GEOSPATIAL ECONOMICS

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

JUSTIN ANDREW JOHNSON

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Stephen Polasky, Carlisle Ford Runge

June, 2014

Acknowledgements

I have many to thank and acknowledge for the wonderful guidance and support given since the beginning of my graduate work.

To my advisors, Steve Polasky and Ford Runge, who showed me that economics can be rigorous and enlightening while remaining useful to policy application.

To the rest of my committee members, Ben Senauer and Terry Roe, who offered thoughtful criticisms and important support throughout the entire process.

To my colleagues for friendship and comradery, including (but certainly not limited to): Aine McCarthy, Ali Bittinger, Brandt Richardson, Brooke Krause, Giovann Alarcon, Jaya Jha, Jose Pacas, Martha Rogers, Nadia Yayitra, Nikhil Joglekar, Nina Yagodkina, and Travis Smith.

To my parents, Jeff and Carol, and my sister, Christen, for years of understanding and inspiration.

And finally, to Lindsey Johnson, my best friend, life partner and wife, for ceaseless care, advice, perspective and support. Thank you.

Table of Contents

iii. ..... List of Tables

iv. ..... List of Figures

1 ..... Introduction

5 ..... Chapter 1: Geospatial Optimization of Global food and Agriculture Tradeoffs

26 ..... Chapter 2: Agents on a Landscape: Simulating Interactions in Economic and Ecological Systems

98 ..... Chapter 3: Reciprocity in Commons Dilemmas

127 ..... Conclusion

131 ..... Bibliography

144 ..... Appendix 1

171 ..... Appendix 2

223 ..... Appendix 3

List of Tables

Page	*Number and Title*
19	Table 1.1: Value of Carbon Storage Saved while Producing 100% more Calories
47	Table 2.1: Summary of Data Sources
48	Table 2.2: IGBP Land Use, Land Cover Categorization of MODIS Data
61	Table 2.3: Summary of Relevant Literature Values for Calibration and Validation of Foraging Model
62	Table 2.4: Assumed parameters on travel speed across different LULC types
79	Table 2.5: Eight Alternative Scenarios for Policy Analysis
82	Table 2.6: Firewood Abundance per grid-cell in m3 used in different scenarios from literature values
92	Table 2.7: Demand met and Profit Earned in 9 Scenarios
163	Table A1.1: Sensitivity Analysis Results
167	Table A1.2: Descriptive and Model Statistics
185	Table A2.1: Questions from NPS Used in Simulation
197	Table A2.2: Agent Attributes

List of Figures

Page	*Number and Title*
12	Figure 1.1: Crop Advantage (CA).
14	Figure 1.2: Comparison of extensification in the optimal solution versus BAU.
16	Figure 1.3: Crop advantage and extensification in optimal and BAU simulations for the U.S. Corn Belt (left) and S.E. Asia (right).
18	Figure 1.4: Net Carbon Storage Change
33	Figure 2.1: Example Geospatial Grid-Cell Network
36	Figure 2.2: Agents Defined on a Grid-Cell Network
41	Figure 2.3: Conceptual Model of Household Production and Consumption Decisions
43	Figure 2.4: Generating a Production Function from a Net Profit Geospatial Grid-Cell Matrix given no Competition
44	Figure 2.5: Max-Marginal-Gain Choices between Leisure, Production and Wage Earning over Multiple Iteration Steps
50	Figure 2.6: MODIS LULC Data Processed for Tanzania
51	Figure 2.7: MODIS LULC Data Processed near Mount Kilimanjaro
53	Figure 2.8: Population per Grid-Cell from WorldPop
55	Figure 2.9: Wage Surfaces
57	Figure 2.10: Location of Maasai Firewood Collection Study (Berin et al. 2004)

58 Figure 2.11: Per capita Firewood Consumption versus Residents per Household

63 Figure 2.12: Creating Traversal Cost from Land Cover and Elevation

64 Figure 2.13: Input Data for Population Downscaling

66 Figure 2.14: Downscaled Population Data

68 Figure 2.15 Population per 450 meter Grid-Cell in Tanzania

69 Figure 2.16: Population per 450 meter Grid-Cell near Mount Kilimanjaro

71 Figure 2.17: Traversal Cost Grid-Cell Network

72 Figure 2.18: Calculation of Arrival Cost Maps

73 Figure 2.19: Illustration of Combined Arrival and Traversal Cost Array

75 Figure 2.20: Calculating Profit Available from Firewood Abundance and Travel Time

78 Figure 2.21: Results Summary Map

83 Figure 2.22: Original Supply

84 Figure 2.23: Supply Left

85 Figure 2.24: Supply Taken

87 Figure 2.25: Original Labor Endowment

88 Figure 2.26: Labor Used to Purchase Fuel Substitute

89 Figure 2.27: Labor used to Collect Firewood

90 Figure 2.28: Profit Earned by Each Agent (national)

91 Figure 2.29: Profit Earned by Agent (near Mt. Kilimanjaro)

93 Figure 2.30a: Demand Met and Profit Earned

93 Figure 2.30b: Profit Deviation from Baseline

96 Figure 2.31: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)

113 Figure 3.1: Effect of increasing intensity of reciprocity

116 Figure 3.2: Effect of increasing intensity of reciprocity

117 Figure 3.3: Effect of Emissions Target on Reciprocally Optimal Abatement

118 Figure 3.4: Effect of Initial Emissions on Reciprocally Optimal Abatement

119 Figure 3.5: Simultaneous Changes of Beta and t on Reciprocally Optimal Abatement

120 Figure 3.6: Simultaneous Changes of Beta and t on Direct Utility

122 Figure 3.7: Stock of Emissions over Time

123 Figure 3.8: Abatement Choices over Time

125 Figure 3.9: Beta Vector by Reciprocal Distance

148 Figure A1.1: Maize Proportion of Grid-Cell Harvested.

149 Figure A1.2: Maize per Ha Yield (tons).

150 Figure A1.3: Maize per Grid-Cell Yield (tons dry weight).

151 Figure A1.4: Maize per Grid-Cell Caloric Yield (kcal).

152 Figure A1.5: All Crop per Grid-Cell Caloric Yield (kcal).

153 Figure A1.6: Carbon Storage in Crops, Assuming Complete Extensification of Grid-Cell at Current Crop Proportions (tons).

154 Figure A1.7: Carbon Storage in Potential Natural Vegetation (tons)

155 Figure A1.8: Carbon Storage in Soils (tons per Ha)

156 Figure A1.9: Marginal Effect of Extensification on Carbon Storage (tons C per grid-cell).

157 Figure A1.10: Crop Advantage (caloric yield per ton carbon storage loss).

158 Figure A1.11: Change in Caloric Yield (post Optimization Simulation).

159 Figure A1.12: Change in Caloric Yield (post BAU Simulation).

160 Figure A1.13: Net Land Protected (Proportion Harvested in BAU less Proportion Harvested in the Optimal Simulation).

161 Figure A1.14: Net Carbon Storage Change (tons per grid-cell).

165 Figure A1.15: Carbon Saved by Optimizing as a Function of Future Calorie Demand

166 Figure A1.16: Value of Carbon Saved as a Function of Extensification Percent

169 Figure A1.17: Grid-cells compared to satellite imagery

174 Figure A2.1: Example Cell with Notation

175 Figure A2.2: Example of an Agent Applied to Minnesota with Interstate Highway Network Links

176 Figure A2.3: Example Geospatial Cell Network using States as Zones and Interstate Highways as Links.

180 Figure A2.4: Process for Defining Cell Network and Normalized Matrices

183 Figure A2.5: Process for Defining Geospatial Grid-Cell Network and Agents based on Tanzanian Data

186 Figure A2.6: Household Survey Reported Average Wages (TSH)

188 Figure A2.7: Example of Problems from Using Village Population Values

191 Figure A2.8: N-Optimal Agent Algorithm Applied to Tanzania

192 Figure A2.9: Sum of Heuristic from N-Optimal Agent Identification Algorithm for different N

193 Figure A2.9: Results from Representative-Agent Simulation

195 Figure A2.10: Results from Representative-Agent Simulation

197 Figure A2.11: Normalized Matrices Corresponding to Elements in the Attribute Table

199 Figure A2.12: Traversal Cost Grid-Cell Network

201 Figure A2.13: Euclidean Distance Arrival Cost Grid-Cell Matrix

204 Figure A2.14: Identification of the Optimal Preference for Directness in the Myopically Optimal Route-Finding Algorithm

205 Figure A2.15: Identification of the Optimal Preference for Directness in the Myopically Optimal Route-Finding Algorithm

207 Figure A2.16: Combination of Myopically Optimal Arrival Cost Calculation with A-Star Algorithm

208 Figure A2.17: Initial Profit Available

209 Figure A2.18: Action Order of First 100,000 Agents

210 Figure 4A2.19 : Total Profit Taken for Varying Trips per Step

213 Figure A2.20: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)

214 Figure A2.21: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)

216 Figure A2.22: Validation using Contour Surfaces (National)

217 Figure A2.23: Validation using Contour Surfaces (near Mt. Kilimanjaro)

218 Figure A2.24: 250 Optimal Agent Zone-Point Pairs

219 Figure A2.25: Naïve Approach to Zone Assignment: Mean or Zero

221 Figure A2.26: Algorithmic Approach to Zone Filling

226 Figure A3.1: Definition of Agent Based Simulation for Forestry Management

228 Figure A3.2: Application of Agent Based Simulation to Forest Management with Reciprocal Utility

Introduction

Understanding how economic systems interact with ecosystems requires models that include geospatial heterogeneity. Integration of economic and ecological systems is vital for sustainable development, especially with respect to climate change, food security and the management of common property resources. Advances in remote sensing and geographic information systems have created a wealth of data applicable to economics, but it is challenging to incorporate high-resolution, global data in existing economic models. In this thesis, I integrate geospatial data with economic theory to analyze important environmental problems.

The next three chapters describe the techniques I use for modeling geospatially-explicit economic systems and apply them to current environmental challenges. Chapter one addresses the tradeoffs between food production and environmental protection. I address the question of how we can optimally feed a growing population (requiring a 100% increase in calorie production by 2050) while minimizing the loss of carbon storage (which is important for mitigating climate change). I use high-resolution, gridded global data to give geospatial specificity to the results of the optimization. The framework I present in this chapter includes only one production choice and one environmental good, but it is more broadly applicable to multiple goods and multiple ecosystem services. This chapter also shows how using geospatial data can increase the policy relevance of an analysis. For example, instead of claiming that tropical forests are *in general* very valuable, using geospatial data allows for the more precise claim that *this specific 10 kilometer patch* of forest is better kept as forest than cultivated. Spatially explicit information like this can help construct more specific policies, such as food-for-carbon swaps or identifying which parcels ought to be protected first given a limited conservation budget.

Chapter two presents a microeconomic model of spatial foraging that addresses how humans gather goods on a spatially heterogeneous landscape when transport costs are non-negligible. The general problem of foraging arises when multiple agents located in space compete for resources that are characterized by their location. I use agent-based simulation methods to account for agents that must move over the landscape subject to terrain and road networks, depletion by rival agents and spatial heterogeneity with a large number of agents (10 million). The model is applicable to several topics in environmental economics, including fuelwood collection and fisheries management, but also more general economic topics such as housing, employment search, transportation, pollution and urban economics.

Another contribution of chapter two is that it makes several methodological advances that allow for spatially-explicit agent-based simulation on extremely large systems. These advances are of two types: first, I present data creation methods that allow for high resolution data to be created globally, relying on satellite-derived data products and spatial downscaling techniques to estimate environmental and social indicators, such as population density or spatially defined wages. Second, I identify computational methods (and implement them in a software application) that allows for fast calculation of agent interactions and movement of space. I discuss the data storage types necessary for this along with a method of vectorizing the calculations to enable computation of extremely large systems (with as many as 10 billion agents).

Chapter two concludes an application of the spatial foraging model in which I assess how villagers in Tanzania gather firewood from forests. Firewood collection is a useful example because the need for a spatially explicit model is clear. Transportation costs of firewood are very high relative to their value and firewood in almost always collected by agents foraging for their own consumption. For instance, Fabe and Grote (2013) report

that 97.5% of household in Tanzania use firewood as their main fuel, but only 13% of firewood is purchased. I define behavior rules in the simulation based on a microeconomic agricultural household production model (Singh et al. 1986, Bardhan and Udry 1999) apply them to high-resolution geospatial data for Tanzania. I iteratively simulate individual agents' foraging actions and observe the value of firewood obtained. The estimates obtained from this method match existing estimates of firewood collection while providing more detail about where the firewood is collected.

The empirical application presented in chapter two is useful to practitioners of ecosystem serve estimation. Calculating the ecosystem service value of non-timber forest products (NTFPs) has been difficult in practice due to computational and theoretical problems. Existing estimation approaches identify the abundance of NTFPs on different forest types but do not explicitly state how agents gather the firewood. The approach I use accounts for how these gathering decisions interact with the abundance of NTFPs to determine the ecosystem service value.

The final chapter in this thesis presents a theoretical model that that incorporates reciprocity in a utility maximization model to analyze common-property resource dilemmas. This model, which I refer to as the commons reciprocity utility model, allows agents to make interpersonal comparisons of utility in order to reward cooperators and punish detractors. Extending traditional utility theory in this way is useful to describe the wide-spread observation that individual economic agents do not always free-ride and do not always fall for the tragedy of the commons (Hardin 1968; Ostrom 1990). Although this chapter is primarily focused on theory, I provide two environmental examples to illustrate how it can be applied (including full details in Appendix 2). First, I discuss how international negotiations on climate change can be modeled in this framework by describing each nations' decisions to meet their emissions abatement targets in reciprocal

terms. Specifically, nations will reward other nations who do meet their abatement goals and will punish those who do not. I provide a numerical example of this situation that shows increased levels of abatement, higher than the prediction of strong free-riding. Second, I apply the commons reciprocity utility model to a forest commons to explore how agents' decisions to engage in sustainable forestry or to clear-cut the forest depends on the reciprocal relationships of nearby agents. This example is preliminary, but shows how the model can be applied to the agent-based simulation techniques introduced in chapter two.

At the deepest level, the goal of this thesis is not to present a complex system of models, but is to answer the question of "how ought we live?" As humans harness an ever-greater portion of available energy and focus it into ever-more complex arrangements, the question of understanding our place in our environment grows more challenging. It requires modeling economic behavior in conjunction with the geospatial landscape on which we act. It is my hope that the methods presented here help condense the nebulous connections among economic and ecological systems into useful bits of truth.

Chapter 1: Geospatial Optimization of Global Food and Carbon Tradeoffs

Justin Andrew Johnson[a,1], Carlisle Ford Runge[a,b], Benjamin Senauer[a], Jonathan Foley[b] and Stephen Polasky[a,b,c,]

Classification: Social Sciences

Minor Category: Sustainability Science

Keywords: ecosystem services, land use change, cropland expansion, food security, carbon storage.

[1] To whom correspondence should be addressed. E-mail: joh07536@umn.edu; Mailing Address: Department of Applied Economics, Ruttan Hall, 1940 Buford Ave., St. Paul, MNB 55108.
[a] Department of Applied Economics, University of Minnesota, St. Paul, MN 55108
[b] Institute on the Environment (IonE), University of Minnesota, St. Paul, MN 55108
[c] Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108

Abstract

Feeding a growing and increasingly affluent world will require expanded agricultural production, which may require converting grasslands and forests into cropland. Such conversions can reduce carbon storage, habitat provision, and other ecosystem services, presenting difficult societal tradeoffs. In this paper we use spatially-explicit data on agricultural productivity and carbon storage in a global optimization procedure to find where agricultural extensification should occur to meet growing demand for food while minimizing carbon emissions from land-use change. Optimal extensification saves approximately 6 billion metric tons of carbon compared to a business-as-usual (BAU) approach with a value of $1.06 trillion (2012 U.S. dollars) using recent estimates of the social cost of carbon. This type of spatially-explicit geospatial optimization can be expanded to include other ecosystem services and other industries to analyze how to minimize conflicts between economic development and environmental sustainability.

Introduction

One of the primary challenges of the 21st century will be to meet growing demand for agricultural output while preserving essential ecosystem processes on which both long-term agricultural production and human well-being depend. Growing demand for food, feed, fuel, and fiber has led to conversion of natural grasslands and forests and reduced the flows of many important non-marketed ecosystem services, such as carbon storage, water filtration, and habitat provision (MA 2005). In the tropics, which are especially important for carbon storage and habitat for biodiversity, 55% of new agricultural land came from conversion of forests (Gibbs et al. 2010). Agriculture is responsible for 92% of the annual global water footprint (Fallenmark 2003; Hoekstra et al. 2012) and uses 90% of rock phosphate. Despite this, nearly one billion people are food insecure, meaning they regularly fail to consume enough calories to lead an active healthy life (FAO 2012). Due to rising population and incomes, the FAO projects global food demand from to 2000 to 2050 to grow by approximately 70% (Alexandratos and Bruinsma 2012), while others have projected as much as 100-110% (Tilman et al. 2011).

Agricultural production can be increased through intensification (higher yields with more fertilizer, pesticide and water inputs, multiple cropping, shorter fallow periods and improved seed varieties) and extensification (expanding on to more hectares). Intensification is expected to play the major role in meeting expanded demand. FAO forecasts that 80% of the future increase in global agricultural production will be met by intensification (70% in developing countries) (Alexandratos and Bruinsma 2012, p. 126). It is possible in biophysical terms that all of the increase in demand could be met by intensification, especially through closing "yield gaps" between high productivity regions (e.g., North America) and low ones (e.g., Sub-Saharan Africa) (Foley et al. 2011). Others argue that intensification has already brought us to "peak farmland," and that we can

begin returning land to a natural state (Ausubel et al. 2012). However, numerous social, political and economic constraints make it unlikely that all of the increase in demand will be met through intensification. Low yield regions often suffer from political instability, lack of infrastructure, and the inability of poor farmers to invest in fertilizers, equipment and other inputs, all of which constrain intensification. Moreover, the rate of increase in crop yields has been declining. While the average annual increase in global yields between 1961-2007 was 2.92% for wheat, 1.91% for rice, and 2.47% for maize, the FAO predicts yield increases of only 0.86% for wheat, 0.63% for rice, and 0.83% for maize between 2005/2007 and 2050 (Alexandratos and Bruinsma 2012, p.106). Climate change may also reduce future yields (Lobell et al. 2011, Schlenker and Roberts 2009). Even when it is possible to intensify, it may be more profitable for farmers to extensify instead.

In this paper we use geospatial optimization techniques to identify where extensification should occur globally to minimize the negative impacts of extensification on the provision of ecosystem services. We illustrate the approach with an analysis of tradeoffs between extensification and carbon storage because we have readily available global data on carbon. The general approach can be extended to include other forms of natural capital or ecosystem services with the main constraint being the availability of suitable global data. Our results show that optimal extensification, taking into account both food production and carbon storage, preserves dramatically more carbon storage than business-as-usual (BAU) extensification scenarios.

Prior spatially-explicit studies analyze tradeoffs between agricultural production and multiple ecosystem services at local or regional scales (e.g. Egoh et al. 2008, Goldstein et al. 2012, Nelson et al. 2008, 2009, Polasky et al. 2008, 2011), or national scales (Ando et al. 1998, Bateman et al. 2013, Lawler et al. 2013). Fewer studies analyze spatially-explicit tradeoffs globally (Nelson et al. 2010, West et al. 2010). We extend the West et al. (2010) global analysis of tradeoffs between crop production and carbon storage by

using an optimization approach capable of estimating the maximum possible amount of carbon stored consistent with increased crop demand. We then value the increased storage using estimates of the social cost of carbon. In addition, we translate production of 175 different crops into production of consumable calories rather than utilizing dry harvest weight. Our work provides a spatially-explicit counterpart to global agricultural analyses using national level data (e.g. Hertel and Warren 1999, Taheripour et al. 2007, Hubert et al. 2009, Rosegrant et al. 2001, 2008, 2009).

We use global high-resolution spatial data for 5x5 minute grid-cells (approximately 10x10km near the equator) on crop cultivation (Monfreda et al. 2008; Ramankutty et al. 2008) and carbon storage (Ruesch and Gibbs 2008) to identify and locate the tradeoffs between crop cultivation and carbon storage. We derive a biophysical indicator of crop advantage (CA) by calculating the ratio of total calories produced to the loss of carbon stored for each grid-cell with extensification: $CA = \frac{CY}{\Delta C}$ where CY represents caloric yield per grid-cell aggregated over 175 crops using the current mix of crops grown (Monfreda et al. 2008), and ΔC is the tons of carbon storage lost (including aboveground, belowground and soil carbon) per grid-cell when a cell is converted from grassland or forest into cropland. To calculate carbon storage loss per unit area, we compare carbon storage in potential natural vegetation to carbon storage in crops. Carbon storage in potential natural vegetation and the methods for calculating crop carbon are from West et al. (2010) (see Methods and Supporting Information for details).

We use the CA score for grid-cells to minimize the loss of carbon storage while meeting increased food demand. The global optimization routine ranks all grid-cells by crop advantage (CA) and extensifies crop production in the cell with the highest CA score subject to constraints on feasibility of extensification. We continue to extensify in the highest ranked remaining cell until future food needs are met (see Methods and Supporting Information for details).

We limit where cultivation can expand by excluding grid-cells in which less than 5% or over 95% of the area is cultivated. Grid-cells above 95% are assumed to be fully utilized for crop production at present. Grid-cells below 5% typically include areas not suited to crop production such as deserts without irrigation, high altitude areas, latitudes too far north or south to grow crops, and protected natural areas. Areas such as the Amazon or Congo Basin have grid-cells that would be productive but currently have no observed cultivation due to lack of infrastructure, access to markets or other factors. However, these areas are extremely rich in carbon and therefore have low CA. We ran sensitivity analyses that allowed expansion into these cells, but they were not chosen for crop production by our optimization routine.

In grid-cells between 5% and 95% cultivation, we increase the proportion cultivated according to the following rules. If the current proportion cultivated is less than 15%, we extensify by a factor of 4. Otherwise, we close 75% of the gap between current cultivation and 95% cultivation. The reason we do this is because we lack spatial resolution to reflect heterogeneity within the grid-cell. Even in grid cells where the land is highly productive and generates high crop yields, there will be portions of the grid-cell that have steep slopes, poor soil, or are developed land, and therefore are unsuitable for crop expansion. The Supplemental Information provides an example of how this type of spatial heterogeneity could limit extensification on an example grid-cell. Our analysis assumes the mix of crops chosen on each grid-cell will stay constant into the future. We make this assumption both to avoid data and computation problems in calculating crop-specific effects, but also because our analysis is focused on the general tradeoff between agricultural production and carbon storage. In the supplement, we provide results from our sensitivity analysis that tests these assumptions.

The results of this optimization procedure are compared to a business-as-usual (BAU) simulation. We define BAU as increasing the share cultivated of each grid-cell by the

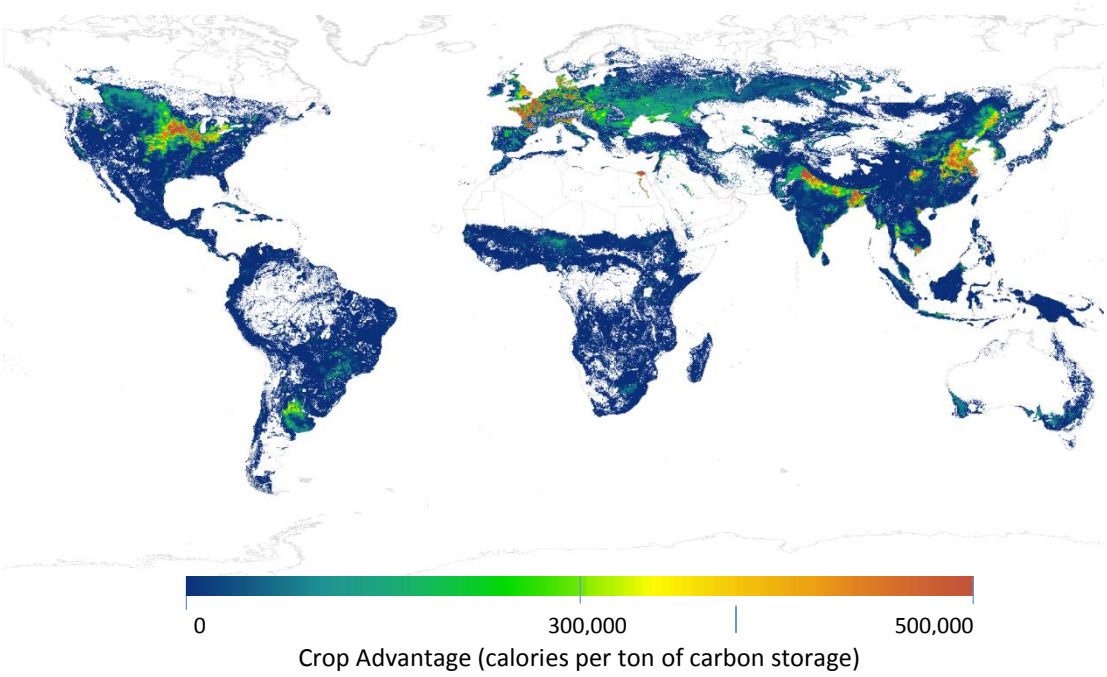
percent necessary to meet increased food demand. BAU increases are subject to the same feasibility constraints and limits as in the optimal simulation but ignore carbon storage. We explore different definitions of BAU in the Supporting Information.

We focus our analysis on a future scenario in which we must produce 100% more calories than in 2000 (in line with estimates from Tilman et al. 2011) with 25% coming from extensification and 75% coming from intensification gains. We also analyzed changes in overall demand for crops and the proportion of increased production coming from extensification versus intensification (see Supporting Information).

Results

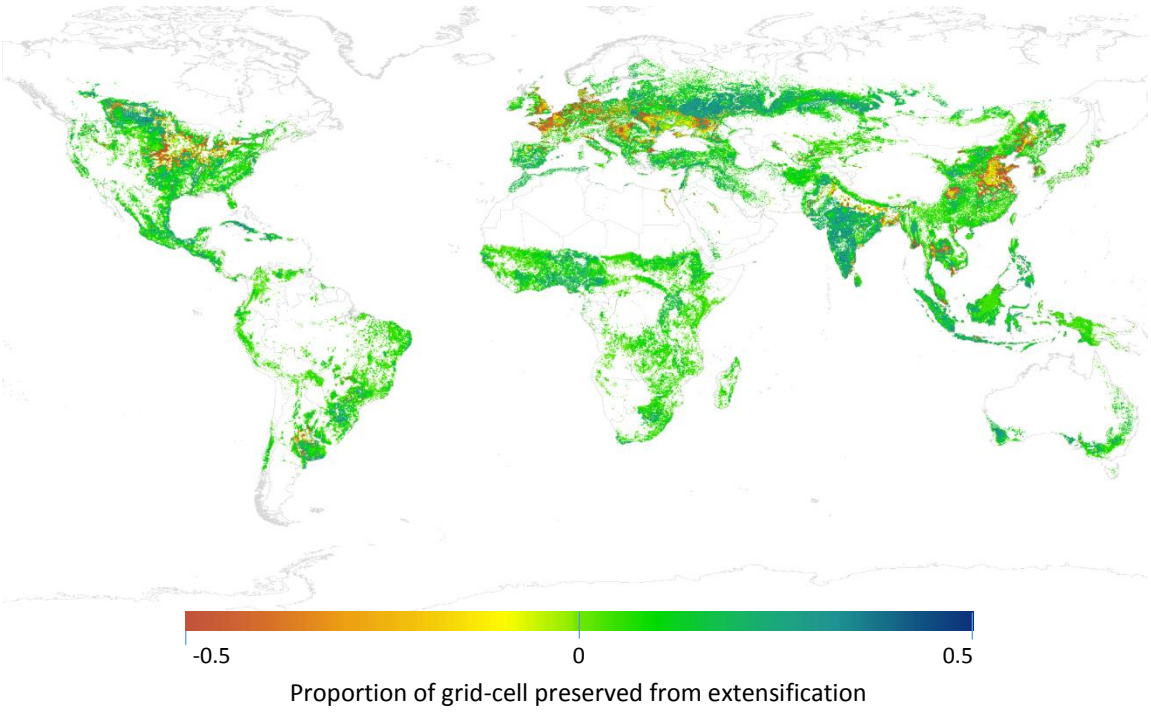
We identify areas with the largest crop advantage (Fig. 1.1). Grid-cells with the highest CA score produce 300,000 calories per ton of carbon storage lost with crop expansion. Areas that are currently heavily farmed, including the Corn Belt of the U.S. Midwest, parts of Western Europe, the Nile Valley, the Ganges River Plain, and much of eastern China, have very high CA values. Much of the tropics have relatively low CA both because of low crop yields and high carbon storage values. Areas with no color have no observed cultivation.

Figure 1.1: Crop Advantage (CA). Ratio of aggregate calories produced divided by carbon storage on each 5x5 minute grid-cell. Red values indicate areas where crop cultivation is comparatively advantaged over carbon storage.



We then identify which grid-cells are optimal for extensification to meet expanding demand for crops while conserving as much carbon storage as possible. We compare the optimal solution with the BAU solution to highlight areas in which it is optimal to concentrate agricultural expansion (Fig. 1.2). Many areas with the highest CA values are already heavily cultivated and have little available land for further extensification. In the optimal solution, extensification increases at the edges of currently intensively farmed areas. The optimal solution has greater extensification on the edges of the U.S. Corn Belt, parts of Western Europe and eastern China. The center of the U.S. Corn Belt, the Nile River Valley and much of the Ganges River Plain are little changed because little land is still available for extensification.

Figure 1.2: Comparison of extensification in the optimal solution versus BAU. Both the optimal and the BAU simulation produce 100% more calories and assume 25% of the calories come from extensification. Blue and green shading indicate areas where less extensification would occur under the optimal solution compared to BAU. Red and yellow shading indicates areas where more extensification would occur under the optimal solution compared to BAU.

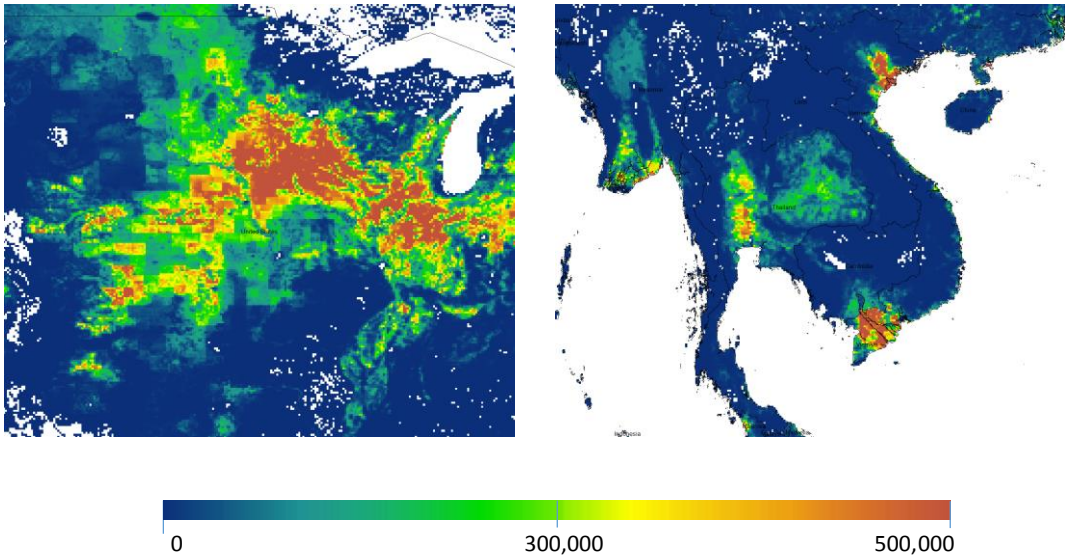


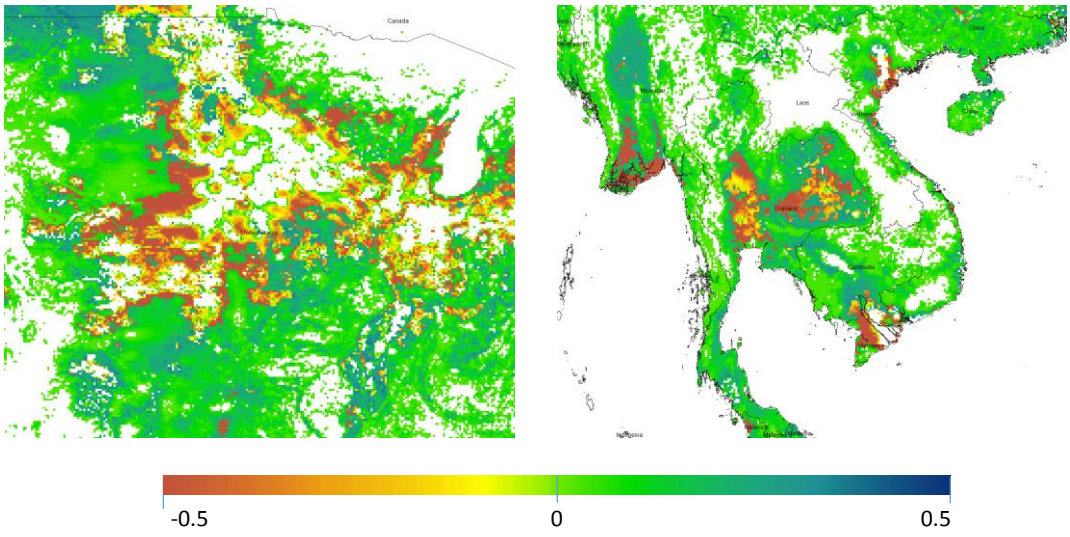
Parts of Eastern Europe, the Ukraine, Russia, and several pockets in Southeast Asia are extensified more heavily in the optimal solution than in BAU. Much less extensification occurs in the Philippines, Indonesia, Southern India, parts of Sub-Saharan Africa and Central America where crop advantage (CA) is low.

An advantage of our geospatial optimization approach is that it can assess land-use changes at many levels of aggregation from the global scale down to individual 5x5 minute grid-cells. To illustrate more detailed regional patterns, we show CA and optimal extensification for two specific regions: the U.S. Corn Belt and Southeast Asia (Fig. 1.3).

At higher resolutions, we see in more detail that optimal extensification occurs along the edges of currently intensively cropped areas in the U.S. Corn Belt and several rich river valleys in Southeast Asia, such as the Mekong Delta and Red River in Vietnam, the Irriwaddy River Basin in Myanmar, and the Chao Phraya River Basin in Thailand. Conversely, fewer new hectares are cultivated and more carbon is stored in most other areas in the optimal solution compared to BAU. These are areas where soils are less productive or the topography is less suited to cropping.

