique has been applied in a sparse sampling algorithm for finding near-optimal solutions for stochastic optimization problems [57]. In this method for each available action, a set of future states for H-step ahead are generated using N random drawings from the (uniform) transition probability distribution. At each time step, the optimal value is found by taking the maximum value of the immediate reward plus the discounted average value of the next state

$$\hat{V}_t(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \frac{1}{N} \sum_{S'_{t+1}} \hat{V}_{t+1}(S'_{t+1}) \right) \tag{21}$$

Although the size of each H-step mini-optimization problem is shown to be independent of the size of the original problem [57], there are some shortcomings to this method. First, the computational complexity grows exponentially with H, the depth of lookahead algorithm. In fact, the running time for finding the best action at any given state is $O((N|A|)^H)$, where $|A|$ is the size of the action space. Second, uniform sampling might not be suitable for all models. In the case of an integrated assessment model for climate change, for example, the uncertainty comes from a parameter with a heavy-tailed distribution [58]. Averaging the random samples from a heavy-tailed distribution has been shown to be misleading in certain applications [15].

As an alternative we can adopt a lookahead technique to approximate the value function and use this approximation in a greedy search algorithm. There are several well-

known methods for approximating the value function; for an exhaustive survey see [43]. One widely used method uses a parametric function (basis function) of the state variables to construct the approximate value function [59]. However by using this method one faces a very fast growth of the number of basis functions in the approximation scheme [19]. Here we introduce a novel way of combining both offline and online techniques. Online methods such as conventional rolling horizon (as discussed above) are widely used in countable state-action spaces [60]. However, when facing continuous state-action spaces, we have to use different approximation techniques to estimate the value of a particular state. The on-line methods lack the updating capabilities of off-line techniques, and therefore fall short of closing the gap between the true and estimated values of a particular state [61]. To address this problem, we combine these two methods in a framework that we call "H-step-ahead value iteration". Another advantage of this method over conventional methods is that since it is based on the functions in the model it does not introduce new functional forms to the model. The idea behind the deterministic rolling horizon or receding heuristics is to look a finite number of steps into the future and solve a smaller problem than the original one [44]. In this case, the approximated value function is obtained by moving forward in time with depth H and finding the value of the next state by adopting a subset of predefined actions. The approximation of the value function will be the linear combination of H-step discounted reward similar to the basis function extraction [62] used in value function approximation:

$$\bar{V}_{t+1}(S_{t+1}, a_t) = \mathcal{F}(S_{t+1}, \dots, S_{t+H} | S_t, a_t, W_t) = \sum_{t'=t+1}^{t+H} \theta_{t'} U_{t'}\big((S_{t'}|a_t, W_t), a_{t'}\big) \quad (22)$$

where $\mathcal{F}(.)$ is the general approximation function, $\theta_{t'}$ is the linear coefficient of the value function approximation and $a_{t'}$ is from a subset of the available actions. Note that the approximation of the value $\bar{V}_{t+1}$ depends on the action a_t and therefore, we can define and approximate the Q function as in the "$Q\ learning$" technique [63]:

$$Q_t(S_t, a_t) = U_t(S_t, a_t) + \bar{V}_{t+1}(S_{t+1}, a_t) \tag{23}$$

$$\hat{V}_t(S_t) = \max_{a_t} \left(Q_t(S_t, a_t)\right) \tag{24}$$

However unlike our case, "$Q\ learning$" is mainly applied to finite state space and uses lookup tables for a limited number of state-action pairs. Combining Equations (22)-(24) we obtain a new equation for calculating the value function

$$\hat{V}_t^n(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \sum_{t'=t+1}^{t+H} \theta_{t'}^{n-1} U_{t'} \left((S_{t'} | a_t, W_t^n), a_{t'}^* \right) \right) \tag{25}$$

The exogenous information W_t^n is drawn at n-th iteration from the probability distribution of the uncertain parameter and the process iterates for M runs. There are two important differences between this new equation and Equation (20): first, instead of finding the optimal actions for all $H + 1$ states in this setup, we only need to find the optimal action for the first state a_t; other actions are predefined in a way that will be discussed later. Second, the estimation of the value of terminal state $\bar{V}_{t+H}(S_{t+H})$ is not present in this new equation and instead, it uses the immediate reward in the terminal state from taking the predefined action $a_{t'}$. The running time for finding the best action at each iteration at any given state is $O(H|A|)$. Since the focus of this paper is on finite horizon (episodic) problems with large or continuous state spaces, it is important to note that the value function approximation $\bar{V}_t(S_t)$ is estimating the value of state S_t. We define an 'episodic state' to be a representation of all states at the time epoch t and therefore $\bar{V}_t$ can be considered as the mapping from episodic state space to value space. The overall running time of the algorithm in the finite setting is $O(H|A|T)$ for the entire time horizon. The only challenge in this new algorithm is to find $a_{t'}^*$, the predefined action. If the reward function $U_t(S_t, a_t)$ is concave in a_t, we can pick $a_{t'}^* = a^* = max_{a_{t'}} U_{t'}(S_{t'}, a_{t'})$ which is the greedy action maximizing the immediate reward for all H steps ahead. The corresponding states are defined by the current exogenous information process W_t^n and

a^*, the one-step optimal action. In the absence of any uncertainty, the value function approximation can be represented as a linear combination of immediate rewards from taking the action a_t at the present time and a^* in any subsequent step.

Table 2 shows the value iteration algorithm for approximate dynamic programing with lookahead value estimation. In each state S_t the exogenous information is fixed at W_t^n level and a candidate action a_t is taken once and then the predefined action a^* is taken H times to produce the next H states with corresponding rewards U at each state. Looking forward into the future and taking into account the utilities of the H next states ahead will capture the tradeoff between a myopic policy to maximize the value of the current state only and a lookahead policy which maximizes the value of current and future states as a single function. This technique provides a fast and robust solution for both deterministic and stochastic cases. The rate of learning (convergence) in this method depends on the updating scheme and we can employ a simple stochastic gradient algorithm to update this coefficient similar to the one shown in Equation (19).

Table 2: Value Iteration algorithm for ADP with ***H***-step-ahead value function approximation

Step 0: Initialize parameters for value function approximation

Step 1: Do for $n = 1$ to $n = N$

Step 2: Generate a sample path from the distribution function of the stochastic parameter

Step 3: Do for each time epoch from $t = 0$ to $t = T$

Step 4: Do for each time epoch from $t' = t + 1$ to $t' = t + H$

$$\bar{V}_{t+1}^{n-1}(S_{t+1}) = \sum_{t'=t+1}^{t+H} \gamma^{t'-(t+1)} \theta_{t'}^{n-1} U_{t'}\left((S_{t'}|a_t, W_t^n), a^*\right)$$

$$\hat{V}_t^n(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$

$$a_t^*(S_t) = \underset{a_t}{argmax} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$

Step 5: Update the value function approximation parameters

$$\bar{\theta}_t^n = \bar{\theta}_t^{n-1} - \alpha_{n-1} \left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right) \frac{\partial \bar{V}_t^{n-1}}{\partial \theta_t^{n-1}}$$

Step 6: Return $\bar{\theta}_t^N$ for all t

Optimal step size and convergence

The conventional rolling horizon algorithms always perform suboptimally since they ignore the rest of the steps after H. The bounds on the performance of a rolling horizon algorithm in the infinite case (Equation (20)) is defined by equation below [64]:

$$0 \leq V^*(S) - \bar{V}(S) \leq \frac{U_{max}}{1-\gamma}\gamma^H \tag{26}$$

where V^* is the true value and U_{max} is the maximum possible reward. In the finite case, however, the upper bound will be adjusted and the equation can be rewritten as:

$$0 \leq V^*(S_t) - \bar{V}(S_t) \leq U_{max}\left(\frac{1-\gamma^{T-t-H}}{1-\gamma}\right)\gamma^H \qquad for\ 0 \leq t \leq T-H \tag{27}$$

In problems with large scale or infinite state space where approximation is devised, the upper bound needs to be adjusted again to reflect the approximation [53]. As the number of lookahead steps (H) increases the value function approximation defined in Equation (22) gain more flexibility in adaptation to changes in the underlying value $V^*(S_t)$ and therefore the approximation error shrinks consequently. In the H-step-ahead algorithm, we first estimate the sum of all future values using the H step ahead rewards and then update those estimates iteratively. We can show that this approximation mapping is nonexpansion and therefore the value iteration algorithm is converging to the true values of each episodic state.

Theorem 1: Let T be the value iteration operator defined by Equation (16), and let F be the H-step-ahead value function approximation. The value iteration algorithm defined in Table 2 converges to $V^*(S_t)$, the true value of the episodic state S_t for every $0 \leq t \leq T-H$ with step size $\alpha_n^i \leq \frac{\delta}{HU_i{}^2}$ where $\delta \in (0,1)$ and $i \in (t+1, t+H)$.

The proof is presented in Appendix B. Once the algorithm converges, the optimal policy (optimal values of parameter $\bar{\theta}_i$) can be determined for each episodic state S_t.

Numerical Example

In this section we apply the H-step-ahead value function approximation to the problem of finding the optimal greenhouse gas (GHG) emissions abatement level. The deterministic version of this problem has been discussed in [31] and the model parameters and equations are represented in Appendix A. The global climate-economy system can be defined as a state with six continuous variables: T_{at} is atmospheric temperature (degrees Celsius above preindustrial), T_{lo} is lower ocean temperature (degrees Celsius above preindustrial), M_{at} is atmospheric concentration of carbon (Giga Tons of Carbon, GTC), M_{up} is concentration in biosphere and upper oceans (GTC), M_{lo} is concentration in deep oceans (GTC), and K is capital ($trill). At each time step, an abatement action (control rate) a_t is taken which indicates the percentage reduction of GHG emissions. It imposes a cost to the economy but prevents the future damage costs of having high temperature due to the increase in the emissions. Taking action a_t at any given state will determine the next state deterministically. We can introduce uncertainty into this system by modeling the atmospheric temperature dynamic as a random process. We define a probability distribution for ΔT the climate sensitivity parameter (i.e. the equilibrium increase in mean global surface temperature due to doubling of atmospheric CO_2 compared to the pre-industrial era [32]). The objective is to maximize the social utility which is a function of economic output and costs.

To calibrate the model and find the appropriate number of steps for lookahead algorithm we run this model under deterministic assumptions and compare the results for different values of H. Figure 1a shows that larger values of H give more flexibility to value function approximation and improve the optimal path calibration. In Figure 1b we show a boundary case of $H = 2$, where the parameter θ_1 is kept at zero. This case is comparable with $H = 1$, since only the value of one of the future states is used for estimating the current value function. For this boundary case (2-step-ahead

approximation with $\theta_1 = 0$) the value function is approximated by projecting the values of states in the next two time periods. The values are calculated under a deterministic forecast and brought back to the present time using an artificial and tunable discount rate (parameter θ_t).

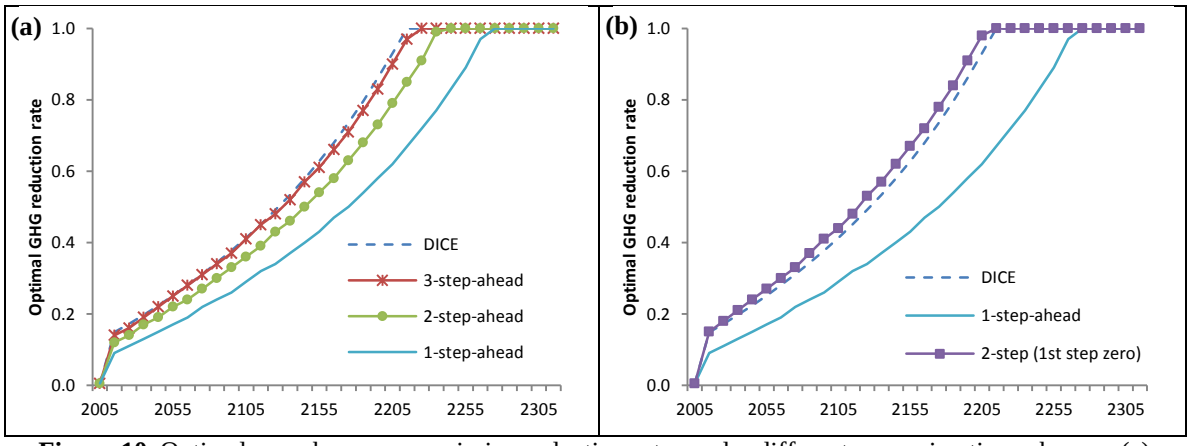


Figure 10: Optimal greenhouse gas emission reduction rates under different approximation schemes, (a) comparison of different values of $\boldsymbol{H}$ for the $\boldsymbol{H}$-step-ahead algorithm, (b) comparison of 1-step-ahead and boundary case ($\boldsymbol{\theta_1 = 0}$) of 2-step-ahead algorithms

We can also investigate the effect of discounting on the approximation of the optimal solution. In the original DICE model the default value of the decadal discount factor is $\gamma = 0.862$ and the optimal path reaches its peak at year 2215 (Figure 10a). Figure 10 shows the optimal path and approximation with different values of the lookahead depth H. As shown here, although the optimal rates vary significantly under different discount factors (i.e. higher discount factor induces faster convergence of the optimal path to its maximum rate of 1.0), the 3-step-ahead approximation follows the optimal path very closely. In general, comparing H-step-ahead approximations with different values of H, one can see that higher values of H provide more flexibility and better approximation of the value function.

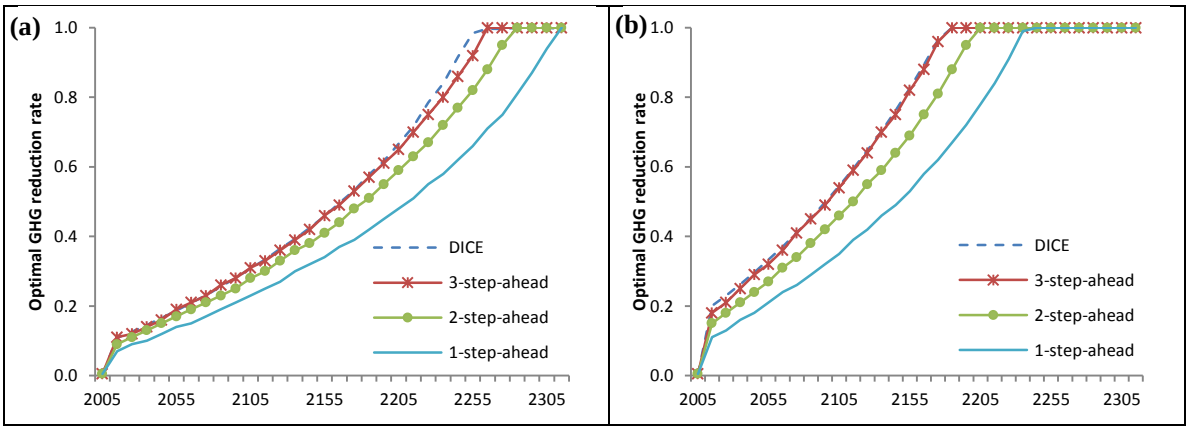


Figure 11: Optimal greenhouse gas emission reduction rates under different discount factors, (a) $\boldsymbol{\gamma = 0.75}$, (b) $\boldsymbol{\gamma = 0.95}$

To demonstrate the algorithm, consider a simple model with four states S_0 to S_3 as shown in Figure 12. The initial values of S_0's state variables as well as the initial realization of climate sensitivity ($\Delta T_1 = 3$℃) are provided. As an example taking the initial action $a_0 = 0.5\%$ at state S_0 will take us to the state S_1 with the state variables shown in the figure. To find the next optimal action a_1^* we deploy our two-step-ahead algorithm. First, under deterministic assumption from the previous state, the value of the current state S_1 will be calculated by taking any candidate action a_1 and two consecutive null actions to obtain two post-decision states S_1^a and S_2^0 and with immediate rewards of $U_1(S_1, a_1)$, $U_1(S_1^a, 0)$, and $U_1(S_2^0, 0)$ at node S_1 and two post-decision nodes after that. The optimal action is the one that maximizes the value of the current sate:

$$a_1^*(S_1) = \underset{a_1}{argmax}\ \left(U_1(S_1, a_1) + \bar{V}_1(S_1^a)\right)$$

As discussed earlier, $\bar{V}_1^{n-1}$ is the value approximation of the post-decision state S_1^a. We consider a very simple function approximation with only one parameter $\bar{V}_1(S_1^a) = \theta_1 \times U_1(S_2^0, 0)$, where θ_1 is the tunable parameter of the value function approximation and defines the "policy". The initial value of this parameter is assumed to be one and it is updated at the end of each iteration. Table 3 demonstrates the value of this approximation for selected actions. The value of state is calculated from $\hat{V}_1(S_1) =$

$max_{a_1}\ \left(U_1(S_1, a_1) + \bar{V}_1(S_1^a)\right)$. Once the optimal action is found ($a_1^* = 10\%$), a realization of the uncertain parameter is drawn from the sample path and the values of state variables of the next state S_2 is calculated accordingly. For this simple example, we assume that the climate sensitivity stays at $\Delta T_1 = 3$℃ level for the next time period. In this case, the optimal action is found to be around 12% and the optimal value is $\hat{V}_2(S_2) = 140939.3$, this value is used to update the approximation function that was used to estimate the value of the post-decision state S_1^a using the following stochastic gradient algorithm: $\theta_1^{new} = \theta_1^{old} - \alpha \times \left(\bar{V}_1 - \hat{V}_2\right) \times U_1(S_2^0, 0)$

The step size α is chosen as $[U(S_2^0, 0)]^{-2}$ to simplify the updating equation and guarantees the convergence. Therefore the new coefficient for the next iteration is calculated as

$$\theta_1^{new} = 1 - \frac{\left(\bar{V}_1 - \hat{V}_2\right)}{U_1(S_2^0, 0)} = \frac{\hat{V}_2}{U_1(S_2^0, 0)} = 1.987$$

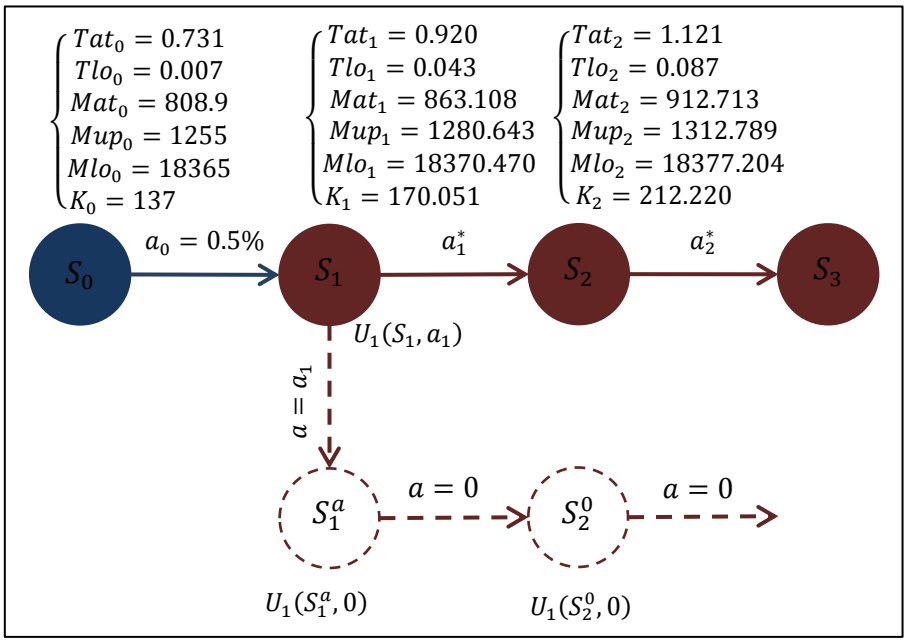


Figure 12: An example of the two-step-ahead algorithm for DICE model

The new value function generates new optimal actions and this process continues until the policy (the coefficients of the value function approximation) converges to its optimal value.

Table 3: value function approximation for different actions and the optimal value of the first two states.

action	$\bar{V}_1(S_1^a) = \theta_1 \times U_1(S_2^0, 0)$	$U_1(S_1, a_1)$	$\hat{V}_1(S_1)$ $(a_1^* = 10\%)$	$\bar{V}_2(S_2^a) = \theta_2 \times U_2(S_2^0, 0)$	$U_2(S_2, a_2)$	$\hat{V}_2(S_2)$ $(a_2^* = 12\%)$
$a = 0\%$ *(No abatement)*	70928.1	61924.8		73918.1	67020.9	
$a = 50\%$	70927.3	61855.4		73919.2	66962.5	
$a = 100\%$ *(Full abatement)*	70895.3	61418.8		73894.8	66597.4	
a^**(Optimal abatement)*	70929.1	61924.1	132853.2*	73919.5	67019.8	140939.3*

Figure 13 shows how fast the two-step-algorithm converges in this case. The error in the early stages of approximation vanishes as the model learns the optimal policy (θ_{10}) and consolidates around its optimal value.

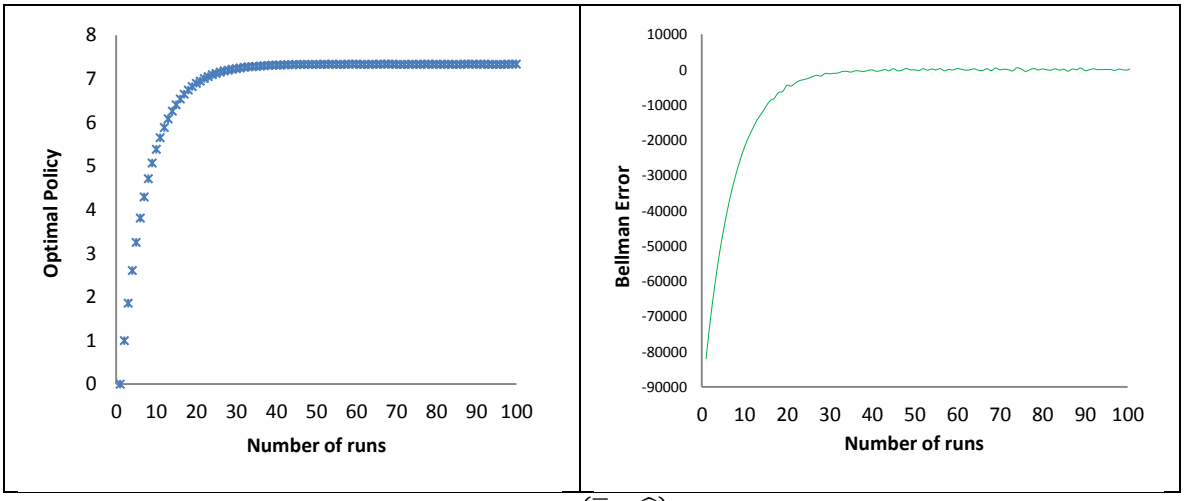


Figure 13: Optimal Policy ($\boldsymbol{\theta_{10}}$) and Bellman error $\left(\bar{\boldsymbol{V}} - \hat{\boldsymbol{V}}\right)$ at time period $\boldsymbol{t = 10}$ using a two-step-ahead algorithm

In summary, our model uses an on-line estimate of the value function by forecasting the two-step ahead states and it also stores and updates the tunable parameters of the value function approximation in an off-line fashion through the value iteration algorithm.

CHAPTER 4

BAYESIAN APPROXIMATE DYNAMIC PROGRAMMING

Climate sensitivity is a single-valued variable whose value can be expected to become better known through time and observations. In the case of tipping points on the other hand, we assume that there is always a chance (however very small) of hitting the tipping point at any atmospheric surface temperature increase greater than 1℃ ($T_{at} > 1℃$) and therefore, it can be considered as a random variable. These two random parameters (climate sensitivity and tipping point event) are not independent. As can be traced through equations A11 and A17 provided in Appendix A, the climate sensitivity parameter is used in calculating the CO_2 doubling coefficient which defines the mean global temperature in the next time step (Equation A17) and therefore affects the probability of having an extreme event in the next time period. To put it in statistical language, we expect to have a different probability distribution for climate sensitivity after each observation of climate status. This leads us to introducing a novel approach for integrating a Bayesian modification into our ADP algorithm, as a mechanism for updating the probability distribution of climate sensitivity after each observation of climate status.

The shape of the probability distribution is initially estimated to be a fat-tailed distribution as discussed in Chapter 2. In the future, new information on the behavior of the climate, based on mean global temperature data, estimates of climate damage, and other types of observations, can be used to update the climate sensitivity probability distribution. For the purpose of this illustration, we limit our attention to the set of realizations of extreme events as an integrated representation of all other climate system observations in each period. Obviously such a proxy is oversimplifying and does not take into account many vital components of the climate system. However, the observation of

extreme events can be modeled within the simulation and therefore is a practical proxy for illustrative purposes. The purpose of this approach is not to show numerically to what extent an observation about a climate extreme event can update our knowledge of climate sensitivity, but rather to demonstrate its capability to do so and a statistical approach for implementation.

The status of the climate after any pre-tipping point state (χ) can be modeled as a binomial distribution with p being the probability of having an extreme event (tipping point) in next time step. As discussed in the previous section, this probability itself is a function of mean global temperature T_{at} which directly depends on the estimate of climate sensitivity ΔT. These relationships can be expressed in a standard Bayesian format as:

$$Pr(\chi_{t+1}|\Delta T, \chi_t) = \chi_t\chi_{t+1} + (1-\chi_t){p_t}^{\chi_{t+1}}(1-p_t)^{1-\chi_{t+1}} \quad (28)$$

$$\Pr(\Delta T) \sim \text{truncated Lognormal distribution} \quad (29)$$

where χ_t is the status of the climate at time epoch t and p_t is the probability of having a tipping point event at time epoch $t+1$, as defined in Equation (5). We consider the binary values of $\chi_t = 0$ for non-extreme (pre-tipping point) states and $\chi_t = 1$ for extreme (post-tipping point) states. The posterior distribution of climate sensitivity can be obtained from prior and conditional distributions.

$$Pr(\Delta T|\chi_{t+1}, \chi_t) = \frac{Pr(\chi_{t+1}|\Delta T, \chi_t)\,\Pr(\Delta T)}{Pr(\chi_{t+1}|\chi_t)} \quad (30)$$

where $Pr(\chi_{t+1}|\chi_t)$ represents the probability of the climate status in next time step given the current status of the climate. This probability can be expressed as the integral of the conditional probability in Equation (28) over different values of climate sensitivity. To calculate $Pr(\chi_{t+1}|\chi_t)$ the climate sensitivity interval $[0,20]$ is split into one hundred equally spaced segments and the integral is computed numerically as the following summation.

$$Pr(\chi_{t+1}|\chi_t) \approx \sum_{i=1}^{100} Pr(\chi_{t+1}|\Delta T_i, \chi_t)\,\Pr(\Delta T_i) \tag{31}$$

The value iteration algorithm discussed in Table 2 can be revised to reflect the Bayesian update of the uncertain parameter.

Table 4: Value Iteration algorithm for BADP with ***H***-step-ahead value function approximation

Step 0: Initialize parameters for value function approximation

Step 1: Do for $n = 1$ to $n = N$

Step 2: Generate a sample path from the distribution function of the stochastic parameter

Step 3: Do for each time epoch from $t = 0$ to $t = T$

Step 4: Do for each time epoch from $t' = t + 1$ to $t' = t + H$

$$\bar{V}_{t+1}^{n-1}(S_{t+1}) = \sum_{t'=t+1}^{t+H} \gamma^{t'-(t+1)} \theta_{t'}^{n-1} U_{t'}\left((S_{t'}|a_t, W_t^n), a^*\right)$$

$$\hat{V}_t^n(S_t) = \max_{a_t} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$

$$a_t^*(S_t) = \underset{a_t}{argmax} \left(U_t(S_t, a_t) + \gamma \bar{V}_{t+1}^{n-1}(S_{t+1})\right)$$

Step 5: Update the value function approximation parameters

$$\bar{\theta}_t^n = \bar{\theta}_t^{n-1} - \alpha_{n-1} \left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right) \frac{\partial \bar{V}_t^{n-1}}{\partial \theta_t^{n-1}}$$

Step 6: Update the probability density of the uncertain parameter

$$Pr(\Delta T|\chi_{t+1}, \chi_t) = \frac{Pr(\chi_{t+1}|\Delta T, \chi_t)\,\Pr(\Delta T)}{Pr(\chi_{t+1}|\chi_t)}$$

Step 7: Return $\bar{\theta}_t^N$ for all t

Table 4 shows a general value iteration algorithm for Bayesian approximate dynamic programing (BADP). In each run and at any time step t, the status of the climate is observed and the status of climate for the next time step $t + 1$ is calculated according to Equation (28). For the next time step, a random value for climate sensitivity is drawn from the posterior distribution given in Equation (30). As discussed in chapter 2, the prior distribution is a lognormal distribution truncated from the left and right sides. The posterior distribution shows a profound shift from higher values of climate sensitivity (associated with higher risk of tipping point events) to lower values with higher

probability. Utilizing Bayesian update provides a learning mechanism to reduce uncertainty in the climate sensitivity parameter. As we proceed in time and more observations of climate behavior are collected, the variability of this parameter shrinks and if no tipping point occurs, the density concentrates around the lower bound of the distribution.

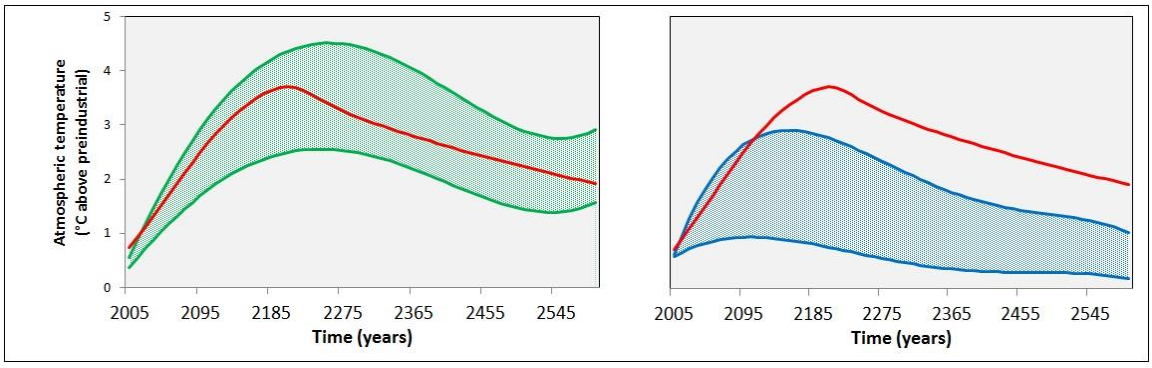


Figure 14: The 5% and 95% percentiles of the global mean temperature under optimal Bayesian stochastic actions with two uncertainties from the climate sensitivity parameter and the probability of tipping point events. The left graph shows the temperature range when the tipping point happens at $\boldsymbol{t = 10}$ and the right panel shows the temperature range when no extreme event is observed throughout the modeling horizon. The red line in both graphs shows the global mean temperature under the optimal deterministic policy.