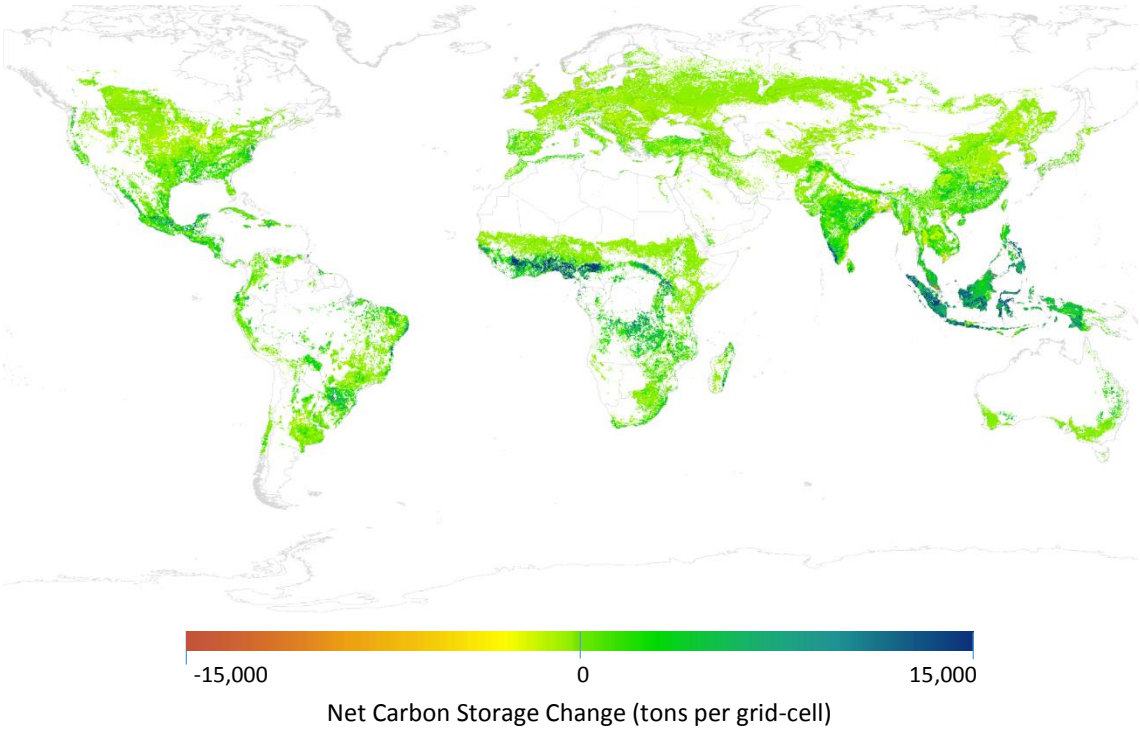
Figure 1.3: Crop advantage and extensification in optimal and BAU simulations for the U.S. Corn Belt (left) and S.E. Asia (right). Fig. 3a and 3b: Crop Advantage. Fig. 3c and 3d: Difference in extensification in optimal versus BAU simulations.





By concentrating extensification in areas with high crop advantage, much more carbon storage occurs under the optimal solution as compared to BAU extensification. Large amounts of carbon storage are preserved in Indonesia and other parts of South East Asia, India, Sub-Saharan Africa, and Central America. Areas with greater extensification under the optimal solution show reduced carbon storage, but the losses are far less than the gains elsewhere (see Supporting Information for detailed comparison). On a global level, optimal extensification results in preserving 5.89 billion metric tons of carbon compared to BAU. This figure rises if more demand must be met through extensification. For example, with 50% of demand met through extensification, optimization results in 12.08 billion metric tons of carbon saved (see Supporting Information).

Figure 1.4: Net Carbon Storage Change. Tons carbon storage preserved per grid-cell under optimal solution versus under BAU. Blue and green indicate areas where larger amounts of carbon storage occur under the optimal solution versus BAU while yellow indicates that less carbon is stored under the optimal solution (areas of greater extensification).



Economic Valuation

Increasing carbon storage in terrestrial systems can reduce the amount of atmospheric CO_2 and potentially reduce damages from climate change. Using results of a survey of 232 published estimates of the social cost of carbon for different discount rates (Tol

2009), we find that the value of the additional carbon stored in the optimal versus the BAU scenario ranges from $0.44 trillion to $1.30 trillion in 2012 U.S. Dollars (Table 1.1).

Table 1.1: Value of Carbon Storage Saved while Producing 100% more Calories

Pure Rate of Time Preference	0%	1%	3%
Social Cost of Carbon (2012 dollars)	$221	$181	$75
Value Saved in Base Scenario (in trillions 2012 dollars)	$1.30	$1.06	$0.44

Values for the social cost of carbon are the mean value for the fitted distribution in Tol (2009), adjusted to U.S. $2012.

Discussion

Given the large projected increases in demand for agricultural crops, it is likely that at least some of this increase will have to be met by expanding the amount of land devoted to agricultural production. Agricultural extensification comes at the expense of natural habitats (forests and grasslands) that provide carbon storage and many other ecosystem services. In this paper we show that by finding the best locations to extensify and the best locations to conserve natural habitats, we can meet increased crop demand while maintaining far higher levels of carbon storage than following a business-as-usual expansion. To minimize the loss of stored carbon with extensification, the expansion of

cultivated hectares should be concentrated on the extensive margin of areas that are currently heavily cultivated, as these areas tend to have the highest crop advantage (i.e. the greatest increase in crop production per unit loss of stored carbon). Following the optimal strategy would conserve an estimated 5.89 billion metric tons of carbon in natural environments by 2050, with an estimated social value of $1.06 trillion at $181/ton C, compared to a business-as-usual scenario.

In this paper, we considered only agricultural extensification and one form of natural capital, carbon storage, but the optimization principles we employed are general and can be extended to include multiple types of natural capital and ecosystem services, as well as considerations of agricultural intensification. Modeling the tradeoffs from intensification requires estimating the increase in yield with intensification and the impact on natural capital and ecosystem services. For example, how does increased application of nitrogen fertilizer affect yields, water quality, and greenhouse gas emissions? Modeling approaches for inclusion of multiple ecosystem services has advanced rapidly over the past few years (Kareiva et al. 2011). Analysis of the impact of land-use choices to maximize a bundle of ecosystem services has been done at regional (e.g., Goldstein 2012) and national scales (e.g., Bateman et al. 2013). To date, lack of consistent global data sets has hindered application of high-resolution analysis at global scales. But work on global data sets also is advancing rapidly (e.g., Mueller et al. 2012). Our approach could be expanded to incorporate other data sets on ecosystem services and integrated with models such as InVEST (Natural Capital Project 2013) that calculate estimates of the provision of a number of ecosystem services as a function of land use and land management choices. In general, it should be possible to solve for the combination of optimal choices of extensification and intensification to find how best to increase agricultural production while maintaining the highest valued bundle of ecosystem services. Doing so would require information on grid-cell costs of intensification as well as the implicit values that society places on the relative importance of various ecosystem services. If a broader

array of ecosystem services and a broader set of actions (intensification and extensification) are analyzed, it is likely that the total social value from optimization compared to BAU would be many times greater than found here.

This analysis makes a number of assumptions about the rate of growth in future demand for crops and the proportion that can be met by intensification and extensification, as well as how much extensification can occur in various grid cells due to sub-grid-cell heterogeneity. Each of these assumptions can be altered and the optimization recalculated (we illustrate the effect of changes in many of these assumptions in the Supporting Information). Moreover, our analysis does not model how optimal extensification choices depend on broader economic factors such as transport systems and trade barriers. Our approach is, however, easily able to incorporate such complexities by defining additional scenarios or constraints for analysis. Even without explicit inclusion of economic factors such as these, our model does implicitly account for them because the observed production data we used reflect agent decisions based on transport, trade and other such factors. Additionally, future climate change will likely influence yields and the provision of other ecosystem services and change the optimization results. Each of these changes will affect the quantitative results but not the overall conclusion of the importance of optimizing extensification choices.

Nonetheless, showing what is possible and actually achieving it are not the same thing. Like West et al. (2010), Foley et al. (2011), and others, this paper shows what is feasible in biophysical terms. We show how careful consideration of both carbon storage and crop yield can maximize carbon storage while meeting agricultural production goals, subject to assumptions about sub-grid-cell heterogeneity that may limit extensification options. Moving closer to desirable outcomes—ones closer to the frontier—requires attention to institutional, political, social and economic factors, because billions of people must change what they are doing. These changes will require recognition by political leaders

and the general public of the value of carbon storage (and other ecosystem services). Otherwise, there will be little push for carbon policies such as establishing a price for carbon storage, and therefore less incentive for landowners to incorporate carbon or value of other ecosystem services into their decision-making. Without this, we are likely to see a trajectory much closer to business-as-usual than the optimal path.

As an example of national policy redirection, Brazil has incorporated the value of preventing deforestation in the Amazon and elsewhere into its national Forest Code. The rate of deforestation in Brazil has been reduced by 83% since 2004 (Tollefson 2013). This reduction was achieved primarily by the creation of new protected zones and stricter enforcement of land-use regulations. Our analysis can help build on such successes by more precisely identifying areas that are good candidates for protected status.

Methods

Defining Crop Advantage

The crop advantage measure for each grid cell is defined as the marginal benefit of extensifying land in different locations and is defined for each 5' by 5' grid-cell with geospatial coordinates (x, y), as:

$$CA_{xy} = \frac{CY_{xy}}{\Delta C_{xy}} \qquad (1)$$

where CY_{xy} is the per-hectare calorie yield in each grid-cell and ΔC_{xy} is the per-hectare carbon storage loss that would occur if the grid-cell was converted from forest or grassland to cultivation. CY_{xy} was calculated by combining data from the EarthStat dataset (Ramankutty et al. 2008, Monfreda et al. 2008) with FAOSTAT (2013) values on

caloric content of each food group. We calculated the per-hectare calorie yield of each xy-th grid-cell, CY_{xy} as:

$$CY_{xy} = \sum_{i=1}^{175} Y_{ixy} * A_{ixy} * C_i \qquad (2)$$

where Y_{ixy} is the dry weight in tons per hectare of the i-th crop, A_{ixy} is the fraction of crop area planted to crop i, and C_i is the caloric content of the i-th crop per ton. C_i is calculated as:

$$C_i = \left(\frac{S_i * 365}{Q_i}\right) \qquad (3)$$

where S_i is the variable from FAOSTAT's Food Balance Sheet dataset named "Food supply (kcal/capita/day)" and Q_i is FAOSTAT's "Food supply quantity (kg/capita/yr)". Finally, we increase CY_{xy} on each grid-cell according to our assumption of how much intensification will increase yields. This process created a gridded map of worldwide per-hectare calorie yield. We calculated per-grid-cell calorie yield by multiplying the per-hectare calorie yield by the amount of hectares present in each grid-cell, which we used for calculating aggregate calorie production. When summed globally, per-grid-cell calorie yield matches the FAO's estimate of total caloric production.

To calculate the change in carbon storage (ΔC_{xy}) with extensification, we use the method from West et al. (2010). We subtract the amount of carbon storage (aboveground and belowground) in potential natural vegetation ($PNVC_{xy}$) and one-quarter of the soil carbon associated with potential natural vegetation (SC_{xy}) from crop carbon (CC_{xy}) that would exist on the grid-cell if it was fully extensified:

$$\Delta C_{xy} = CC_{xy} - PNVC_{xy} - .25\left(SC_{xy}\right) \qquad (4)$$

Data on potential natural vegetation carbon comes from West et al. (2010), which used carbon values from the tier 1 methodology of the Intergovernmental Panel on Climate Change (Ruesch and Gibbs 2008), applied to potential natural vegetation data (Ramankutty and Foley 1999). To estimate soil carbon loss, we used gridded data on global soil organic carbon density (measured as kg Carbon/m2 to a depth of 1 meter) from IGBP (1998) interpolated to match the resolution of the data from Monfreda et al. (2008) and Ramankutty et al. (2008).

To calculate the carbon stored in each of the 175 crops, we assumed crop carbon storage of annual herbaceous crops is equal to their annual net primary productivity (West et al. 2010), calculated as:

$$CC_{xyi} = \frac{DF_i * C * Y_{ixy}}{HI_i * R_i} \qquad (5)$$

Where $CC_{xyi}$is the crop carbon of the i-th crop on the xy-th grid-cell, Y_{ixy} is the yield of the i-th crop on that cell, DF_i is the proportion of dry matter of the yield for crop i, C is the carbon content of dry matter (0.45 g C per g dry matter), $HI_i$is the proportion of aboveground biological yield that is economically valued for crop i, and R_i indicates the proportion of the i-th crop that is belowground biomass. Carbon stocks in woody crops was calculated as in Gibbs et al. (2008). Summation over each of the 175 crops gives C_{xy}, the total carbon that that would be stored in the grid-cell's crop cover if the grid-cell was fully converted to cultivation (assuming the same proportional crop mix as in 2000). Finally, we converted ΔC_{xy} to be the change in carbon per hectare extensified.

Assuming that annual NPP is equal to a crop's biomass likely overstates the amount of stored carbon in crops because the biomass only stored carbon part of the year. In the context of identifying which areas are better left natural, this assumption makes our conclusions and estimation of saved carbon conservative. Accounting for crop carbon,

however, has very a very small impact on the overall results because the amount of carbon able to be stored in crops is much less than the amount of natural carbon storage in most locations.

Chapter 2: Agents on a Landscape: Simulating Interactions in Economic and Ecological Systems

Introduction

Understanding how human behavior affects ecosystems, and vice-versa, requires consideration of spatial heterogeneity. Models of human behavior in economics, however, are not easily applied to situations with many heterogeneous locations and a large number of individuals. Non-spatial approaches give powerful insight into human behavior on a broad set of topics, such as modeling financial markets or intellectual property, but can fail when spatial aspects are a key part of the phenomenon in question. Environmental problems, for example, often require an explicitly spatial approach to understand because environmental impacts occur at a particular place.

In this chapter, I present a model that uses agent-based simulation to identify how individuals act on a heterogeneous landscape to gather spatially-defined goods. I consider the case where both agents and goods are defined by high-resolution, gridded spatial data, and where agents must move over a non-uniform landscape characterized by terrain and road networks to gather the goods. In this context, agent decisions depend on the configuration of the landscape, resulting in divergent behavior even at extremely small spatial resolutions (such 500 meters in the data I use). To account for this, I present a method that uses high resolution data and simulation with many (millions) agents, all of which have different opportunities and outcomes depending on their specific location on the landscape at each moment in time. Modeling this level of complexity results in a system of equations that is difficult to solve (even with numeric solution, this level of

complexity often cannot be solve with current computing power and existing solution methods). Thus, I also present computational methods capable of handling this modeling environment by using efficient computation.

I root the agent-based simulation in this chapter on microeconomic theory by deriving behavior rules from the production and utility maximization decisions individuals make in a household production model (such as in Bardhan and Udry, 1999). However, each of the agents in the simulation are allowed to have different parameters on their production function, determined both by their physical proximity to the gathered good, but also by the depletion of that good from nearby agents.

I illustrate the general foraging model with an application that uses data on how villagers in Tanzania collect firewood from forests. Careful identification of location is particularly important to understanding firewood collection decisions of rural households in Tanzania. Firewood in Tanzania is consumed by 97.5% of household, yet only 13% of firewood is purchased (Fabe and Grote 2013). The majority of firewood is supplied by subsistence foraging behavior and is rarely bought or sold on the market. Additionally, as will be discussed in the data section, collection of firewood is very time-consuming, and thus it is important to consider the opportunity costs of time spent foraging. The primary reason there does not exist a complete market for firewood is that firewood has very low value per kilogram and thus high transport costs make it uneconomical to transport very far to reach markets.

To describe and apply the model, this chapter begins by describing the many existing approaches to modelling spatially-explicit behavior and how the model in this chapter builds on these approaches. Second, I will describe the general modeling framework, including how I define the agent, the passage of time, behavior rules and the definition of

equilibrium in this type of model. Third, I will discuss the data I use for the empirical application of the model with emphasis on how data can be created for any location on the globe. Fourth, I describe particular methodological approaches I used to make the simulation computationally feasible with many agents and high resolution. Finally, I conclude by showing detailed results on foraging behavior in Tanzania, along with analysis of alternate policies and consideration of how the results are sensitive to modeling parameters. In addition to the work presented in this chapter, I include an extensive supplemental methods section in Appendix 3. This appendix formally defines the mathematical notation I use to describe utility-maximizing agents that move over a network of grid-cells. Additionally, I discuss in more depth the algorithms I used to calculate optimal travel routes and other computationally intense aspects of the simulation.

Existing Literature and Contribution

This chapter builds on a broad base of existing work, including literature on spatial competition of firms, urban and regional economics and agent-based economics (along with their applications in ecology, economics, geography and land-use change modeling). One of the first analyses of economic behavior with consideration of explicit space is the Hotelling model of spatial competition (Hotelling 1929). Hotelling analyzed the strategic interaction of firms when they choose a location in linear space and when their revenue is based on their relative closeness to consumers. Many subsequent analyses made important extensions on Hotelling's modeling, including different types of negotiation models (Hamilton et al. 1989), agglomeration (Gabszewicz and Thisse 1986; Gupta et al. 1997) and transportation costs with non-convexity (Stahl 1982). These analyses incorporate microeconomic theory into space-based conceptions of competition, which is a fundamental aspect of this chapter. More generally, Hotellings' and subsequent work

provides a theoretical basis for understanding economic behavior in which the location of agents can be endogenous and affects the agents' choices.

Analyses in urban economics that identify how production decisions are affected by proximity to spatially defined goods (or other spatial aspects, such as information externalities) provide a more direct theoretical basis for this chapter. The seminal publications in urban economics are Alonso (1964), Mills (1967), Mills (1972), Anas et al. (1998), which are themselves based on the von Thunen's (1826) central place theory. Closely related research in economic geography also considers location while considering the connections between cities, regions and international trade (Krugman 1998; Fujita et al. 2001; Fujita and Krugman 2004).

A subset of these literatures that is particularly relevant to this chapter addresses the value of open space and how open space arises in different spatial configurations. Specifically, Yang and Fujita (1983) model the formation green-belts around cities, while Anderson and West (2006) consider how this affects residential property values. Recent analyses along these lines typically employ sorting models (as in Klaiber and Phaneuf 2009) and are solved by maximizing spatially located objective functions subject to constraints derived from the search model, employing mathematical optimization techniques (and solved in GAMS), such as Tajibaeva et al. (2008, 2014). One advantage of this literature is that many analyses are explicitly 2-dimensional in their choice sets. Previous analyses, such as Krugman (1995), use 1-dimensional linear descriptions of space. While this approach is extremely useful and is simple enough to allow for analytic solutions, it is difficult to apply directly the insights from 1-dimensional models to specific policy questions, such as green space management.

Another subset of urban economics that directly informs my research focuses on land-use change. Recent work (Robertson et al. 2007; Irwin 2010; Evans et al., 2011) assesses a variety of economic factors that drive land-use change, including how agriculture affects land-cover change. Other work in this area directly considers how projected land-use changes might impact ecosystem services (Polasky et al. 2011; Lawler et al. 2014), or how we can identify the optimal configuration of land-use given tradeoffs between development and conservation (Polasky et al. 2008). These models, as well as bioeconomic models (Sanchirico and Wilen 1999; Costello and Polasky 2008), are useful when describing how human systems and ecological systems interact on a specific landscape.

The solution method I use, agent-based simulation, has been employed before in many applications, including many analyses that simulates human behavior in order to predict land-use change (Robinson et al. 2007, Filatova et al. 2007; Filatova et al. 2009). I draw methods from these and other publications in ecological modeling and geography on this toping, including An et al. (2001, 2005); Berman et al. (2004), Crooks and Castle (2012), and Cabrera et al. (2012).

Agent-based simulation methods have been applied to topics more broad than just land-use change. Economics has long used agent-based modeling as an approach to incorporate heterogeneity or special specificity into models (as reviewed by Tesfatsion and Judd, 2006). More recently, Magliocca et al. (2012) construct an agent-based model to analyze how suburban development districts and agricultural land owners interact to predict which areas will be developed. Some of the newest work utilizes agent-based simulation to assess human-ecosystem interactions between hunting and subsistence agriculture (Iwamura et al. 2014). One of the best example of an agent-based simulation that addresses environmental and economic issues is from Schreinemachers and Berger

(2011). In their work, the present a software package called Mathematical Programming-based Multi Agent Systems (MP-MAS). These approaches are summarized in Matthews et al. (2007). Finally, this chapter is based on literature that has combined many of the methods above to simulate how human behavior is affected by the provision of ecosytstem services and sustainability (e.g. Berman et al. 2004), including analyses that use 2-dimensional gridded data on non-timber forest products (Damania et al. 2005).

Aside from explicitly spatial models, I also base my analysis on economic models of household behavior that allow for individual households to produce their own product, which they may consume themselves (such as the agricultural household production model of Bardhan and Udry, 1999).

My work contributes to these literatures by modeling environmental-economic interactions that include many heterogeneous agents, spatially-dependent decisions that require high-resolution data to specify along with computational and data methods necessary to enable this work. I include several methodological approaches that increase realism when defining how agents move through space, such as calculating travel routes not by distance, but with optimal route-finding algorithms. Additionally, I present several methods based on interpolation to create high-resolution demographic data in locations where this type of data does not exist. I use a 2-dimensional representation of space that both covers a large area (all of Tanzania) and derives its spatial detail from remote-sensing data sources. Thus, although this work is primarily theoretical, the data used and the outputs created (such as those shown in the empirical application) are directly relevant to land-use managers and conservation planners.

Model Description

The foraging model I present in this section defines the four elements that comprise my agent-based simulation model (additionally, each element is defined formally in Appendix 3): the landscape network, the passage of time, the agent and behavior rules. This section describes each of these parts of the general model and incorporates examples from non-timber forest product (NTFP) collection in Tanzania to illustrate the importance of each element.

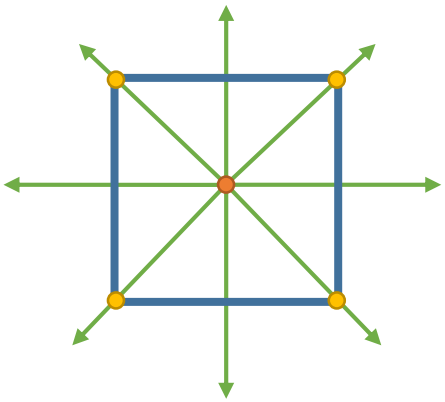
The Landscape Network

The landscape definition I use augments standard data on land-use, land-cover (LULC) by combining the data with a travel network and foraged-good regrowth functions on each grid-cell. Additionally, the network definition I use allows for connections between multiple functional units within the landscape. For example, grid-cells in this framework can be defined for sub-state units, such as counties or voting precincts, but also non-politically defined units such as farms, factories, watersheds, markets or households. In the Tanzanian example of the foraging model, each grid-cell is characterized by a travel network (defined from road, river, terrain and other inputs) as well as the abundance of firewood.

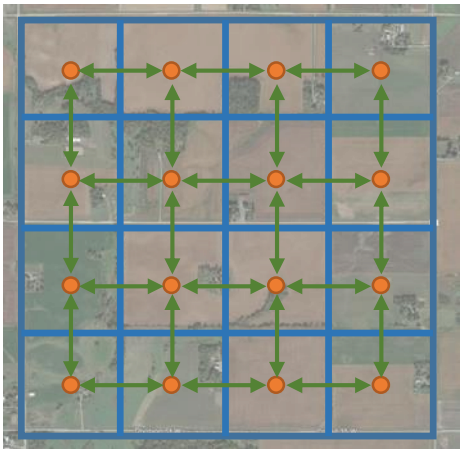
Network definitions of landscapes similar to this have been used extensively in transportation and civil engineering literature (for instance, Rodrigue et al. 2013) or in physics applications (Barthelemy 2011). Additionally, air pollution distribution models for PM2.5 (WRF/Chem, for instance), use variable grid-cell definitions that can be formalized as grid-cell networks as described here. In addition to defining the network, I also present (in Appendix 3) a method for normalizing the landscape network into nested

matrices to allow for vectorization of computation and efficient computation. Fig. 2.1 illustrates the definition of grid-cells and their application to a landscape. The image on the bottom right of Fig. 2.1 represents a grid-cell network for a farm landscape in Iowa.

Figure 2.1: Example Geospatial Grid-Cell Network



I define the landscape by creating a set of cells (such as the cell to the left) connected in a network that covers the landscape.



Two Conceptions of Time

Throughout this analysis, I refer to the passage of time with two different concepts: the iteration step (subscripted s) and the time period (subscripted t). One time period comprises many iteration steps. The reason for this differentiation is to distinguish between the actual passage of time and the iteration method I use to make the simulation possible. Within this framework, cells may only change when transitioning from time t to $t + 1$, except as the result of an agent action, (e.g. deforesting the cell or planting a crop) while the agent may change between either steps or time periods. An example of a change that a cell may make over time without the action of an agent is the regrowth of trees in a forest cell.

Additionally, I define the *action order*, a list of agents ordered by when they can make an action relative to other agents. During each time period, the first agent in the action order makes one action in step s, followed by the second agent in the action order in step $s + 1$. This process continues with new steps until the end of the action order list is reach, at which point the simulation loops through the full action order list again, repeating until no valid agent actions exist.

Having a time-step smaller than the full time period is necessary to simulate agent competition and also is a mechanism to increase fidelity of the simulation (though at the cost of computation speed). It is possible that biases arise based on how the action order is defined, but in most cases the bias approaches zero as the size of the step decreases. For example, if we define an action order in which agents may forage firewood from a forest, if we choose a large step size in which each agent is allowed to satisfy their full demand on their first action, then the results will be very sensitive to the order in which

actions happen. However, we limit the amount an agent can forage during each step to ever smaller quantities, the results become less sensitive to the initial ordering.

I denote these two conceptions of time with sets S and T. Set S contains the action order and the definition of what may be done within one iteration step. Set T assumes time progresses from $t = 1$ to $t = T$ and includes the full set of what may change between time periods (forces exogenous to agent actions, such as forest regrowth).

The Agent

The set of agents, A, is described by an attribute table consisting of a agent identity column (unique for each agent) and corresponding columns of agent attributes. This data type is different than the cell networks because the agent identity is not tied to a fixed location as it is for a cell. Rather, the cell network is a space through which agents move.

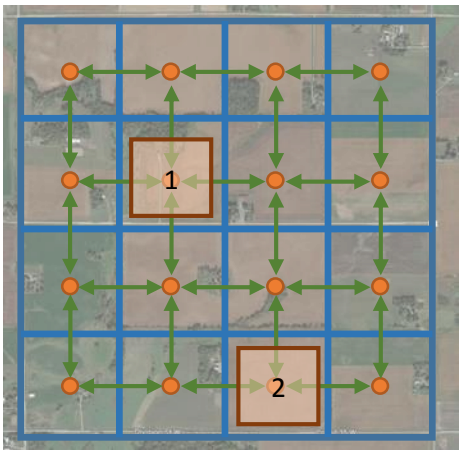
Although agents do not have fixed locations, agent locations and attributes can be aggregated, *at a given time and step*, to create a matrix that describes a static moment (this process is at the heart of the computational methods used in the simulation). To be a valid agent, $a_{i \in I}$, the agent attribute table must denote the agent's geographic location at every time period and iteration step defined by reference to a cell within a valid geospatial cell network and steps within the time structure. Additionally, the agent attribute table may include additional attributes, such as per-agent demographics, current assets, other location references like *work location* and *house location*, or relationship status with other agents.

Fig. 2.2 below illustrates the definition of two example agents using the cell network we defined above. In the top image of Fig. 2.2, each agent is plotted on top of the corresponding grid-cell that defines their location along with the agent ID used in the

agent attribute table in the bottom of Fig. 2.2. The example in Fig. 2.2. describes only one moment in time because the agents will move through the network and will have agent attributes that will change (such as the attribute for how much firewood they are currently carrying).

Figure 2.2: Agents Defined on a Grid-Cell Network

Agents placed on cell network by location



Agent attribute table

Agent ID	Location Cell	Agent Attribute 1 (labor endowment)	Agent Attribute 2 (production efficiency)
1	(1, 1)	6	.5
2	(3, 2)	4	.75

Behavior Rules: Household Production and Utility Maximization

At each iteration step in each time period, agent actions are determined by a set of behavior rules, R. The simulation approach I use in this section derives behavior rules

from traditional economic analysis. Specifically, my approach is similar to the agricultural household model used extensively in development economics (Singh et al. 1986, Bardhan and Udry 1999). I allow each agent at each iteration step to choose to allocate their labor to leisure, foraging or wage work. I extend the agricultural household model by making their marginal gains from forage and wage (and thus also their labor choices) depend on their location within the network and the depletion actions of nearby rivals.

I solve the model iteratively, allowing each agent at each iteration step to make their utility maximizing choice. Specifically, the agent follows a max-marginal-gain behavior rules that specifies how agent considers the marginal gain they would get from expending one step's worth of labor on each of their possible labor choices. The agent then does whichever action maximizes this gain. For example, if the marginal gain from foraging is higher than from wage work or leisure, then the agent will spend the full simulation step foraging (note that simulation steps will be defined sufficiently small such that this assumption does not affect the results). If after foraging from one grid-cell the agent still has available labor, they may spend it on another action, continuing until the labor they have in the current step is depleted.

Ignoring the per-step decisions for a moment, the choice each agent faces in each full time period is depicted in the maximization below. Intuitively, the agent maximizes their utility in the time period by choosing to allocate their labor to gathering firewood, to a wage generating activity (which generates money to purchase a firewood substitute) or to leisure. Their choice is constrained by the foraging production possibilities, a budget constraint and the amount of labor they have available to allocate on each activity.

$$\max_{L_i^g, L_i^w, L_i^l} u_i(c_i, L_i^l) = \alpha_i c_i^{\beta_i^c} L_i^{l\,\beta_i^l}$$

$$Subject\ to: \qquad g_i = f(L_i^g | N_i, R_{j\in\sim i})$$

$$c_i = c_i^g + c_i^b - c_i^s$$

$$g_i = c_i^g + c_i^s$$

$$p_b c_i^b = p_w L_i^w + p_s c_i^s$$

$$L_i^g + L_i^w + L_i^l \leq L$$

In this framework, each agent is endowed with labor, L, which they may allocate to gathering a spatially defined good, to wage work and to leisure, L_i^g, L_i^w, L_I^l respectively. Each agent's utility function is calibrated with a scalar, α_i, and Cobb-Douglass exponents $\beta_i^c + \beta_i^l = 1$ so that utility is concave in both consumption firewood and in leisure. Each agent is able to use labor, L_i^g, to produce the gathered good, g_i, which is affected by the agent's location in a network, N_i, as well as the gathering actions of other agents, $R_{j\in\sim i}$. We see here one of the main ways my approach differs from traditional economics insofar as the production function defined is determined by the spatially heterogeneous distribution networks unique to each agent and the actions of rival agents to change the production function. These components in $f(\cdot)$, N_i and $R_{j\in\sim i}$, will be discussed in more depth below. Once produced, the gathered good can either be consumed, c_i^g, or sold, c_i^s. Agents may purchase the gathered good as an alternative or supplement to their gathering

behavior, such that total consumption of the gathered good, c_i, is the sum of c_i^g and the quantity purchased of firewood and firewood substitute, c_i^b. The purchased good is bought at price p_f^b from the money income of the agent, w_i. Income is earned either by selling the gathered good at price p_f^s or working for wages at price p_w.

I simplify the model further by assuming that agents do not sell any of the firewood they have gathered due to transport costs and distance to markets (but I retain that they may buy firewood or its substitute), so that $p_f^s = 0 \; and \; c_i^s = 0$. Additionally, assume agents spend their full budget each time period and all of the endowed labor is allocated to one of the three possible choices. As discussed in the introduction to this chapter (which will be expanded in the data section), the assumption that agents do not sell their firewood holds quite broadly in Tanzania. With the simplifications above, the problem becomes:

$$\max_{L_i^g, L_i^w, L_i^l} u_i(c_i, L_i^l) = \alpha_i {c_i}^{\beta_i^c} {L_i^l}^{\beta_i^l}$$

$$Subject\ to: \qquad c_i^g = f\left(L_i^g | N_i, R_{j \in \sim i}\right)$$

$$c_i = f\left(L_i^g | N_i, R_{j \in \sim i}\right) + c_i^b$$

$$c_i^b = \left(\frac{p_w}{p_b}\right) L_i^w$$

$$L_i^g + L_i^w + L_i^l \leq L$$

Which simplifies further to:

$$\max_{L_i^g, L_i^w, L_i^l} \alpha_i \left(f\left(L_i^g | N_i, R_{j \in \sim i}\right) + \left(\frac{p_w}{p_b}\right) L_i^w \right)^{\beta_i^c} {L_i^l}^{\beta_i^l}$$

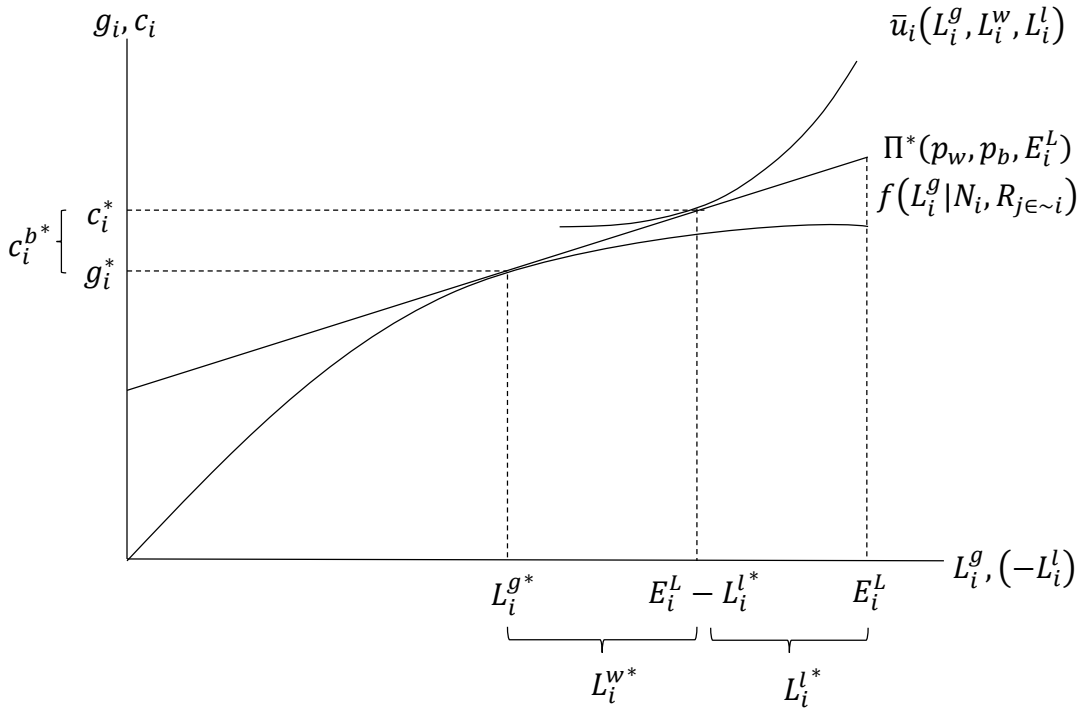
$$Subject\ to: \qquad L_i^g + L_i^w + L_i^l \leq L$$

The full profit function the individual faces combines the wages they receive and the firewood they gathered, valued at the price of the firewood substitute:

$$\Pi^*(p_w, p_b, E_i^L) = p_b f\left(L_i^g | N_i, R_{j \in \sim i}\right) + p_w L_i^w$$

Fig. 2.3 below graphs one example agent's labor allocation choice and specifies an example production function, profit function and a utility indifference curve at the optimized labor allocation choice. In this particular case, the indiviual forages for the majority of their fuel needs (g_i^*) but supplements firewood with purchased fuel ($c_i^{b^*}$), shown as quantities on the vertical axis. The labor decisions made by this agent are shown on the horizontal axis. Note that the horizontal axis also plots the negative value of leisure so that utility is represented with a flipped indifference curve between the gathered good and leisure.