Figure 14 shows the model results for global mean temperature for two cases, of having an extreme event at time $t = 10$ (year 2095) and observing no extreme event over the entire modeling time horizon. The left graph shows a pattern indicating the slow rate of learning in the Bayesian model when the tipping point happens at time $t = 10$. However, in the case of no extreme event (right panel), the temperature peak and most of its range are below the deterministic case as a result of the increase in probability of observing lower climate sensitivity values. Another interesting observation is that the number of cases with no extreme event is about 20% in the Bayesian ADP model compared to only 7% without Bayesian learning. This shows the power of adaptation in this case: with the observation of ``no tipping point'' made in a time step t, as modeled there will be less chance of having a higher climate sensitivity, which diminishes the probability of observing an extreme event in the next period.

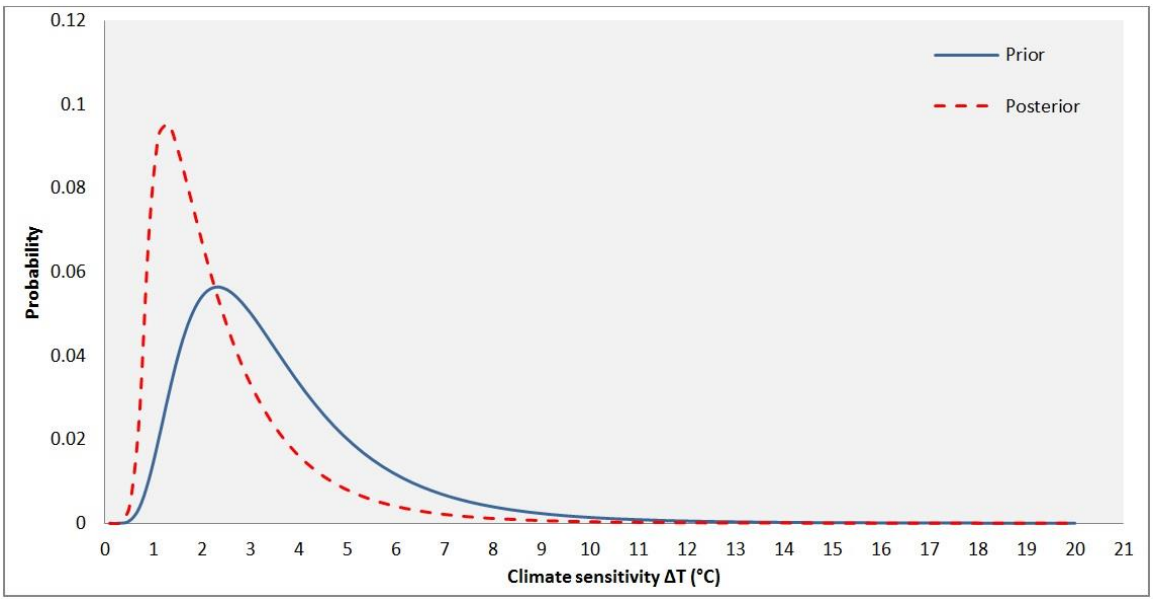


Figure 15: A posterior distribution (dotted) of climate sensitivity when no extreme event is observed throughout the modeling horizon. The solid lined graph shows the prior distribution which is a truncated lognormal distribution between **0**℃ and **20**℃.

Figure 15 shows that in the case where no extreme event is observed, the probability distribution of climate sensitivity is updated and brings a new posterior distribution with a thinner tail compared to the prior distribution. It should be noted that the climate sensitivity will be updated after observing an extreme event as well.

The BADP algorithm was deployed in both cases in this paper: in the stochastic case and in the Bayesian stochastic with tipping point case. We perform a numerical comparison between these two models with a focus on time epoch $t = 10$ (i.e. 100 years into the future). The results of the optimal policy and optimal action from the value iteration algorithm in different cases are presented in Figure 16.

Deterministic case: In the deterministic case, the optimal "policy" (parameter in the value function approximation) is $\hbar_{10} = 7.329$ and the optimal action is $a_{10}^{*} = 0.41$. In other words, the final value function approximation for t = 10 has the form $\bar{V}_{10}(S_{10}^{a}) = 7.329U_{10}(S_2^0, 0)$, where S_2^0 is the second state constructed by deploying action a and action zero accordingly having the climate sensitivity parameter level fixed at its $\Delta T = 3$℃ level. The optimal action $a_{10}^{*} = 0.41$ from solving $a_{10}^{*} = max_{a_{10}}\ (U_{10}(S_{10}, a_{10}) +$

$7.329U_{10}(S_2^0, 0)\big)$. These values are shown by the dotted red line in Figure 16 and compared with the distribution of optimal values under uncertainty.

Stochastic case: In this case, the value of ΔT, the climate sensitivity, is drawn from a truncated lognormal distribution. The distribution of optimal policies forms a narrow distribution with the mean around the optimal policy under the deterministic case. Although the optimal values seem to have minor variations, the distribution is skewed to the right. Since the climate change parameter has a heavy-tailed distribution with low probabilities of high impact values, the policy response distribution skewedness to the right reflects a precautious approach that favors future gains to present utility. The optimal action is found from solving $a_{10}^* = max_{a_{10}}\ \big(U_{10}(S_{10}, a_{10}) + \hbar_{10}U_{10}(S_2^0, 0)\big)$. As shown in the top right panel, the mean of the optimal action distribution falls on the left side of the deterministic value. This may seem to suggest a lower required level of abatement in the stochastic case compared to the deterministic case. However, it is worth noting that the deterministic values were obtained using $\Delta T = 3℃$ as in the DICE 2007 model. However for the stochastic case, we use the updated distribution built on the IPCC 2013 assumptions with the mode less than 3℃. Therefore, the difference between the mean value of the optimal action in the stochastic model and the deterministic action is due to the fact that the updated climate sensitivity distribution is used in the stochastic model.

Bayesian stochastic tipping point case: In this case, the prior distribution of ΔT is modified after each observation of an extreme event. There is a risk of hitting an extreme event that is modeled in the climate damage function; in each decade with a small probability that directly depends on the observed global mean temperature, an extreme climate event is expected to affect the natural and economic system of the earth in an irreversible manner. To demonstrate the effect of such uncertainty on the optimal solution, we consider an exaggerated case of economic loss of 25% in the case of such extreme event. As shown in Figure B2, the mode of the optimal policy ($\hbar_{10} = 7.35$) is greater than the deterministic policy ($\hbar_{10} = 7.329$). The higher value of the optimal

policy ($\hbar_{10}$ the coefficient of the approximation function) indicates the higher weight of future states in the optimal decisions for the current state. The optimal action as before, can be calculated from solving $a_{10}^{*} = max_{a_{10}}\ (U_{10}(S_{10}, a_{10}) + \hbar_{10} U_{10}(S_2^0, 0))$. In the case of having two optimal values for $\hbar_{10}$, depending on the realization of the climate sensitivity parameter, the optimal action may switch between higher values (a_{10}^{*} =0.54) in pre-extreme cases and lower values which are scattered around (a_{10}^{*} =0.33) in post-extreme cases. Due to the small probability of extreme events, the higher actions comprise about 80% of the cases.

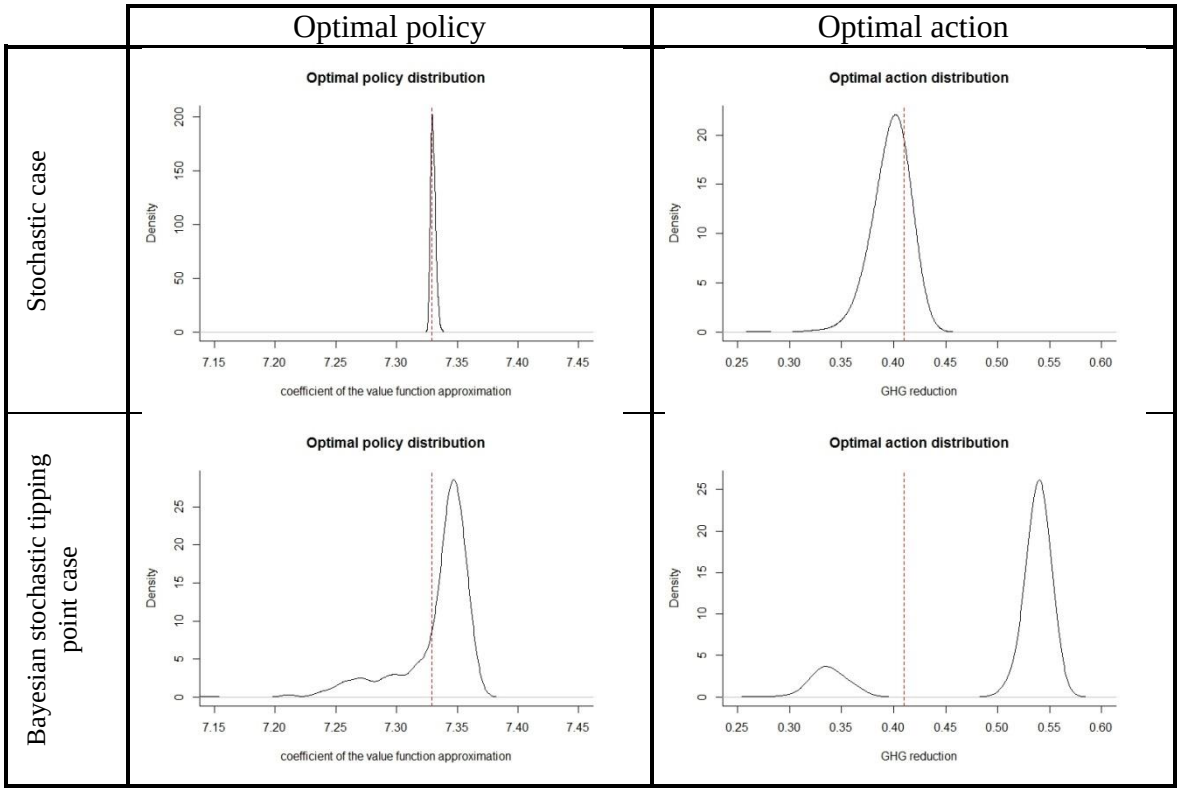


Figure 16: The results from 1000 runs of the stochastic case (upper charts) and the Bayesian stochastic with tipping point (lower charts); the vertical dotted lines show the results of the deterministic model. In the stochastic case the optimal policy remains relatively close to its optimal deterministic value although the corresponding optimal GHG reduction action ranges from about 35% to 42%. In the Bayesian stochastic tipping point case, the optimal policy's variation is larger due to the impact of tipping points and it clusters around a higher value compared to the deterministic case. The optimal action in this case however, forms two distinct peaks corresponding to higher abatement actions before a tipping point happens and lower abatement action after a tipping point happens.

CHAPTER 5

ACTIVE LEARNING IN POWER GENERATION EXPANSION PLANNING PROBLEM

Background

Limiting long-term climate change would require substantial energy system transformation including the fast decarbonization of the electricity sector [65], [66]. Current short term international pledges to reduce greenhouse gas emissions are significantly insufficient for closing the gap to reach emission levels consistent with the often stated 2℃ climate target [67]–[69]. This gap could be closed if the greenhouse gas emission reduction potentials in different sectors are fulfilled in the short term [70]. Several studies have looked at overall energy system transformation to attain long term climate goals and have highlighted the role of availability and deployment of certain lower-GHG-emission (LGE) technologies such as bioenergy and carbon capture and storage (CCS) after 2030 to compensate the likely delayed emission mitigation [71]. For the power sector, it is estimated that there will be 2.2 to 3.9 Gt CO_2 emission reduction potential per year in 2020 and 2.4-4.7 $GtCO_2$ in 2030 [72], [73]. This potential reduction is likely to include the complete phase-out of conventional coal-based power plants (without CCS) as well as lower or negative emission technologies [74], [75].

In this study, we investigate the optimal transition path from the current mix of power generation resources to a state with minimum overall cost of generation and climate change impact. We develop an integrated assessment model for electricity generation and use it to explore the impact of two sets of strategies (i.e. generation cost minimizing versus damage cost minimizing) for power expansion in the near-term. In particular, the paper is focused on the role of learning-by-doing technology improvement

that reduces not only the capital cost of power generation with a certain type of technology, but also the emissions related to the build-out phase of such technologies. The objective of this study is to better understand the relationship between power generation strategies and climate change objectives in the near-term assuming different rates of learning for generation technologies. We recognize that the plant operators have a myopic objective of minimizing the generation cost that is not aligned with long-term climate goals and therefore it is important to investigate the difference in generation technology mix and associated costs between achieving the cost minimizing objective (without any stringent climate policy) and achieving the climate damage minimizing objective. The rest of this chapter is structured as follows: In the next section we introduce baseline projections of fuel costs, generation technology costs, total electricity generation growth, and generation technology choices. Then, we introduce our optimization model, its state variables, and the decision variables. After that, the results of the optimization models with different objective functions are compared. We end this chapter with the discussion and applications of this method in integrated power generation expansion planning and assessment (IPGEPA).

Introducing Climate Change into Energy System Models

Power generation expansion planning (PGEP) models are used by utilities and government agencies to find the optimal mix of renewable and fossil fuel power plants in order to meet the expected demand under a range of generation and environmental regulation and policies [17]. The structure of these models varies greatly by their objectives, constraints, and decision variables. The PGEP objectives may include minimizing the total generation and transmission cost, maximizing the total net profit including sale of electricity, maximizing the reliability of the power network, and minimizing the pollution emitted from the electricity generation. Integrating environmental considerations into traditional PGEP models is done either by introducing

emission control measures as new constraints to the model, or alternatively assigning a secondary objective function for minimizing the generation costs and emissions iteratively.

One of the shortcomings of current PGEP models is that even when a multi objective function is adopted to take the environmental impacts into account, it is generally restricted to calculating the CO_2 emissions from different power plants while the link between the emissions and climate change is missing. Most of these models treat the environmental cost as the emission trading costs under certain GHG reduction policies [76]. The results of these studies suggest that carbon pricing is an important signal for achieving appropriate generation in the long term [77], [78]. However and in the absence of a global market for carbon trade, the emissions remain unaccounted for in the economic analysis of power generation. Therefore, there is a need for introducing a physical measure of climate change for policy making in the global scale. The change in global mean temperature that each alternative technology would produce under various schedules of deployment can provide a measurable metric for energy policy making and climate change modeling [79]. This metric has been used in studying the transition from fossil fuel power generation to an LGE power system [80]. We use the global mean surface temperature as a proxy for climate impact to calculate the damage costs and to compare different scenarios. The calculation is presented in Appendix C.

While transferring to LGE technologies seems inevitable under the long-term climate goal, the pace and quality of this transition is less certain. Since these technologies are emission intensive during the building period, their rapid deployment could in the short term substantially increase emissions and, consequently, raise the global mean surface temperature [80]. We take into account both the costs of building and operating the power plants, as well as the contingent environmental damages from the increase in the global mean temperature.

Baseline Electricity Generation Projections

Power plant costs and emissions

We design a model to find and allocate the generation for each type of power plant in every year. The model assumes a uniform decommissioning and construction rate for old and new power plants. We limit our study to eight major types of power generation: coal, natural gas, nuclear, solar PV, wind, solar thermal, coal with carbon capture and sequestration (CCS), and hydroelectric. The analysis will take into account the two stages of construction and operation for each power plant. The construction costs and associated emissions in LGE power plants are typically higher than the fossil fuel type power generation technologies, while the latter plants have higher fuel cost and emissions during their operation life. We quantify the cost of emissions through a climate change damage function. The emissions contribute to the GHG concentration in the atmosphere and the increase in the global radiative forcing. The change in the global mean surface temperature is linked to the radiative forcing through a simple energy-balance model. Information about the cost and emissions for each power plant type is given in tables C1 and C2 in Appendix C.