Figure 2.3: Conceptual Model of Household Production and Consumption Decisions



This basic model could be solved for equilibrium by taking first order conditions, as in the traditional approach.

I now consider the addition of spatially heterogeneous production and competition. Consider first the production decisions an agent faces when on a heterogeneous landscape but while temporarily assuming there are no competing agents. Above, I denoted the production function as subject to a term N_i that represents the distribution grid-cell network as viewed from the i-th position. Using the network and normalization methods discussed earlier, N_i can also be represented as a set of geospatial grid-cell matrices. Two

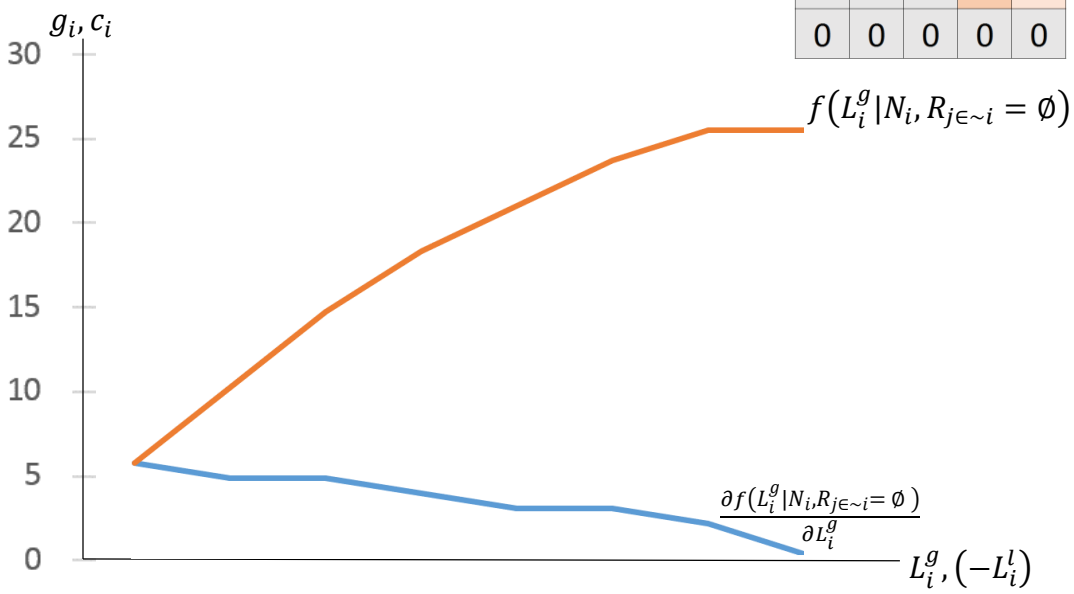
key matrices in this set are the firewood abundance matrix and the net-profit matrix. In the methods section, I will define precisely how net profit is defined, but for now it is useful to simply say that the net-profit depends both on how much firewood is present in each grid-cell and also the travel costs incurred getting to that cell from the center of the matrix.

In the upper-right of Fig. 2.4, I show an example net-profit matrix. In this example, there is a high-quality forest in the northwest and a lower-quality forest in east-southeast. In both cases, the forests have diminishing net profit available on the sides furthest from the center of the matrix (the agent's location) because foraging from these cells will incur additional travel costs. When it is the i-th agent's iteration step, the max-marginal-gain behavioral rule discussed above implies the agent will gather from grid-cell $N_{1,1}$ where the net profit is 6. After a cell has been foraged, assume its remaining value is zero. Assume further that the step-size in this example is defined so that an agent may only deplete one grid-cell per step. Thus, in this case, the agent gets a marginal value on this step equal to 6. This value is plotted in the blue line, which represents the marginal product of foraging, and also the orange line, which represents the production function. On the agent's next turn, they will choose to forage either on $N_{0,1}$ or $N_{1,0}$, will gain a marginal value equal to 5, and will increase total production to 11. This process will iteratively continue, one action per iteration step, until the agent has no remaining grid-cells with positive net profit (zero-profit, market clearing conditions will be discussed below).

Figure 2.4: Generating a Production Function from a Net Profit Geospatial Grid-Cell Matrix given no Competition

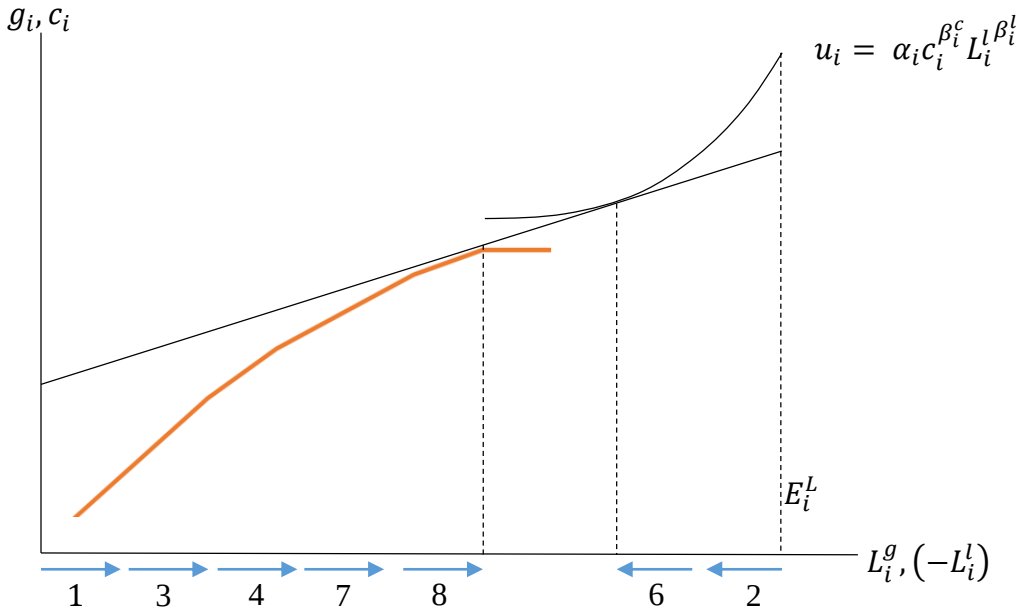
$N_i =$

4	5	0	0	0
5	6	0	0	0
0	0	0	0	3
0	0	0	3	2
0	0	0	0	0



The production function described here ignores the non-production activities the agent may do. At each iteration step in the full decision framework, the agent chooses to allocate labor to L_i^g, L_i^w or L_i^l, based on whichever has the highest marginal value. The marginal value from L_i^g is derived from N_i as described above, while the marginal values from leisure and purchased fuel depend on the shape of the agent's utility function and the prices of wage labor and fuel. Intuitively, given concave production and concave utility from leisure, this means in the initial iteration steps of a time period, the agent will alternate between producing firewood and enjoying leisure. Eventually, at least in the example above, diminishing returns will cause the agent to switch to wage labor for the final units of fuel consumed. This process is depicted in Fig. 2.5 where each blue arrow indicates the action chosen on the s-th iteration step.

Figure 2.5: Max-Marginal-Gain Choices between Leisure, Production and Wage Earning over Multiple Iteration Steps



The iterative process by which the agent moves towards the maximized point allows for events to happen in between each decision step and agent takes. The main type of inter-step event I include is depletion of the forest by other agents. The production function defined above assumed no competing agents were present, but we can extend the process to multiple agents (see Appendix 3).

Situations in which agents compete with each other may lead the max-marginal-gain decision rule to suboptimal allocations. For instance, agent 1 may have predicted that agent 2 would want to forage from the contested cell and would thus choose to forage there earlier. This type of strategic behavior can easily be included in agent simulation by

using modified decision rules (such as doing the best-response action in a Nash equilibrium). In many applications though, this becomes computationally impossible when millions of agents are considered simultaneously and does not present large deviations from the naïve max-marginal-gain decision rule. Thus, in this project, I assume there is no strategic interdependence between agents and instead address the issue by comparing base results to sensitivity analyses that do address interdependence.

Definition of Equilibrium

The results of this simulation are characterized by inter-temporal and intra-temporal equilibria. Because the application used in this chapter focuses on sustainable harvest rates, I describe here only the intra-temporal equilibrium, using a conception similar to the Ramsey (1928) approach.

To define equilibrium, first define a zero-profit zones for each agent $B_i = \{N_i \; \forall \; i \; where \; f_i(E_i^L | N_i, R_{j \in \sim i} = \emptyset) > 0\}$.

The zero-profit boundary B_i is the set of all grid-cells where the agent has positive net-profit assuming no other agents deplete the grid-cells and assuming the agent spends their entire endowment of labor on foraging. Outside the zero-profit boundary, the agent has no profitable cells even in the best circumstances, so ignoring these cells has no impact on the outcome.

An intra-temporal equilibrium is characterized by allocation choices of labor $(L_i^g, L_i^w, L_i^l) \; \forall \; i \in A$ and a geospatial grid-cell matrix of local prices P_k for each k-th good such that each agent satisfies the following conditions:

1. There does not exist any grid-cell within the i-th agent's zero-profit zone B_i where $U_i\big(N_i(net\ profit)\big) > U_i\big(L_i^l\big)$.
2. There does not exist any possible purchase where $U_i\left(\left(\frac{p_w}{p_b}\right)L_i^w\right) > U_i\big(L_i^l\big)$.

One useful attribute of the agent-based simulation approach to economic analysis is that models can easily be extended. Although I have specified a model that is based only on one good (produced two ways) and leisure chosen with a single decision rule, it is very easy to extend the analysis to include additional goods, additional behavior rules, or a very wide set of heterogeneity among agents. The reason this approach is more capable of extension than approaches based on constrained optimization is that we are not solving equations but rather testing rules. This means we do not have to limit ourselves to models with analytic solutions or models that can be approximated by systems on differential equations. As long ago as Leibniz in the 17th century, we know that nonlinearity in variables leads to a broad set of models to be unsolvable analytically. In the next sections, I show how this can be done with data on firewood foraging from Tanzania.

Data

The data requirements for agent-based simulation of economic systems and ecosystems are considerably higher than for traditional economic analyses. In this section, I describe the data sources I used along with several new methods I created to manage the varying data needs. First, however, I summaries the overall data needs for the foraging model to work in Table 1.

Table 2.1: Summary of Data Sources

Data Type	Use	Data Source
Land Use, Land Cover (LULC)	Used in many applications, including identification of agent locations, creation of objective abundance and calculation of traversal cost.	MODIS*
Spatial Population	To use the pixel-agent specification described in the methods section, data on the per-pixel population is necessary.	AfriPop*
Traversal Cost Factors	I used road networks, river networks, land cover, and elevation (converted to ruggedness), but other locations may require different sets, such as political crossings or other transit modes.	OpenStreetProject, HydroSHEDS, MODIS, Digital Elevation Model, Ruggedness
Household Behavior and Demographics	I use local wages as a proxy for opportunity cost. Given that wages vary considerably within Tanzania, I use household survey information from the National Panel Survey (2008) to construct a wage map. Many other household-specific data may be used, either for initial parameterization or model validation. For example, I discuss how I used household data on hours spent foraging for firewood to validate the model.	Tanzanian National Panel Survey (NPS) *
*** indicates data source discussed in a separate section.**		

Satellite Based Data

Due to the spatial focus of my work, the majority of my data come from remote sensing methods, most frequently from the Moderate-resolution Imaging Spectroradiometer

(MODIS) instrument on the Terra (EOS AM) and Aqua (EOS OM) satellites from the National Aeronautics and Space Administration (ORNL DAAC, 2011). MODIS measures radiation spectral bands from 0.4 to 14.4 micrometers at varying spatial resolutions. A very large literature exists to study remote sensing tools like MODIS, for instance Justice et al. (1998), Barnes et al. (1998), Lefsky (2010), and Toller et al. (2013). Within the large set of MODIS data products, I use the Land Cover Type dataset (NASA Data Product MCD12Q12), but future analyses may incorporate additional products for burned areas (MCD45A1) vegetation indices (MxD13C1). The raw MODIS radiation data are algorithmically categorized into five different classification schemes, of which I use the IGBP global vegetation classification scheme (Friedl et al. 2010). Table 2 presents the IGBP classification system with land cover types and corresponding MODIS classification indices.

Table 2.2: IGBP Land Use, Land Cover Categorization of MODIS Data

Class	IGBP (Type 1)
0	Water
1	Evergreen Needleleaf forest
2	Evergreen Broadleaf forest
3	Deciduous Needleleaf forest
4	Deciduous Broadleaf forest
5	Mixed forest
6	Closed shrublands
7	Open shrublands
8	Woody savannas
9	Savannas
10	Grasslands
11	Permanent wetlands
12	Croplands
13	Urban and built-up
14	Cropland/Natural vegetation mosaic
15	Snow and ice
16	Barren or sparsely vegetated

254	Unclassified
255	Fill Value

Using MODIS data is not easy. The datasets are very large, span different time periods with different time steps (ranging from minutes to days), are broken into smaller data blocks (to avoid memory constraints on computers) that need to be stitched together correctly while re-projecting the different spherical projections used in different zones, are provided in file types not commonly used in typical GIS applications (HDF format), and in their raw form are provided as non-categorized integer values.

To overcome the challenges discussed above, I used the HDF-EOS To GeoTIFF Tool (HEG), provided by NASA to work with MODIS data and other satellite data streams (ASTER, MISR, AIRS, and AMSR-E), described in Dwyer and Schmidt (2006). Additionally, because the process for using HEG was very time intensive when applied to the large set of files I needed, I wrote a python script to automate the process (available on request).

Overcoming the challenges of using MODIS are well worth it because MODIS allows for high-frequency time series data from 2000 to the present with high resolution on dozens of different variables at a global level. One of the goals of this dissertation is to create a method that can easily be replicated in any location, and thus, using a source like MODIS is particularly fitting.

Fig. 2.6 below shows the results of processing the MODIS data for Tanzania while Fig. 7 presents the same data for a subset near Mount Kilimanjaro.

Figure 2.6: MODIS LULC Data Processed for Tanzania

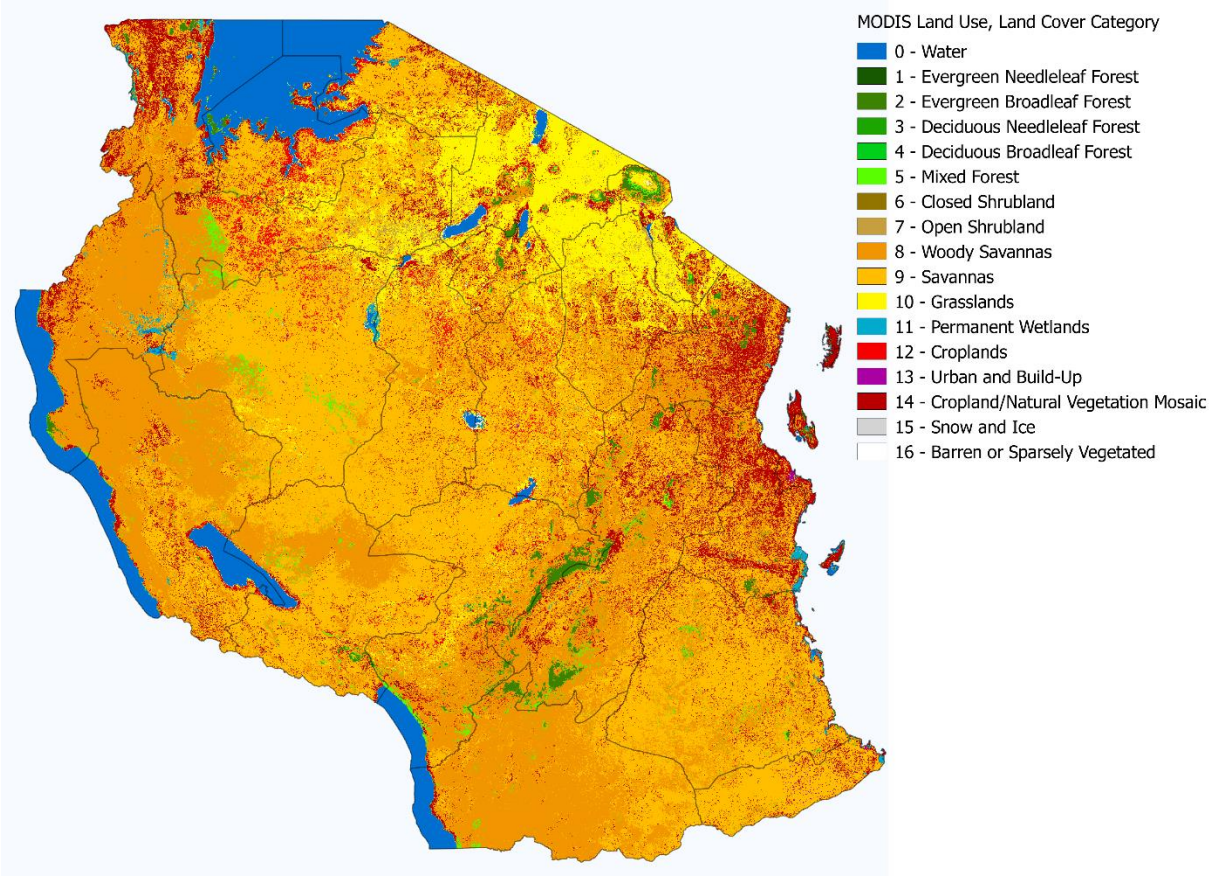
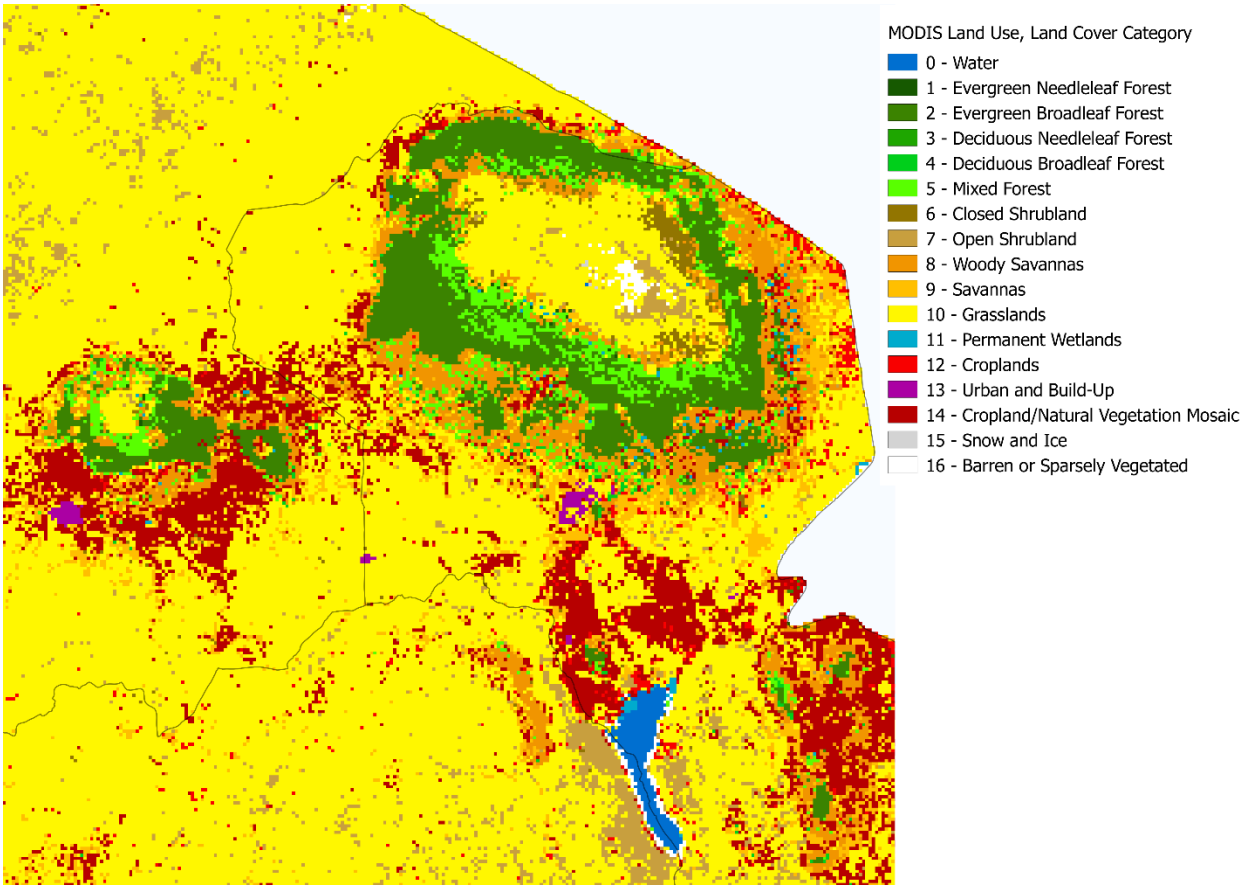


Figure 2.7: MODIS LULC Data Processed near Mount Kilimanjaro



Household Survey Data

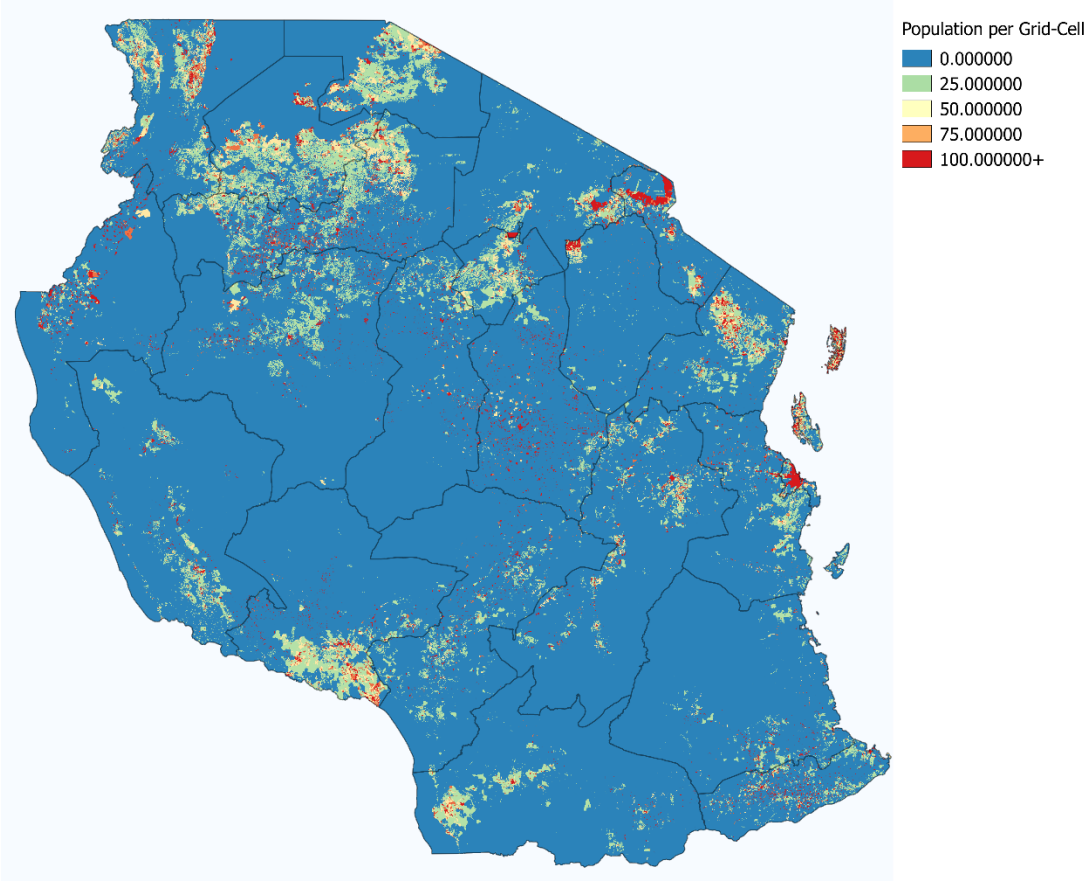
Economic and demographic information cannot easily be obtained from satellites, so I supplement the remote sensing data with traditional household surveying methods. Data from the Tanzanian National Panel Survey (NPS) provides information on wages, hours spent collecting firewood and a wide variety of other household- and individual-level statistics that can be used to assess other determinants of foraging behavior, such as

opportunity cost. Over 16,000 individuals in nearly 4,000 individuals were included in the survey, which asked over 200 questions. Care must be taken when using data from the NPS because: 1.) many questions have separate look-up tables for coded responses; 2.) questions can be specific to the individual, household, firm, or a variety of other entities (which presents difficult aggregation issues); and 3.) four different index systems were used for different components of the survey. The NPS is especially useful in this study because every household has a recorded latitude and longitude location, allowing for combination of ecosystem data and economic data in many new ways. Unfortunately, In the publically available data these coordinates are obfuscated for privacy reasons by 3 to 10km and lumped into enumeration area groups, but requests to the World Bank can be made, especially for projects that already show results with the obfuscated data. For this analysis, I was unable to obtain the private data and so had to use the obfuscated coordinates. Because the obfuscation was done randomly, it will introduce greater standard errors in predicted relationships but will not produce biased results, but it prevents robust validation, as will be discussed in a later section. See Appendix 2 for more discussion of the NPS and other data sources.

Population Data

I will discuss population interpolation in the methods section of this chapter for situations when high resolution data is not available. Fortunately, high resolution data *is* available for Tanzania through the WorldPop project. The methodology uses remote sensing data to identify where settlements are located and then allocates population Figures from survey information to these settlements (and other land use classes that have positive population density). See Gaughan et al. (2013) for more information on this methodology. Fig. 2.8 shows the WorldPop population density Figures for Tanzania.

Figure 2.8: Population per Grid-Cell from WorldPop



Calculation of Wage Surface

Firewood collection is a time consuming activity. Because firewood can often be collected from common property areas, the main input cost is determined by what the individual could have done with their time. National wages for Tanzania are not difficult to collect, but in the context of firewood collection, national aggregates do not provide enough resolution to reflect the extreme heterogeneity of wages across the country. The household survey information confirms this story. Unfortunately, the surveys are only taken in a few hundred physical locations (after obfuscation, which merges multiple

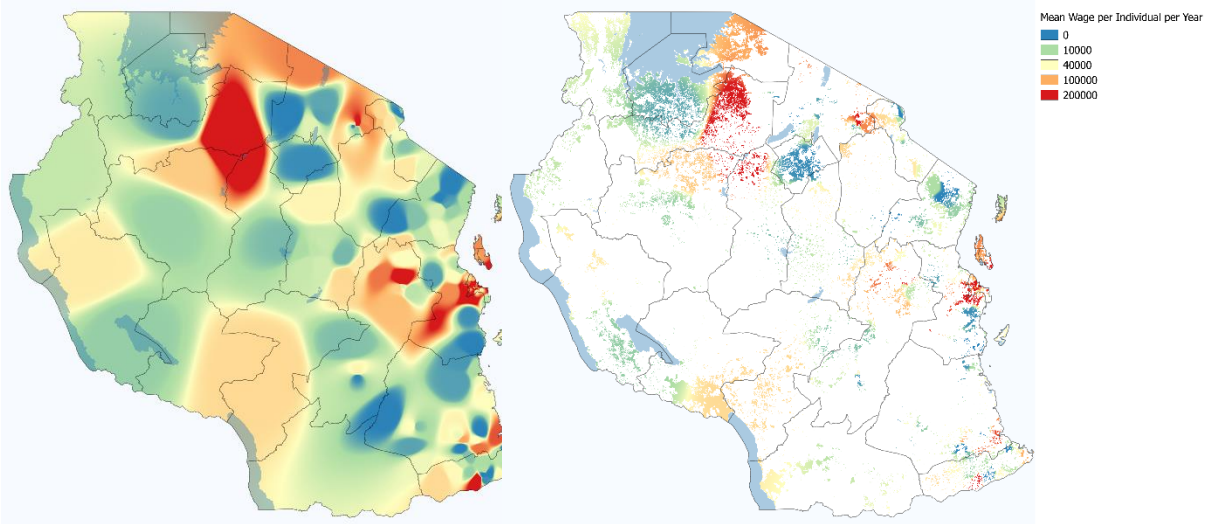
observations onto a single point) and thus provide only a sparse data input in which the vast majority of grid-cells have no observed wage. To overcome this problem, I present and then utilize a method for creating a wage surface from these sparse points. The process by which I convert point-based data to data surfaces is useful in a broad set of situations in addition to creation of wage surfaces. Although I succeed in applying the method to a data-sparse environment, many of the challenges I will discuss below are more easily overcome when there are more observation points.

First, using the reported wage values on the NPS, I plotted the wage points on a sparse agent-point raster with the z-value indicating the wage. With this, I used an inverse distance weighted interpolation algorithm to create a continuous surface based on the household survey data (see Zimmerman et al. 1999 or Mueller et al. 2004 for a discussion of different interpolation techniques relevant to this type of problem). This interpolation process converts the sparse point data to smooth surfaces where the value of non-observed grid-cells changes continuously between observed points according to the interpolation assumptions chosen. The map on the left of Fig. 2.9 shows the wage surface this method created. I apply the information we know about population density per grid-cell (discussed above) to calculate agent-specific wage figures (shown in the map on the right of Fig. 2.9).

This wage surface is useful because it identifies with more detail how wages, and thus the opportunity cost incurred when collecting firewood, vary with geographic location. More relevant to the analysis here, this wage surfaces provides a wage value for all grid-cells within the country so that we can simulate the behavior of individuals who are not near an NPS observation point. A criticism of this approach is to argue that it is not legitimate to infer that the wages in grid-cells unobserved by the NPS are systematically related to their (inverse weighted) distance from observed cells. I will address this in future work

that uses the inverse distance weighted interpolation technique with a multivariate regression to capture the non-distance effects. For now, however, it is worth noting that the assumption I use is superior to the assumption that there is no geographic component to wages (this is the assumption that would implicitly be used if one were to simply use national wage averages). Additionally, the contribution I am working to present here is a new method for simulating spatial economic activity and that any different set of assumptions on wage are consistent with the agent-based approach I use and can readily be incorporated into my model.

Figure 2.9: Wage Surfaces



Literature Values on Foraging Behavior and Firewood Abundance

One of the most difficult and important aspects of simulating firewood foraging was identifying the abundance of firewood available on each grid-cell. Identifying the cubic meter yield of each grid-cell in a small subset of Tanzania would constitute an ambitious

biology project, so approximations must be made to make nation-wide estimations possible. For this task, I survey the literature on land use types in Tanzania and estimate values of firewood density for the 16 MODIS land use classes. Additionally, I use literature estimates for other parameters in my model and for validation, including average consumption Figures, likelihood of purchasing fuel from a market, hours spent on foraging and a variety of other factors.

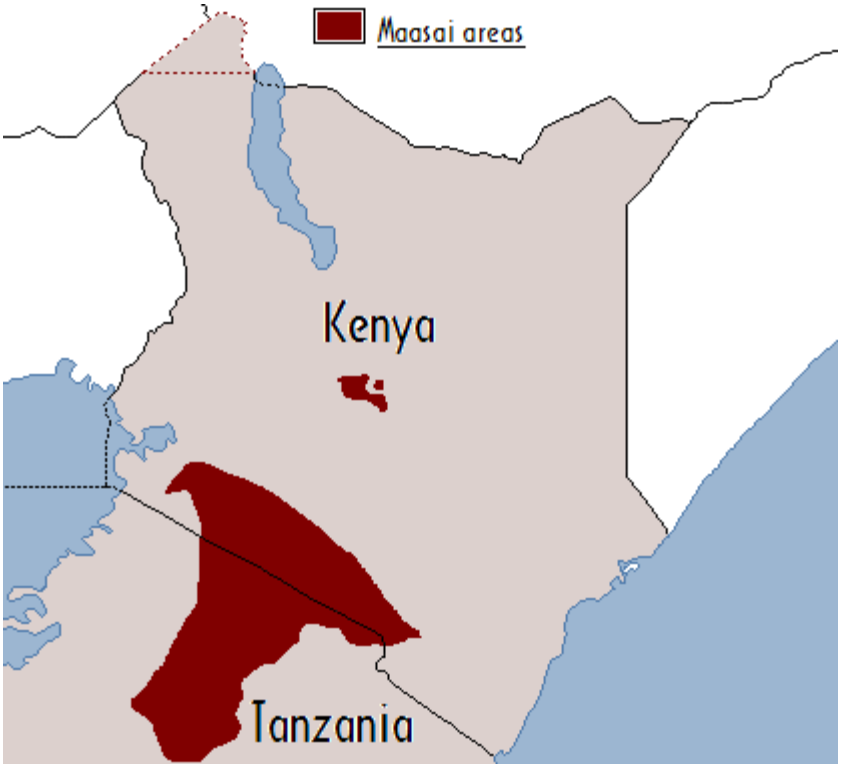
Firewood consumption in Tanzania has been well studied, including sources as early as Fleuret (1978), Berio (1984) and many newer studies. Of particular use to our simulation tasks, Berio studied how villagers in the four sub-Saharan Africa region allocated their time to household tasks, including wood collection. Berio estimates the weight of firewood consumption per person per day at 1.1, 1.0, 1.3 and 1.0 for the four regions studied.

After the studies conducted by Berio, other studies have focused primarily on the relationship between firewood and charcoal collection. For instance, Hosier et al. (1993), van der Plas (1995) and Johnsen (1999) found that the industry for charcoal creation poses large risks of deforestation, but that direct firewood collection rarely poses a threat to forests. The reason for the lower threat of firewood collection is that villagers prefer to collect dead wood because it is easier to gather and burns better. Charcoal, conversely, can be produced using fresh-cut trees and thus does result in deforestation.

Biran et al. (2004) studied firewood foraging behavior in two sub-Saharan Africa locations, which included the Maasai community in Northeastern Tanzania (see Fig. 2.10) and the area around Lake Malawi National Park (which shares a border with Western Tanzania and has very similar conditions). Biran reports that the mean length of a wood collection trip in the national part was 241 minutes and was undertaken every 3.8

days, resulting in a mean daily collection time of 63 minutes. For the Maasai, conversely, reported times again were 10 minutes per day per capita for women and 30 for girls.