Technology Learning

New technology market penetration depends on several factors including the research, development and demonstration (RD&D) investment and the relative price of technologies in the electricity supply market. Technological learning is a process of gaining knowledge in manufacturing and production of certain type of technologies combined with the impact of the economies of scale, resulting in lower capital investment cost [81]. Learning in the early stages of technology development in renewable and LGE energy technologies has been faster than learning in mature and conventional technologies such as coal and natural gas [82]. However, the limitations of learning

curves such as large uncertainties in future cost development, should be taken into account when these curves are used for energy policy purposes [83]. One remedy is to determine the learning at a component level so that the overall learning of a technology can be explained as the sum of its components [84]. Using this concept, the National Energy Modeling System (NEMS) has incorporated endogenous learning into its cost calculations for power plants [85]. We use the learning mechanism devised in NEMS to calculate the cost and emission reduction in each technology over time. The current power generation technologies are classified into one of the three stages of learning: Radical Technologies that are new and untested are assigned high rates of learning, Incremental Technologies are considered to have potential for significant learning and commercialization and are assigned moderate learning rates, and Mature Technologies are well-known with an established market and are assigned low learning rates. Cost reduction per doubling of capacity is based on maturity of the technology or vintage.

Demand Projection

The International Energy Outlook 2013 (IEO 2013) is an assessment by the U.S. Energy Information Administration (EIA) of the current and future state of international energy markets through 2040 [86]. In the IEO 2013 Reference Case, future legislation or policies that might affect energy markets are not considered. We use the Reference Case projections to build our baseline projection. According to these projections, total net power generation will increase by 93% from 20.2 trillion kWh in 2010 to 39.0 trillion kWh in 2040. The increase is not uniform across different power generation technologies. While the share of natural gas and wind power plants in total generation experience a dramatic increase, coal's share continues to decline. The share of other renewables and nuclear power will have a steady but slow growth. Figure 17a shows the IEO 2013 Reference Case projections for different technologies. In Figure 17b we show eight selected technologies and their new installed capacity to meet the projected demand in

Figure 17a. The generation deficit for each type of power plant is calculated as the difference between the projected generation in the next time step and the carried-over generation after adding new power plants and subtracting the decommissioned plants.

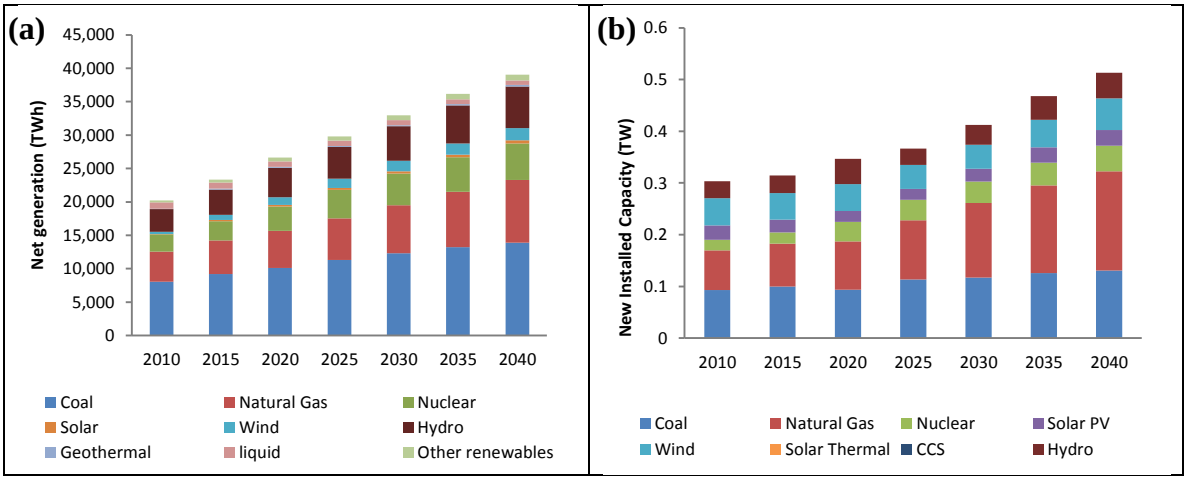


Figure 17: IEO2013 projections for different technologies from 2010 to 2040 (a) Net generation, (b) New installed capacity for selected technologies.

Discount Rate

Achieving sensible and applicable results from any PGEP model depends on the choice of proper discount rate for the future costs. The International Energy Agency (IEA) uses two discount rates, 5% and 10%, to account for the investing risks in electricity generation market around the world [87]. On the other hand, the EIA uses 7% as the real discount rate for evaluating energy efficiency investments [88]. Other approaches have been introduced, including declining discount rates for long-term projects [89], [90]. Following the IEA's methodology, we apply two discount rates, 5% and 10% to our analysis of the optimal allocation of generation technologies.

Optimization Model

Modeling power generation expansion planning is often a large scale mixed integer programming optimization problem. The decision variables are the type of power plant (binary) and the amount of power needed to be generated or purchased from each

type of power plant. The objective is to find the minimum cost of generation expansion through a finite time horizon. The optimization problem is subject to some operational and physical constraints that may include, but are not limited to, power demand constraints (demand at any given time must be met with enough power generation/purchase), generation constraints (total generation of any power plant at any given time cannot exceed the available generation of that power plant), thermal energy availability constraints (power generation at any type of power plant is bound to an operational limit defined by its capacity factor), and environmental constraints (*e.g.* a cap on GHG emissions from power generation).

Generation Modeling

We design an optimization model to find the best mix of new power generation technologies to 2040 to generate electricity sufficient to meet demand in the IEO 2013 Reference Case projection. The generation at each time step t can be modeled as a state variable with four dimensions: new under-construction generation S_t^1, carried-over under-construction generation S_t^2, new operation generation S_t^3, and carried-over operation generation S_t^4. We assume uniform initiation and retirement for each type of power plant. The relationships among the state variables are shown in the Appendix C. The decision variable X in each time step t is the percentage of the new generation that should be designated to type i power plant. The objective function to minimize the overall cost of electricity generation over the finite time horizon $\mathrm{t} = 1 \ldots \mathrm{T}$ is shown in Equation (32) below,

$$\underset{X_t^i}{\mathrm{Min}} \sum_{t=1}^{T} \gamma^{t-1} \left\{ \sum_{i=1}^{8} \left(\frac{c^i}{h^i} \left(S_t^1 X_t^i + S_t^{2,i} \right) + \frac{f^i}{h^i} \left(S_t^{3,i} + S_t^{4,i} \right) + 8760 v^i \left(S_t^{3,i} + S_t^{4,i} \right) \right) \right\} \tag{32}$$

where c^i, f^i, and v^i are construction cost, fixed operation cost, and variable operation cost. h^i is the capacity factor. $S_t^{2,i}$, $S_t^{3,i}$, and $S_t^{4,i}$ denote the carried-over under-

construction, new operation, and carried-over operation capacities of type i power plant at time t. The discount factor γ is chosen to reflect the time preference throughout the modeling horizon.

An alternative objective function is designed to minimize the cost associated with the increase in the mean global surface temperature due to emissions from power generation. In this case, we adopt an energy balance model to calculate ΔT_t, the mean surface temperature increase at each time period t. Although the choice of damage function is open, many integrated assessment models of climate change use either an exponential or power function of ΔT_t to estimate the economic damage of the increase in global mean surface temperature [91]. Here we take an exponential as our damage function and therefore the damage minimizing objective function can be expressed as

$$\underset{X_t^i}{\text{Min}} \sum_{t=1}^{T} \left(\gamma^{t-1} e^{\theta \times \Delta T_{t+1}}\right) \tag{33}$$

where θ is the scaling coefficient of the damage cost function. The emissions are calculated from the construction of the new power plants and operation of old power plants.

Constraints

The generation constraints define the feasibility of optimal solution to the above minimization problem. The optimization model of the global power system is constrained by physical and technological constraints. By definition, the decision variable X at every time step and for each power plant type is a non-negative real number between zero and one. Furthermore, the total new generation at any given time should be equal to the demand deficit from adding new generation and retiring old power plants. This will be guaranteed by constraining the sum of X_t^i at each time step t to one: $\sum_{i=1}^{8} X_t^i = 1$. We also keep the hydropower generation below 10% of the total new generation at every time step.

Results

The optimization model was solved for two different objective functions. To capture the long-term impacts of newly planned power plants, the models were constructed and solved for 50 years but the results are reported for the first 30 years for comparison with the baseline projection. Each objective function was run in two cases, with and without learning functions. Scenarios are identified by a combination of their learning assumption and discount rate (e.g., NL-10), as summarized in Table 5.

Table 5: Summary of scenario abbreviations

Scenario	Learning	Discount rate
NL-10	No	10%
L-10	Yes	10%
NL-5	No	5%
L-5	Yes	5%

The optimization model results from each scenario are compared with the baseline projection (IEO 2013 Reference Case). The results show significant difference among the optimal mix of technologies under different objective functions.

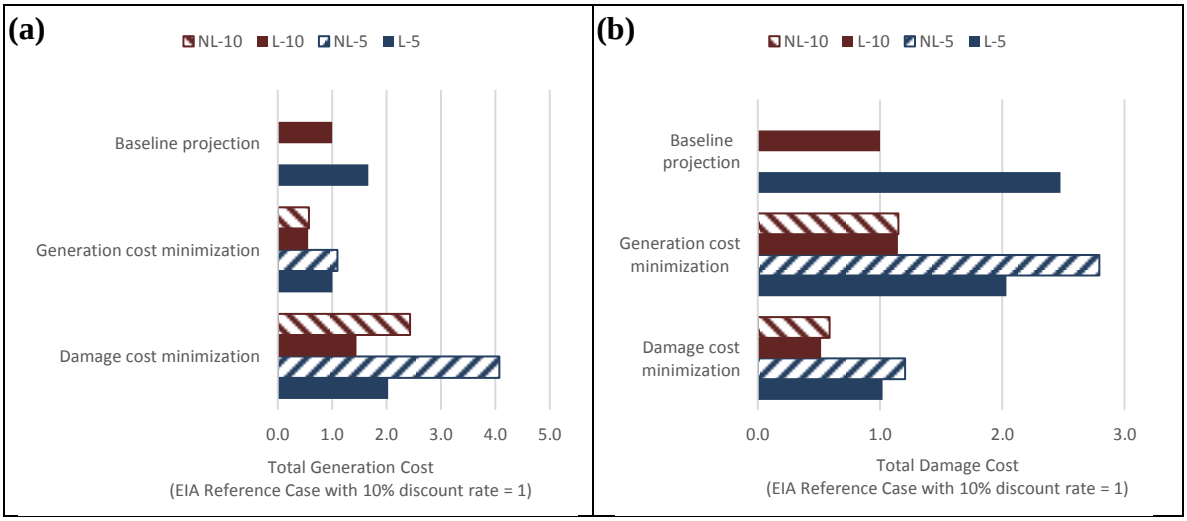


Figure 18: Comparison of costs (a) Generation costs (b) Damage costs.

As expected, generation cost minimization models show a significantly lower net generation cost compare to their counterpart damage cost minimization models (Figure 18a). However, the generation costs of the damage minimizing models are higher than

those in the baseline projection. While higher discount rate sharply reduces the net present value in all models, the impact of learning is not homogenous across different models.

Figure 19 and Figure 20 show the optimal portfolio and the new installed capacity in different scenarios. Since the renewable technologies have relatively lower capacity factors (i.e. the ratio of the actual output to the potential output), the optimal portfolios with renewables require larger new installed capacity than those with fossil fuel technologies (e.g. compare Figure 20b with Figure 20a). In the generation cost minimization scenarios, learning helps lower the costs but does not change the optimal portfolio in the scenarios with the higher interest rate (Figure 19a-d and Figure 20a-d). In damage cost minimization scenarios (Figure 19e-h and Figure 20e-h) learning changes the costs very early on by introducing wind (in L-5) or solar thermal (L-10) to the optimal portfolio. Natural gas is the overall cheapest available technology and as expected dominates the portfolio in all generation cost minimization scenarios. Solar thermal is the only new technology that enters the generating cost minimizing portfolio in the learning case with lower discount rate (Figure 19b and Figure 20b), however this happens only later in the future and therefore its cost impact is curbed through discounting. On the other hand, learning changes the optimal portfolio early on in the damage cost minimization scenarios (Figure 19f, Figure 19h, Figure 20f, and Figure 20h). Although solar PV is the dominant technology in this set of scenarios, in L-5 wind power enters the portfolio at the beginning but later on is replaced by solar thermal. Similarly in L-10, solar thermal starts to grow fast and will replace solar PV towards the end of the modeling horizon. In both cases the cost structure changes in the early years and therefore the overall generation cost is significantly different from those cases without learning.

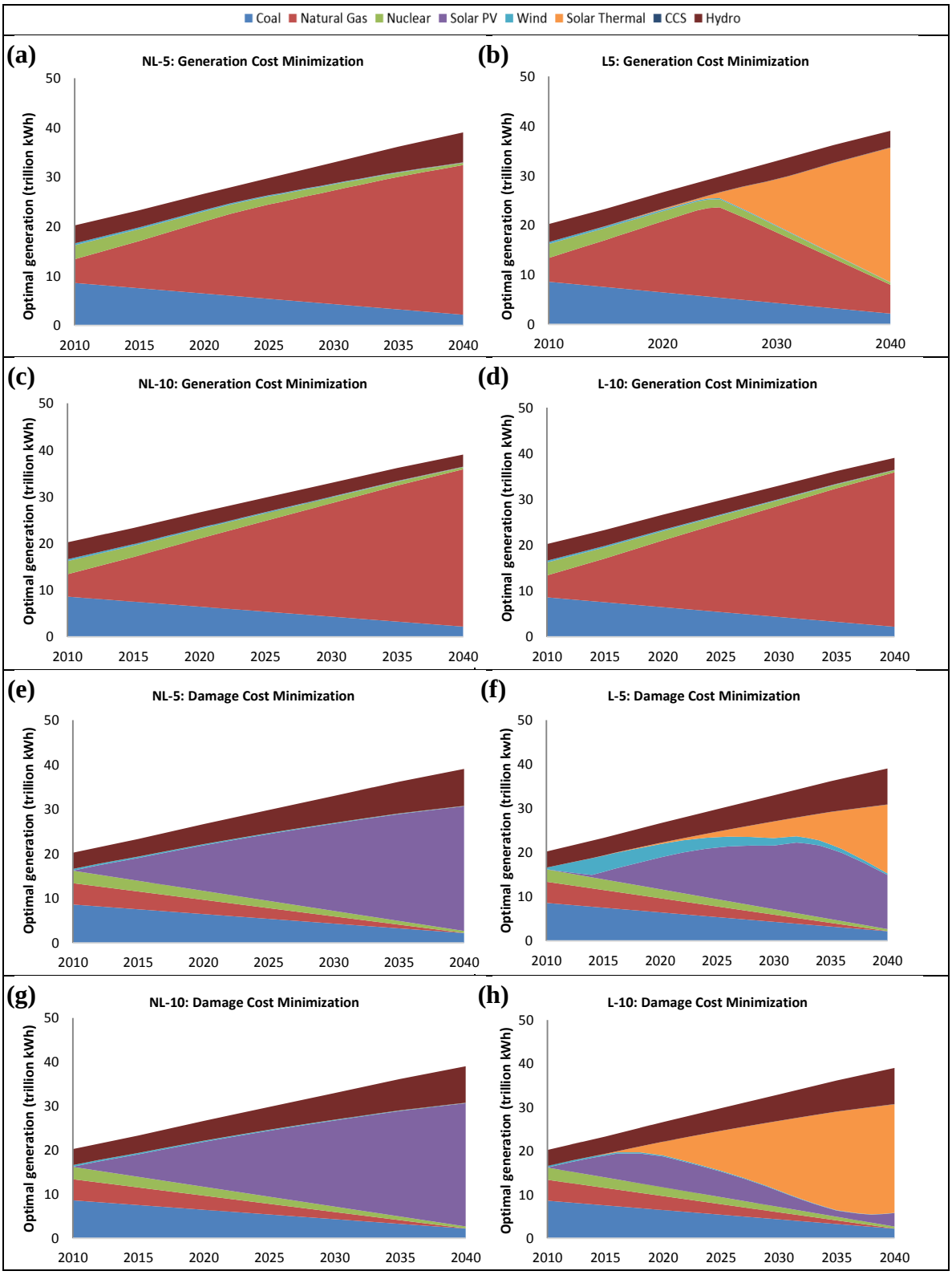


Figure 19: Optimal portfolios: scenarios without learning are on the right and those with learning are on the left. (a)-(d) show generation cost minimization scenarios and (e)-(h) show damage cost minimization scenarios. (a), (b), (e), and (f) are scenarios with 5% discount rate while (c), (d), (g), and (h) are scenarios with 10% discount rate.

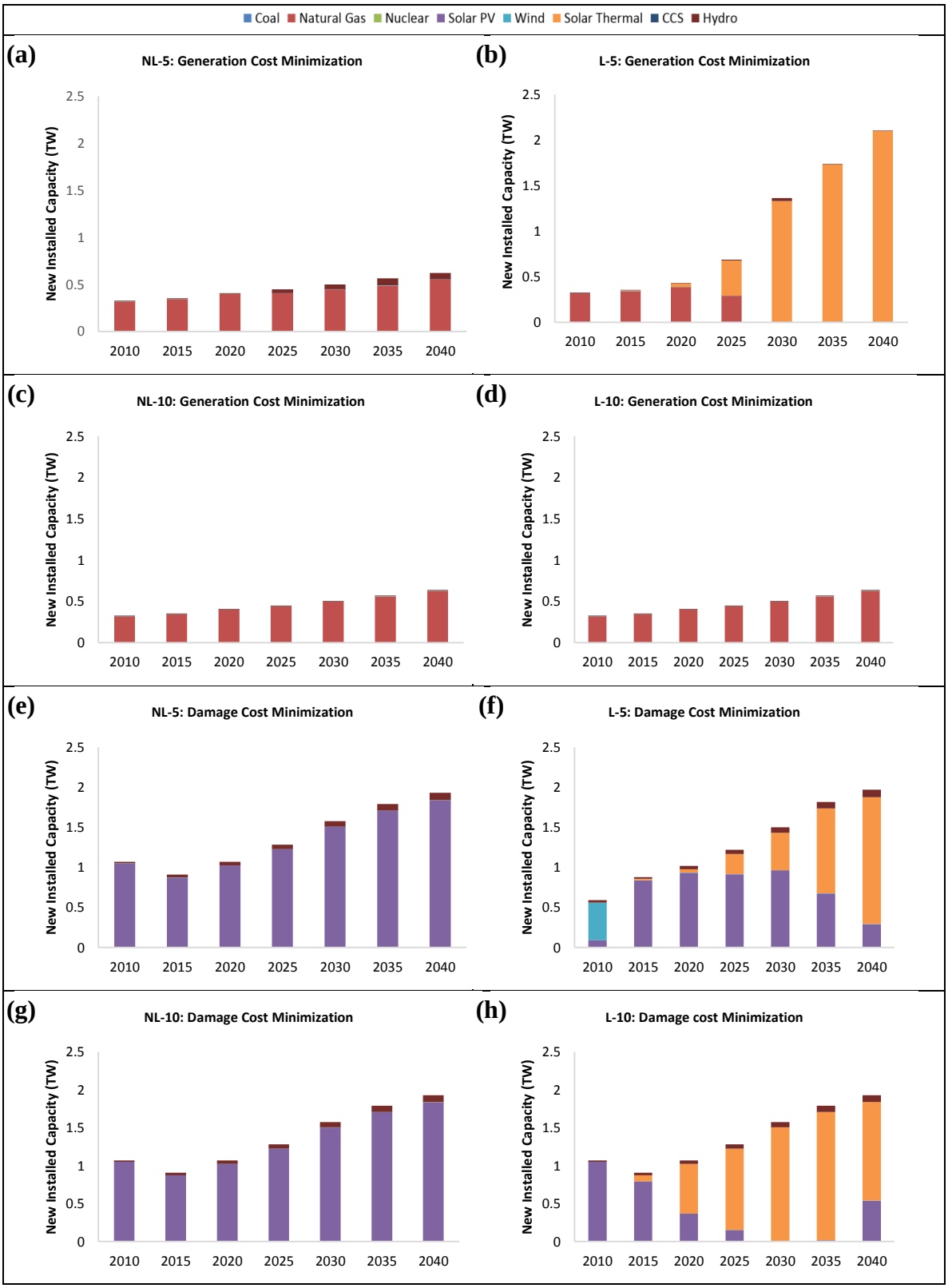


Figure 20: New installed capacity: scenarios without learning are on the right and those with learning are on the left. (a)-(d) show generation cost minimization scenarios and (e)-(h) show damage cost minimization scenarios. (a), (b), (e), and (f) are scenarios with 5% discount rate while (c), (d), (g), and (h) are scenarios with 10% discount rate.

In the climate damage minimizing scenarios, the cost difference between scenarios with and without learning is not as profound (Figure 18b). While generation cost minimization scenarios in general demonstrate a comparable damage cost as the baseline projection, the damage cost in damage cost minimization models is almost half of that in baseline projection. Although such difference reflects the structural changes in generation portfolios (damage cost minimizing models are dominated by renewable technologies), the quantitative results relate to how we defined the damage function and its relationship to the global mean surface temperature in Equation (33).

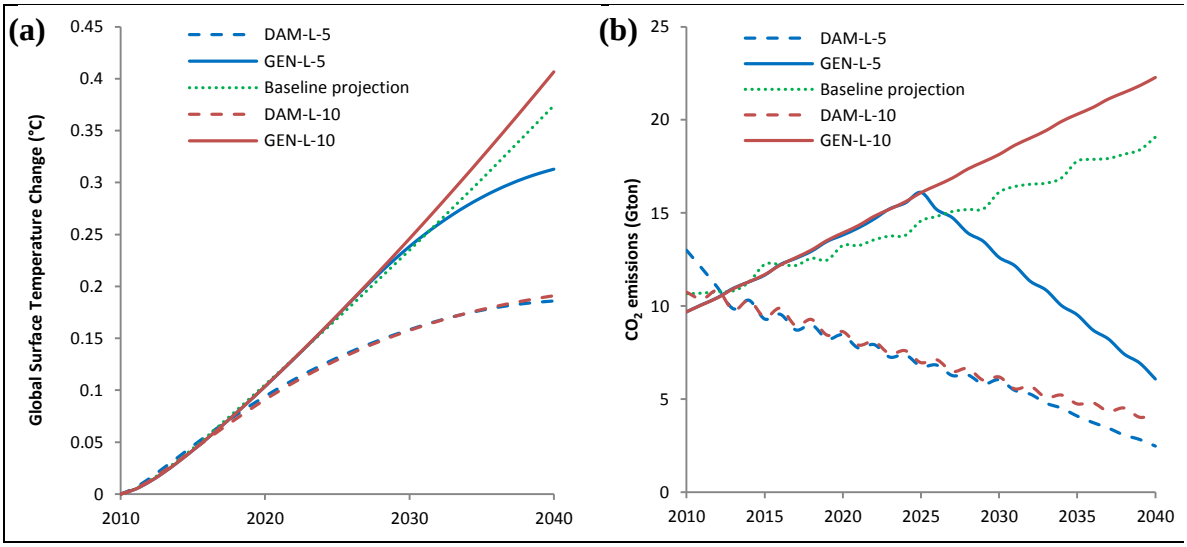


Figure 21: The results from the baseline and the optimization models with learning only. (a) Global mean surface temperature change compared to the 2010 level (b) CO_2 emissions. Damage cost minimization models (DAM) demonstrate lower emissions and temperature increase than generation cost minimization models (GEN).

The change in the global mean surface temperature is tied to the choice of the generation mix. Low greenhouse gas (LGE) technologies such as solar PV, wind, and solar thermal are more favored in the damage cost minimization problem despite the high initial emissions during their construction. Figure 21 shows the different trajectories for the CO_2 emissions and the global mean surface temperature change under the optimal power generation expansion planning. Switching from natural gas to solar thermal in the last decade in the generation cost minimization model GEN-L-5 model (Figure 19b and

Figure 20b) will sharply reduce the CO_2 concentration to a level close to the one from damage cost minimization portfolios (Figure 21b). However, it fails to fully compensate for early emissions made by the choice of natural gas that set the tone for the upward trend in temperature change (Figure 21a).

The analysis of cost structure in different models is presented in Figures C3 and C4 in Appendix C. In cost minimizing scenarios (Figure C3) natural gas dominates the portfolio and therefore fuel cost is the major component of the total cost. In damage cost minimization models (Figure C4) however, renewable technologies such as solar and wind are dominant and therefore construction cost makes up the majority of the total cost. In the learning case, the construction costs fall as the total cumulative investment in a certain technology increases. As shown in Figure 18a, while damage minimizing models have a different portfolio than the baseline projection, the total cost of achieving the minimum temperature increase is not largely different from the baseline costs. This has a significant implications for climate policy since it shows that the overall cost under stringent policies cannot be considered as a major obstacle in devising such policies.

Discussion and Conclusion

In this chapter we introduced a novel approach to the global power generation expansion planning problem, by incorporating an endogenous learning mechanism to update the construction cost and emissions as a function of cumulative built capacity for each technology. We also included an energy balance system to translate the emissions from construction and operation phases into the change in the global mean surface temperature. The optimization was performed with two objective functions – minimizing generation cost or minimizing climate damage – with or without considering the learning effect of investment in each technology. To investigate the sensitivity of our analysis to the time value of investments, we also applied two different discount rates for calculating the net present value of investment in each scenario. In total, eight scenarios were

developed, optimized, and analyzed based on these model variations. Comparing the results of optimal mix shows that the optimal solutions differ under learning assumptions particularly for the damage cost minimizing objective.

The analysis shows the impact of discounting on the optimal mix of technologies. In the case of generation cost minimization, with a lower discount rate, the costs can be felt more in today's investment calculations and therefore more attempts are made to have lower cost technologies in the future by early investment in radical technologies such as solar thermal. While natural gas dominates the optimal generation portfolio without learning in the generation cost minimization problem, renewable power plants will gradually replace the conventional coal-fired power plants in the optimal portfolio under the learning assumption. On the other hand, the damage cost minimization portfolio demonstrates different behavior under the learning assumption. While solar PV is the main contributor to the optimal portfolio without the learning assumption, solar thermal adds to it when learning is taken into account in the model. These results were compared against the IEO 2013 Reference Case as the baseline.

Achieving the minimum temperature increase is insensitive to the choice of the discount rate; that is, the damage minimizing scenarios DAM-L-5 and DAM-L-10 in Figure 21a have almost the same temperature increase trajectory, although they have different optimal mixes of power generation technologies.). This establishes a lower bound for a near-term climate policy target for electricity generation that meets the projected demand, and in the absence of pre-mature retirement of existing power plants. The corresponding lower bound on the CO_2 emissions is also established the same way in Figure 21b. It also shows that although lower emissions are achieved even under generation cost minimization alone, this requires sufficient investment in early stages and even then the temperature increases substantially. In other words, the temperature path is "stickier" than the emission path.

APPENDIX A

DICE OPTIMIZATION PROBLEM

Total factor of Productivity (TFP)

A1 $$A_g(t) = A_{g0} \times e^{[-A_{gd} \times 10 \times (t-1)]}$$

A2 $$A(t) = \frac{A(t-1)}{1 - A_g(t-1)}$$

Labor

A3 $$L(t) = L_0 \times \Big(1 - L_g(t)\Big) + L_A \times L_g(t)$$

A4 $$L_g(t) = \frac{e^{[L_{g0} \times (t-1)]} - 1}{e^{[L_{g0} \times (t-1)]}}$$

Capital

A5 $$K(t) = I(t-1) + (1 - \delta_K)^{10} \times K(t-1)$$

A6 $$I(t) = s(t) \times Q(t)$$

Production

A7 $$Y(t) = Y_0 \times \left(\frac{A(t)}{A_0}\right) \times \left(\frac{K(t)}{K_0}\right)^{\beta} \times \left(\frac{L(t)}{L_0}\right)^{1-\beta}$$

A8 $$Q(t) = \big(\Omega(t) - \Lambda(t)\big) \times Y(t)$$

A9 $$Q(t) = C(t) + I(t)$$

Climate Change Damage

A10 $$\Omega(t) = \frac{1}{1 + \Psi_1 \times T_{at}(t) + \Psi_2 \times T_{at}^{\Psi_3}(t)}$$

A11 $$T_{at}(t) = T_{at}(t-1) + \xi_1 \times \{F(t) - \xi_2 \times T_{at}(t-1) - \xi_3 \times [T_{at}(t-1) - T_{lo}(t-1)]\}$$

A12 $$T_{lo}(t) = T_{lo}(t-1) + \xi_4 \times [T_{at}(t-1) - T_{lo}(t-1)]$$

A13 $$\mathrm{F}(t) = \Delta R_f \times \log_2\left(\frac{M_{at}(t)}{M_{at}(1750)}\right) + F_{ex}(t)$$

A14 $$M_{at}(t) = E(t-1) + \phi_{11} \times M_{at}(t-1) + \phi_{21} \times M_{up}(t-1)$$

A15 $$M_{up}(t) = \phi_{12} \times M_{at}(t-1) + \phi_{22} \times M_{up}(t-1) + \phi_{32} \times M_{lo}(t-1)$$

A16 $$M_{lo}(t) = \phi_{23} \times M_{up}(t-1) + \phi_{33} \times M_{lo}(t-1)$$

A17 $$\xi_2 = \frac{\Delta R_f}{\Delta T}$$

A18 $$E(t) = E_{ind}(t) + E_{land}(t)$$

A19 $$E_{ind}(t) = \sigma(t) \times [1 - a(t)] \times Y(t)$$

A20 $$E_{land}(t) = E_0 \times 0.9^{t-1}$$

A21 $$\sigma(t) = \frac{\sigma(t-1)}{1 - \sigma_g(t)}$$

A22 $$\sigma_g(t) = \sigma_{g0} \times e^{[-10 \times \sigma_{gd} \times (t-1) - 10 \times \sigma_{ga} \times (t-1)^2]}$$

A23 $$F_{ex}(t) = F_{2000} + (F_{2100} - F_{2000}) \times (t-1)$$

Climate Change Abatement

A24 $$\Lambda(t) = \pi^{(1-\theta_2)}(t) \times \theta_1(t) \times a^{\theta_2}(t)$$

A25 $$\pi(t) = \varphi^{1-\theta_2}(t)$$

A26 $$\theta_1(t) = \left(\frac{P_b \times \sigma(t)}{\theta_2}\right) \times \left(\frac{P_r - 1 + e^{[-P_d \times (t-1)]}}{P_r}\right)$$

Utility

A27 $$U(t) = u(t) \times L(t)$$

A28 $$u(t) = \frac{c^{1-\alpha}(t)}{1-\alpha}$$

A29 $$c(t) = \frac{C(t)}{L(t)}$$

A30 $$R(t) = \frac{R(t-1)}{(1+\rho)^{10}}$$

APPENDIX B

PROOF OF THE MAIN THEOREM

This appendix outlines the proof of the main theorem from chapter 3:

Theorem 1: Let T be the value iteration operator defined by Equation (16), and let F be the H-step-ahead value function approximation. The value iteration algorithm defined in Table 2 converges to $V^*(S_t)$, the true value of the episodic state S_t for every $0 \le t \le T - H$ with step size $\alpha_n^i \le \frac{\delta}{HU_i^2}$ where $\delta \in (0,1)$ and $i \in (t+1, t+H)$.

Proof: To prove this theorem, we first need to show that the H-step-ahead value function approximation $\bar{V}_t^n(S_t)$ is an averager (i.e. the weighted average of target values $\hat{V}_t^i(S_t)$ for $i = 1,..,n$).