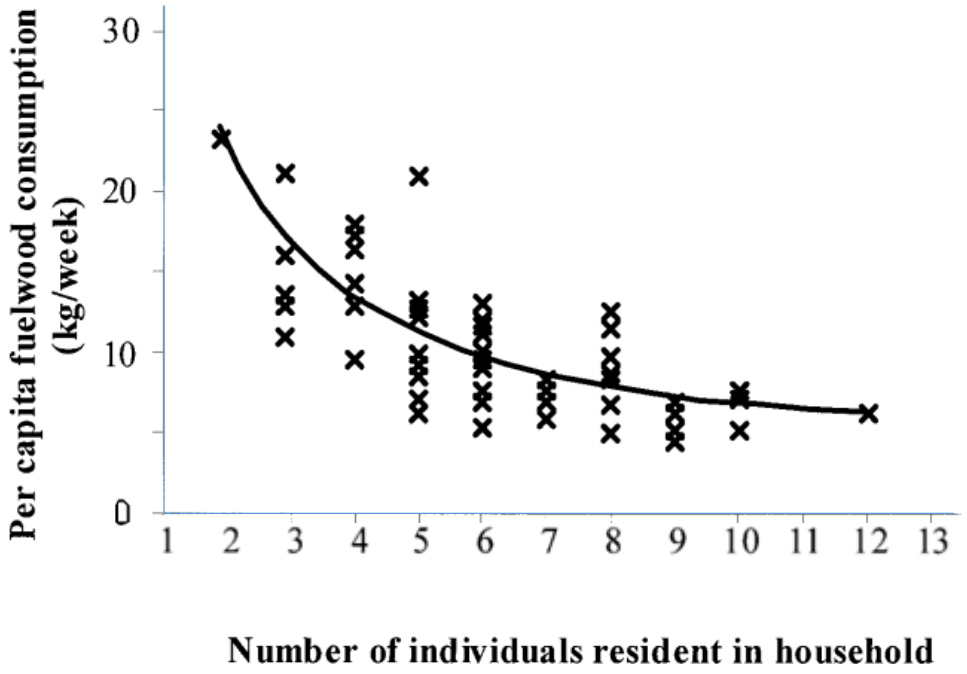
Figure 2.10: Location of Maasai Firewood Collection Study (Berin et al. 2004)



Of the length of a firewood collection trip, only 123 minutes were spent actually in the forest, implying that travel to the collection location consumed a significant portion of the overall time required.

Another useful value that Biran presents is the relationship between number of residents in a household and the per capita fuel wood consumption Fig. 2.11 shows the relationship Biran found.

Figure 2.11: Per capita Firewood Consumption versus Residents per Household



The average household size in Tanzania is approximately six, which means, according to Biran's relationship, the per capita firewood consumption is 12.5 kg/week. Due to population growth and other factors, firewood demand in Tanzania has grown since Biran's report.

Another area of work on this topic comes from Luoga et al. (2000, 2002 and 2004), based on previous work from Maliumbwi et al. (1994). This research focuses on the Miombo woodlands, in central and eastern Tanzania. This region is especially important because it contains very high populations and relies heavily on the surrounding woodlands for non-timber forest products, but also wood products that require clearing the forest (charcoal

and timber). Their work showed that harvesting levels in public lands were not sustainable (6.38 $m^3 ha^{-1}$ versus the mean annual increment value of 4.35 $m^3 ha^{-1}$). These harvest levels were provided along with volumes of standing stocks, reported as 47 $m^3 ha^{-1}$ in forest reserves, but only 16.7 $m^3 ha^{-1}$ on public lands. These values help us identify the firewood abundance map, which is one of the primary inputs to this model.

Luoga's work also provides us with two qualitative results that we can test to see if our model validates with his observations:

1. "Harvesting intensity decreased with increasing distance from village settlements and reserve boundaries but the pattern had no significant overall impact on standing stocks of wood. The Dar-es-Salaam–Morogoro highway, which bisects the study area is the major axis of disturbance."
2. "Harvesting intensity tended to decline with increasing distance from the village or forest reserve boundaries (r^2=0.19 and 0.16 for "old" and "all stumps", respectively, in public lands (linear fits), with all other r^2<0.1 irrespective of the model used)."

Luoga et al. (2002) focused attention on the economics of firewood and charcoal production. They noted that most of the value of firewood collection lied in the "domestic and subsistence uses within households" (based on previous work from Shackleton, 1993; Campbell et al., 1997; Shackleton et al., 1999) and noted the "very limited market opportunities" further from cities.

More recently yet, Fabe and Grote (2013) discussed the importance of firewood to household subsistence, noting that only 13% of total firewood consumption comes from purchased firewood. They also report that 97.5% of households use firewood for their heating and cooking needs. Together, this shows that there clearly is not a complete

market for firewood and thus it is important to understand the non-market aspect of firewood collection. Additionally, Fabe and Grote report that each household member consumes 368 kg of firewood per year. This supports previous estimates (Zahabu et al. 2005, Malimbwi and Zahabu 2009), including those that use consumption Figures to calculate national demand (for instance, Felix and Gheewala (2011) report that average national consumption of firewood was 55.5 million m^3.

One of the most useful sources for estimating firewood abundance and firewood demand is Zahabu et al. (2005) and Malimbwi and Zahabu (2009). In their analysis of the sustainability of fuelwood production systems in the Miombo woodlands in Southern Tanzania, they further support the contention that "subsistence fuelwood collection rarely affects the natural Miombo woodlands since only the dead wood or wood cut for other purposes is collected." Additionally, they report stand densities of woody plants. They note that the basal areas range from 7 to 25 m^2 per ha and stand densities of 380 to 1400 stems per ha. They calculate the mean annual volume increment (MAI) of mature woodland species at 0.58 to 3.0 $m^3 ha^{-1} year^{-1}$. The consumption values they report is in line with previous estimate 1.5 $m^3 capita^{-1} year^{-1}$. Malimbwi and Zahabou also report that most firewood is collected within 500 meters of the household (from Bandeira et al.,1996).

The focus of this thesis is on the modeling techniques not the field-level observations of firewood abundance. Other disciplines are better suited to biophysically estimate what the expected firewood abundance per hectare is. Nonetheless, the discussion above presents a way to use literature estimates to create a range of parameters that can be used in the foraging model. I summarize the specific value or range of values I use as parameter inputs into my model in Table 3.

Table 2.3: Summary of Relevant Literature Values for Calibration and Validation of Foraging Model

Parameter Name	Value (or range) Used in Model or Validation
Average travel time per trip:	118 minutes
Average Growth rate of trees	15 kg per tree
Mean annual volume increment	0.58 – 3.0 $m^3 ha^{-1} yr^{-1}$
Trees per Hectare	33
Mean harvestable volume per hectare	14 – 117 $m^3 ha^{-1}$
Per capita firewood consumption	1.5 m^3
National firewood consumption	25.8 to 55.5 million m^3

Creating Traversal Cost

The final input data required to run the model is the cost of traversing each grid-cell to populate the travel-cost network in the landscape data. This is another input that can be chosen changed by the user to match the information known about a region. The process I use starts with the assumption that humans walk at 1.38 meters per second on paved surfaces. This is the median value reported in Levine and Norenzayan (1999), who studied walking speeds in 31 countries. I then account for slower travel over different land types by following adjusting the assumed walking speeds using values in Table 4.

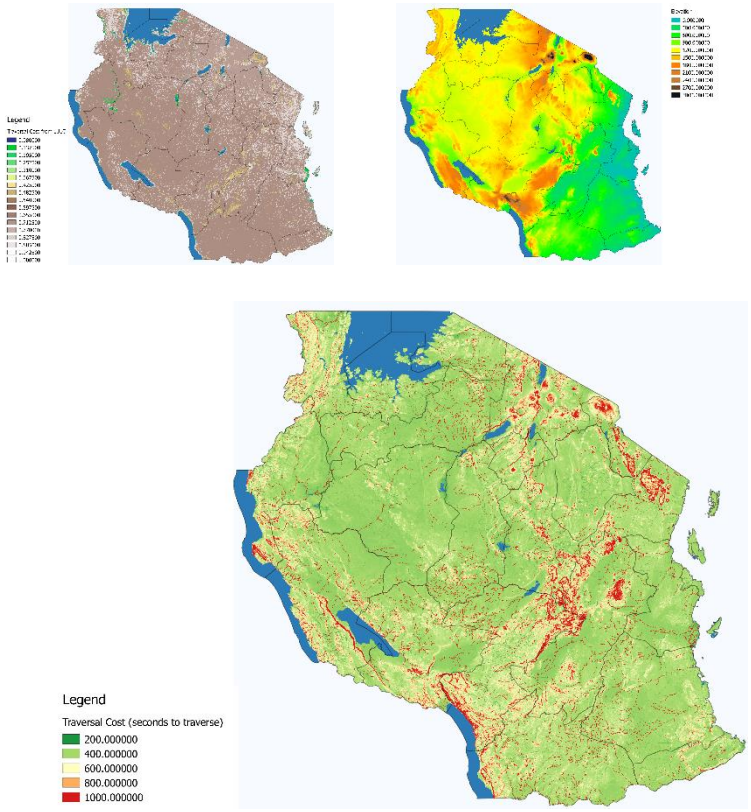
Table 2.4: Assumed parameters on travel speed across different LULC types

MODIS LULC Class	MODIS Name	Times slower than road travel	Implied Travel Speed (meters per second)
0	Water	15	0.092592593
1	Evergreen Needleleaf forest	3	0.462962963
2	Evergreen Broadleaf forest	3	0.462962963
3	Deciduous Needleleaf forest	3	0.462962963
4	Deciduous Broadleaf forest	3	0.462962963
5	Mixed forest	3	0.462962963
6	Closed shrublands	2	0.694444444
7	Open shrublands	2	0.694444444
8	Woody savannas	2	0.694444444
9	Savannas	2	0.694444444
10	Grasslands	1.9	0.730994152
11	Permanent wetlands	9	0.154320988
12	Croplands	1.6	0.868055556
13	Urban and built-up	1	1.388888889
14	Cropland/Natural vegetation mosaic	1.6	0.868055556
15	Snow and ice	9	0.154320988
16	Barren or sparsely vegetated	9	0.154320988

As with many of the parameters discussed above, these are ad-hoc assumptions necessary for the simulation to be well-defined. However, these values are best interpreted as user-inputs that can be defined with the specific foraging location in min. Applying the values in table 4 to the MODIS LULC data, I created the map shown in the upper left of Figure 12, which is the minutes necessary to traverse across a given grid-cell based on the LULC information alone. I then combine this time with additional information derived from the elevation and ruggedness of the terrain (shown in the upper right) as well as network data

on roads and rivers. The result is a traversal map, shown in the bottom center of Figure 12. I will revisit the use of this traversal map extensively in the methods section and use it to calculate optimized-route arrival times.

Figure 2.12: Creating Traversal Cost from Land Cover and Elevation



Methods

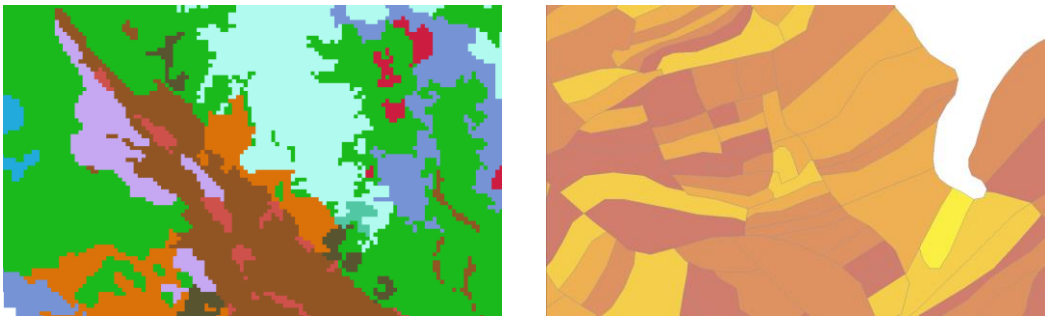
In this section, I discuss specific methods I used to run the simulation in Tanzania. Many of these methods are broadly applicable to all applications of spatial foraging. I will discuss an implementation that identifies 10 million agents in Tanzania, but I also provide

an example of how to define a model with fewer (250) representative agents in Appendix 3.

Population Downscaling

In many applications, there does not exist high resolution data on population density such as I have used here. Thus, I present a method that combines data at multiple resolutions to create high-resolution population density data. A common situation faced when modeling spatial economic phenomena includes having access to high resolution LULC data, but having population figures available only for larger administrative districts. Fig. 2.13 shows these two types of data in the upper left (for LULC data) and upper right (for administrative data).

Figure 2.13: Input Data for Population Downscaling



Different colors in the land-use map correspond to different types of development, for which I define different population densities. I assume, for instance, that land-cover categorized as "Urban" has higher density than "Agricultural," which I assume has higher density than "Forest" categories. These assumptions create a vector of weights, w_i, for each land use type according to the density that the land use has relative to other land

uses. I apply this vector of weights to the LULC map to create a new map with the weight values defined on each grid-cell. Next, I convert the administrative data into a grid-cell network with the same size and extent as the weights grid-cell network. Each administrative zone is given an integer ID value which is used to populate the values in the administrative ID matrix so that all grid-cells corresponding to a zone have the stored the same administrative ID. Finally, I calculate p_{rc}, the population density for each grid-cell, starting with the first administrative ID and incrementing upwards. I divide the sum of weights in the i-th zone by the population in the zone, then multiply this ratio by the value in the rc-th grid-cell for all grid-cells, as shown below:

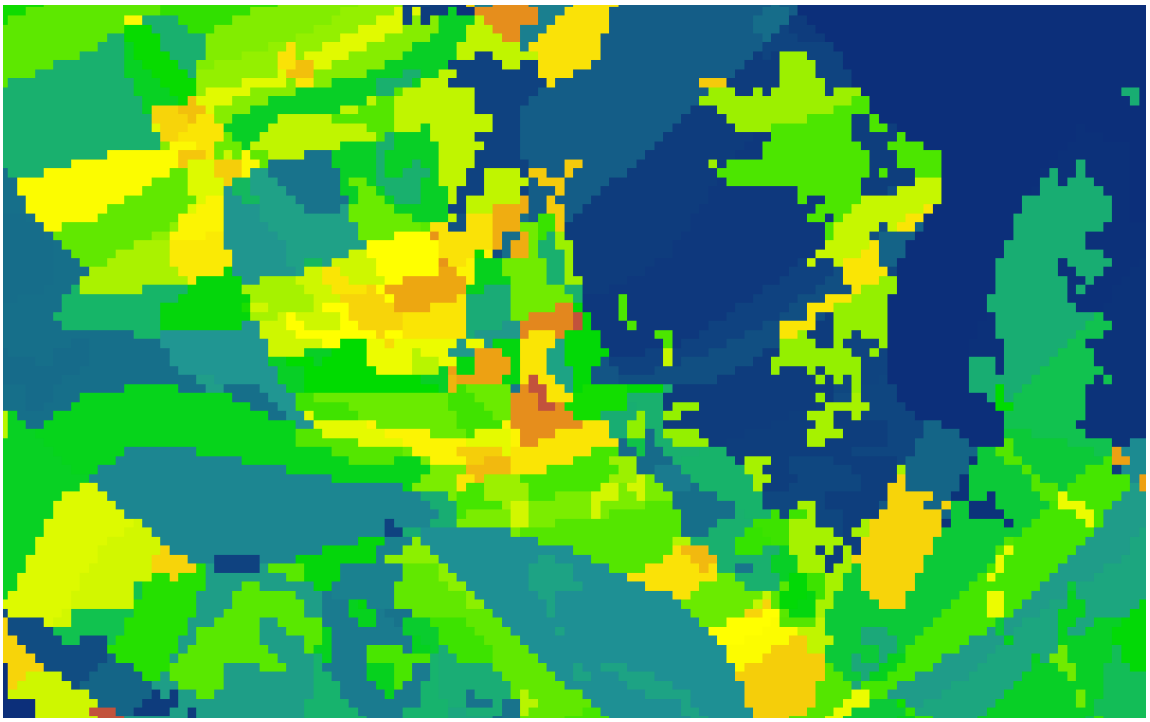
$$p_{rc} = \left(\frac{\sum_{i=0}^{I} w_i}{p_i} \right) w_{rc} \; \forall \; r, c \in i$$

This process identifies a per grid-cell population density. It has the useful attribute that summing up all grid-cells within an administrative zone will exactly equal the population listed by the zone's survey, but it also leverages what we know about land cover to interpolate likely values in all locations. Results from this method are shown in Fig. 2.14. Downscaling methods like this (which have been used extensively in other applications though with slightly different methods from mine[2]) are sensitive to the assumptions one makes about which LULC types have what relative proportions of population. However, many of these assumptions will be broadly correct even where they are not precisely calibrated. For instance, the assumption that urban land cover has higher population density than forests is not difficult to support. Further research in this area can be conducted that validates the process by making predictions, selecting a sample from the

[2] AsiaPop and AfriPop are the best examples of other applications of this general approach.

prediction and then testing those points against observed micro-data. In contrast to the lack of calibration on the assumptions used here, it is important to note that non-downscaled data also make an assumption about what the population density is on different land-use types, namely, that it the population is uniformly distributed everywhere in the administrative zone.

Figure 2.14: Downscaled Population Data



This downscaling method allows for analysis of phenomena that are sensitive to higher-resolution effects than are available in traditional population density datasets. The next section will discuss how higher-resolution data like this allow for definition of the pixel-agent approach.

Defining the Agents

In some applications, it is possible to make a model with relatively few representative agents. However, with firewood in Tanzania, this is difficult because individuals rarely forage more than five kilometers away from their home, yet village boundaries can be many time wider than this distance. Thus, I define my agent using the highest-resolution population density available. Fig. 2.15 presents the population present in every 450 meter grid-cell. Below that, Fig. 2.16 shows the same data, focusing on the area around Mount Kilimanjaro. This subset of the country around Mount Kilimanjaro will be used in much of the rest of this analysis because it allows individual pixels to be seen. Keep in mind, however, that everything shown for this sub-region is calculated in full for e very pixel in the broader scope of Tanzania.

Figure 2.15 Population per 450 meter Grid-Cell in Tanzania

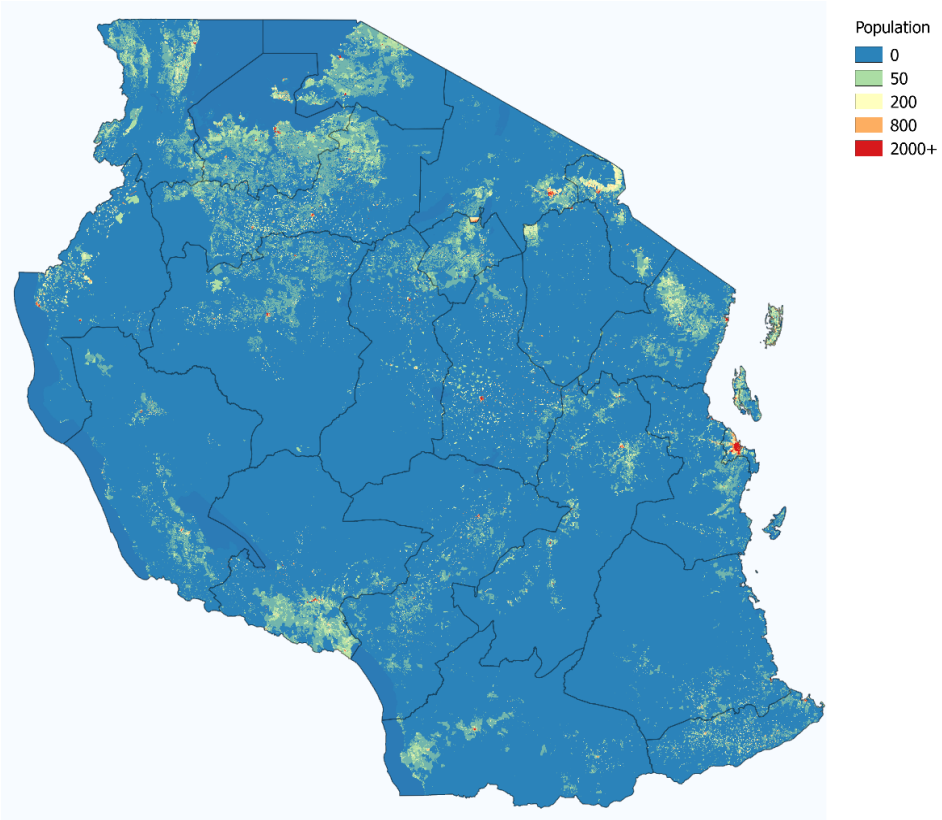


Figure 2.16: Population per 450 meter Grid-Cell near Mount Kilimanjaro

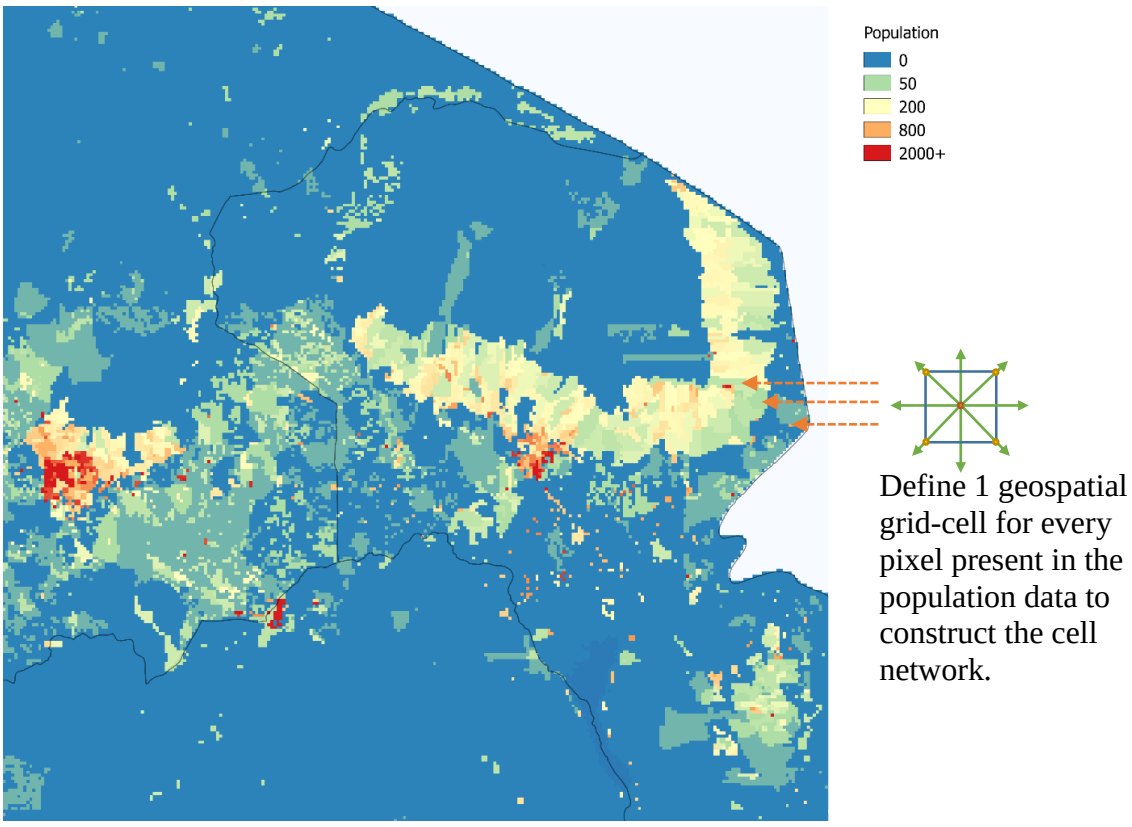


Fig. 2.16 above also shows how the population data are used to construct an agent set that defines one agent for every grid-cell in the network. This agent set can also be converted to a normalized matrix, which allows for mathematic matrix operations directly between the input data, cell networks or other sets of normalized agents (as described in Appendix 3.

The framework presented here is computationally feasible with 10-million agents, and in certain circumstances, can be extended to as many as 10 billion agents (thus allowing for a definition of agents as identical to individual humans). The next section considers

methods capable of defining actions, specifically travel of agents over landscapes and road networks, while maintaining the same degree of computational feasibility.

Calculating Travel Cost

One of the most basic calculations necessary when simulating agent behavior on a landscape is to identify the cost of traveling from one location to another. Even when using matrix views as discussed in the previous section, finding the optimal route can become computationally infeasible, especially if the destination is remote from the origin. In this section, I identify a method that allows for route-finding even in large data sets. I do this by identifying four different route-finding algorithms (presented in Appendix 3) and then create a composite algorithm that combines all four approaches hierarchically to apply the correct algorithm (defined as maximizing accuracy while minimizing computational intensity). This approach is able to provide a large computation speed increase because many route finding problems are quite simple (for instance, traversing a uniformly flat field) and the simpler algorithms identify the optimal route in a faction of the computational time necessary for more complex route-finding algorithms.[3]

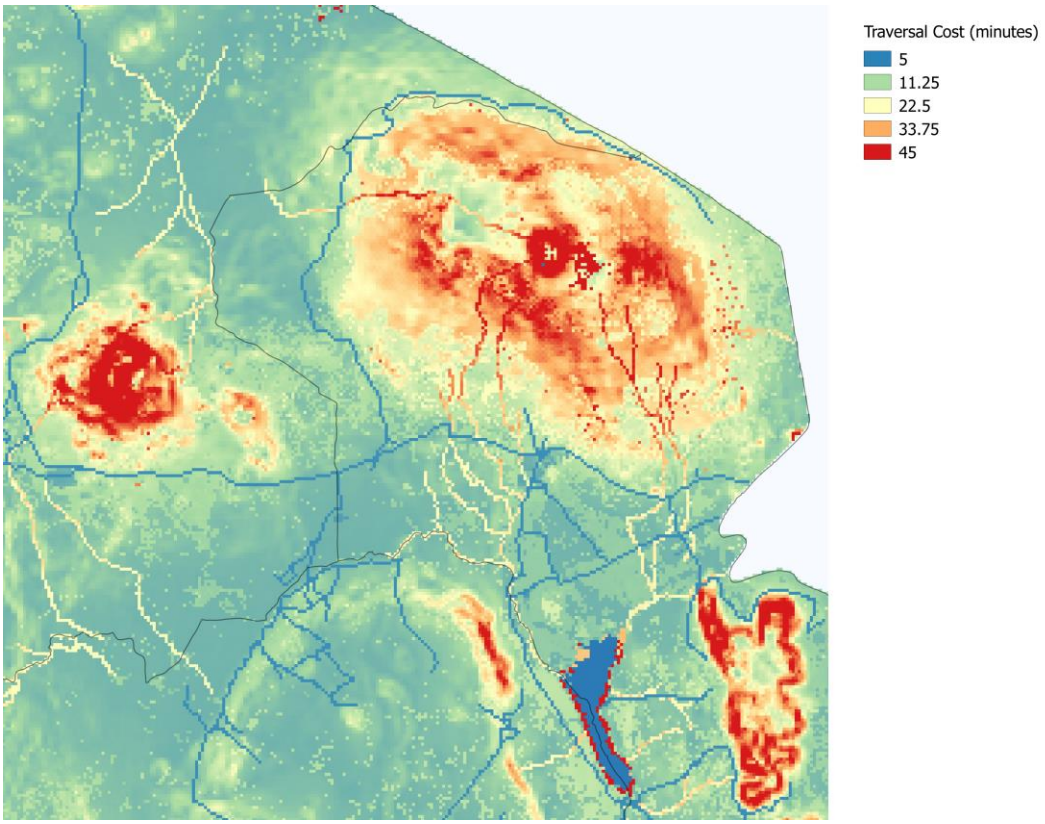
Constructing the Traversal Cost and Arrival Cost Grid-Cell Networks

I distinguish between two terms relevant to travel cost in this section that have unavoidably similar terms. The first term is *traversal cost*, which denotes the amount of time it takes an agent to traverse from one side of the grid-cell to the other when traveling

[3] In the course of this work, I also created an algorithm that works well even when extremely long routes must be calculated. This algorithm works by solving the hierarchical search algorithm at varying levels of resolution for the traversal cost data and then combining the results to calculate the full route. I have omitted it from this analysis due to time constraints.

on one of the cardinal axes. Fig. 2.17 shows the traversal cost matrix for the Northeast of Tanzania.

Figure 2.17: Traversal Cost Grid-Cell Network

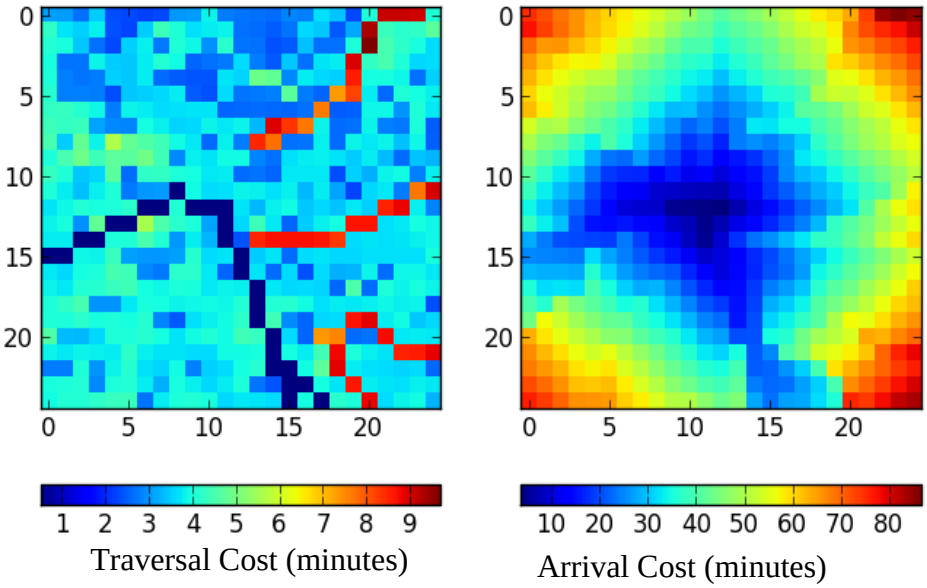


This matrix was constructed by using literature values (discussed in the data section) on how fast humans can walk over different terrain on average. The map above takes into account slower travel over different land cover types, increased speed from roads, difficulty crossing rivers, and finally, a ruggedness index based on topographical data for slope and elevation. In future work, this traversal cost will be modified to allow for multi-modal transit (such as cars and boats), but for now I am limiting analysis to walking

speed only. I expect this assumption not to influence the results unduly because walking remains the dominant mode of transport when foraging in forests and other rough terrain.

The second term is *arrival costs*, which is the cumulative traversal cost incurred crossing all cells necessary to reach a destination while taking the optimal route. Fig. 2.18 shows an example of calculating arrival cost. The left image shows the traversal costs as calculated above, while the right image shows the arrival cost, defined by traveling from the centermost grid-cell in the image to each of the other grid-cells.

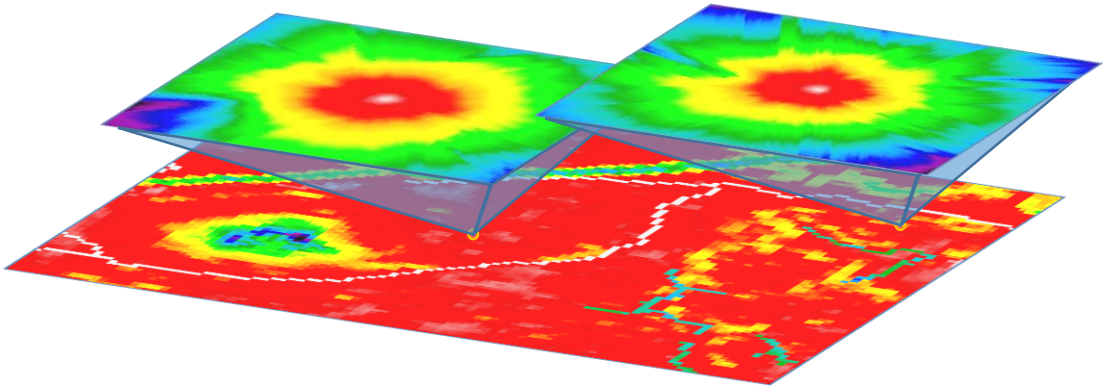
Figure 2.18: Calculation of Arrival Cost Maps



Because this map is unique to which pixel is the starting location, it must be recalculated for every grid-cell on the landscape. Additionally, to ensure that the equilibrium can be identified, each arrival cost map must be calculated for areas large enough to cover each

grid-cells' zero-profit zone. This is computationally challenging as it involves generating a large amount of data. In the computational approach I use, the arrival map for each grid-cell is described in a 4-dimensional matrix, which Fig. 2.19 illustrates with a traversal cost (the base map) and two representative arrival cost maps.

Figure 2.19: Illustration of Combined Arrival and Traversal Cost Array



Once created, this array is particularly useful in the foraging simulation when calculating the net profit available to the agent at each step and each time.

Defining Profit-Maximizing Production

At each iteration step in the simulation, the agent chooses to do their max-marginal-gain action. With the route-finding methods discussed in the appendix applied to the arrival

cost matrices, we can now calculate how the agent identifies which grid-cell yields the highest foraging profit. Profit in the context of firewood collection is calculated as:

$$\pi_i = p_b c_i^g - p_w m_i$$

where p_b and p_w are prices for buying firewood and for wage work, respectively, c_i^g is the minimum value of either the amount available to be gathered or the max amount an individual can gather in one trip, and m_i is the minutes necessary to travel to the foraging location. In this version of the model, all non-travel costs are assumed to be identical and thus are included in the value of the firewood.

Fig. 2.20 illustrates the calculation of π_i for an example forest patch in Tanzania. The image on the left shows the abundance of firewood, measured in cubic meters. This value, as discussed in the literature section, is drawn from a survey of forestry literature specific to Tanzania. In this particular patch, we see a forest of high firewood quality in the middle and lower-middle, with a patch of medium quality forest on the middle and upper sections. The areas in red are low value shrub lands where some, but very little, firewood can be gathered. The village in this case is exactly in the center of the plot, just outside the forest. The middle map shows the arrival time from the center cell to every other cell using the nested search algorithms described above. The traversal time map is omitted in this example, but see the figure after this one for an example including traversal cost. Finally, the figure on the right uses the first two maps in the profit equation to calculate the available profit on each cell. The available profit map shows which action would be consistent with the max marginal gain behavior rule (gather in the black cell in this case). Note that this map does not define on its own how much profit the agent will eventually collect. The reason is that this net profit map is a snapshot in time, calculated at exactly the iteration step in which the agent acts, but other agents will get to act before

the agent in question gets to gather a second time. Competition from other agents may deplete the firewood in the cells with positive net profit, making them no longer available for the agent to gather.

Figure 2.20: Calculating Profit Available from Firewood Abundance and Travel Time

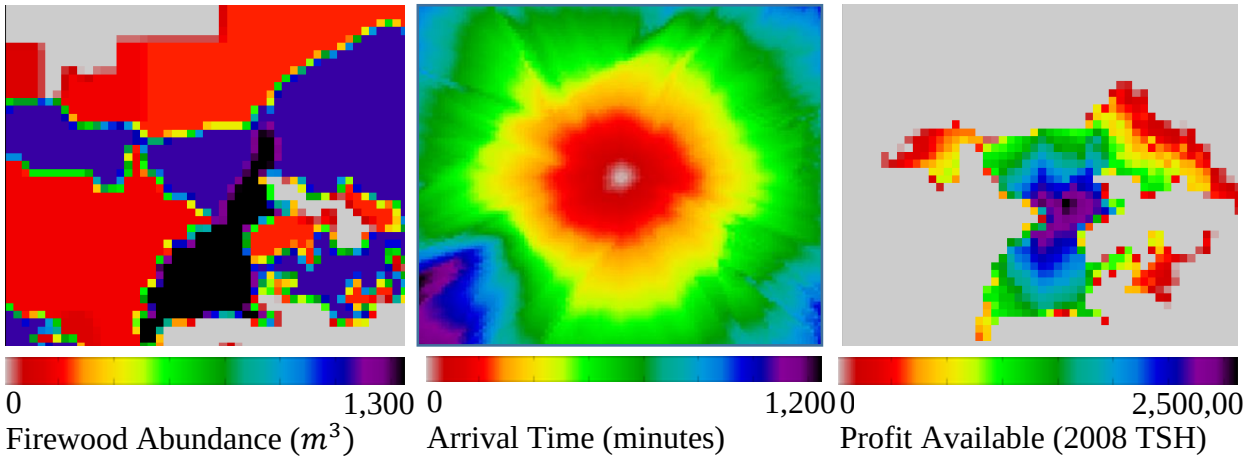


Fig. 2.20 represents the profit map specific to a single grid-cell (defined as the center of the map). As with the arrival cost maps, the net profit map is calculated for every grid-cell. Unlike the arrival cost maps, however, this array changes over time as agents make their choices on which cell they will forage. The simulation iterates through all iteration steps (in action order defined in Appendix 3), repeating until there all agents have reached an equilibrium (defined in the model description section). At each iteration step, the profit array is updating and recalculating itself to reflect the foraging behavior that happened on the previous step. Additionally, the profit array will update itself again as it moves to the next time period with, for instance, forest regrowth.

Results

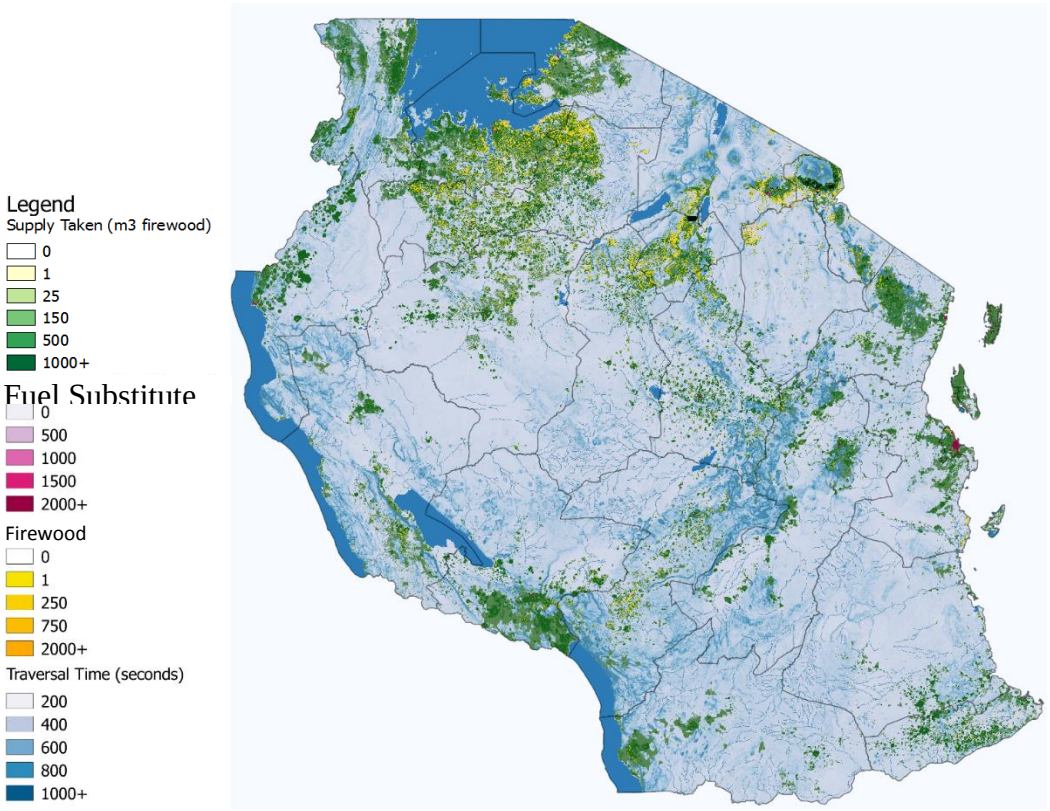
The results of the model described above fall into two general types: results that show where and how agents value firewood; and, results that show how the results change when different forest conservation policies are considered. In typical ecosystem service valuation analyses (at which this method is ultimately aimed), the second type of results require the majority of effort and yields the most interesting results. In this case, however, most of my results focus on the first type, describing in detail what behavior this model predicts on the current landscape. Once the method is established and verified, future analyses can focus on analysis of alternate policies, but for now, I focus primarily on a single outcome (the equilibrium that results from the landscape in existence in 2012, ignoring potential deforestation from non-foraging phenomena) and describing the arrangement of profit and abundance evident in this equilibrium. The other scenarios that are presented are not indicative of actual policies being considered, but instead are presented to show how the resulting equilibria change as our assumptions about the landscape or the agents change.

In this section I will first describe an overview figure that combines many of the important components of the overall result that I want to express. After that, I will work step-by-step through the components that enabled creation of the overview figure. I reverse the typical order of exposition because it is difficult to understand how individual pieces fit together without understanding the overall goal.

Fig. 2.21 provides an overview of the results for Tanzania. Four different variables are plotted with colors where intensity of the color indicates the value of the variable. Green represents supply of firewood taken by agents in cubic meters. Note that these pixels are the location of the firewood that was taken but not necessarily the location where the

agent lives. This is because the agent travels from their home to the foraging location. Also note that there is much more firewood present on the landscape than what is shown here because forests that experience no foraging are not plotted. Magenta represents the amount of fuel demand per agent grid-cell that was not met with foraged firewood. In these cases, the agent satisfies their fuel demand with the firewood substitute purchased with returns from labor. As expected, areas where the fuel substitute is purchased most is the areas furthest from forests and the densest in population (urban areas, such as Dar es Salaam on the east coast). Yellow represents the amount of fuel demand per agent that was met with firewood. In line with observed national statistic, this model predicts that approximately 97% of fuel demand is met with firewood. The areas that benefit the most from foraging are high-population areas near high-valued forests (such as the southern slopes of Mount Kilimanjaro). Finally, blue represents the traversal time in minutes to cross each grid-cell. Technically it is arrival cost, not traversal cost, that directly affects the production decision, but it is not physically possible to represent arrival costs on a single map because there are 10 million unique arrival cost maps (one for each pixel) for the map below. This map presents an overview of multiple aspects of the model simultaneously. This was achieved with color transparency by making visible only the variable that had the highest value for a pixel.

Figure 2.21: Results Summary Map



The overview map here can also be summarized by a single sentence: *yellow pixels gather firewood from green pixels taking into account blue pixel traversal costs while red pixels fail to meet their demand*. In the next section, I will break down this overview map and overview sentence into their constituent parts to give a more full description.

Scenario Definition

The results shown above are for the baseline scenario that I define. As discussed above, most of the contribution of this analysis is creating new methodological approaches to

describe foraging behavior with more detail and more realism. However, simple description of the behavior alone is not enough to answer questions about what policies are optimal. Thus, I also present eight alternative policy scenarios, described in table 5. Each of these scenarios describe a change in the landscape, and thus analyzing the difference in value collected from firewood in these scenarios versus in the baseline scenario identifies the value gained or lost by pursuing the policy that would lead to the scenario landscape.

Table 2.5: Eight Alternative Scenarios for Policy Analysis

Scenario Name	Description
Forests as Shrubs	Changes the land use type from forest to shrubland.
Half Shrubs Half Forest	Reduces the amount of firewood available by 50% on shrubland, savannah and forests.
Quarter Shrubs Half Forest	Reduces the amount of firewood available by 50% in forests and by 75% in shrubland and savannah.
Half Shrubs	Reduces shrubland and savannah firewood availability by 50%.
Quarter Shrubs	Reduces shrubland and savannah firewood availability by 75%.
No Shrubs	Eliminates all firewood from shrublands and savannah. Only forests have firewood.
Forests as Shrubs	Changes the land use type from forest to shrubland.
Half Shrubs Half Forest	Reduces the amount of firewood available by 50% on shrubland, savannah and forests.
Lower Wage	Reduced average wage by 13.1%. This reduces the opportunity cost of travel and thus makes foraging less expensive.
Higher Wage	Increased average wage by 13.1%.

I plan to incorporate newly released data on forest cover from Hansen et al. (2013) that uses higher resolution (30 meters) to create additional scenarios. The Hansen data set also

charts change in forest cover from 2000 to 2012, so it will also be possible with this data set to analyze how firewood collection was affected by specific deforestation actions. These data were only made available in 2014 and require an enormous amount of processing time to parse, thus they were excluded from the current analysis.

Calculating Supply

Supply of firewood is calculated by combing LULC data with literature values on firewood per hectare on different land types. As discussed in the data section, the LULC data I use comes from MODIS, a land sensing and classification project by NASA. MODIS identifies 16 land use categories, of which five are forest, two are shrublands and two are savannah. I apply the literature values discussed in the data section and present the per grid-cell abundance values for firewood in each of these 16 land categories.

Table 6 presents these values for the baseline scenario and the six policy analysis scenarios discussed above. The third column shows the values for baseline scenario that is used to generate most of the maps and results in this paper.

Because the results of this model are sensitive to how we parameterize firewood abundance, I use the six policy analysis scenarios to test how and where results change when we make different assumptions about how much firewood is available for each land type. Literature values for abundance typically focus on two broad categories, distinguishing between forests and non-forest vegetation (shrubland and savanna). The first scenario tests what happens when we do not assume this differentiation and instead apply the same value to forests as we do for shrubland. Note that the six scenarios presented here are just endpoints on a continuous spectrum of tests I ran. In this example, that means that I tested many of the values between forests having the same abundance as shrublands and having the higher, forest specific value.

The next scenarios test what happens when shrubs and forests have less firewood available than in the baseline. The second scenario tests what happens when both are at 50% of original values, while the third scenario tests with 50% forests but 25% shrubs. The fourth through sixth scenarios show what happens when forests are kept at their original values but shrublands are reduced to 50%, 25% and 0% respectively. Table 2.7 below presents the assumptions used in each of these scenarios.

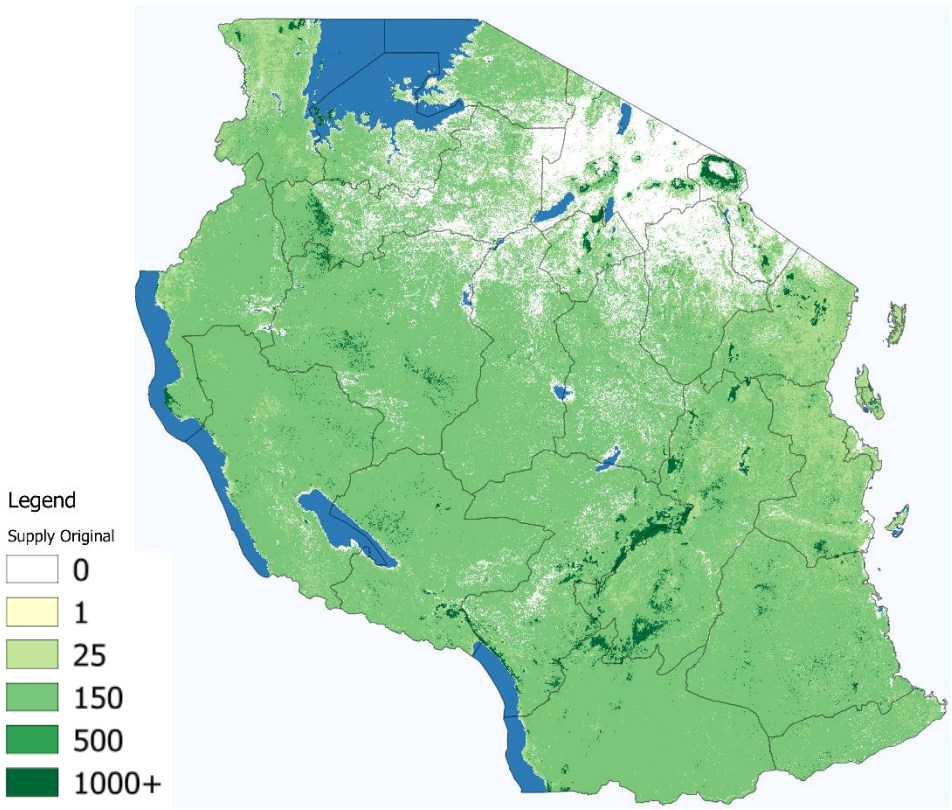
Table 2.6: Firewood Abundance per grid-cell in m3 used in different scenarios from literature values

LULC Name	Baseline Scenario	Forests as Shrubs	Half Shrubs Half Forest	Quarter Shrubs Half Forest	Half Shrubs	Quarter Shrubs	No Shrubs
Evergreen Needleleaf forest	1269.39	106.896	634.695	634.695	1269.39	1269.39	1269.39
Evergreen Broadleaf forest	1336.2	106.896	668.1	668.1	1336.2	1336.2	1336.2
Deciduous Needleleaf forest	1202.58	106.896	601.29	601.29	1202.58	1202.58	1202.58
Deciduous Broadleaf forest	1149.132	106.896	574.566	574.566	1149.132	1149.132	1149.132
Mixed forest	1256.028	106.896	628.014	628.014	1256.028	1256.028	1256.028
Closed shrublands	106.896	106.896	53.448	26.724	53.448	26.724	0
Open shrublands	106.896	106.896	53.448	26.724	40.086	20.043	0
Woody savannas	106.896	106.896	53.448	26.724	46.767	22.7154	0
Savannas	106.896	106.896	53.448	26.724	33.405	16.0344	0
Grasslands	0	0	0	0	0	0	0
Permanent wetlands	0	0	0	0	0	0	0
Croplands	0	0	0	0	0	0	0
Urban and built-up	0	0	0	0	0	0	0
Cropland/Natural vegetation mosaic	53.448	53.448	26.724	13.362	6.681	2.6724	0
Snow and ice	0	0	0	0	0	0	0
Barren or sparsely vegetated	0	0	0	0	0	0	0

Although the values used in the baseline scenario for firewood abundance are well supported by the literature, it is important to note that this analysis could easily be re-run with new values for abundance. For example, many locations have done field surveys specific to the land in question, and thus values like that can easily be substituted in to the model presented here.

Fig. 2.22 below shows the supply of firewood available before the first iteration of gathering begins. The majority of land is in shrublands and savannah, though there exist areas of dense forest around Mount Kilimanjaro and in the natural reserve land in the south east and North West. The values presented here for abundance are biophysical observations only. The agent considers the value of the forests, taking into account both the value of fuel, their local wage and the travel related collection costs.

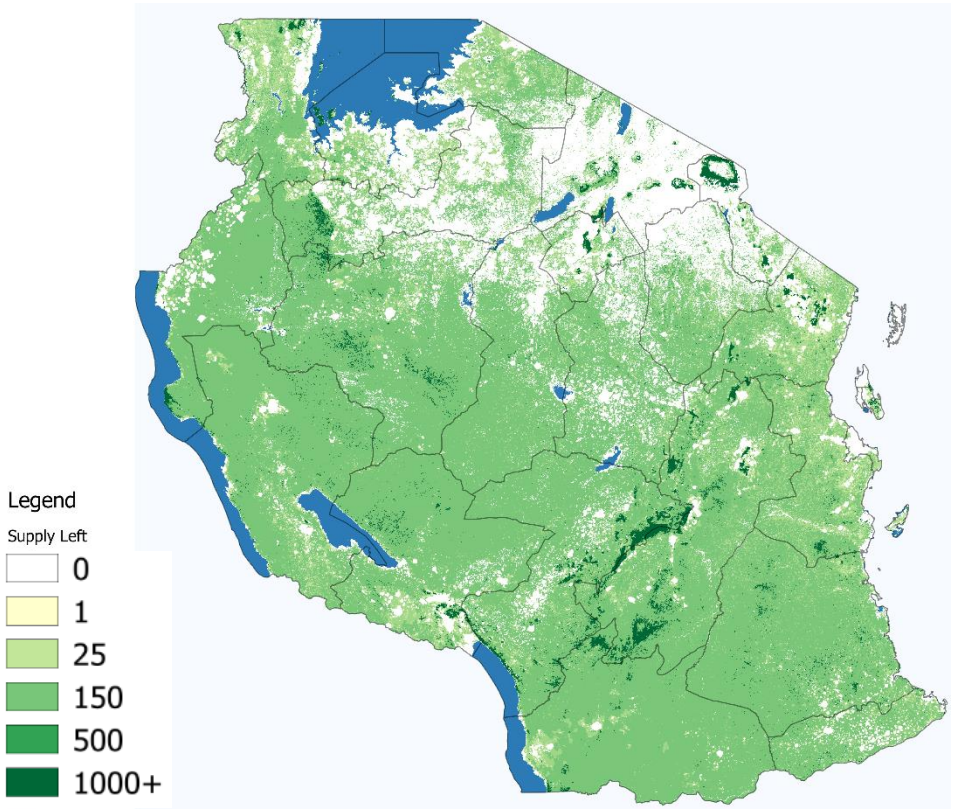
Figure 2.22: Original Supply (m^3 firewood per grid-cell)



After the simulation has run, forests near agents who find it profitable to forage will be depleted. However, due to the remoteness of many natural areas, there exist many

locations that go un-foraged. This is shown in Fig. 2.23, which plots the amount of firewood that is still present on the landscape after the simulation has finished all iterations in the first time period.

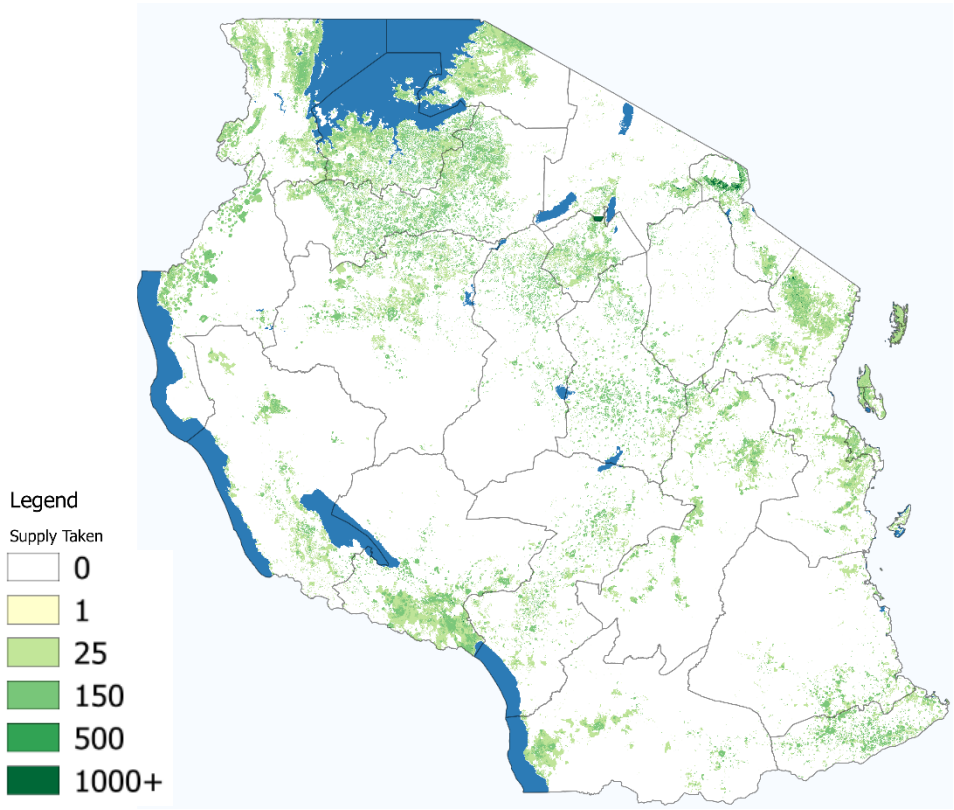
Figure 2.23: Supply Left (m^3 firewood per grid-cell)



It is visually difficult to see the differences between Figs. 22 and 23 above, so Fig. 2.24 presents the difference between the two, which defines firewood supply taken. This map is a key part of this model's results. The areas shown as foraged are determined by where

demand is the greatest, where supply is the greatest and where costs make collection relatively more valuable than the fuel substitute. Some key locations shown here are the high degrees of foraging on the southern slopes of Mount Kilimanjaro, in the shrublands to the south of Lake Victoria and the arterial transportation links coming out of Dar es Salaam.

Figure 2.24: Supply Taken (m^3 firewood per grid-cell) (difference between Supply Original and Supply Left)

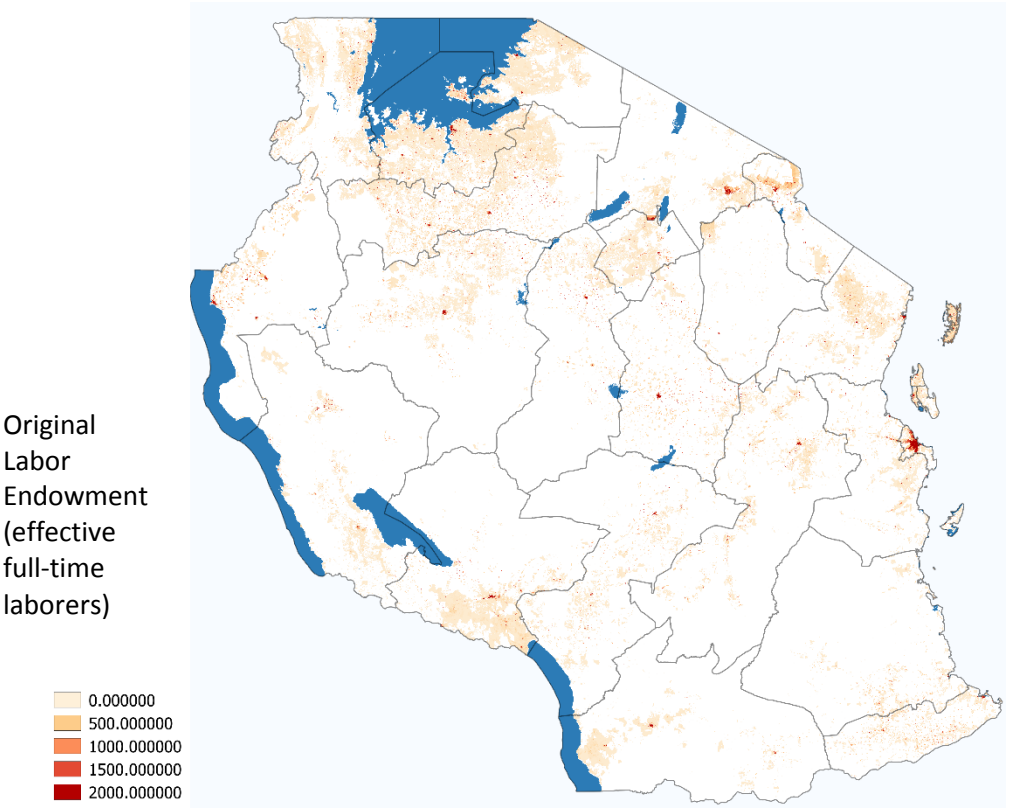


Calculating Consumption and Demand

The images above only present half of the foraging action. For every cubic meter of firewood gathered, there is a corresponding amount of firewood that arrives at an agent's household location and is consumed. This value can be described multiple ways, including by noting the value of firewood that accrues to each agent after travel costs are accounted for, by noting the amount of labor spent by each agent on foraging, or by observing the raw amount of firewood that arrives at a given location. The last measure, while very easy to define, is very difficult to interpret because demand is determined endogenously by the agent's household utility maximization decisions. Even if we sum firewood and the firewood substitute to arrive at aggregate fuel needs, this also is not taken as exogenous because the marginal rate of substitution between heating and leisure does not have to be fixed (and probably is not fixed where there are wage differentials). Thus, care must be taken when stating how much firewood is "demanded" by different agents.

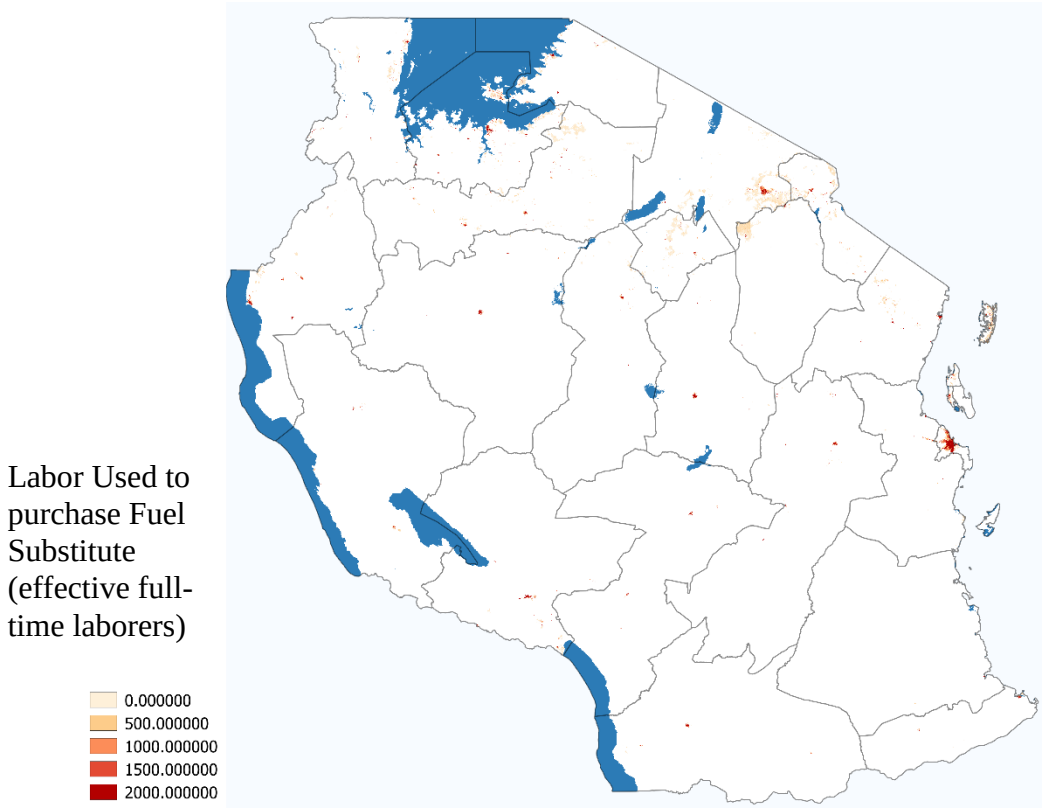
The labor endowment of each agent, however, is taken to be exogenous in this model. Each agent in this model represents the full population of the 450m by 450m grid-cell. Thus, I define the labor endowment of an agent as the population of the grid-cell multiplied by the labor participation rate. Future work can easily extend this analysis to match location-specific labor participation rates, but this was not available for Tanzania. Fig. 2.25 presents the total labor endowment available at the beginning of the simulation.

Figure 2.25: Original Labor Endowment



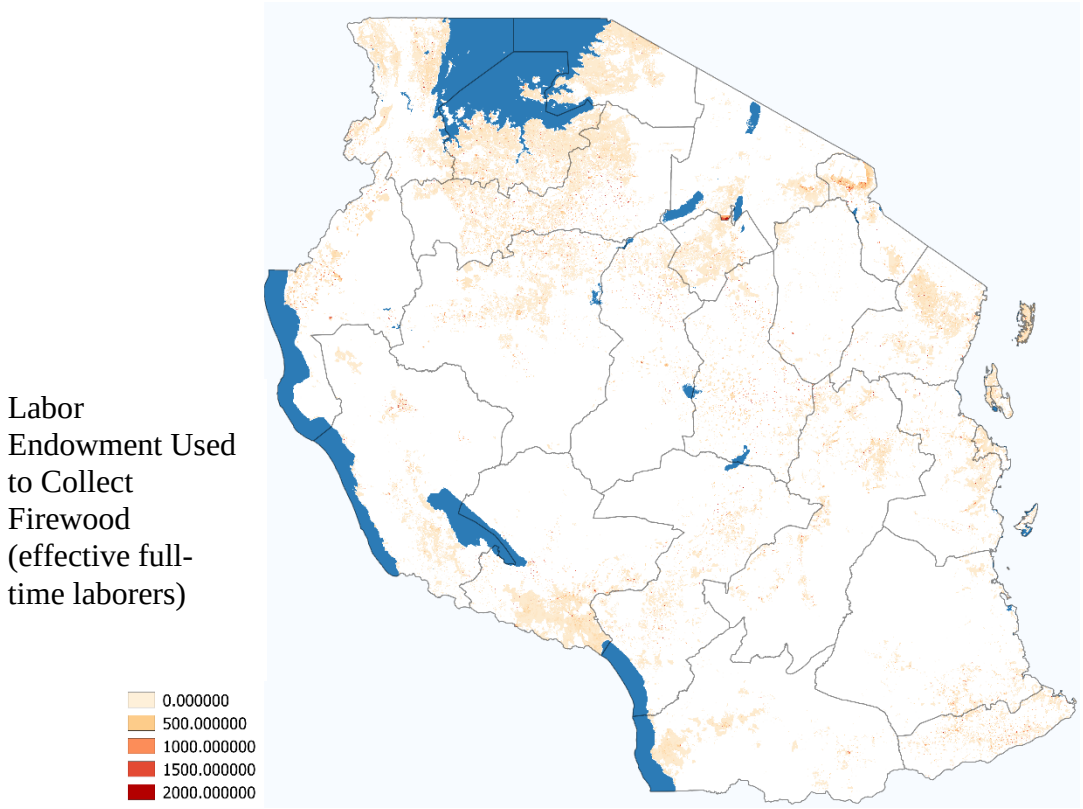
The above map was exogenous but endogenously determines labor allocation choices. Fig. 2.26 presents the amount of labor on each grid-cell that was used to purchase the fuel substitute (usually kerosene). A clear result of this model is that locations with high population density purchase the majority of the fuel substitute. This is because the agents on the edges of the urban centers deplete all nearby firewood and thus everyone on the interior of the dense population area has no profitable firewood available to forage.

Figure 2.26: Labor Used to Purchase Fuel Substitute



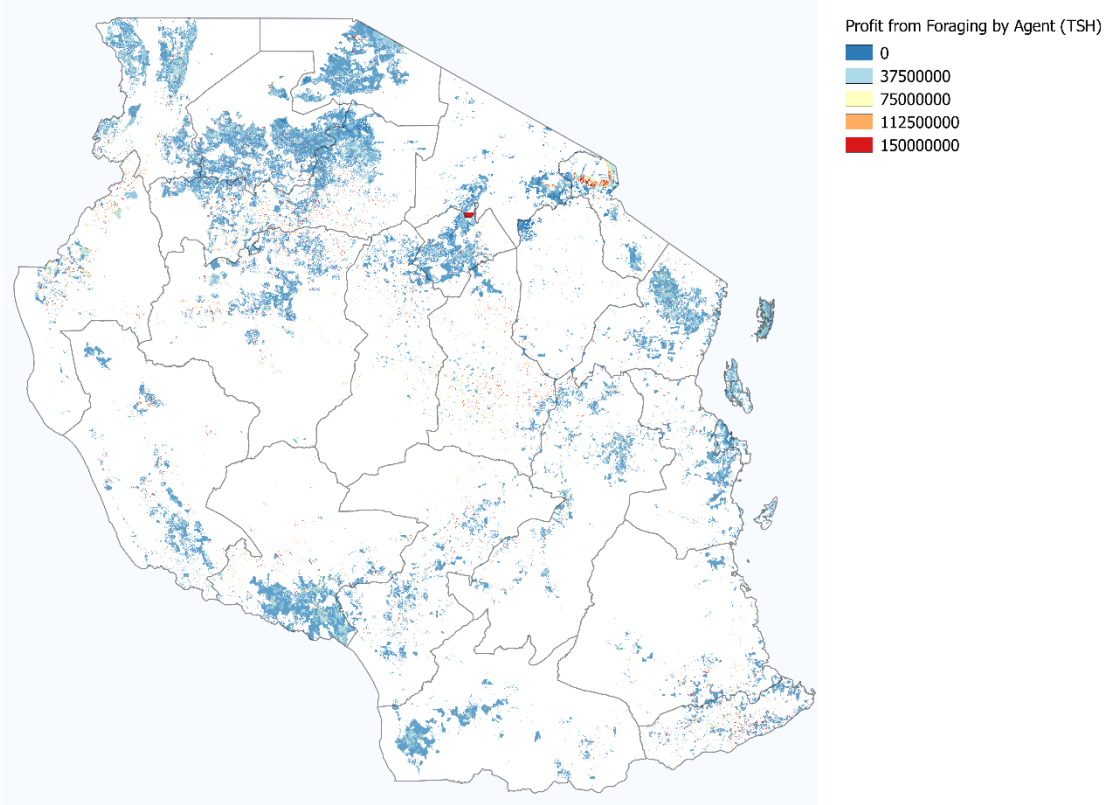
The majority of labor used to satisfy heating demand was dedicated to collecting firewood, as shown in Fig. 2.27. The values shown in this figure reflect agents choosing optimal travel routes and making the utility maximizing choice among foraging, laboring and leisure.

Figure 2.27: Labor used to Collect Firewood



Accounting for labor usage and supply of firewood taken, we can now calculate the profit (or if desired, utility) that each agent generates. Fig. 2.28 plots the value in 2008 TSH that each agent collects from foraging.