Recall from the direct gradient update Equation (18), for $i \in (t+1, t+H)$ we have:

$$\bar{\theta}_i^n = \bar{\theta}_i^{n-1} - \alpha_n^1 \left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)\frac{\partial \bar{V}_t^{n-1}}{\partial \theta_i^{n-1}} = \bar{\theta}_i^{n-1} - \frac{\delta}{HU_i^2}\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)\frac{\partial \bar{V}_t^{n-1}}{\partial \theta_i^{n-1}}$$

Using Equation (22) we can calculate $\frac{\partial \bar{V}_t^{n-1}}{\partial \theta_t^{n-1}} = U_i$ and substitute it in the above equation will give:

$$\bar{\theta}_i^n = \bar{\theta}_i^{n-1} - \frac{\delta}{HU_i}\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right) \rightarrow U_i\bar{\theta}_i^{n-1} - U_i\bar{\theta}_i^n = \frac{\delta}{H}\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)$$

$$\sum_{i=t+1}^{t+H}\left(U_i\bar{\theta}_i^{n-1} - U_i\bar{\theta}_i^n\right) = \sum_{i=t+1}^{t+H}\frac{\delta}{H}\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right) = \delta\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)$$

$$\sum_{i=t+1}^{t+H} U_i\bar{\theta}_i^{n-1} - \sum_{i=t+1}^{t+H} U_i\bar{\theta}_i^n = \delta\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right)$$

$$\bar{V}_t^{n-1}(S_t) - \bar{V}_t^n(S_t) = \delta\left(\bar{V}_t^{n-1}(S_t) - \hat{V}_t^n(S_t)\right) \rightarrow \bar{V}_t^n(S_t) = (1-\delta)\bar{V}_t^{n-1}(S_t) + \delta\hat{V}_t^n(S_t)$$

$$\bar{V}_t^n(S_t) = (1-\delta)\bar{V}_t^{n-1}(S_t) + \delta\hat{V}_t^n(S_t) = (1-\delta)\left((1-\delta)\bar{V}_t^{n-2}(S_t) + \delta\hat{V}_t^{n-1}(S_t)\right) + \delta\hat{V}_t^n(S_t)$$

$$\bar{V}_t^n(S_t) = (1-\delta)^n\bar{V}_t^0 + \delta\sum_{k=0}^{n-1}(1-\delta)^k\hat{V}_t^{n-k}(S_t)$$

To show that $\bar{V}_t^n(S_t)$ is an average we need to show that the sum of the coefficients on the right hand side is one:

$$(1-\delta)^n + \delta\sum_{k=0}^{n-1}(1-\delta)^k = (1-\delta)^n + \delta\frac{1-(1-\delta)^n}{1-(1-\delta)} = (1-\delta)^n + 1 - (1-\delta)^n = 1$$

Since we can assume any initial constant value for $\bar{V}_t^0$, we showed that the approximation function $\bar{V}_t^n(S_t)$ is the average of target values $\hat{V}_t^i(S_t)$ for $i = 1,..,n$. Using Theorem 3.2 from [92] we can prove that F is a nonexpansion in max norm:

Let S_t and S_t' be two episodic states with target values $\hat{V}_t(S_t)$ and $\hat{V}_t(S_t')$ and let $\|.\|$ denote max norm,

$$\begin{aligned}
|\bar{V}_t^n(S_t) - \bar{V}_t^n(S_t')| &= \left|\left((1-\delta)^n\bar{V}_t^0 + \delta\sum_{k=0}^{n-1}(1-\delta)^k\hat{V}_t^{n-k}(S_t)\right) - \left((1-\delta)^n\bar{V}_t^0 + \delta\sum_{k=0}^{n-1}(1-\delta)^k\hat{V}_t^{n-k}(S_t')\right)\right| \\
&= \left|\delta\sum_{k=0}^{n-1}(1-\delta)^k\left(\hat{V}_t^{n-k}(S_t) - \hat{V}_t^{n-k}(S_t')\right)\right| \\
&\leq \max_k\left(\hat{V}_t^k(S_t) - \hat{V}_t^k(S_t')\right) \\
&= \left\|\hat{V}_t(S_t) - \hat{V}_t(S_t')\right\|
\end{aligned}$$

That means the approximation $\bar{V}_t^n$ is nonexpansion in max norm and since Equation (16) is a contracting mapping, the value iteration algorithm defined in Table 2 is converging.

APPENDIX C

INTEGRATED POWER GENERATION EXPANSION PLANNING MODEL

The dynamics of the integrated global power generation expansion planning (PGEP) problem can be better understood by defining the generation state and its components. For each technology i we define the generation state as $S_t^i = (S_t^{1,i}, S_t^{2,i}, S_t^{3,i}, S_t^{4,i}, S_t^{5,i})$, where $S_t^{1,i}$ is the new under-construction generation, $S_t^{2,i}$ is the carried-over under-construction generation, $S_t^{3,i}$ is the new operation generation, $S_t^{4,i}$ is the carried-over operation generation, and $S_t^{5,i}$ is the cumulative built capacity of type i power plant until time t. The PGEP optimization problem tries to find the optimal values of $S_t^{1,i}$ over the planning horizon in order to minimize costs associated with power generation including construction, operation, and environmental costs. The relationships among these components are demonstrated in the equations below.

The projected generation D_{t+1} determines how much new generation is needed for the next time period. Let m^i and n^i be the lead time (construction time) and operation life time of the type i power plant respectively. Given the past operation generations $(S_{1\dots t}^{3,i}, S_{1\dots t}^{4,i})$ and carried-over under construction generations $S_{1\dots t}^{2,i}$ up to the current time t, new operation generation, carried-over operation generation and therefore the generation deficit at time $t+1$ can be calculated.

$$O_{t+1}^i = \frac{1}{m^i} \sum_{j=t+1-m^i}^{t-1} S_j^{1,i}$$

$$S_{t+1}^{4,i} = \sum_{j=t+2-n^i}^{t} \left(1 - \frac{t+1-j}{n^i}\right) S_j^{3,i}$$

$$G_{t+1}^{\mathrm{i}} = \sum_{i=1}^{8} \left(O_{t+1}^i + S_{t+1}^{4,i}\right)$$

$$\mathrm{DEF}_{t+1} = D_{t+1} - \sum_{i=1}^{8} G_{t+1}^{\mathrm{i}}$$

where $O_{t+1}^{i}$is the new operation generation from construction of the power plants before time t. G_{t+1}^{i} is the gross operation generation (not including time t's generation). DEF_{t+1} is the generation deficit that needs to be fulfilled by new generation. The decision variable X_t^i is the portion of this deficit which will be assigned to the type i power plant. Therefore, the current new construction generation capacity will be calculated as:

$S_t^{1,i} = X_t^i \, m^i \, \mathrm{DEF}_{t+1}$

Since only $1/m^i$ of this capacity will be available for operation at time $t+1$, the reminder will be carried over as the under-construction capacity:

$$S_{t+1}^{2,i} = \frac{m^i - 1}{m^i}\left(S_t^{2,i} + X_t^i \, m^i \, \mathrm{DEF}_{t+1}\right)$$

The new operation generation will be

$$S_{t+1}^{3,i} = \frac{1}{m^i}\left(S_t^{2,i} + X_t^i \, m^i \, \mathrm{DEF}_{t+1}\right)$$

Cumulative built capacity is updated for the next time step as

$S_{t+1}^{5,i} = S_t^{5,i} + S_t^{3,i}$

Generation Cost

The cost analysis for different power plants is based on estimations from the U.S. Energy Information Administration (EIA) Annual Energy Outlook. These costs are used as an idealization of global cost parameters; variations in access to fuels increases costs in some locations; adoption of US cost parameters represents an idealization of global fuel access. Table S1 shows the cost parameters for different technologies. Generation expenses are divided into four main categories of Capital Cost, Fixed O&M Cost, Variable O&M Cost, and Fuel Cost. The Capital Cost includes civil, structural, mechanical, and electrical material and installation in addition to indirect and owner's cost. Fixed O&M expenses including staffing and administrative expenses are those costs

that do not vary significantly with the output of the power plant. Variable Costs on the other hand are directly related to the generation level and availability of the plant. To calculate the Fuel Costs for the coal-fired and natural gas-fired power plants, we use the EIA's fuel prices projection to 2040 [88]. For nuclear power plants a constant value for the uranium oxide price including a nuclear waste fee is added to the variable O&M cost [93].

Table C1: Power plants cost characteristics [94]

Generation Type	Lead Time (years)	Life Time (years)	Heat Rate (Btu/kWh)	Capital Cost (USD/kW)	Fixed O&M Cost (USD/kw year)	Variable O&M Cost (USD/MWh)	Fuel Cost 2011 (USD/mil Btu)
Coal	4	40	8800	3246	37.8	4.47	2.38
Natural Gas	3	30	7050	917	13.17	3.60	4.80
Nuclear	5	35	0	5530	93.28	2.14+5.60	0
Solar PV	2	20	0	3873	24.69	0	0
Wind	4	30	0	2213	39.55	0	0
Solar Thermal	2	20	0	5067	67.26	0	0
CCS	4	40	12000	5227	80.53	9.51	2.38
Hydro	4	100	0	2936	14.13	0	0

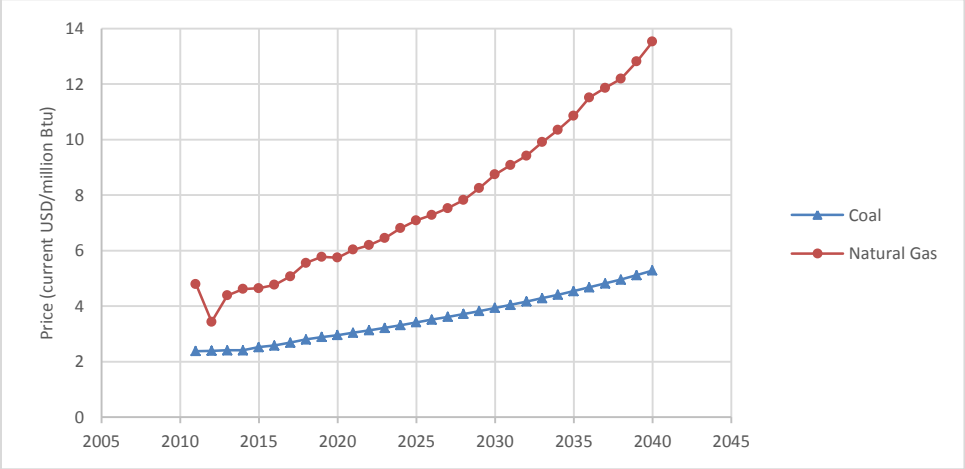


Figure C1: Coal and natural gas prices projection [88]

Learning curves

We implant an endogenous learning mechanism into the model to update the construction cost and emissions of each type of power plant based on its total generation. Therefore, the construction of any power plant type i will contribute to reduction in its construction cost and emissions in future. We define the nonlinear learning function as:

$$cc_{t+1}^{i} = a\left(S_t^{5,i}\right)^{-b}$$

where a is the baseline normalization $\left(a = cc_1^i\left(S_0^{5,i}\right)^b\right)$ and b is the learning parameter equal to $-[ln(1-LR)/ln(2)]$ and cc_{t+1}^{i} is the capital cost of type i power plants at time $t+1$. For each power plant type i, the reduction in capital cost for every doubling of operating generation (LR) is an exogenous parameter. The construction emissions possess similar learning behavior. As a technology is being installed and operated more, the capital cost and emissions of that technology will decline. Three learning rates (LR_1, LR_2, and LR_3) are considered to distinguish between different learning stages as a new technology is introduced into the market. While new untested technologies see high rates of learning initially, more conventional designs have very limited learning potential. The status of each technology along with the corresponding learning rates are presented in table C2. Radical technologies have a fast learning at the first stage but after 3 doublings (eight times the initial installed capacity) their learning rates fall in the second stage. In the last stage when the technologies are mature, a flat rate of 1% applies to all technologies.

Table S2: Power plant learning characteristics (from table 8.3 Electricity Module [88])

Generation Type	Technological Status	Learning rate LR_1	Learning rate LR_2	Learning rate LR_3	Period 1 doublings	Period 2 doublings
Coal	Mature	-	-	1%	-	-
Natural Gas	Mature	-	-	1%	-	-
Nuclear	Incremental	5%	3%	1%	3	5
Solar PV	Incremental	-	-	1%	-	-
Wind	Mature	-	-	1%	-	-
Solar Thermal	Radical	20%	10%	1%	3	5
CCS	Radical	20%	10%	1%	3	5
Hydro	Mature	-	-	1%	-	-

Figure C2 shows how the price of each technology is projected to drop at different stages after a total of 9-fold increase in installed capacity. Mature technologies show a steady rate of decline while the revolutionary technologies including Solar Thermal and CCS demonstrate a sharp decline in the early stages of deployment.

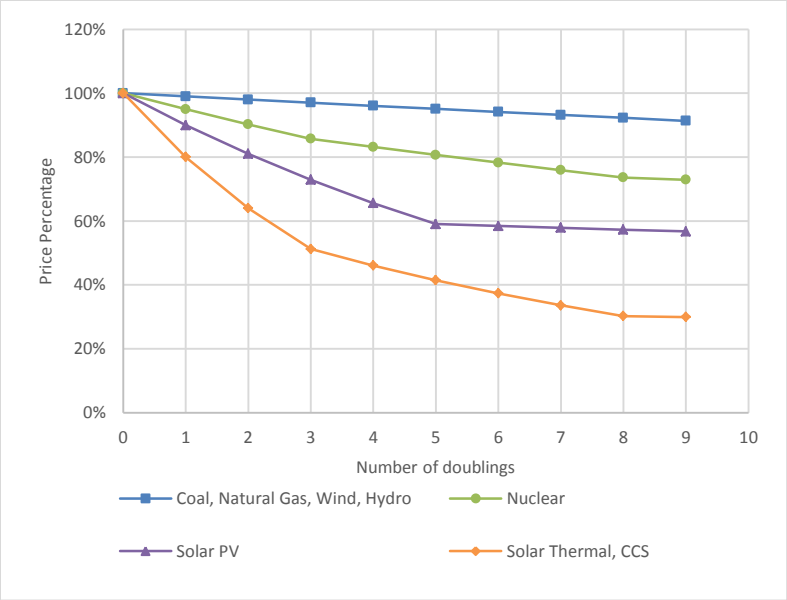


Figure C2: Learning curves for different technologies [88]

GHG emissions and concentration

We consider three types of GHG emissions for this model. The emissions from each type of power plant are calculated using the values provided in Table C3.