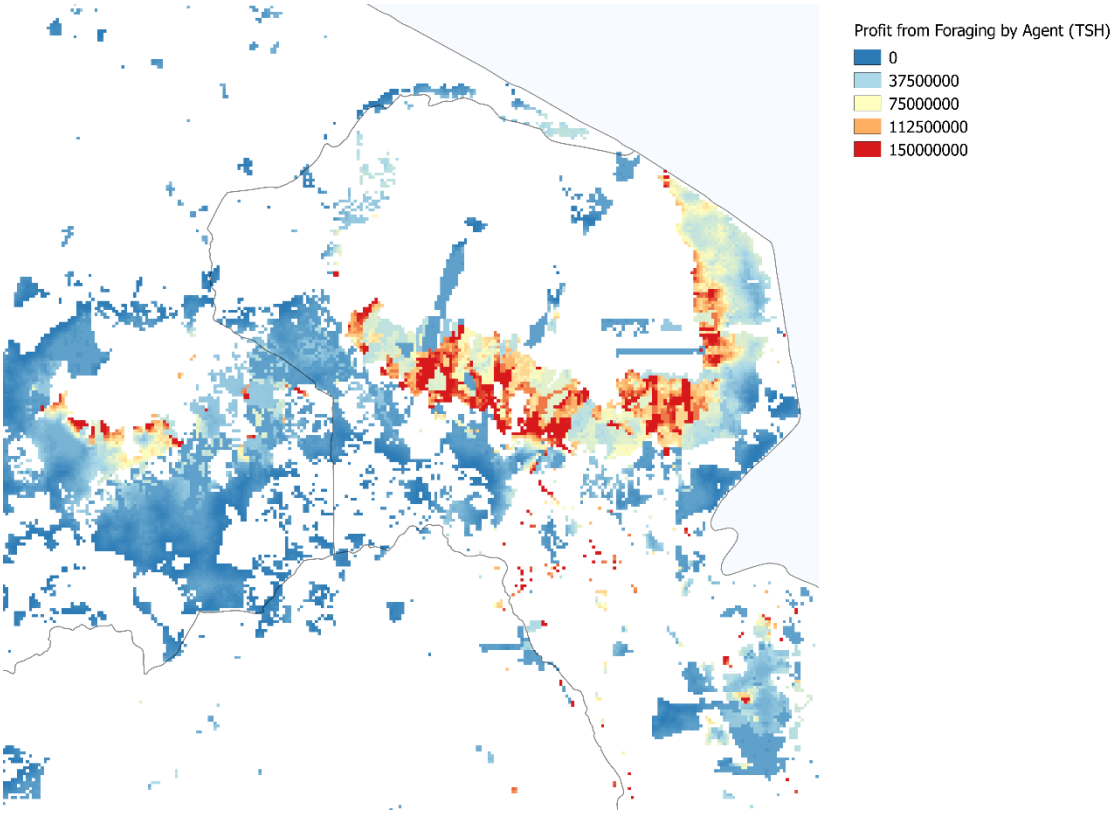
Figure 2.28: Profit Earned by Each Agent (national)



It is difficult to see the results when viewed at a national level so I present the area around Mt. Kilimanjaro in Fig. 2.29. Here we can see the complexity of behavior that arises as a result of the competition. In the western area of this subplot, for instance, we can observe agents right on the border of a high quality forest who obtain a very positive profit from foraging with a decreasing gradient of profit obtained as we move south west away from the forest boundary. The gradient descent arises because the agents farther to the south have to travel ever further to get the firewood. Note that the agents plotted with yellow or blue (lower values of profit) may well gather the same (or more) firewood than

the agents right on the border of the forest due to tradeoffs implied by their household optimization decisions.

Figure 2.29: Profit Earned by Agent (near Mt. Kilimanjaro)



The results that the previous two figures show can also be considered in aggregate. Aggregation by village boundary can be done if policy makers prefer to know how political units are impacted on average rather than how specific grid-cell agents are affected. In Table 7 below I present a national level aggregation of cubic meters of

firewood foraged and profit generated from firewood collection, along with a comparison of change in profit compared to the baseline scenario.

Table 2.7: Demand met and Profit Earned in 9 Scenarios

Scenario	Demand Met	Profit Earned	Profit Deviation from Baseline
Baseline	6.67E+07	1.48E+13	n/a
Half Forest, Half Shrubs	5.54E+07	1.16E+13	-3.24E+12
Half Forest, Quarter Shrubs	4.14E+07	8.06E+12	-6.76E+12
Reduced Shrubs	4.63E+07	9.11E+12	-5.71E+12
Minimized Shrubs	3.28E+07	5.90E+12	-8.93E+12
No Shrubs	1.07E+07	1.55E+12	-1.33E+13
No Forests	6.37E+07	1.43E+13	-5.74E+11
Lower Wage	7.38E+07	1.51E+13	2.99E+11
Higher Wage	6.56E+07	1.45E+13	-3.46E+11

Figure 30 charts the values found for each of these scenarios, while Fig. 2.31 shows the deviation from the baseline.

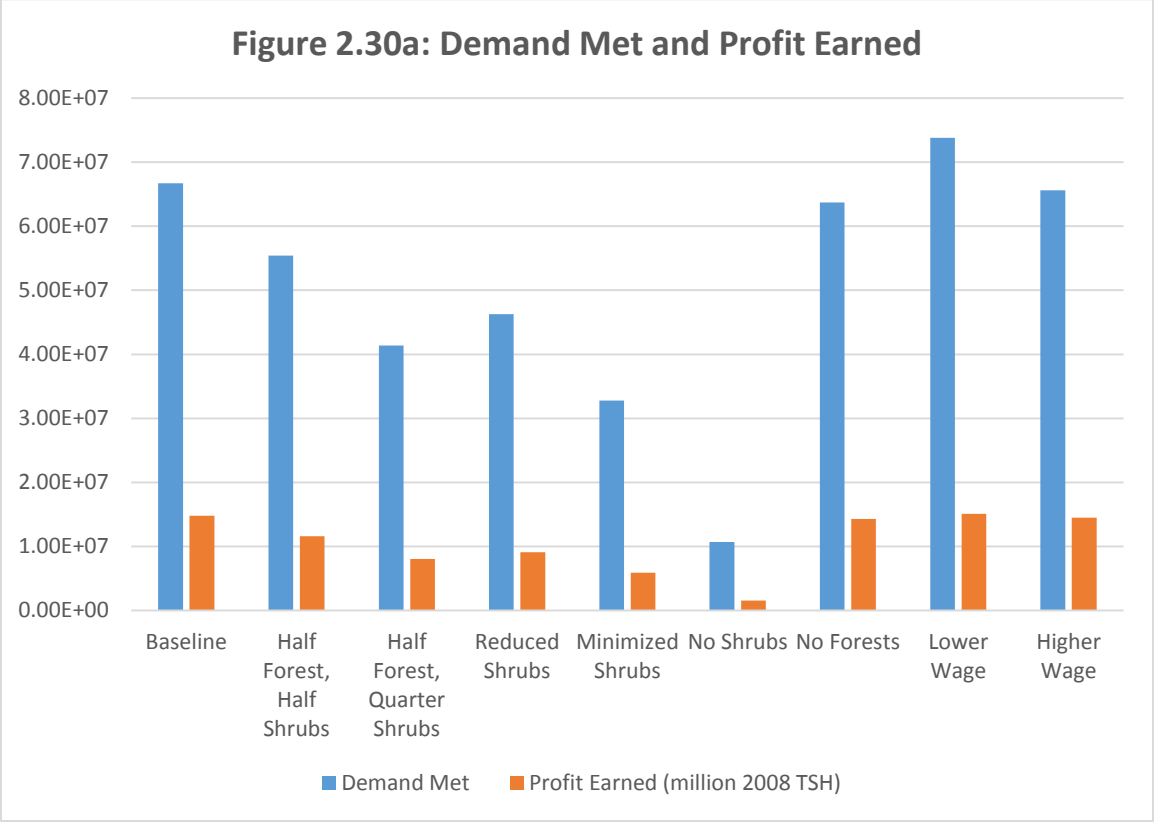


Figure 2.30a: Demand Met and Profit Earned
8.00E+07
7.00E+07
6.00E+07
5.00E+07
4.00E+07
3.00E+07
2.00E+07
1.00E+07
0.00E+00
Baseline
Half Forest, Half Shrubs
Half Forest, Quarter Shrubs
Reduced Shrubs
Minimized Shrubs
No Shrubs
No Forests
Lower Wage
Higher Wage
Demand Met
Profit Earned (million 2008 TSH)

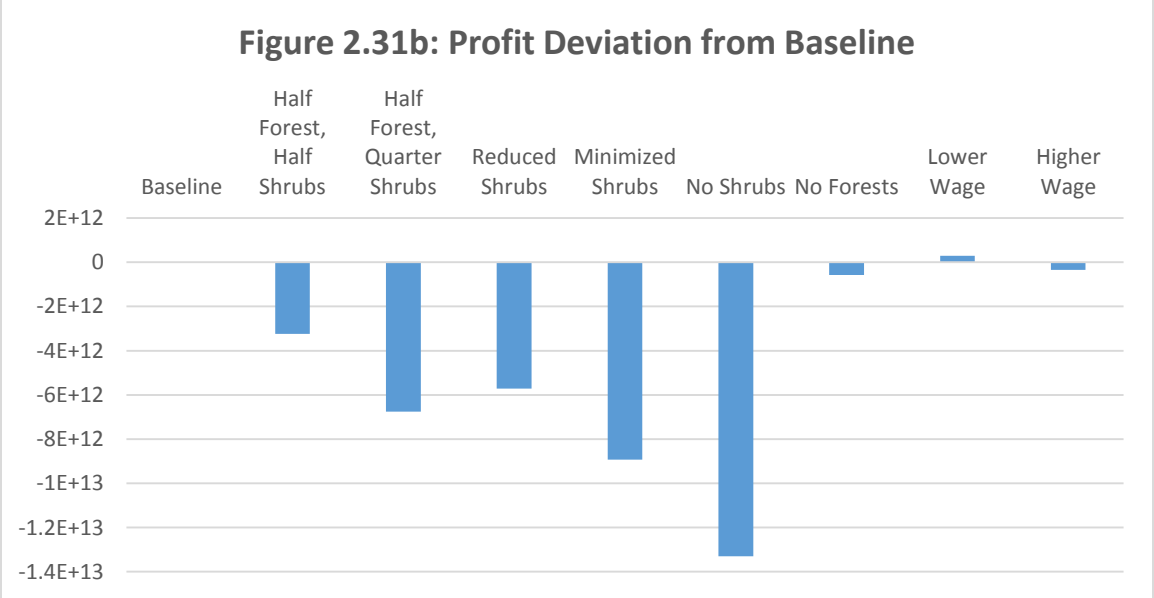


Figure 2.31b: Profit Deviation from Baseline
Baseline
Half Forest, Half Shrubs
Half Forest, Quarter Shrubs
Reduced Shrubs
Minimized Shrubs
No Shrubs
No Forests
Lower Wage
Higher Wage
2E+12
0
-2E+12
-4E+12
-6E+12
-8E+12
-1E+13
-1.2E+13
-1.4E+13

The results here are presented merely to be suggestive of what future policy application may look like. The reason why shrub land and savannah has such a larger impact than forests is simply because much less land covered by forests in Tanzania. Analyzing an area where forests represent a greater percentage of the land cover, however, results in values for forests that are much higher than for shrub lands. In reality, the scenarios presented here are too broadly defined to be of specific policy use, though they could be used to identify the marginal value of forest quality in different locations. Instead, the value of the results shown here must be seen in the context of providing spatially explicit, falsifiable statements about who gathers what from where based on microeconomic theory but aggregated to a national level without losing spatial heterogeneity or inter-agent competition. There exist many additional outputs of the model that can be considered results, such as predicting how many hours each agent will devote to foraging. The next section will present several of these additional results while discussing how the foraging model can be validated.

Validation

Validating an agent based model is critically important because the results are sensitive to assumptions on a variety of variables. This section summarizes the validation results and potential future methods for validation. The full validation exercise can be found in Appendix 3.

Given that demand for firewood is not given as an input parameter but instead is calculated endogenously as a function of the landscape and the agents' behavior rules, we can start validating the model by comparing predicted firewood demand with literature estimates on firewood demand. Values from existing studies suggest that demand is between 25.8 and 55.5 million cubic meters. The baseline scenario predicts 66.7 million

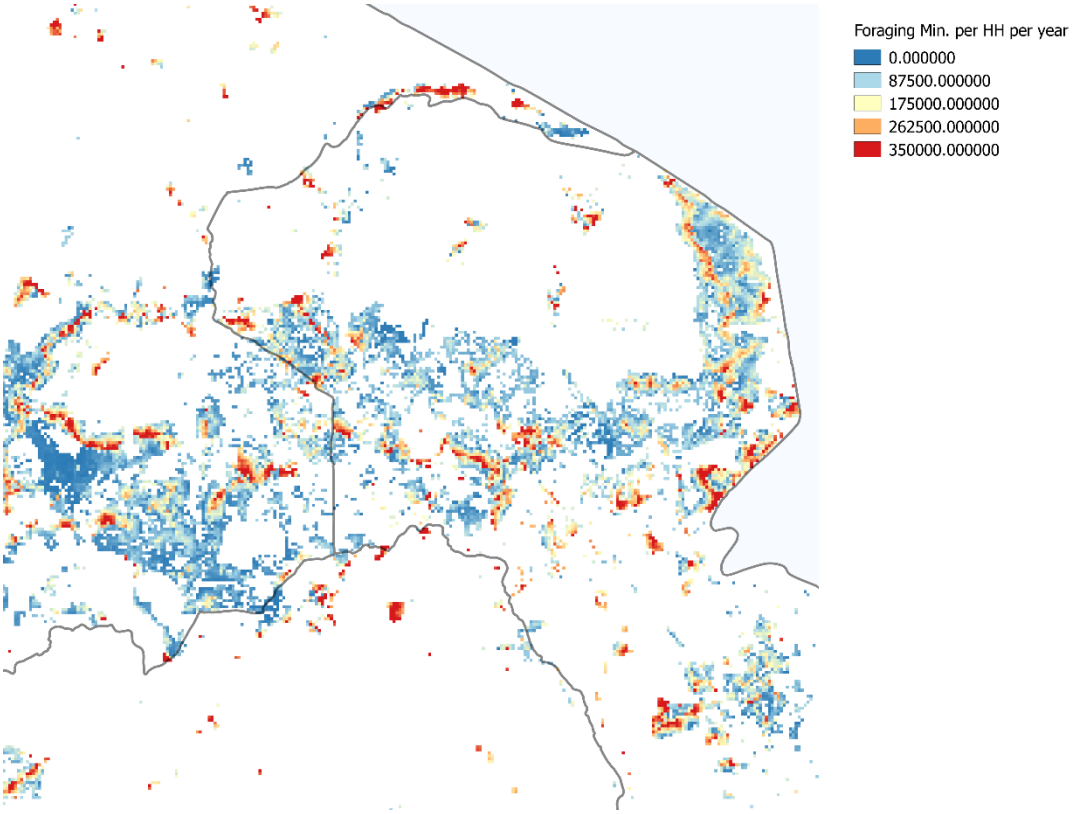
cubic meters of firewood will be gathered. This is above, but within the order of magnitude reported in the literature. The overestimated value here is very likely a result of wrong assumptions on mean harvestable volume per hectare. The range of values I drew from the literature (14 – 117 $m^3 ha^{-1} yr^{-1}$) is based on a relatively small sample of land which was primarily covered in woodlands and newer estimates suggest the value should be closer to the minimum value of this distribution than to the mean (which is what I used). Additionally, given that the majority of Tanzania is scrublands and savannah, which may have lower densities of firewood, it is not surprising that using the woodlands estimates would result in overstating the national effect. For this reason, the sensitivity analyses (presented in Table 7 in the results section) may be especially important because they show cases in which land is assumed to have less firewood abundance. For instance, the "half-forest, quarter-shrubs" scenario predicts approximately 41.4 million cubic meters, which is comfortably within the range of estimated consumption Figures.

Future work will be able to devote more time to calibration that focuses on original work in the project area. For instance, I have already secured research-time with partners in Cambodia to make a new data-set of field observations on firewood abundance and gathering behavior. This study will be able to identify individual foragers' home location and foraging location. Pairs of home and foraging location pairs will allow for more precise validation and testing.

Although I can show that the model validates on an aggregate level, it remains difficult to validate at the grid-cell level due to the precise locations of village data from the NPS are obscured by 10 km for privacy reasons. If these data were not obscured, it would be straightforward to validate on the grid-cell level by running the model and making predictions about, for instance, how many hours each agent will forage for firewood, and

then compare these predictions to household survey data on firewood collection. Figure 31 shows the travel spent for agents near Mount Kilimanjaro. Note that there is extreme variation at small scales. This is an expected outcome of the model given that firewood is not transported very far and that it has high transportation costs.

Figure 2.31: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)



Given the hyper-local production decisions made with firewood gathering and the high degree of spatial variability in the figure above, it is not surprising a comparison with the NPS failed to validate.

Conclusion

This chapter has presented an approach for applying agent-based simulation to firewood collection in Tanzania. The main contribution of this approach is that it identifies an approach to solving two difficulties that have hindered understand of non-timber forest product foraging: geospatial heterogeneity and inter-agent competition for foraging goods. There is much demand among practitioners of ecosystem service valuation to include non-timber forest products in their valuation exercises but these problems have prevented widespread application of non-timber forest product valuation to conservation planning. The modeling details discussed in this chapter have primarily been theoretical; however, the software used to instantiate the method has already been slated for inclusion in InVEST, the software application of The Natural Capital Project. Initial applications of InVEST version of this tool include analyzing bush-meat foraging in Cambodia (in partnership with the World Wildlife Fund of Cambodia) and extension to fruit and grass gathering in Tanzania. By providing a detailed statement of how to implement agent-based simulation to understand subsistence markets like this, I hope to have broken through the barrier that has prevented the valuation of non-timber products in forests, and thus, provided a method for more accurately understanding values of forests that are very often missed.

Chapter 3: Reciprocity in Commons Dilemmas

Publishing note: This chapter was written concurrent to the creation of the published paper:

> *Runge, C. F., & Johnson, J. A. (2014). Are we in this together? Risk bearing and collective action. Journal of Natural Resources Policy Research, 6(1), 71-76.*

The published version is similar to this chapter, but here I extend discussion of theoretical aspects of the model and apply it to agent-based simulation

Introduction

This chapter explores the theoretical aspects of collective action in commons dilemmas when individuals' utility is linked via reciprocity. We present a model that formalizes how reciprocity affects contribution to public goods in a way that matches evidence from behavioral economics that agents will not always free-ride in commons dilemmas. Our model is broad enough to include strong free-riding, socially optimal behavior and all intermediate choices in a single modeling framework.

Our model shows that inclusion of reciprocity results in partial cooperation where the level depends on how intensely individuals care about reciprocity. It bears emphasis that this model relies on making interpersonal comparisons of utility. Although inclusion of interpersonal comparisons of utility is not typically employed in modern economics, our conception of this is similar to what Arrow (1977) termed "extended sympathy," and is consistent with current neurobiological research (eg., Meltzoff and Decety 2010).

Moreover, we believe this use of utility echoes the original insights of Bentham, Hume and Adam Smith.

The theoretical implications of reciprocity and fairness have been explored before in the literature (Sugden 1985, Runge 1981, 1984; Fehr and Schmidt 1999; Rabin 1993).[4] In these models, the formulation of reciprocity is exogenous or chosen prior to the players' strategy choices. Other approaches, (Marchiori 2010, Dufwenberg and Kirchsteiger 2003) have expanded or endogenized the concern for reciprocity and shown that equilibrium levels of cooperation are above the strong free-riding prediction. In the next section, we present our reciprocal utility model and discuss how the model challenges traditional welfare economics and conceptions of economic value. Throughout the literature on reciprocity and fairness, it is clear that institutions and norms are critical in determining how reciprocity affects strategy choices. Our approach explicitly links reciprocity to the norms, expectations and institutional arrangements that are involved in collective action.[5] More generally, our model provides a method for disentangling the complex strategies that players pursue based on these norms and institutions. To make this possible, we define a behavioral response function and incorporate it in a Nash equilibrium utility model.

After presenting the general form of our model, we then apply it to an example commons dilemma: global climate change. For this example, we discuss how the response function can be constructed to match observed phenomena and institutions. In the case of climate

[4] Some relevant models use the terms fairness while others use reciprocity. We believe that reciprocity best describes the type of behavior in our model, though often the reciprocal response is that which is considered fair. In our usage, fairness is the more general term, relevant to a broad set of moral axioms, while reciprocity is a specific form of fairness.

[5] This framework is applicable to many collective action dilemmas. Our model is generally applicable to situations where individuals gain utility from observing other individuals' levels of utility.

change, the institution we model is binding emissions reduction targets in which the countries' reciprocal behavior depends on if other countries do or do not meet their emissions abatement goals. We show that in this context, a type of equilibrium exist in which some agents meet their goal while other agents pursue a token level of contribution (still higher than strong free-riding levels). This outcome matches the current state of climate change negotiations in which some countries are pursuing aggressive and binding policies while other countries are pursuing weaker, voluntary mitigation approaches. Finally, the last section of this chapter includes a discussion of how this model may be better estimated with new data and extended with future research.

The Commons Reciprocity Utility Model

In a survey of the role of interpersonal comparisons of utility in economic theory, Hammond (1991) notes the strong (and confining) hold that disallowing interpersonal comparisons of utility has had in welfare economics. To avoid these confines, we use a definition of reciprocity in which players do indeed make interpersonal comparisons of utility. Agents in our model view reciprocally beneficent players as friendly cooperators and receive positive utility from seeing them do well (similar to Valavanis 1958). Conversely, they see reciprocally harmful players as rival competitors and receive utility from seeing their rival do poorly. The idea of friendly reciprocators and unfriendly punishers matches with psychological theory of reciprocity (Cialdini et al. 1975, Regan 1971) as well as behavioral game theory evidence (Falk and Fischbacher 2006, Fehr and Gächter 2000).

Much of economic theory avoids making interpersonal utility comparisons due to the difficulty in using cardinal utility in place of ordinal utility. Nearly as soon as these difficulties were established, however, others noted that such comparisons are essential

for understanding bargaining, social welfare and collective decision making (Shapley 1967, for example). More recently, Ng (2003) argues that without making comparisons of cardinal utilities, welfare economics is too narrow and ought to be expanded, while Elster and Roemer (1993) point out the incompatibility of using interpersonal comparisons in many Nash axioms and the inability to define common strategy choices, such as "split the difference." Parallel to these arguments have been the fast growth of the behavioral economics literature. The complex human behavior that this literature has uncovered seems incongruent with the assumption that individuals only consider their own direct utility.

Fundamentally, the problems herein are focused on how we assign value. Traditional utility approaches may be used to account for caring about others' welfare by including the value an individual gets from their other-regarding behavior directly in their utility function.[6] However, this approach presents two problems: first, many findings in welfare economics require separable utility functions among individuals, which often is violated in models that include others' utility as a direct gain; second, this modeling approach does not easily incorporate complex reciprocal behavior, such as contingent strategies based on inequity aversion (similar in conception to Fehr and Schmidt, 1999); third, it is not well understood how preferences are formed on indirect consumption of a good.

In a broader sense, the description of caring about others as an individual consumption choice may be part of the reason that non-economists quickly bristle at the implications of economic rationality (even though these may often be based on misunderstanding).

---

[6] See Dellarocas, C., Fan, M., & Wood, C. (2004). Self-interest, reciprocity, and participation in online reputation systems; or see Ribar, D. C., & Wilhelm, M. O. (2002). Altruistic and Joy- of- Giving Motivations in Charitable Behavior. Journal of Political Economy, 110(2), 425-457. For examples.

Indeed, the implications derived from defining care for others as purely selfish preference leads some to describe "*homo-economicus* [as] a sociopath," (Stout 2006, Henrich et al. 2001, and expanded in Gintis 2000) or are autistic. While many of these criticisms may be unfair or miss the point, they do highlight the perception outside of economics that having too narrow of a definition of what humans value may result in strange or distorted models.

Model Definition

We present a model in this section that includes reciprocity between agents. Suppose that each player in a commons dilemma must choose between two options: allocating income to a private consumption good or a public investment good. In the context of climate change, for example, investing in the public good is defined as pursuing abatement activity that lowers the amount of GHG that a player emits. As another example in the context of harvesting non-timber forest products, one could define private consumption is defined as consuming a product taken from a public forest while investing in the public good takes the form of conservation management strategies or limiting extraction. Utility in our models is assumed to be increasing in consumption but can be decreased by actions of other agents' private consumption. Unlike traditional economic models, we do not assume each player's utility is separable from the other players' choices insofar as the commons is affected by decisions of all players, and thus utility for each agent depends on each other agent.

To formalize this concept, we must first define the archetypical commons dilemma utility framework on which we will build our extensions. Define player i's payoff as $u_i^D\left(c_i, a_i, \sum_{j\neq i}^N a_j\right)$, where c_i is consumption, a_i is the i-th player's contribution to the public good and $\sum_{j\neq i}^N a_j$ is the sum of all other players' contributions. We use the

superscript D to indicate that this is the direct utility an individual receives. Direct utility, in our definition, is identical to that used in traditional economics[7] and assumes a rational preference set and utility that is separable among players. Given these payoff functions, together with prices of consumption and investment of the public good, p_c and p_a, and wealth w_i, the i-th individual solves:

$$\max_{c_i,\, a_i} u_i^D\left(c_i\,, a_i, \sum_{j\neq i}^{N} a_j\right) \qquad (1)$$

$$Subject\ to\ p_c c_i + \ p_a a_i \ \leq w_i$$

This is a traditional commons problem, and the solution will result in under-contribution to the public good due to strong free-riding. If the price of consumption is normalized to unity and there are two identical players, i and j, we can re-express consumption in terms of wealth and contribution to the public good, which reduces the problem to one choice variable, a_i. This type of model has been well-studied (Varian 2004, Runge 1984, Hirshleifer 1983, Marwell and Ames 1979), and results in under-provision of effort towards the public good because each player is only considering their own marginal benefit from abatement and the resulting allocation in equilibrium has less abatement than the Samuelson condition, $\sum_{i=1}^{N} \frac{\partial u_i^D}{\partial a_i} = p_a$. In the literature, many institutions or policies have

[7] $u^D(\cdot)$ is a continuous utility function representing a locally non-satiated, rational preference relation $\succsim$ that is homogeneous of degree zero in prices and wealth, strictly increasing in prices, quasi-convex in prices, and continuous. Given the preference relation $\succsim$, $u^D(\cdot)$ is such that for all goods $x, y \in X$, $u^D(x) \geq u^D(y)$ if and only if $x \succsim y$.

been proposed to address this form of under-provision, such as the Vickrey-Clarke-Groves mechanism (Clarke 1971, Groves 1973), although others have pointed to shortcomings in these approaches (for example, Rothkopf 2007).

We extend the model above by augmenting direct utility, $u^D(\cdot)$, with an interpersonal utility component, $u^I(\cdot)$. The interpersonal utility function describes the gain or loss of utility that player i experiences from observing a change in another player's utility (player j henceforth). The basic assumption underlying the indirect component of utility is that humans enjoy seeing benefit come to those who are friends and may enjoy seeing harm come to those who are foes. Cikara et. al (2011), along with many others in the psychology literature, explore the evidence for this, noting that "The failures of an in-group member are painful, whereas those of a rival out-group member may give pleasure—a feeling that may motivate harming rivals." Our interpersonal utility component seeks to identify and correctly parameterize this type of reciprocity.

To account for the variety of ways in which individuals react to interpersonal exchanges, we define a response function $R_{ij}(\cdot)$ that represents how player j's actions (a_j) determine if player i will regard player j as a friend or as a foe. Thus, we allow the relationship alignment (friend/foe) to be determined by the exchanges that take place between pairs of players. An important component in determining this alignment is the extent to which players believe other players have acted reasonably, justly, or fairly (as in Rabin 1993). In this case, a higher value of $R_{ij}(\cdot)$ implies that player i believes player j has acted fairly and thus will want to reward player j with reciprocally beneficent behavior. Similarly, a lower value for $R_{ij}(\cdot)$ implies that player i believes player j has not acted fairly and so will want to punish player j. Rewarding beneficent behavior and punishing malevolent behavior is what we refer to as reciprocity. Our model also allows players to have different intensities of reciprocal response towards different individuals, defined by $\beta_{ij} \in$

(0,1), based on how close the pair of players are or how frequently they interact. Thus, we define interpersonal utility for player i as:

$$u_i^I(\cdot) = \beta_{ij}\, R_{ij}(a_j)\, u_j^D(a_i, a_j)$$

In this formulation, if player i views player j as a friend ($R_{ij}(a_j) > 0$), then player i will receive an increase in interpersonal utility when player j receives an increase in direct utility.

The ways in which interpersonal behavior is assessed are not simple (and thus cannot easily be included into the traditional economic utility framework embodied in U_i^D). Rather, reciprocal behavior depends on a complex set of norms and institutions. By separating the interpersonal component of utility from traditional utility and by endogenizing the relationship orientation between players, our model provides a framework that more easily incorporates norms and institutions that often violate assumptions required for traditional utility functions to work.

We combine our interpersonal utility expression with the direct utility expression to identify the total utility a player receives as the sum of these components. We label this function as reciprocal utility:

$$u_i^R(a_i, a_j) = \underbrace{u_i^D(a_i, a_j)}_{\substack{Player\ i's \\ direct\ utility}} + \underbrace{\beta_{ij}\, R_{ij}(a_j)\, u_j^D(a_i, a_j)}_{\substack{Player\ i's \\ interpersonal\ \ utility,\ u_i^I}} \qquad (3)$$

In the context of the climate dilemma, the term on left of equation three represents the direct utility player i receives from consumption after damages from climate change have been subtracted. These damages are a function of total abatement among all players. The

term on the right, conversely, represents the additional utility that player i receives from reciprocal assessment of player j's behavior. Assuming that player i views abatement action positively, she will receive more utility when high-abating players also receive utility. This term represents the player's utility from their perception of reciprocity, but not the utility they gain from the other players' actual investment in the public good (this portion of utility is expressed in the direct utility term, as a part of the direct enjoyment of the public good). In this framework, individuals maximize their reciprocal utility by choosing abatement levels taking into account both impacts from climate change, but also any potential gains or losses from how other players perceive their helpfulness. The results will differ from the direct-utility framework because abatement affects the best-response functions of each player in both the direct and the interpersonal terms.

Different values of β define a spectrum of outcomes between the traditional commons dilemma and the socially optimal outcome. Two benchmark cases are $\beta = 0$, which converts the equations to be identical to the myopic free-rider problem discussed above, while $\beta = \frac{N}{N+1}$ results in an allocation identical to the socially optimal solution. See Appendix 1 for further definition of these cases and derivation of the benchmark values of β. Choosing $\beta < 0$ implies the player rewards behavior that decreases his or her utility. Choosing $\beta > \frac{N}{N+1}$ implies that the player cares more about helping or hurting the other player than they do about themselves. The latter case results in lower aggregate utility due to over-abatement. We will not focus on the over-abatement situation in this paper, but it is worth noting that the general effect of reduced aggregate utility from too much reciprocity is quite useful. For instance, it describes cases of extravagant gift-giving culture (Carmichael and MacLeod 1997) as well as cycles of revenge (Eisenberger et al. 2004).

To solve for optimal private and public consumption, we must find first-order conditions for each player, identify best response functions, and solve for individual players' abatement levels as a function of the parameters in $u_i^D(\cdot)$, $u_j^D(\cdot)$, $R_{ij}(\cdot)$, $R_{ji}(\cdot)$ and β. When players are identical, Nash equilibrium strategies are such that $a_i = a_j$. With heterogeneous players, an analytical solution becomes difficult for most parameterizations of utility and R_{ij}, but can still be found numerically. The next section presents and solves a specific parameterization of a commons dilemma with several possible response functions.

Application to the Climate Commons

This section identifies two ways that the norms and institutions surrounding climate change can can be expressed in response functions. The first response function we identify is based on meeting or not meeting emissions reduction goals. The response function for how agent i reacts to agent j's abatement activity is labeled $R_{ij}(\cdot)$ and is based abatement targets, t_j, and abatement levels, a_j. This response function considers only the binary observation of if agents met their goals or not. Equation 4 presents this function.

$$R_{ij}(a_j, t_j) = \left\{ \begin{array}{cc} 1 & if\ a_j \geq t_j \\ -1 & if\ a_j < t_j \end{array} \right\} \qquad (4)$$

This response function implies that players view other players as friendly reciprocators so long as they meet their abatement target, but view them as a rival otherwise. This response function is solved with a normal form game and results in different equilibria depending on the value of β.

We also provide an extension of this response function that accounts for reciprocity from partial compliance. Define $\hat{a}_j$ as the level of abatement the j-th player would pursue in the individually myopic game (the case where $\beta = 0$). We assume also that t_j is set equal to the socially optimal level of abatement for player j, defined as a_j^s. To account for partial compliance and the possibility that player's have continuous responses to other player's abatement, we redefine our response function to consider what portion of the goal an agent met, $\frac{a_j}{t_j}$, and how much total abatement they did compared to what they would have done in autarky, $t_j - \hat{a}_j$, and is shown in equation 5.

$$R_{ij}(a_j, t_j) = \left(\frac{a_j}{t_j}\right)(t_j - \hat{a}_j) \qquad (5)$$

In this version, player i views player j as a friendly reciprocator to the extent that player j approaches their abatement target. Additionally, player i responds more positively if j's target is ambitious (that is, set high relative to $\hat{a}_j$). With this response function, each i-th player's maximization problem becomes:

$$\max_{a_i} \; u_i^D(\, a_i,\, a_j; X) + \; \beta_{ij}\left(\frac{a_j}{t_j}\right)(t_j - \hat{a}_j) u_j^D(\, a_i,\, a_j; X) \qquad (6)$$

The functions defined here are just two of many possible response functions but are emblematic of reciprocal behavior based on norms created by abatement target focal points.

The inclusion of reciprocity creates two new ways in which utility is affected by the choice of abatement in addition to direct benefit from abatement. First, players gain utility from rewarding or punishing other player, which we refer to as direct reciprocity.

Second, players gain utility by causing other players to reciprocally increase their abatement. We refer to this as strategic reciprocity. Strategic reciprocity is more easily analyzed in traditional economics via game theory, but the direct reciprocity portion does not easily fit in the framework. By stating that agents gain or lose their own utility by observing their friends or rivals utility, we depart from traditional economics because it requires interpersonal comparisons of utility.

Numeric Example in a Climate Commons

The inclusion of reciprocity identifies optimal abatement strategies that are higher than the strong free-rider but below that of the socially optimal level. This section uses a specific functional form for utility and the continuous response function above to show how reciprocally optimal strategies arise given various parameters. We use a relatively standard direct utility function, equation 7, which is a monotonically increasing, concave function of consumption (expressed as $w - a_i$ as before) and monotonically decreasing, convex function of GHG stock, X.

$$u_i^D(a_i, a_j; X) = v_1(w - a_i) - v_2(w - a_i)^2 - \gamma\left(X - \sum_{j=1}^{N_a} a_j\right)^2 \qquad (7)$$

The parameters $v_1, v_2,$ and γ affect the degree of concavity and relative utility impacts from consumption versus GHG damages. Equation 7 is used as the direct utility component of our maximization in equation 6. We chose values for these parameters and solved the reciprocity model in terms of β, X, t_i and $\mathrm{a_i}$ to see how behavior changes under different assumptions in the response function. Using these parameters, we solve

for three cases: the strong free-rider solution (where $\beta = 0$), the socially optimal solution ($\beta = \frac{N}{N+1}$) and a range of reciprocally optimal values in between these benchmarks.

Strong Free-Riding Solution

Using direct utility alone, we define the strong free-rider solution as the maximization of:

$$\max_{a_i} v_1(w - a_i) - v_2(w - a_i)^2 - \gamma\left(X - \sum_{j=1}^{N_a} a_j\right)^2 \qquad (8)$$

The solution to this is such that the marginal benefits from consumption are equal to the marginal benefits of abatement that results from the individual agent (not from the aggregate level of abatement). Solving the first order condition for player i's abatement gives us:

$$\frac{\partial u}{\partial a_i} = -v_1 + 2v_2(w - a_i) + 2\gamma(X - a_i - a_j) \qquad (9)$$

This can be solved for a_i, which gives us:

$$a_i = \frac{2v_2 w + 2X\gamma - v1 - 2\gamma \sum_{j\neq i}^{N_a - 1} a_j}{2(v_2 + \gamma)} \qquad (10)$$

Assuming that players are identical in their abatement activities such that:

$$\sum_{j\neq i}^{N_a - 1} a_j = (N_a - 1)a_i \qquad (11)$$

We get:

$$a_i = \frac{2v_2 w + 2X\gamma - v1 - 2\gamma(N_a - 1)a_i}{2(v_2 + \gamma)} \qquad (12)$$

Solving this for a_i:

$$a_i = \frac{2v_2 w + 2X\gamma - v_1}{2(v_2 + N_a\gamma)} \qquad (13)$$

To add a level of concreteness to these expressions, we use a specific set of parameters so that the optimal conditions can be graphed. The parameters chosen here and throughout are $X = 15, w = 4,\ \gamma = .35, v_1 = 20, v_2 = 2,$ and $N_a = 2.$ With these parameters, the individual-Nash non-cooperative level of abatement as $a_i = 1.203.$ This represents the myopic prisoner's dilemma baseline solution with no conception of reciprocal fairness included.