Table C3: Power plants emission characteristics

Generation Type	Construction Period Total			Operation Period		
	CO_2 (kg/MWe)	CH_4 (kg/MWe)	N_2O (kg/MWe)	CO_2 (kg/MWe/yr)	CH_4 (kg/MWe/yr)	N_2O (kg/MWe/yr)
Coal	2.02E+05	6.17	9.34	6.59E+06	1.21E+03	5.82E+02
Natural Gas	7.62E+05	0	0	5.30E+06	0	0
Nuclear	3.67E+07	0	0	6.66E+05	0	0
Solar PV	1.74E+07	5.33E+02	4.78E+02	0	0	0
Wind	3.25E+07	0	0	0	0	0
Solar Thermal	3.73E+07	0	0	0	0	0
CCS	4.63E+06	0	0	1.75E+06	1.32E+04	7.36E+03
Hydro	7.88E+05	5.78E+01	4.51E+01	0	0	0

The concentration from unit emission of each greenhouse gas is calculated using the concentration equations below (from Table 2.14 of [95]) :

$$M_t^{CO_2} = 0.217 + 0.259e^{-t/172.9} + 0.338e^{-t/18.51} + 0.186e^{-t/1.186}$$

$$M_t^{CH_4} = e^{-t/12}$$

$$M_t^{N_2O} = e^{-t/114}$$

To calculate the concentration of GHG emissions in the atmosphere we need to account for the methane breakdown in the atmosphere that ultimately forms CO_2 and will

add to the concentration of it. Let E_t be the GHG emission at time t and N_t be the concentration. We can write the relationships between emissions and concentration as:

$$N_t^{CO_2} = \sum_{u=1}^{t} \left(M_u^{CO_2} \left\{ E_{t-u+1}^{CO_2} + (1 - M_u^{CH_4})\, E_{t-u+1}^{CH_4} \right\} \right)$$

$$N_t^{CH_4} = \sum_{u=1}^{t} \left(M_u^{CH_4} E_{t-u+1}^{CH_4} \right)$$

$$N_t^{N_2O} = \sum_{u=1}^{t} \left(M_u^{N_2O} E_{t-u+1}^{N_2O} \right)$$

Radiative Forcing

The concentration of GHG emissions further will be translated into changes in radiative forcing. Using the Intergovernmental Panel on Climate Change's (IPCC) assessments we can derive the radiative forcing from each GHG concentration using the equations below [95]:

$$F_t^{CO_2} = 3.35 \left(f_1\left(p_{CO_2}\right) - f_1(400) \right)$$

where $f_1(x) = Ln(1 + 1.2x + 0.005x^2 + 1.4 \times 10^{-6} x^3)$ and $p_{CO_2} = 400 + N_t^{CO_2}/7820$

$$F_t^{CH_4} = 0.36\left(\sqrt{p_{CH_4}} - \sqrt{1800}\right) - f_2\left(p_{CH_4}, 320\right) + f_2(1800, 320)$$

$$F_t^{N_2O} = 0.12\left(\sqrt{p_{N_2O}} - \sqrt{320}\right) - f_2\left(1800, p_{N_2O}\right) + f_2(1800, 320)$$

where $f_2(x, y) = 0.47\, ln(1 + 2.01 \times 10^{-5} (xy)^{0.75} + 5.31 \times 10^{-15} x(xy)^{1.52})$,

$p_{CH_4} = 1800 + N_t^{CH_4}/2.844$ and $p_{N_2O} = 320 + N_t^{N_2O}/7.820$

Total change in radiative forcing can be obtained from adding radiative forcing changes corresponding to each GHG emission.

$$F_t = F_t^{CO_2} + F_t^{CH_4} + F_t^{N_2O}$$

Temperature change

In order to estimate the change in the global mean surface temperature (ΔT) from power generation emissions, we adopt a one-dimensional "slab ocean" model:

$$\frac{\partial \Delta T}{\partial t} = k \frac{\partial^2 \Delta T}{\partial z^2}$$

The boundary conditions will guarantee the existence of a numerical solution for above equation. These conditions include:

$$\left.\frac{\partial \Delta T}{\partial z}\right|_{z=0} = \left.\frac{(\lambda \Delta T - F)}{\rho k \vartheta c}\right|_{z=0}$$

$$\Delta T|_{t=0} = 0$$

$$\left.\frac{\partial \Delta T}{\partial z}\right|_{z=z_{max}} = 0$$

Choosing $z_{max} = 4000$ meters as the maximum depth of the ocean to be modeled, $\rho = 1030\ kg/m^3$ the density of seawater, $c = 4000\ J/(kg\ °K)$ the heat capacity of seawater, $k = 5.5 \times 10^{-5}\ m^2/s$ the ocean vertical thermal diffusivity, and $\vartheta = 0.7$ as the fraction of the earth covered by ocean, we will be able to calculate the temperature change due to the change in the radiative forcing. The results depend also on the choice of the climate sensitivity related parameter λ. This parameter is the ratio of the radiate forcing change to the change in ΔT from doubling of the CO_2 concentration in the atmosphere.

Results

Each optimization model is coded and solved in MATLAB R2013b. The total generation cost at each time period comprises four components namely Construction cost, Fixed O&M Cost, Variable O&M Cost, and Fuel Cost. Figures C3 and C4 show the changes in the total generation cost over time for different models.

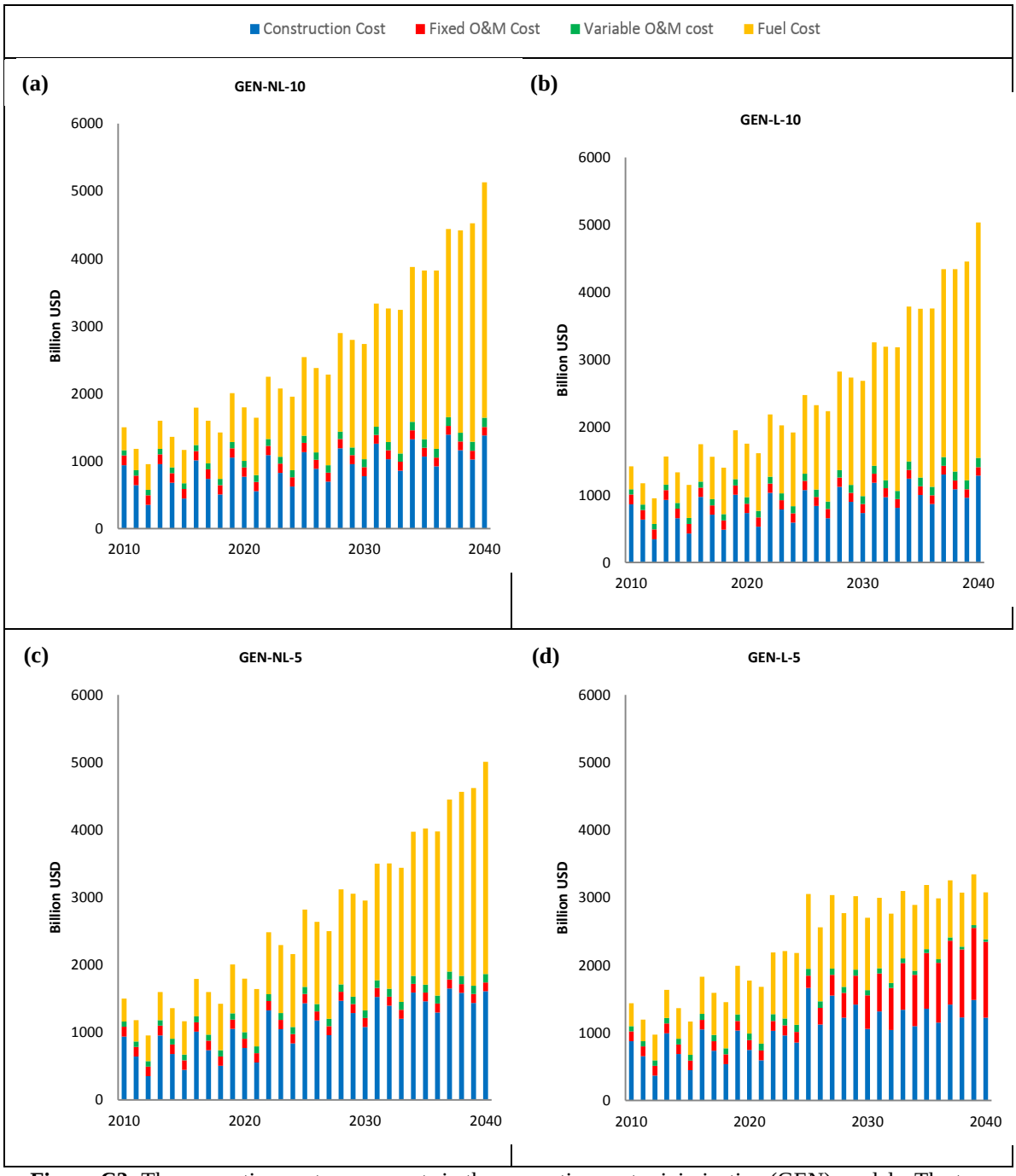


Figure C3: The generation cost components in the generation cost minimization (GEN) models. The top row (a and b) shows the model with 10% discount rate and the bottom row (c and d) shows the model with 5% discount rate. The right panels (b and d) are the cases with learning while the left panels (a and c) are without learning.

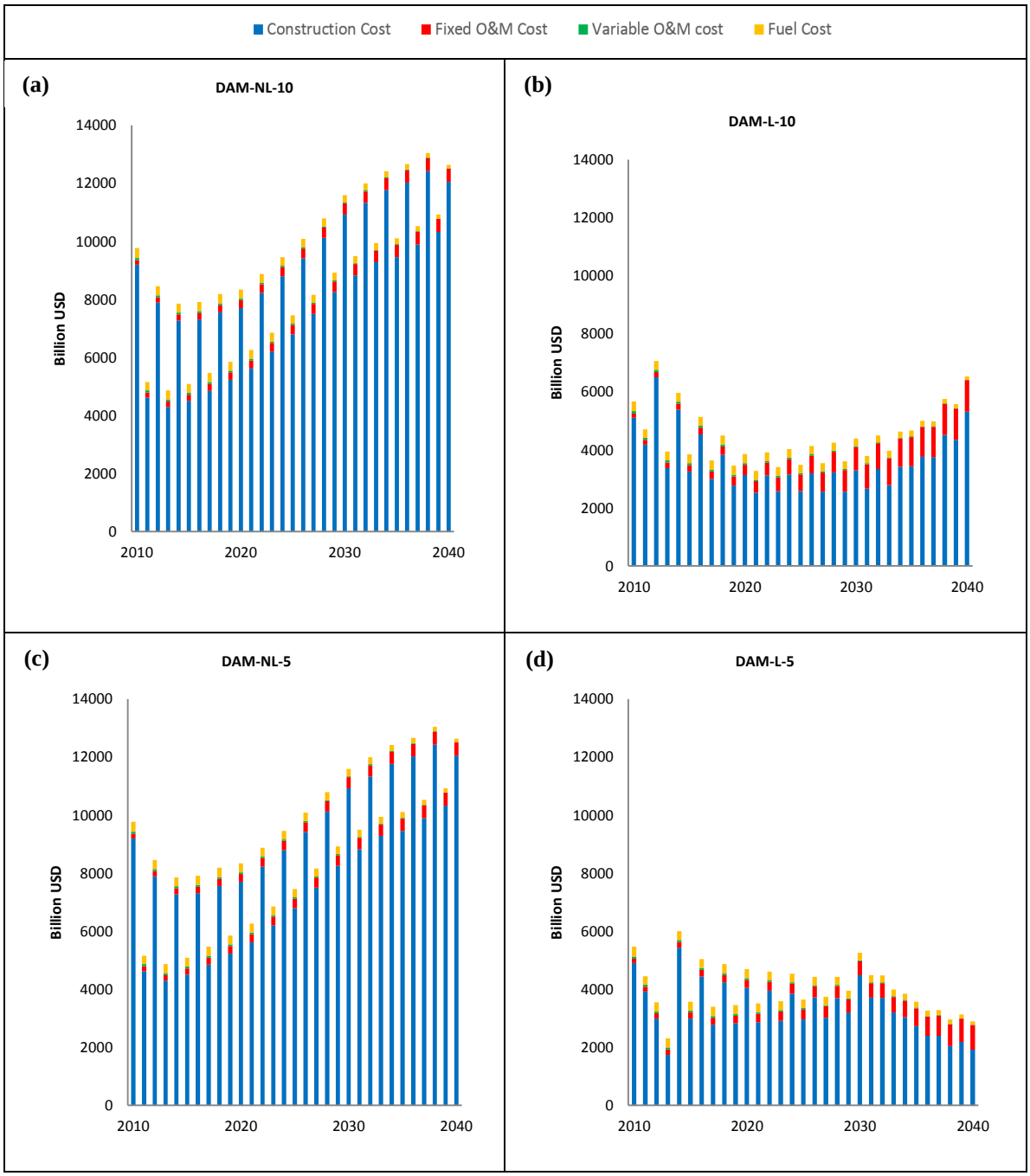


Figure C4: The generation cost components in the damage cost minimization (DAM) models. The top row (a and b) shows the model with 10% discount rate and the bottom row (c and d) shows the model with 5% discount rate. The right panels (b and d) are the cases with learning while the left panels (a and c) are without learning.

REFERENCES

[1] F. O. Hoffman and J. S. Hammonds, "Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability," *Risk Anal.*, vol. 14, no. 5, pp. 707–712, 1994.

[2] T. Aven and E. Zio, "Some considerations on the treatment of uncertainties in risk assessment for practical decision making," *Reliab. Eng. Syst. Saf.*, vol. 96, no. 1, pp. 64–74, Jan. 2011.

[3] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *ArXiv Prepr. Cs9603104*, 1996.

[4] D. L. Kelly and C. D. Kolstad, "Integrated assessment models for climate change control," in *International yearbook of environmental and resource economics*, 1999/2000 ed., Edward Elgar, 1999, pp. 171–197.

[5] E. A. Stanton, F. Ackerman, and S. Kartha, "Inside the integrated assessment models: Four issues in climate economics," *Clim. Dev.*, vol. 1, no. 2, pp. 166–184, Jul. 2009.

[6] S. H. Schneider, "Integrated assessment modeling of global climate change: Transparent rational tool for policy making or opaque screen hiding value-laden assumptions?," *Environ. Model. Assess.*, vol. 2, no. 4, pp. 229–249, 1997.

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

[8] A. Rezai, "Recast the Dice and Its Policy Recommendations," *Macroecon. Dyn.*, vol. 14, no. Supplement S2, pp. 275–289, 2010.

[9] D. Popp, "ENTICE: endogenous technological change in the DICE model of global warming," *J. Environ. Econ. Manag.*, vol. 48, no. 1, pp. 742–768, Jul. 2004.

[10] S. M. N. Islam, P. Sheehan, J. Gigas, H. Trinh, and F. Puno, "Climate change and endogenous growth theory: forecasting by the ADICE model," *Int. J. Environ. Pollut.*, vol. 8, no. 1/2, pp. 122 – 133, 1997.

[11] T. Kosugi, K. Tokimatsu, A. Kurosawa, N. Itsubo, H. Yagita, and M. Sakagami, "Internalization of the external costs of global environmental damage in an integrated assessment model," *Energy Policy*, vol. 37, no. 7, pp. 2664–2678, Jul. 2009.

[12] K. C. de Bruin, R. B. Dellink, and R. S. J. Tol, "AD-DICE: an implementation of adaptation in the DICE model," *Clim. Change*, vol. 95, no. 1–2, pp. 63–81, Jan. 2009.

[13] Y. Cai, K. Judd, and T. Lontzek, "The Social Cost of Stochastic and Irreversible Climate Change," National Bureau of Economic Research, Working Paper 18704, Jan. 2013.

[14] G. Engstroem, "Assessing Sustainable Development in a DICE World," Nov. 2009.

[15] M. L. Weitzman, "On Modeling and Interpreting the Economics of Catastrophic Climate Change," *Rev. Econ. Stat.*, vol. 91, pp. 1–19, Feb. 2009.

[16] C. J. Costello, M. G. Neubert, S. A. Polasky, and A. R. Solow, "Bounded uncertainty and climate change economics," *Proc. Natl. Acad. Sci.*, vol. 107, no. 18, pp. 8108 –8110, May 2010.

[17] S. C. Newbold and A. Daigneault, "Climate Response Uncertainty and the Benefits of Greenhouse Gas Emissions Reductions," *Environ. Resour. Econ.*, vol. 44, no. 3, pp. 351–377, May 2009.

[18] G. H. Roe and M. B. Baker, "Why Is Climate Sensitivity So Unpredictable?," *Science*, vol. 318, no. 5850, pp. 629 –632, Oct. 2007.

[19] F. Ackerman, E. A. Stanton, and R. Bueno, "Fat tails, exponents, extreme uncertainty: Simulating catastrophe in DICE," *Ecol. Econ.*, vol. 69, no. 8, pp. 1657–1665, 2010.