Socially Optimal Solution

The socially optimal solution can be obtained similarly, except we now assume that a benevolent social planner chooses all levels of abatement simultaneously to maximize aggregate welfare:

$$\max_{a_1 \ldots a_{N_a}} \sum_{i=1}^{N_a} \left(v_1(w - a_i) - v_2(w - a_i)^2 \quad - \gamma \left(X - \sum_{j=1}^{N_a} a_j \right)^2 \right) \qquad (10)$$

Solving this for the first order condition as above gives us:

$$a_i = \frac{-v_1 + 2v_2 w + 2N_a X\gamma}{2(v_2 + N_a^2\gamma)} \qquad (11)$$

Using the same parameters as in the myopic case, we find the socially optimal level of abatement is $a_i = 2.5$, significantly higher than the individual solution. These benchmarks will be used as comparisons to the fairness model.

Reciprocal Solution

We now add the behavioral response function, $R_{ij}(a_j, t_j) = \left(\frac{a_2}{t_2}\right)(t_2 - \hat{a}_2)$, into our above functions. Assuming again that there are two players, player 1's problem becomes:

$$\max_{a_1} \underbrace{v_1(w - a_1) - v_2(w - a_1)^2 - \gamma\left(X - \sum_{j=1}^{N_a} a_j\right)^2}_{Player\ 1's Direct\ Utility} + \beta_{12}\underbrace{\left(\frac{a_2}{t_2}\right)(t_2 - \hat{a}_2)}_{Response\ function}\underbrace{\left(v_1(w - a_2) - v_2(w - a_2)^2 - \gamma\left(X - \sum_{j=1}^{N_a} a_j\right)^2\right)}_{Player\ 2's Direct\ Utility}$$

The key term here is $\beta_1\left(\frac{a_2}{t_2}\right)(t_2 - \hat{a}_2)$, recalling that β_i parameterizes how much each individual cares about reciprocity. The response function in this formulation is the mechanism by which players will gain positive value from increasing the other player's utility when the other player is helpful, and similarly, gain positive value from decreasing the other player's utility when the other player is unhelpful. Intuitively, the overall result of including the response function is that if someone abates little relative to their abatement target, their rivals will want to punish them with less reciprocal abatement.

Similarly, if a player abates closer to their abatement target, their impact causes a multiplying effect on the aggregate abatement effort where all players are rewarding other players reciprocally.

At this point, we are begging the question of how t_2 is established. For the time being, we are keeping it at an arbitrary level, in this case set at the socially optimal level of abatement. In an upcoming paper we will endogenize the abatement targets by developing a nested-commons dilemma in which international negotiations set national abatement targets, and conversely, national abatement choices affect the reciprocal response of other nations at the international level.

To solve our algebraic example above, we take the first order conditions as above, and assuming that we have an interior solution and that players are identical, we find the optimal abatement for each i-th individual is:

$$a_i =$$

$$\frac{t(X\beta\gamma - v_2 - 2\gamma) - X\beta\gamma + \sqrt{4(t-1)t\beta\gamma(2\mathrm{v}_2 w + 2X\gamma - v_1) + (X\beta\gamma + t(\mathrm{v}_2 + (2 - X\beta)\gamma))^2}}{4(t-1)\beta\gamma}$$

Using our parameters from above, we this solution identifies a range of optimal abatement values depending only on the value chosen for β. For any positive value of β the reciprocally optimal abatement level is higher than the individual solution. The next section will analyze the shape of the relationship between β and reciprocally optimal abatement.

Results for the Climate Commons

Note that it is possible for the reciprocal model to predict *more* abatement than in the socially optimal case, but this only occurs if people care more about rewarding and punishing their rivals than they do about their direct utility, as in gift giving examples. Fig. 3.1 shows the relationship between β and aggregate utility. In this figure we see that the socially optimal level of abatement is met when $\beta = \frac{2}{3}$, as expected.

Figure 3.1: Effect of increasing intensity of reciprocity (β) on abatement

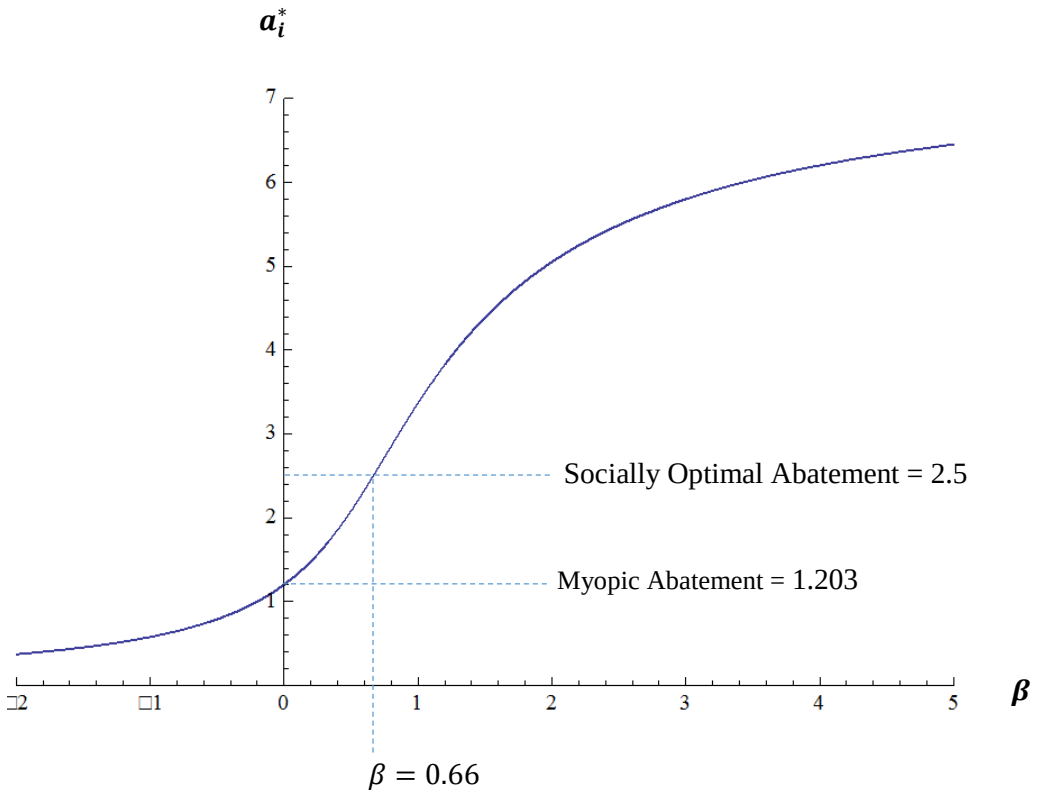
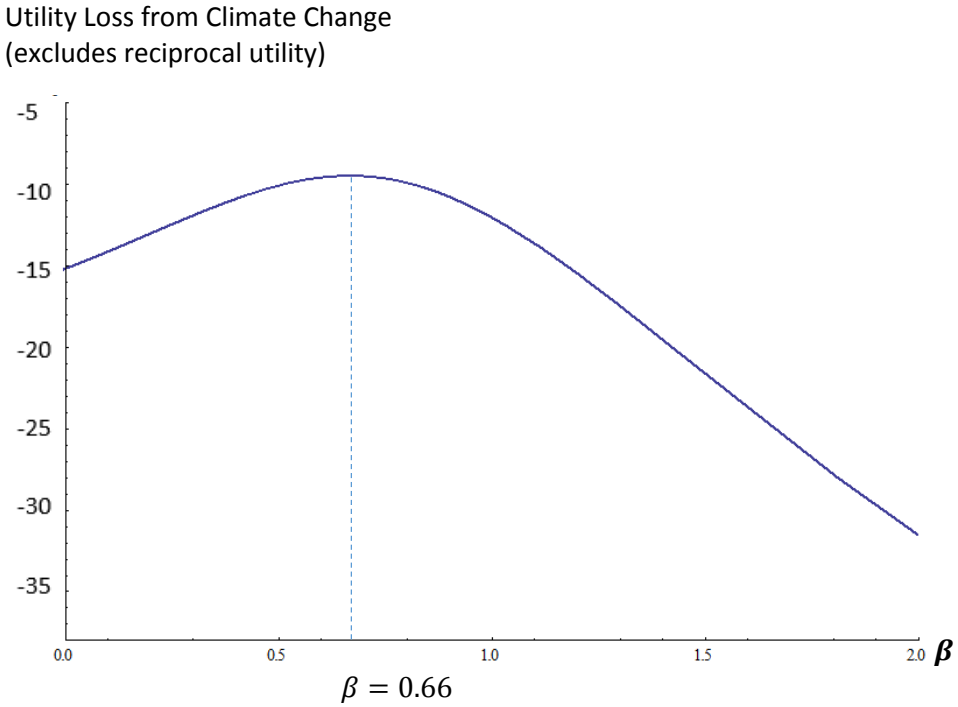


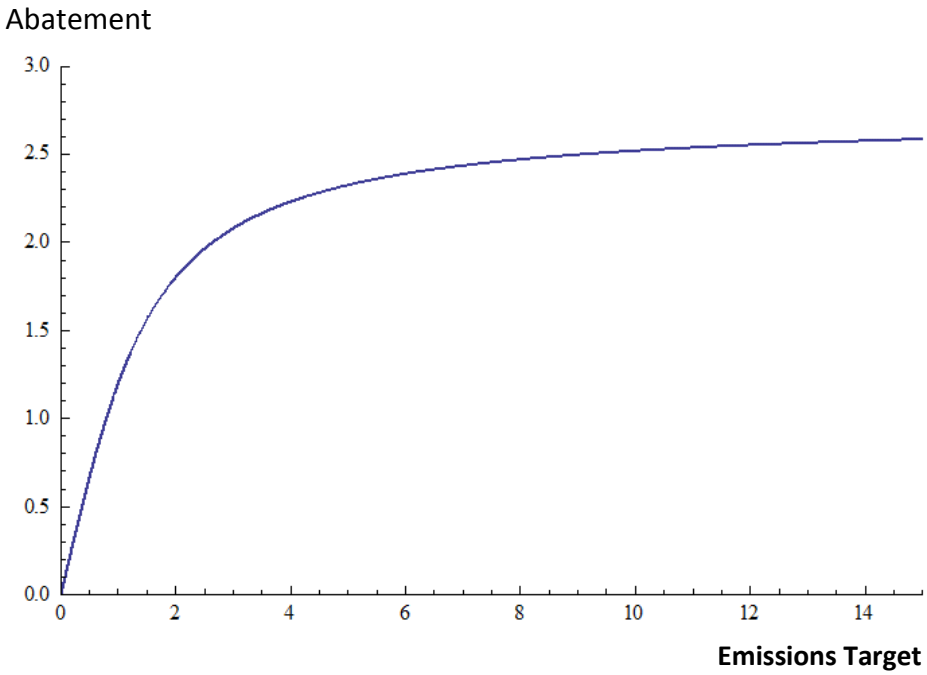
Fig. 3.2 illustrates the effect that changing the target has on the optimal abatement level. If one sets the target only slightly above what individuals would do out of their own self-interest, the behavior response function implies that neither side would care much about achieving this goal because it was not a very ambitious or impressive goal. Conversely, with a higher emissions target, each player's response function changes so that movement towards the emissions target garners more respect, and therefore more reciprocal response. This is intuitive insofar as players respond with more positive reciprocity to other players' actions that constitute more effort. Notice that no matter how high one sets the target, individuals do not abate past what would have been socially optimal. This limit holds for any value of β, but the rate of convergence increases as β rises.

Figure 3.2.: Effect of increasing intensity of reciprocity (β) on utility



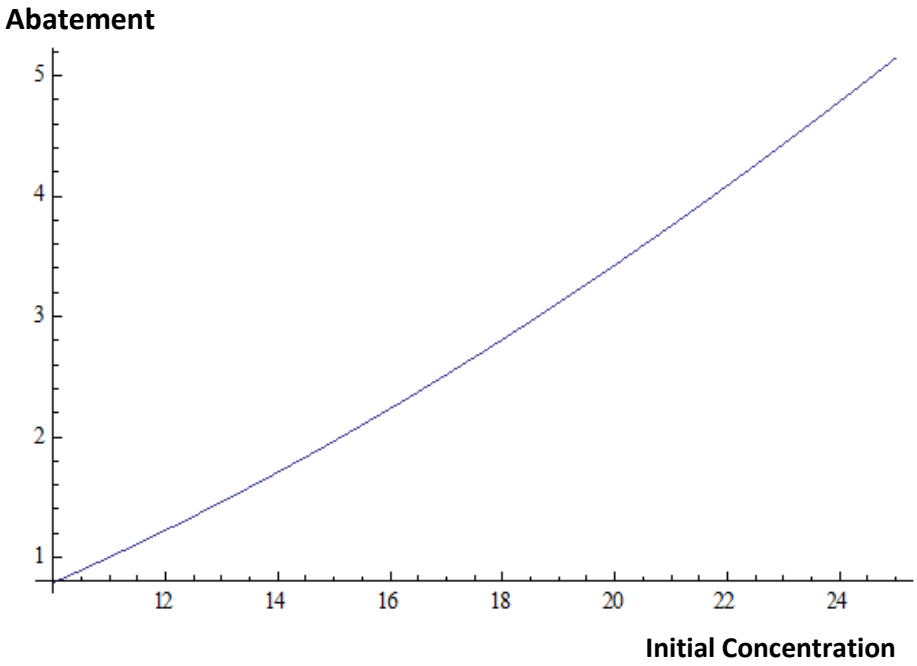
The optimal abatement in the reciprocity model approaches zero as the target approaches zero (as in Fig. 3.3). While this may seem contradictory, as it would seem to imply that setting a goal would result in a lower level of abatement than selfish players would achieve, it is quite intuitive. Any target set below the individually rational level would constitute an abatement-reducing target (which increases emissions). In this context of climate change, this would not seem to be an interesting case, but the flexibility here suggests that the reciprocity model is robust to writing equations in terms of reducing a public bad (emissions) and increasing a public good (abatement).

Figure 3.3: Effect of Emissions Target on Reciprocally Optimal Abatement



Finally, we can see from Fig. 3.4 that increasing the initial stock of GHG increases the optimal level of abatement. This is because the harm caused by emissions is a convex function while returns to consumption are concave. Other parameters in the model, such as γ, which measures the harm of emissions relative to the benefits of consumption, show a similar expected result.

Figure 3.4: Effect of Initial Emissions on Reciprocally Optimal Abatement



Simultaneous changes of β and the target increase the equilibrium abatement, shown in Fig. 3.5. One interesting result of the model is seen here. If the intensity of reciprocity (β) is high, even a slight increase in the emissions target has a large impact. This is due to players rewarding the reciprocally beneficent actions of other players meeting the new goal.

Figure 3.5: Simultaneous Changes of β and t on Reciprocally Optimal Abatement

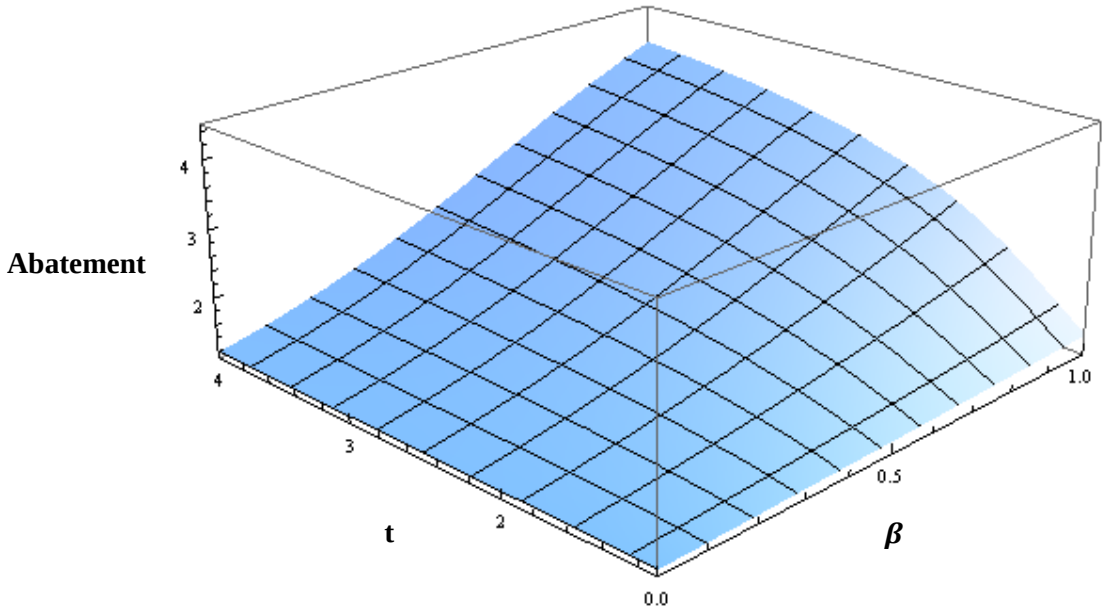
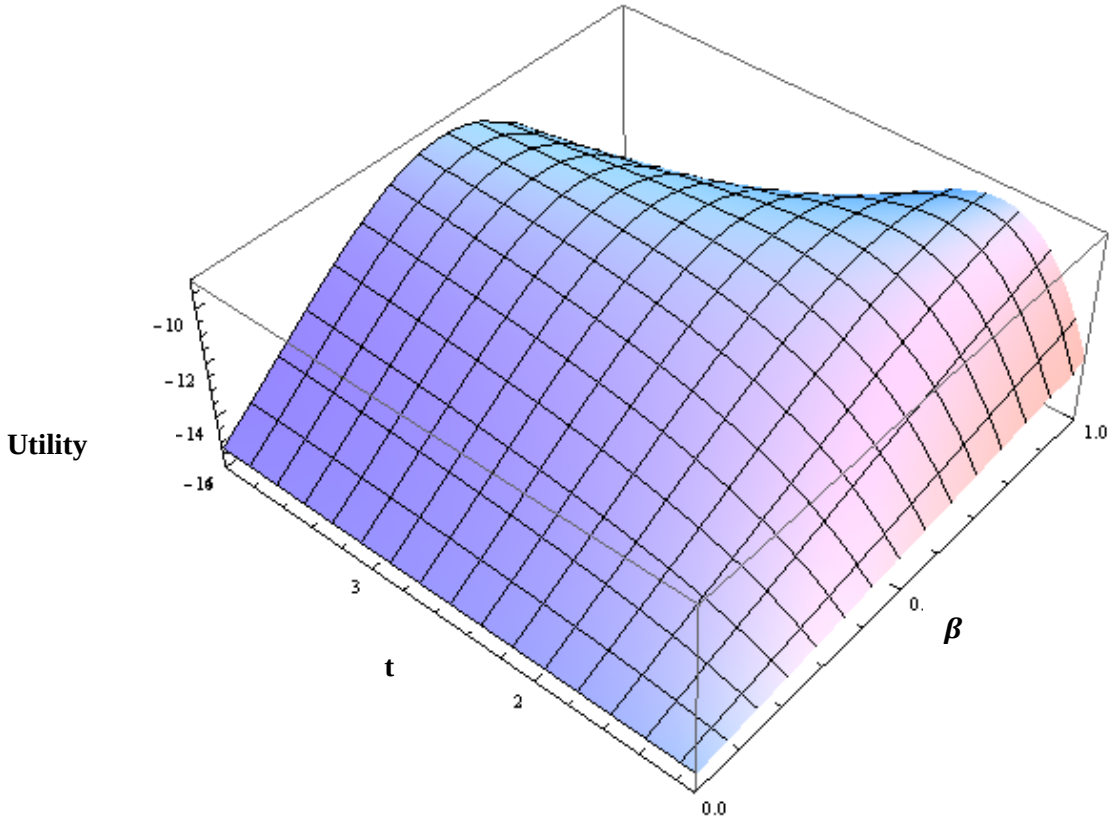


Fig. 3.6 shows the effect that changing the target and β have on aggregate utility. For values of β between zero and $\frac{N}{N+1}$, increasing the target increases total utility. If players do not care about reciprocity at all, changing the target changes nothing, as expected. For values of β higher than $\frac{N}{N+1}$, it is still possible to achieve the socially optimal level of abatement, but only be setting a correspondingly lower target.

Figure 3.6: Simultaneous Changes of β and t on Direct Utility



The non-monotonicity present here is another useful result. For lower values of β and t, utility is increasing in a straight-forward, concave way. But, when both β and t become too high, there is a precipitous drop in overall direct[8] utility (I refer to this as the over-

[8] If one includes both direct and interpersonal utility, the effect becomes much more negative. This, however, is due to the additional loss of utility persons experience through their response functions (they are mad at their rivals and thus unhappy). I focus results discussion on the direct utility component alone because the policy recommendations and interpretation tasks become difficult if we consider more than direct utility.

commitment pit). The reason for the drop is that if the goal is too difficult to achieve, the previously virtuous feedback effect (in which partial abatement caused other players to respond with more abatement) becomes a negative feedback effect where each player is focused on punishing their rivals who failed to meet the challenging goal.

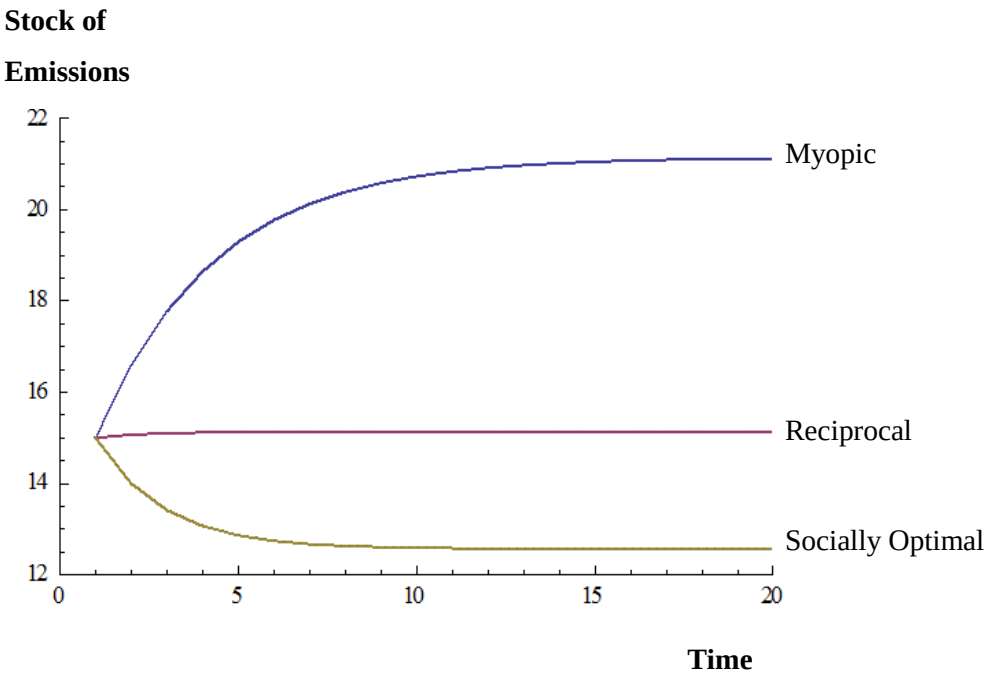
Extending of the Climate Example to Multiple Time Periods

The above solution relies on a static solution concept, however, it is reasonable to wonder what would happen if this problem played itself out over time. Using the continuous behavioral response function outlined above, we add a simple equation of motion to the stock of emissions:

$$X_{t+1} = X_t + \varphi_1\left(w - \sum_{i=1}^{N_a} a_i\right) - \varphi_2\left(\sum_{i=1}^{N_a} a_i\right) \qquad (16)$$

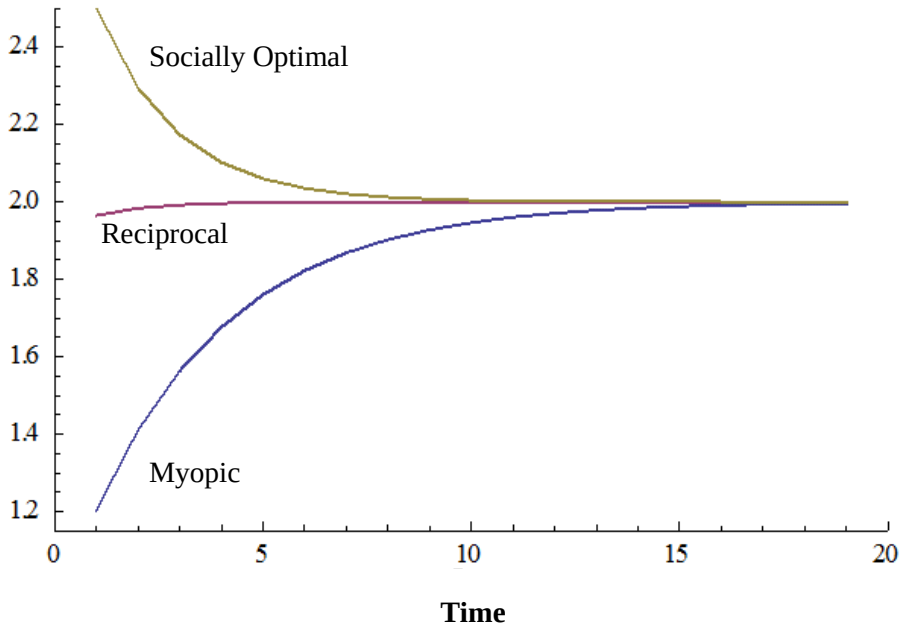
Where consumption increases the emissions stock at a rate of φ_1 and abatement reduces the emissions stock at a rate of φ_2. For simplicity, assume $\varphi_1 = \varphi_2 = 1$, and solve period-by-period for the equilibrium path and levels of abatement and utility. Charting this out over time (shown in Fig. 3.7), we can see how the equilibrium paths differ over time between the non-cooperative solution and the fairness solution. The reciprocity model results in more abatement activity, limiting the increase in GHG concentration growth to a very low level. As expected, the GHG stock is much higher at every time period in the non-reciprocity case, and is much lower in the socially optimal case, beginning with drastic reductions in the first period.

Figure 3.7: Stock of Emissions over Time



The optimal abatement choices at each time are depicted in Fig. 3.8. As we might expect given the figure for GHG stocks in Fig. 3.7, the socially optimal abatement path involves dramatic reductions in emissions right away, the reciprocity model predict a moderate and increasing level of abatement, and the non-reciprocity model predict a very significant delay before the abatement levels increase (as a result of significantly higher harm from climate change).

Figure 3.8: Abatement Choices over Time



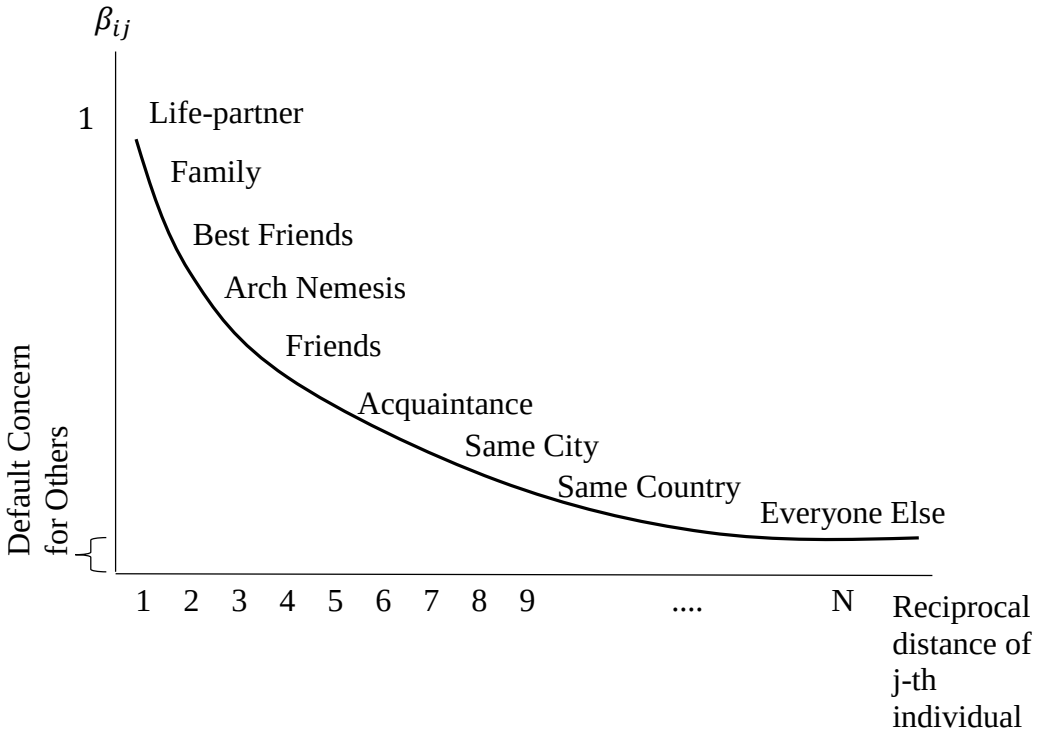
Discussion and Conclusion

The application discussed above provides an example of how to define, solve and interpret a climate commons while allowing for partial free-riding due to reciprocity. Although the parameters used here are for example purposes only, one of the main contributions of this model it explicitly states a functional model that can explain situations in which humans do not completely free-ride. Given the strong evidence from behavioral economics that observed human actions are not always typified by perfect rationality, it is important to develop models that can predict partial free-riding (or other implications of the extended utility framework) as a direct result of the model. Once

created, such models can be used to generate specific hypotheses and functional forms that can then be confirmed or refuted with empirical work. I include in Appendix 2 a discussion of the econometric estimation technique I will use. Unfortunately, collecting the data necessary for this was outside the scope and budget of this thesis. Nonetheless, using laboratory observations, similar to Marwell and Ames (1979 and 1981), the model could readily be tested.

Finally, note that the extended utility framework can be extended to consider multiple persons. For example, consider a vector of β_{ij} values sorted by value so that the highest values are at the beginning of the vector. Moving further down this vector means the i-th individual has progressively less intensity of their reciprocal feelings toward the j-th individual in the vector. Fig. 3.9 plots this relationship.

Figure 3.9: β_{ij} Vector by Reciprocal Distance



In this figure, we see that those relationships closest to the i-th individual have much higher reciprocal intensities. Also note that the arch-nemesis has a relatively high value, positive value. The high value of β for the arch-nemesis simply means the i-th individual experiences large changes in utility from observing the nemesis gain or lose. Using a vector of βs like this along with the solution methods above, it is possible to solve for a reciprocal equilibrium for many individuals, such as is expressed in equation 17. The basic approach remains the same, which is to find first order conditions and best response functions for all individuals and solve the system of equations.

$$u_i^R(a_i, a_j) = u_i^D(a_i, a_j) + \sum_{j \neq i}^{I} \beta_{ij} R_{ij}(a_j) u_j^D(a_i, a_j) \qquad (17)$$

The shape of the curve in Fig. 3.9 provides another basis for future empirical work. Specifically, I econometric estimation can help identify the relationship between reciprocal distance and reciprocal intensity.

Finally, the commons reciprocity utility model can be implemented with agent-based simulation methods, as discussed in the previous chapters of this thesis. The reciprocal distance, in the context of forest commons, can be defined as physical distance and/or the extent to which villagers use a common while the extended utility framework provides behavior rules used in the simulation. I present a preliminary implementation of this approach in Appendix 2.

Conclusion

This thesis attempted to address how we can model economic systems and ecological systems in a coherent framework. I presented three specific models to illustrate this point, along with results from various locations. Although these chapters are complete, they also represent starting points down different methodological paths. For each of these topics, and a wide variety of other topics, I will extend these methods with more precision and more examples in future research. This chapter will outline, in cursory form, several of the research ideas most likely to be next on my research agenda.

The work in chapter two on geospatial optimization of global food and carbon tradeoffs attempts to identify a biophysical frontier rather than a bio-economic frontier that incorporates economic behavior and decisions. However, with future work, I believe this is exactly the direction global ecosystem service analysis needs to go. To do this, I will focus on the following three research topics. First, I will extend the ability of the model used in chapter two to more accurately predict where crop extensification is likely to happen using crop models (DSSAT, Lizaso et al. 2011), historical data, suitability data, and economic indicators. Forthcoming data from Deepak et al. (2014, in prep) will provide historical data that can calibrate the extensification predictions, as will time series data from MODIS (2000 to 2012). Another avenue will be to use sub-grid-cell heterogeneity to make predictions about which grid-cells still have extensification potential. For example, if a 10km grid-cell is currently cultivated at 50%, but if we know that 25% of the 30 meter sub-grid-cells in this same location (using higher-resolution data from LANDSAT or MODIS remote sensing sources) then we can hypothesize that the grid-cell cannot be extensified beyond 75%. In chapter two, we avoided making specific extensification assumptions like this and instead limited our analysis to using ad-hoc

assumptions that are accurate in aggregate. However, data are available to make these assumptions specific rather than ad-hoc.

Second, I will extend the analysis methodology used by the Global Landscapes Initiative (GLI, Foley et al. 2011, Mueller et al. 2012) to create the data I used in chapter two. Specifically, the analysis used by GLI relies on a concept called "climate bins" to make hypotheses about where intensification can increase caloric production by closing "yield gaps." This methodology divides the globe into 100 non-continuous areas identified by similar growing degree days (GDD) and precipitation. So for example, if an area in Iowa and Ukraine have the same GDD and precipitation, they are defined to be in the same climate bin, even if Ukraine is observed to have lower yields. The yield gap analysis argues that if these two plots are in the same climate bin, it is reasonable to assume that Ukraine could increase its yields to match that of Iowa through better management and intensification strategies. Although the climate bin methodology is very useful and well-accepted in the peer-reviewed scientific literature, it has several aspects I believe can be extended. Specifically, the climate bin approach can be rephrased as a simple two variable regression that predicts yields based on GDD and precipitation:

$$yield_{potential} = \beta_1 GDD + \beta_2 Precip + \epsilon$$

This is a testable hypothesis and should be thoroughly validated. However, initial regressions I have run suggest that while the simple regression does have predictive power, it has considerable omitted-variable bias, specifically by ignoring management decisions, irrigation, access to markets and fertilization rates. Thus, I believe the climate bin should be extended to a broader multi-variable regression, approximately of the form:

$$yield\_potential_{xy} = \beta_0 + \beta_1 GDD_{xy} + \beta_2 Precip_{xy} + \beta_2 SoilType_{xy} + \beta_2 SoilQuality_{xy} + \beta_2 Slope_{xy} + \beta_2 Fertilizer_{xy} + \beta_2 Irrigation_{xy} + \beta_2 MarketAccess_{xy} + \beta_2 Infrastructure_{xy} + \beta_2 Demographics_{xy} + \beta_2 ManagementMethod_{xy} + \epsilon_{xy}$$

This too is a testable hypothesis, and I believe that formal statistical testing of which equation best predicts yields will be very useful to the analysis presented in chapter two, but also much more broadly.