[20] M. D. Gerst, R. B. Howarth, and M. E. Borsuk, "Accounting for the risk of extreme outcomes in an integrated assessment of climate change," *Energy Policy*, vol. 38, no. 8, pp. 4540–4548, Aug. 2010.

[21] A. Haurie, “Integrated assessment modeling for global climate change : An infinite horizon optimization viewpoint,” *Environ. Model. Assess.*, vol. 8, no. iv, pp. 117–132, 2003.

[22] M. Webster, N. Santen, and P. Parpas, “An approximate dynamic programming framework for modeling global climate policy under decision-dependent uncertainty,” *Comput. Manag. Sci.*, vol. 9, no. 3, pp. 339–362, May 2012.

[23] Y. Cai, K. Judd, and T. Lontzek, “DSICE: A Dynamic Stochastic Integrated Model of Climate and Economy,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1992674, Jan. 2012.

[24] D. P. Bertsekas and J. N. Tsitsiklis, “Neuro-dynamic programming: an overview,” in , *Proceedings of the 34th IEEE Conference on Decision and Control, 1995*, 1995, vol. 1, pp. 560–564 vol.1.

[25] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley & Sons, 2011.

[26] Y. Cai, K. Judd, and T. Lontzek, “Continuous-Time Methods for Integrated Assessment Models,” National Bureau of Economic Research, Working Paper 18365, Sep. 2012.

[27] H. Dowlatabadi, “Integrated assessment models of climate change: An incomplete overview,” *Energy Policy*, vol. 23, no. 4–5, pp. 289–296, Apr. 1995.

[28] M. D. Mastrandrea and S. H. Schneider, “Probabilistic Integrated Assessment of ‘Dangerous’ Climate Change,” *Science*, vol. 304, no. 5670, pp. 571–575, Apr. 2004.

[29] R. N. Jones, “Managing Uncertainty in Climate Change Projections – Issues for Impact Assessment,” *Clim. Change*, vol. 45, no. 3–4, pp. 403–419, Jun. 2000.

[30] M. New and M. Hulme, “Representing uncertainty in climate change scenarios: a Monte-Carlo approach,” *Integr. Assess.*, vol. 1, no. 3, pp. 203–213, Jul. 2000.

[31] W. D. Nordhaus, “Optimal Greenhouse-Gas Reductions and Tax Policy in the ‘DICE’ Model,” *Am. Econ. Rev.*, vol. 83, no. 2, pp. 313–317, May 1993.

[32] J. M. Murphy, D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, "Quantification of modelling uncertainties in a large ensemble of climate change simulations," *Nature*, vol. 430, no. 7001, pp. 768–772, 2004.

[33] J. Rogelj, M. Meinshausen, and R. Knutti, "Global warming under old and new scenarios using IPCC climate sensitivity range estimates," *Nat. Clim. Change*, vol. 2, no. 4, pp. 248–253, 2012.

[34] T. M. Lenton and J.-C. Ciscar, "Integrating tipping points into climate impact assessments," *Clim. Change*, vol. 117, no. 3, pp. 585–597, Apr. 2013.

[35] IPCC, "Climate Change 2013: The Physical Science Basis," 2013.

[36] R. S. Pindyck, "Fat Tails, Thin Tails, and Climate Change Policy," *Rev. Environ. Econ. Policy*, vol. 5, no. 2, pp. 258 –274, Jul. 2011.

[37] C. Russill and Z. Nyssa, "The tipping point trend in climate change communication," *Glob. Environ. Change*, vol. 19, no. 3, pp. 336–344, Aug. 2009.

[38] M. L. Weitzman, "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change," *Rev. Environ. Econ. Policy*, vol. 5, no. 2, pp. 275 –292, Jul. 2011.

[39] E. Kriegler, J. W. Hall, H. Held, R. Dawson, and H. J. Schellnhuber, "Imprecise probability assessment of tipping points in the climate system," *Proc. Natl. Acad. Sci.*, Mar. 2009.

[40] C. B. Field, V. Barros, T. F. Stocker, and Q. Dahe, Eds., *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2012.

[41] K. Zickfeld, A. Levermann, M. G. Morgan, T. Kuhlbrodt, S. Rahmstorf, and D. W. Keith, "Expert judgements on the response of the Atlantic meridional overturning circulation to climate change," *Clim. Change*, vol. 82, no. 3–4, pp. 235–265, Jun. 2007.

[42] K. Zickfeld, M. G. Morgan, D. J. Frame, and D. W. Keith, "Expert judgments about transient climate response to alternative future trajectories of radiative forcing," *Proc. Natl. Acad. Sci.*, Jun. 2010.

[43] M. Geist and O. Pietquin, “Parametric value function approximation: A unified view,” in *2011 IEEE Symposium on Adaptive Dynamic Programming And Reinforcement Learning (ADPRL)*, 2011, pp. 9–16.

[44] K. P. Murphy, “A Survey of POMDP Solution Techniques,” UC Berkeley, Technical Report, 2000.

[45] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[46] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” *ArXiv Prepr. Cs9605103*, 1996.

[47] W. B. Powell, “Merging AI and OR to Solve High-Dimensional Stochastic Optimization Problems Using Approximate Dynamic Programming,” *Inf. J. Comput.*, vol. 22, no. 1, pp. 2–17, 2010.

[48] D. P. Bertsekas, *Dynamic programming and optimal control*, 4th ed. Belmont, Mass.: Athena Scientific, 2012.

[49] S. Gelly and D. Silver, “Combining online and offline knowledge in UCT,” in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 273–280.

[50] S. Davies, A. Y. Ng, and A. Moore, “Applying Online Search Techniques to Reinforcement Learning,” in *In Proceedings of the National Conference on Artificial Intelligence. 753–760*, 1998.

[51] C. Szepesvári, “Algorithms for reinforcement learning,” *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 4, no. 1, pp. 1–103, 2010.

[52] B. Van Roy, D. P. Bertsekas, Y. Lee, and J. N. Tsitsiklis, “A neuro-dynamic programming approach to retailer inventory management,” in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, 1997, vol. 4, pp. 4052–4057.

[53] J. Hu, M. C. Fu, and S. I. Marcus, *Simulation-based algorithms for Markov decision processes*. Springer, 2013.

[54] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*, vol. 414. John Wiley & Sons, 2009.

[55] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, 1988.

[56] L. Baird, "Residual Algorithms: Reinforcement Learning with Function Approximation," in *In Proceedings of the Twelfth International Conference on Machine Learning*, 1995, pp. 30–37.

[57] M. Kearns, Y. Mansour, and A. Y. Ng, "A Sparse Sampling Algorithm for Near-Optimal Planning in Large Markov Decision Processes," *Mach. Learn.*, vol. 49, no. 2–3, pp. 193–208, Nov. 2002.

[58] Soheil Shayegh and Valerie M. Thomas, "Adaptive Stochastic Integrated Assessment Modeling of Optimal Greenhouse Gas Emission Reductions," *Clim. Change*.

[59] J. N. Tsitsiklis and B. Van Roy, "Feature-based methods for large scale dynamic programming," *Mach. Learn.*, vol. 22, no. 1–3, pp. 59–94, 1996.

[60] W. A. Van Den Broek, "Moving horizon control in dynamic games," *J. Econ. Dyn. Control*, vol. 26, no. 6, pp. 937–961, 2002.

[61] J. Boyan and A. W. Moore, "Generalization in reinforcement learning: Safely approximating the value function," *Adv. Neural Inf. Process. Syst.*, pp. 369–376, 1995.

[62] D. P. de Farias and B. Van Roy, "The linear programming approach to approximate dynamic programming," *Oper. Res.*, vol. 51, no. 6, pp. 850–865, 2003.

[63] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3–4, pp. 279–292, 1992.

[64] O. Hernandez-lerma and J.-B. Lasserre, "Error bounds for rolling horizon policies in discrete-time Markov control processes," *IEEE Trans. Autom. Control*, vol. 35, no. 10, pp. 1118–1124, Oct. 1990.

[65] E. Kriegler, J. P. Weyant, G. J. Blanford, V. Krey, L. Clarke, J. Edmonds, A. Fawcett, G. Luderer, K. Riahi, R. Richels, S. K. Rose, M. Tavoni, and D. P. van Vuuren, "The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies," *Clim. Change*, vol. 123, no. 3–4, pp. 353–367, Apr. 2014.

[66] M. Wise, K. Calvin, A. Thomson, L. Clarke, B. Bond-Lamberty, R. Sands, S. J. Smith, A. Janetos, and J. Edmonds, "Implications of Limiting CO2 Concentrations for Land Use and Energy," *Science*, vol. 324, no. 5931, pp. 1183–1186, May 2009.

[67] UNEP, "The Emissions Gap Report 2013," United Nations Environment Programme (UNEP), 2013.

[68] G. Luderer, R. C. Pietzcker, C. Bertram, E. Kriegler, M. Meinshausen, and O. Edenhofer, "Economic mitigation challenges: how further delay closes the door for achieving climate targets," *Environ. Res. Lett.*, vol. 8, no. 3, p. 034033, Sep. 2013.

[69] K. Riahi, E. Kriegler, N. Johnson, C. Bertram, M. den Elzen, J. Eom, M. Schaeffer, J. Edmonds, M. Isaac, and V. Krey, "Locked into copenhagen pledges—implications of short-term emission targets for the cost and feasibility of long-term climate goals," *Technol. Forecast. Soc. Change*, 2013.

[70] K. Blok, N. Höhne, K. van der Leun, and N. Harrison, "Bridging the greenhouse-gas emissions gap," *Nat. Clim. Change*, vol. 2, no. 7, pp. 471–474, Jul. 2012.

[71] J. Eom, J. Edmonds, V. Krey, N. Johnson, T. Longden, G. Luderer, K. Riahi, and D. P. Van Vuuren, "The impact of near-term climate policy choices on technology and emission transition pathways," *Technol. Forecast. Soc. Change*, 2013.

[72] UNEP, "Bridging the Emissions Gap. A UNEP Synthesis Report," United Nations Environment Programme (UNEP), 2011.

[73] B. Metz, O. R. Davidson, P. R. Bosch, R. Dave, and L. A. Meyer, "Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change," 2007.

[74] J. Rogelj, D. L. McCollum, B. C. O'Neill, and K. Riahi, "2020 emissions levels required to limit warming to below 2 °C," *Nat. Clim. Change*, vol. 3, no. 4, pp. 405–412, Apr. 2013.

[75] N. Johnson, V. Krey, D. L. McCollum, S. Rao, K. Riahi, and J. Rogelj, “Stranded on a low-carbon planet: implications of climate policy for the phase-out of coal-based power plants,” *Technol. Forecast. Soc. Change*, 2014.

[76] J.-H. Han and I.-B. Lee, “Development of a scalable infrastructure model for planning electricity generation and CO2 mitigation strategies under mandated reduction of GHG emission,” *Appl. Energy*, vol. 88, no. 12, pp. 5056–5068, Dec. 2011.

[77] R. Doherty, H. Outhred, and M. O. O’Malley, “Generation portfolio analysis for a carbon constrained and uncertain future,” in *Future Power Systems, 2005 International Conference on*, 2005, p. 6 pp.–6.

[78] N. E. Koltsaklis, A. S. Dagoumas, G. M. Kopanos, E. N. Pistikopoulos, and M. C. Georgiadis, “A spatial multi-period long-term energy planning model: A case study of the Greek power system,” *Appl. Energy*, vol. 115, pp. 456–482, Feb. 2014.

[79] K. Caldeira and N. P. Myhrvold, “Temperature change vs. cumulative radiative forcing as metrics for evaluating climate consequences of energy system choices,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 27, pp. E1813–E1813, Jul. 2012.

[80] N. P. Myhrvold and K. Caldeira, “Greenhouse gases, climate change and the transition from coal to low-carbon electricity,” *Environ. Res. Lett.*, vol. 7, no. 1, p. 014019, Mar. 2012.

[81] P. H. Kobos, J. D. Erickson, and T. E. Drennen, “Technological learning and renewable energy costs: implications for US renewable energy policy,” *Energy Policy*, vol. 34, no. 13, pp. 1645–1658, Sep. 2006.

[82] L. Neij, “The development of the experience curve concept and its application in energy policy assessment,” *Int. J. Energy Technol. Policy*, vol. 2, no. 1, pp. 3–14, 2004.

[83] L. Neij, “Cost development of future technologies for power generation—a study based on experience curves and complementary bottom-up assessments,” *Energy Policy*, vol. 36, no. 6, pp. 2200–2211, 2008.

[84] F. Ferioli, K. Schoots, and B. C. C. Van der Zwaan, “Use and limitations of learning curves for energy technology policy: A component-learning hypothesis,” *Energy Policy*, vol. 37, no. 7, pp. 2525–2535, 2009.

[85] E. Gumerman and C. Marnay, “Learning and cost reductions for generating technologies in the national energy modeling system (NEMS),” 2004.

[86] U.S. Energy Information Administration, “International Energy Outlook 2013,” Office of Energy Analysis U.S. Department of Energy, Washington, DC 20585, DOE/EIA-0484, 2013.

[87] International Energy Agency, “Projected Costs of Generating Electricity,” IEA, France, 2010.

[88] U.S. Energy Information Administration, “EIA - Annual Energy Outlook 2014,” U.S. Department of Energy, DOE/EIA-0383ER(2014), May 2014.

[89] K. Arrow, M. Cropper, C. Gollier, B. Groom, G. Heal, R. Newell, W. Nordhaus, R. Pindyck, W. Pizer, and P. Portney, “Determining benefits and costs for future generations,” *Science*, vol. 341, no. 6144, pp. 349–350, 2013.

[90] G. M. Heal and A. Millner, “Agreeing to disagree on climate policy,” *Proc. Natl. Acad. Sci.*, vol. 111, no. 10, pp. 3695–3698, 2014.

[91] R. S. Pindyck, “Climate Change Policy: What Do the Models Tell Us?,” *J. Econ. Lit.*, vol. 51, no. 3, pp. 860–872, 2013.

[92] G. J. Gordon, “Stable function approximation in dynamic programming,” DTIC Document, 1995.

[93] M. van den Broek, A. Faaij, and W. Turkenburg, “Planning for an electricity sector with carbon capture and storage: case of the Netherlands,” *Int. J. Greenh. Gas Control*, vol. 2, no. 1, pp. 105–129, 2008.

[94] U.S. Energy Information Administration, “Updated Capital Cost Estimates for Utility Scale Electricity Generating Plants,” U.S. Department of Energy, Apr. 2013.

[95] IPCC, *Climate Change 2007: The Physical Science Basis*. Intergovernmental Panel on Climate Change, 2007.

VITA

SOHEIL SHAYEGH

Soheil was born in Isfahan, Iran. He received his B.S. and M.S. degrees in Civil Engineering from Iran, and moved to Armenia to get an M.Eng. degree in Industrial Engineering and Management from American University of Armenia in 2007. Before coming to Georgia Tech, he worked as a professional engineer and researcher in multiple developing projects in the Middle East and Africa. Soheil started his PhD at the H. Milton Stewart School of Industrial & Systems Engineering (ISyE) at Georgia Institute of Technology in 2009. As a former fellow of the Sam Nunn Security Program at the School of International Affairs, Soheil has formed his research interests around some of the most daunting global challenges, such as climate change, while gaining extensive experience in science and technology innovation and policy. He is a two-time recipient of the Georgia Tech Research and Innovation Conference award and the recipient of the Southeast Regional Energy Symposium (SERES) award. In 2014 he was awarded the innovation summer fellowship from the Georgia Tech's Enterprise Innovation Institute and the Georgia Tech Program in Science, Technology and Innovation Policy (STIP). In 2011 he received a senior fellowship from the National Iranian American Council (NIAC) in Washington D.C. and interned for U.S. Congressman Rush Holt. While in his PhD program, Soheil was the project coordinator for the Engineers for a Sustainable World (ESW) student club. He is seeking careers in global policy making that relate to his background in science, technology, and policy.