Third, using the results from the previous two points, I will re-run the entire geospatial optimization model of chapter two with parameterized extensification cost functions. Thus, each grid-cell would have a function for marginal production increases of calories. The optimization, then, would choose the grid-cells where extensification maximizes marginal-yield / marginal carbon, rather an the existing indicator of crop advantage.

Finally, I will extend the above analysis points with more policy discussion. Specifically, I will include an in-depth analysis of food-for-environment swaps by identifying ranked pairs of "best exchanges." Analysis of where these pairs exist and how the best exchanges can be made to happen will give specific policy relevance to the global analysis.

The next research topic that I plan to address post-graduation is to fully implement the simulation of common property forest management I presented in chapter five and validate the data with field observations in Cambodia. I have built a relationship with a field office of the World Wildlife Fund in Cambodia that will be able to do household surveys of villagers in the Mekong watershed and collect paired observations of villager

household location and foraging location. Once this data has been collected, it will be used to validate the general method and serve as example data in the public release of the modeling software I developed. The software will be a part of the InVEST modeling tools of the Natural Capital Project.

Finally, the last extension of my research that I will pursue is creating a way to interpolate higher-resolutions of gridded data on socioeconomic indicators. The approach I will used, referred to as downscaling, works by determining the relationship between a coarse data input and a finer input and then using the relationship to impute what the best approximation of the coarse data on the finer grid should be. This approach was discussed in chapter three where it was used to interpolate high-resolution population density data. The approach, however, is much broader and can be used to create a wealth of economic data on a spatial resolution finer than what has been available. The follow-up research that I need to do will be to create the methodology, downscale it for a specific data set, such as the Earthstat data, and then show if (or if not) the higher-resolution data are accurate.

Although the work presented in this these (along with the additional research avenues discussed in this chapter) are primarily methodological and theoretical, I have attempted to focus them on important problems humanity faces as we grow into a more complex and resource demanding society. Climate change, food security and managing our common property resources (such as forests) all will require concerted effort to understand and manage. Moreover, as our societies grow more complex and harness a greater amount of resources, the informational requirements to understanding these systems will also grow. The methodologies presented in this these attempt to extend the scope of insights gained in economics to a broader set of problems. In this way, I hope to help answer the questions of how ought humans live in our environment.

Bibliography

Alexandratos N, Bruinsma J (2012). World agriculture towards 2030/2050: the 2012 revision. No. 12-03. (FAO: ESA Working paper, Rome).

Alfano,F ., and Gerald Marwell (1980). "Experiments on the Provision of Public Goods II: Non-Divisibility and Free Riding in 'Real' Groups." Social Psychology Quarterly 43:300-309.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent.Location and land use. Toward a general theory of land rent.

An, L., Linderman, M., Qi, J., Shortridge, A., Liu, J., 2005. Exploring complexity in a human-environment system: an agent-based spatial model for multidisciplinary and multiscale integration. Ann. Assoc. Am. Geogr. 95 (1), 54e79.

Anas, A., Arnott, R., Small, K.A., 1998. Urban spatial structure. Journal of Economic Literature 36 (3), 1426–1464.

Anderson, S.T. and S.E. West. 2006. Open space, residential property values, and context. Regional Science and Urban Economics 36: 773-789.

Ando A, Camm J, Polasky S, Solow A (1998). Species distributions, land values, and efficient conservation. *Science* 279, no. 5359: 2126-2128.

Andreoni, James (1988). "Why Free Ride? Strategies and Learning in Public Goods Experiments." Journal of Public Economics 37 (Dec.):291-304. Axelrod, Robert. 1981.

Arrow, K.J. (1977). "Extended Sympathy and the Possibility of Social Choice." *American Economic Review* 67(1)(Feb.): 219-25.

Ausubel JH, Wernick IK (2012). Peak Farmland and the Prospect for Land Sparing. *Population and Development Review* 38, no. s1: 221-242.

Bandeira, S. O., Marconi, L., & Barbosa, F. (1996). Preliminary study of threatened plants of Mozambique. In The Biodiversity of African Plants (pp. 306-309). Springer Netherlands.

Bardhan, P., & Udry, C. (1999). Development microeconomics. Oxford University Press.

Barnes, W. L., Pagano, T. S., & Salomonson, V. V. (1998). Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. Geoscience and Remote Sensing, IEEE Transactions on, 36(4), 1088-1100.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1), 1-101.

Bateman IJ, et al. (2013). Bringing ecosystem services into economic decision-making: Land use in the United Kingdom. *Science* 341: 45-50.

Berio, A. (1984). The analysis of time allocation and activity patterns in nutrition and rural development planning. Food and Nutrition Bulletin 6(4): 53-68.

Berkes, Fikret, and Aykut Kence (1987). "Fisheries and the Prisoner's Dilemma Game: Conditions for the Evolution of Cooperation among Users of Common Property Resources." METU Journal of Pure and Applied Sciences 20 (Aug.):209-27.

Berkes, Fikret, David Feeny, Bonnie J. McCay, and James M. Acheson (1989). "The Benefits of the Commons." Nature 340 (6229, July 13):91-93.

Berman, M., Nicolson, C., Kofinas, G., Tetlichi, J., Martin, S., 2004. Adaptation and sustainability in a small arctic community: results of an agent-based simulation model. Arctic 57 (4), 401e414.

Biran, A., Abbot, J., & Mace, R. (2004). Families and firewood: a comparative analysis of the costs and benefits of children in firewood collection and use in two rural communities in sub-Saharan Africa. Human Ecology, 32(1), 1-25.

Cabrera, A.R., Deadman, P., Moran, E., Brondízio, E.S., Vanwey, L.K., 2012. Exploring Demographic and Lot Effects in an ABM/LUCC of Agriculture in the Brazilian Amazon. In: Heppenstall, A.J., Crooks, A.T., See, L.M., Batty, M. (Eds.), Agent-based Models of Geographical Systems. Springer, Heidelberg.

Campbell, B.M., Luckert, M., Scoones, I., 1997. Local level valuation of savanna resources. A case study from Zimbabwe. Econ. Bot. 51, 59–77.

Carmichael, H. L., and W. Bentley MacLeod (1997). *Gift Giving and the Evolution of Cooperation*. International Economic Review. Vol. 38, No. 3. August 1997.

Carraro C, Marchiori C, Sgobbi A (2005). "Advances in negotiation theory: bargaining, coalitions and fairness." World Bank policy research working paper 3642, June 2005

Carraro C, Marchiori C, Sgobbi A (2007). "Negotiating on water: insights from non-cooperative bargaining theory." Environ Dev Econ 12:329–349

Carraro C, Sgobbi A (2007). A stochastic multiple players, multi-issues bargaining model for the Piave River Basin. FEEM working paper 101.2007

Chang, K. T. (2010). Introduction to geographic information systems. New York: McGraw-Hill.

Cialdini, R.B., Vincent, J.E., Lewis, S.K., Catalan, J., Wheeler, D. & Darby, B.L. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology, 31*, 206-215.

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus them social identity shapes neural responses to intergroup competition and harm. *Psychological science*, *22*(3), 306-313.

Ciriacy-Wantrup,S. V., and Richard C. Bishop. 1975. "'Common Property' as a Concept in Natural Resource Policy." Natural Resources Journal 15 (Oct.):713-27.

Clarke, E. (1971). Multipart pricing of public goods. *Public Choice* 8 19–33.

Cordell D, Drangert JO, White S (2009). The story of phosphorus: Global food security and food for thought. *Global environmental change* 19, no. 2: 292-305.

Costanza R, et al. (1997). The value of the world's ecosystem services and natural capital. *Nature* 387, no. 6630: 253-260.
Costello, C., & Polasky, S. (2008). Optimal harvesting of stochastic spatial resources. Journal of Environmental Economics and Management, 56(1), 1-18.
Crooks, A. T., & Castle, C. J. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In Agent-based models of geographical systems (pp. 219-251). Springer Netherlands.
Damania, R., Milner-Gulland, E.J., Crookes, D.J., 2005. A bioeconomic analysis of bushmeat hunting. Proc. Royal Soc. B: Biol. Sci. 272 (1560), 259e266.
Dasgupta, P.S., and G.M. Heal (1979) Economic Theory and Exhaustible Resources (Cambridge: Cambridge University Press)
De Bruyn A, Bolton GE (2004). Predicting the role of fairness in bargaining. Out-of-sample estimates from a social utility model with quantal response. ISBM report 1-2004
Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation, IEEE Transactions on, 6(2), 182-197.
Dufwenberg, Martin and Georg Kirchsteiger (2003). A Theory of Sequential Reciprocity. Games and Economic Behavior. Volume 47, Issue 2, May 2004, Pages 268–298
Dwyer, M. J., & Schmidt, G. (2006). The MODIS reprojection tool. In Earth science satellite remote sensing (pp. 162-177). Springer Berlin Heidelberg.
Egoh B, et al. (2008). Mapping ecosystem services for planning and management. *Agriculture Ecosystems and Environment* **127** (1–2): 135–140.
Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin*, *30*(6), 787-799.
Elster, J., & Roemer, J. E. (Eds.). (1993). *Interpersonal comparisons of well-being*. Cambridge University Press.
Evans, T.P., Phanvilay, K., Fox, J., Vogler, J., 2011. An agent-based model of agricultural innovation, land-cover change and household inequality: the transition from swidden cultivation to rubber plantations in Laos PDR. J. Land Use Sci. 6(2e3), 151e173.
Falk., & U. Fischbacher. (2006). A theory of reciprocity. Games and Economic Behavior, 54(2), 293-315.
Falkenmark M (2003). Freshwater as shared between society and ecosystems: from divided approaches to integrated challenges. *Philos T Roy Soc B* 358(1440) (2003): 2037–2049. doi: 10.1098/rstb.2003.1386.
Farrow, S. (1985) 'Testing the efficiency of extraction from a stock resource.' Journal of Political Economy 93, 452-87

Faße, A., & Grote, U. (2013). The economic relevance of sustainable agroforestry practices—An empirical analysis from Tanzania. Ecological Economics, 94, 86-96.
Feeny, David, Fikret Berkes, Bonnie J. McCay, and James M. Acheson (1990). "The Tragedy of the Commons: Twenty-Two Years Later." Human Ecology 18 (1):1-19.
Fehr E, Schmidt K (1999). A theory of fairness, competition and cooperation. Q J Econ 114:817–868
Fehr, Ernst; and Simon Gächter (2000). "Fairness and Retaliation: The Economics of Reciprocity". *Journal of Economic Perspectives* 14 (3): 159–181. doi:10.1257/jep.14.3.159. ISSN 0895-3309.
Felix, M., & Gheewala, S. H. (2011). A review of biomass energy dependency in Tanzania. Energy Procedia, 9, 338-343.
Filatova, T., Parker, D.C., van derVeen, A., 2007. Agent-based land markets: heterogeneous agents, land prices, and urban land use change. Proceedings of the 4th Conference of the European Social Simulation Associations (ESSA'07), Toulouse, France.
Filatova, T., Parker, D.C., van derVeen, A., 2009. Agent-based urban land markets: agent's pricing behavior, land prices and urban land use change. Journal of Artificial Societies and Social Simulation 12 (13).
Fleuret, P. C., and Fleuret, A. K. (1978). Firewood Use in a Peasant Community: A Tanzanian Case Study. The Journal of Developing Areas 12: 315-322.
Foley JA, et al. (2011). Solutions for a cultivated planet. *Nature* 420: 337-342.
Food and Agriculture Organization (FAO) (2012) *The State of Food Insecurity in the World, 2012*; accessed at www.fao.org on March 24, 2013.
Food and Agriculture Organization (FAO) (2013). Food Balance Sheet: World+(Total)-2009. *FAOSTAT*; accessed at <faostat.fao.org> on March 24, 2013.
Food and Agriculture Organization (FAO) 2012 Agriculture Organization of the United Nations. *FAO statistical yearbook*.
Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., andHuang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168–182.
Fujita, M., & Krugman, P. (2004). The new economic geography: Past, present and the future*. Papers in regional science, 83(1), 139-164.
Fujita, M., Krugman, P. R., & Venables, A. J. (2001). The spatial economy: Cities, regions, and international trade. MIT press.
Gabszewicz, J. J., & Thisse, J. F. (1986). Spatial competition and the location of firms. Fundamentals of pure and applied economics, 5, 1-71.

Gardner, Roy, Elinor Ostrom and James M. Walker (1990). The Nature of Common-Pool Resource Problems. *Rationality and Society* July 1990 *vol. 2 no. 3 335-358* doi: 10.1177/1043463190002003005

Gaughan, A. E., Stevens, F. R., Linard, C., Jia, P., & Tatem, A. J. (2013). High resolution population distribution maps for Southeast Asia in 2010 and 2015. PloS one, 8(2), e55882.

Gibbs HK, et al. (2010) Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences* 107, no. 38: 16732-16737.

Gintis, H. (2000). Beyond Homo economicus: evidence from experimental economics. *Ecological economics*, *35*(3), 311-322.

Goldstein J (2012). Integrating ecosystem service tradeoffs into land-use decisions. *Proceedings of the National Academy of Sciences*, 109(19): 7565-7570.

Greenblatt, S. A. (1996). Wavelets in econometrics (pp. 139-160). Springer Netherlands.

Groves, T. 1973. Incentives in teams. *Econometrica* 41 617–631.

Gupta, B., Pal, D., & Sarkar, J. (1997). Spatial Cournot competition and agglomeration in a model of location choice. Regional Science and Urban Economics, 27(3), 261-282.

Güth W, Schmittberger R, Tietz R (1990). Ultimatum bargaining behaviour: a survey and comparison of experimental results. J Econ Psychol XI:417–449

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. Pervasive Computing, IEEE, 7(4), 12-18.

Hall, R. W. (1986). The fastest path through a network with random time-dependent travel times. *Transportation science*, *20*(3), 182-188.

Hamilton, J. H., Thisse, J. F., & Weskamp, A. (1989). Spatial discrimination: Bertrand vs. Cournot in a model of location choice. Regional Science and Urban Economics, 19(1), 87-102.

Hammond, P.J. (1991). "Interpersonal Comparisons of Utility: Why and How They Should Be Made." European University Institute. In J. Elster and J. Roemer, *Proceedings of the Sloan Conference on "Interpersonal Comparability of Welfare"*, New York: Cambridge University Press.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. science, 342(6160), 850-853.

Hardin, G. (1968). The tragedy of the commons. science, 162(3859), 1243-1248.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001). "In search of homo economicus: behavioral experiments in 15 small-scale societies." *The American Economic Review* 91, no. 2: 73-78.

Hertel T, Verma M, Ivanic M, Rios AR (2011). GTAP-POV: A Framework for Assessing the National Poverty Impacts of Global Economic and Environmental Policies. (Center for Global Trade Analysis Working Paper No. 3731, Department of Agricultural Economics, Purdue University).

Hertel TW, Warren T (1999). *Global trade analysis: modeling and applications*. Edited by Thomas W. Hertel. Cambridge university press.

Hirshleifer, Jack (1983). From weakest-link to best-shot: the voluntary provision of public goods. *Public Choice,* 41:371–86.

Hoekstra AY, Mekonnen MM, Chapagain AK, Mathews RE, Richter BD (2012). Global Monthly Water Scarcity: Blue Water Footprints versus Blue Water Availability. *PLoS ONE* 7(2): e32688. doi:10.1371/journal.pone.0032688.

Hotelling, H. (1929): "Stability in Competition," Economic Journal, 39, 41-57

Hotelling, H. (1931) 'The economics of exhaustible resources.' Journal of Political Economy 39, 137-75

Hubert B, Rosegrant M, van Boeke MAJS, Ortiz R (2010). The future of food: scenarios for 2050. *Crop Science* 50, no. Supplement_1: S-33.

IGBP-DIS (1998). SoilData (V.0) A program for creating global soil-property databases, IGBP Global Soils Data Task, France.

Iwamura, T., Lambin, E. F., Silvius, K. M., Luzar, J. B., & Fragoso, J. (2014). Agent-based modeling of hunting and subsistence agriculture on indigenous lands: Understanding interactions between social and ecological systems.Environmental Modelling & Software.

Justice, C. O., Vermote, E., Townshend, J. R., Defries, R., Roy, D. P., Hall, D. K., ... & Barnsley, M. J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. Geoscience and Remote Sensing, IEEE Transactions on, 36(4), 1228-1249.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Kareiva P, Tallis H, Ricketts TH, Daily GC, Polasky S (2011). *Natural Capital: Theory and Practice of Mapping Ecosystem Services*, Oxford University Press.

Klaiber, H.A. and D.J. Phaneuf. 2010. Valuing open space in a residential sorting model of the Twin-Cities. Journal of Environmental Economics and Management 60: 55-77.

Kohyama, T., Suzuki, E., Partomihardjo, T., Yamada, T., & Kubo, T. (2003). Tree species differentiation in growth, recruitment and allometry in relation to maximum height in a Bornean mixed dipterocarp forest. Journal of Ecology, 91(5), 797-806.

Krugman, P. (1998). What's new about the new economic geography?. Oxford review of economic policy, 14(2), 7-17.

Lambin EF, Meyfroidt P (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences* 108, no. 9: 3465-3472.

Lamm, C., Meltzoff, A. N., & Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *22*(2), 362-376.

Lawler, J., D. Lewis, E. Nelson, A.J. Plantinga, S. Polasky, J. Withey, D. Helmers, S. Martinuzzi, D. Pennington, and V. Radeloff. Projected land-use change impacts on ecosystem services in the U.S. *Proceedings of the National Academy of Sciences. In review.*

Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophysical Research Letters, 37(15).

Lehner, B., Verdin, K., & Jarvis, A. (2006). HydroSHEDS technical documentation, version 1.0. World Wildlife Fund US, Washington, DC, 1-27.

Levine, R. V., & Norenzayan, A. (1999). The pace of life in 31 countries. Journal of cross-cultural psychology, 30(2), 178-205.

Licker R, et al. (2010). Mind the gap: How do climate and agricultural management explain the yield gap of croplands around the world? *Global Ecological Bioengineering* 19: 769-782.

Lizaso, J. I., Boote, K. J., Jones, J. W., Porter, C. H., Echarte, L., Westgate, M. E., & Sonohat, G. (2011). CSM-IXIM: A new maize simulation model for DSSAT version 4.5. *Agronomy Journal*, *103*(3), 766-779.

Lobell DB, Schlenker W, Costa-Roberts J (2011). Climate Trends and Global Crop Production since 1980. *Science* 333: 616-620.

Lopomo G, Ok EA (2001). Bargaining, interdependence and the rationality of fair division. RAND J Econ 32(2):263–283

Luoga, E. J., Witkowski, E. T. F., & Balkwill, K. (2000). Economics of charcoal production in miombo woodlands of eastern Tanzania: some hidden costs associated with commercialization of the resources. Ecological Economics, 35(2), 243-257.

Luoga, E. J., Witkowski, E. T. F., & Balkwill, K. (2002). Harvested and standing wood stocks in protected and communal miombo woodlands of eastern Tanzania. Forest Ecology and Management, 164(1), 15-30.

Luoga, E. J., Witkowski, E. T. F., & Balkwill, K. (2004). Regeneration by coppicing (resprouting) of miombo (African savanna) trees in relation to land use. Forest Ecology and Management, 189(1), 23-35.

Magliocca, N., V. McConnell, M. Walls, and E. Safirova. 2012. Zoning on the urban fringe: Results from a new approach to modeling land and housing markets. Regional Science and Urban Economics 42: 198–210.

Malimbwi, R., Zahabou, E., 2009. The analysis of sustainable fuelwood production systems. In: Rose, S., Remedio, E., Trosseo, M.A. (Eds.), Criteria and Indicators for Sustainable Woodfuels: Case Studies from Brazil, Guyana, Nepal, Philippines and Tanzania. FAO, Rome, pp. 195–228.

Marchiori, Carmen (2010). Concern for Fairness and Incentives in Water Negotiations. Environ Resource Econ (2010). 45:553–571, DOI 10.1007/s10640-009-9328-y

Marpu, P. R., Pedergnana, M., Dalla Mura, M., Benediktsson, J. A., & Bruzzone, L. (2013). Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data. Geoscience and Remote Sensing Letters, IEEE, 10(2), 293-297.

Marwell, G., & Ames, R. E. (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. *journal of Public Economics*, *15*(3), 295-310.

Matthews, Robin B., Nigel G. Gilbert, Alan Roach, J. Gary Polhill, and Nick M. Gotts. "Agent-based land-use models: a review of applications." Landscape Ecology 22, no. 10 (2007): 1447-1459.

McKean, Margaret A (1982). "The Japanese Experience with Scarcity: Management of Traditional Commons Lands." Environmental Re-view 6:63-88. –

Millennium Ecosystem Assessment (MA) (2005). *Ecosystems and Human Well-Being. Synthesis*. (Washington DC: Island Press).

Mills, E.S., 1967. An aggregative model of resource allocation in a metropolitan area. American Economic Review 57 (2), 197–210.

Mills, E.S., 1972. Studies in the Structure of the Urban Economy. Johns Hopkins Press.

Mohtashe, Mojdeh and Lik Mui (2003). Evolution of indirect reciprocity by social information: the role of trust and reputation in evolution of altruism. Laboratory for Computer Science, MIT, 200 Technology Square, Room 419, Cambridge, MA 02139, USA. http://dx.doi.org/10.1016/S0022-5193(03)00143-7.

Monfreda C, Ramankutty N, Foley JA (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles 22*, no. 1.

Mueller ND, et al. (2012). Closing yield gaps through nutrient and water management. *Nature*, *490*(7419), 254-257.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., & Shearer, S. A. (2004). Map quality for ordinary kriging and inverse distance weighted interpolation. Soil Science Society of America Journal, 68(6), 2042-2047.

Naidoo R, et al. (2008). Global mapping of ecosystem services and conservation priorities. *Proceedings of the National Academy of Sciences* 105, no. 28: 9495-9500.

Natural Capital Project (2013). Accessed at www.naturalcapitalproject.org on March 22, 2013.
Nelson E, et al. (2008). Efficiency of incentives to jointly increase carbon sequestration and species conservation on a landscape. *Proceedings of the National Academy of Sciences* 105(28): 9471-9476.
Nelson E, et al. (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment* 7(1): 4–11.
Nelson E, et al. (2010). Projecting global land-use change and its effect on ecosystem service provision and biodiversity with simple models. *PloS one* 5, no. 12: e14327.
Ng, Y. K. (2003). From preference to happiness: Towards a more complete welfare economics. *Social Choice and Welfare*, *20*(2), 307-350.
Nowak, Martin A., Corina E. Tarnita and Edward O. Wilson (2010). The evolution of eusociality. Nature 466, 1057–1062 (26 August 2010).
Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20-36.
Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). 2011. MODIS subsetted land products, Collection 5. Available on-line [http://daac.ornl.gov/MODIS/modis.html] from ORNL DAAC, Oak Ridge, Tennessee, U.S.A. Accessed April 8, 2014.
Oliphant, Travis E. A Guide to NumPy. Vol. 1. USA: Trelgol Publishing, 2006.
Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge university press.
Ostrom, Elinor, Roy Gardner, and James Walker (1994). Rules, Games and Common-Pool Resources. Ann Arbor: University of Michigan Press.
Polasky S, et al. (2008). Where to put things? Spatial land management to sustain biodiversity and economic returns. *Biological Conservation* 141 y(6): 1505-1524.
Polasky S, Nelson E, Pennington D, Johnson K (2011). The impact of land-use change on ecosystem services, biodiversity and returns to landowners: a case study in the State of Minnesota. *Environmental and Resource Economics* 48(2): 219-242.
Polasky, S., E. Nelson, J. Camm, B. Csuti, P. Fackler, E. Lonsdorf, C. Montgomery, D. White, J. Arthur, B. Garber-Yonts, R. Haight, J. Kagan, A. Starfield, and C. Tobalske. 2008. Where to put things? Spatial land management to sustain biodiversity and economic returns. Biological Conservation 141: 1505-1524.
Questining the Assumptions of the Tragedy of the Commons Model of Fisheries - Feeny 1996
R.E Malimbwi, B Solberg, E.J Luoga (1994). Estimation of biomass and volume in miombo woodland at Kitulanghalo Forest Reserve, Tanzania. J. Trop. For. Sci., 7 (1994), pp. 230–242

Rabin M (1993). Incorporating fairness into game theory and economics. Am Econ Rev 83:1281–1302

Ramankutty N, Evan AT, Monfreda C, and Foley JA (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000 *Glob. Biogeochem. Cycles* 22, GB1003.

Ramankutty N, Foley JA (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochem Cycles.* 13,: 997–1027.

Ramsey F.P. (1928), "A Mathematical Theory of Saving," Economic Journal, Vol. 38, No 152, pp. 543 559

Rand, David G., Corina E. Tarnita, Hisashi Ohtsuki, and Martin A. Nowak. "Evolution of fairness in the one-shot anonymous Ultimatum Game." *Proceedings of the National Academy of Sciences* 110, no. 7 (2013): 2581-2586.

Regan, R. T. (1971). Effects of a favor and liking on compliance. *Journal of Experimental Social Psychology,7*,627-639.

Robinson, D.T., Brown, D.G., Parker, D.C., Schreinemachers, P., Janssen, M.A., Huigen, M., Wittmer, H., Gotts, N., Promburom, P., Irwin, E., 2007. Comparison of empirical methods for building agent-based models in land use science. Journal of Land Use Science 2 (1), 31e55.

Rodrigue, J. P., Comtois, C., & Slack, B. (2013). *The geography of transport systems*. Routledge.

Rosegrant M, Paisner MS, Meijer S, Witcover J (2001). *Global food projections to 2020. Emerging trends and alternative futures* Washington, DC: IFPRI Publications. ISBN 0-89629-640-7. (2001) See http://www.ifpri.org/pubs/books/globalfoodprojections2020.htm.

Rosegrant MW, Fernandez M, Sinha A (2009). Looking into the future for agriculture and KST. p. 370–376. In B.D. McIntyre et al. (ed.) IAASTD Global Rep. Island Press, Washington, DC.

Rosegrant MW, Meijer S, Cline SA (2008). International model for policy analysis of agricultural commodities and trade (IMPACT): Model description. (Washington, DC, USA: International Food Policy Research Institute).

Roth AE (1995). Bargaining experiments. In: Kagel J, Roth AE (eds) Handbook of experimental economics. Princeton University Press, Princeton

Rothkopf, Michael H (2007). Thirteen Reasons Why the Vickrey-Clarke-Groves Process Is Not Practical. OPERATIONS RESEARCH, Vol. 55, No. 2, March–April 2007, pp. 191–197. doi 10.1287/opre.1070.0384

Ruesch A, Gibbs HK (2008). New IPCC Tier-1 Global Biomass Carbon Map For the Year 2000. Available online from the Carbon Dioxide Information Analysis Center *[http://cdiac.ornl.gov]*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Runge, C. F. (1984). "Institutions and the Free Rider: The Assurance Problem in Collective Action." *Journal of Politics* 46(1)(Feb.): 154-81.
Runge, C.F. (1981). "Common Property Externalities: Isolation, Assurance, and Resource Depletion in a Traditional Grazing Context." *American Journal of Agricultural Economics* 63(4): 595-606.
Sanchirico, J. N., & Wilen, J. E. (1999). Bioeconomics of spatial exploitation in a patchy environment. Journal of Environmental Economics and Management, 37(2), 129-150.
Sanfey, Alan G., James K. Rilling, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen (2003). "The neural basis of economic decision-making in the ultimatum game." *Science* 300, no. 5626: 1755-1758.
Schlenker W, Roberts MJ (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106(37): 15594-15598.
Schreinemachers, Pepijn, and Thomas Berger. "An agent-based simulation model of human–environment interactions in agricultural systems." Environmental Modelling & Software 26, no. 7 (2011): 845-859.
Shackleton, C.M., 1993. Fuelwood harvesting and sustainable utilization in a communal grazing land and protected area of the eastern Transvaal lowveld. Bio. Cons. 63, 247–254.
Shackleton, C.M., Netshiluvhi, T.R., Shackleton, S.E., Geach, B.S, Ballance, A., Fairbanks, D.F.K.,(1999). Direct Use Values of Woodland Resources from Three Rural Villages,Pretoria: CSIR, Environmentek.
Shapley, L. S. (1967). Utility Comparison and the Theory of Games. Rand Corporation. Santa Monica, CA.
Singh, A. (1989). Review Article Digital change detection techniques using remotely-sensed data. International journal of remote sensing, 10(6), 989-1003.
Singh, I., Squire, L., & Strauss, J. (1986). Agricultural household models: extensions, applications, and policy. Johns Hopkins University Press.
Singh, I., Squire, L., & Strauss, J. (1986). Agricultural household models: extensions, applications, and policy. Johns Hopkins University Press.
Stahl, K. (1982). Location and spatial pricing theory with nonconvex transportation cost schedules. The Bell Journal of Economics, 575-582.
Stout, M. (2006). *The sociopath next door: The ruthless versus the rest of us*. Random House LLC.
Sugden, R. (1985). Consistent conjectures and voluntary contributions to public goods: why the conventional theory does not work. *Journal of Public Economics*, *27*(1), 117-124.
Sugden, Robert (1995). A Theory of Focal Points. *The Economic Journal*, 105 (May), 533-550.

Taheripour F, Birur DK, Hertel TW, Tyner WE (2007). Introducing liquid biofuels into the GTAP database. *GTAP Research Memorandum.* Center for Global Trade Analysis, Purdue University.

Tesfatsion, L., & Judd, K. L. (Eds.). (2006). Handbook of computational economics: Agent-based computational economics (Vol. 2). Amsterdam: North-Holland.

The Economics of Ecosystems and Biodiversity (TEEB) (2010). The Economics of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: A Synthesis of the Approach, Conclusions and Recommendations of TEEB. http://www.teebweb.org.

The World Bank (2014). Tanzania National Panel Survey, 2008. Retrieved online at http://microdata.worldbank.org/index.php/catalog/76/study-description on April 8, 2014.

Tilman D, Balzer C, Hill J, Befort BL (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* 108, no. 50: 20260-20264.

Tol RSJ (2009). The economic effects of climate change. *The Journal of Economic Perspectives* 23, no. 2: 29-51.

Tollefson J (2013). A Light in the Forest. *Foreign Affairs* 92(2): 141-151.

Toller, G., Xiong, X., Sun, J., Wenny, B. N., Geng, X., Kuyper, J., ... & Wu, A. (2013). Terra and Aqua moderate-resolution imaging spectroradiometer collection 6 level 1B algorithm. Journal of Applied Remote Sensing, 7(1), 073557-073557.

Trivers, Robert L (1971). The Evolution of Reciprocal Altruism. The Quarterly Review of Biology. Vol. 46, No. 1, Mar., 1971. Stable URL: http://www.jstor.org/stable/2822435

U.S. Geological Survey (1996). GTOPO30 . Sioux Falls, SD: United States Geological Survey Center for Earth Resources Observation and Science (EROS).

Valavanis, Stefan (1958). "The Resolution of Conflict When Utilities Interact," Journal of Conflict Resolution, 2, 2, 156- 69.

Varian, Hal (2004). System Reliability and Free Riding, in Economics of Information Security Advances in Information Security, 2004, Volume 12, 1-15, DOI: 10.1007/1-4020-8090-5_1 First published in *ICEC2003: Fifth International Conference on Electronic Commerce*, N. Sadeh, ed., ACM Press, 2003, pp. 355–366.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* 16 8–37.

von Thunen, J.H (1826). The Isolated State. Hamburg: Perthes. English translation. Oxford: Pergamon, 1966.

West PC, et al. (2010). Trading carbon for food: Global comparison of carbon stocks vs. crop yields on agricultural land. *Proceedings of the National Academy of Science*: 1-4.

Yang, C.H., Fujita, M., 1983. Urban spatial structure with open space. Environment and Planning A 15, 67–84.
Zahabu, E., Malimbwi, R., & Ngaga, Y. (2005, November). Payments for environmental services as incentive opportunities for catchment forest reserves management in Tanzania. In Paper to the Tanzania Association of Foresters Meeting. Dar es Salaam, Tanzania (pp. 6-9).
Zaks DPM, Ramankutty N, Barford CC, Foley JA (2007). From Miami to Madison: Investigating the relationship between climate and terrestrial net primary production. *Global Biogeochemical Cycles* 21(3), GB3004.
Zimmerman, D., Pavlik, C., Ruggles, A., & Armstrong, M. P. (1999). An experimental comparison of ordinary and universal kriging and inverse distance weighting. Mathematical Geology, 31(4), 375-390.

Appendix 1: Supplemental Information for Chapter 1

In the first section of this supplement, we provide additional details on the steps taken in the optimization. In the second section, we illustrate each of the steps in section one with global maps. In the third section, we present a sensitivity analysis and results from alternate scenarios that illustrate how alternative assumptions influence results. In the fourth section, we discuss heterogeneity within grid-cells and present an example grid cell that illustrates our assumption on heterogeneity.

Optimization Details

Our optimization procedure identifies areas to extensify agricultural production to meet increased demand while minimizing the loss of carbon storage. Formally, we solve the following optimization:

$$\min_{E_{xy}} \sum_{x=1}^{X}\sum_{y=1}^{Y} E_{xy}\left(\sum_{i=1}^{175}\left(\frac{DF_i\, C}{R_i} * Y_{ixy}\right) - PNVC_{xy} - .25\left(SC_{xy}\right)\right)$$

$$S.t. \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{i=1}^{175}\left(Y_{ixy} + \left(E_{xy} * H_{ixy}\right)\right) \geq D \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{i=1}^{175}\left(Y_{ixy}\right)$$

$$0.05 \leq P_{xy} + E_{xy} \leq 0.95 \qquad \forall\, x \in X, y \in Y$$

$$E_{xy} = \min\left(P_{xy} * G_u,\quad \left(.95 - P_{xy}\right) * G_c\right) \quad \forall\, x \in X, y \in Y$$

where $E_{xy} \in \{[0.05,0.95]\ \forall\, x \in [0,4320]$ and y $\in [0,2160]\}$ is the proportion of the xy-th grid-cell to be extensified (this is a set of approximately 10 million choice variables), DF_i is the proportion of dry matter of the yield for crop i, C is the carbon content of dry matter (0.45 g C per g dry matter) and R_i represents the proportion of the crop that leaves the farm (rather than remaining on the field or belowground). Tons of economically valued dry crop yield, measured per xy-th grid-cell for each of the 175 i-th crops, is denoted as Y_{ixy}, while H_{ixy} is the per hectare yield. To calculate the amount of

carbon that would be contained either in natural land or in the crop chosen for extensification, we identify the potential natural vegetation carbon ($PNVC_{xy}$) and soil carbon (SC_{xy}). D represents the required increase in calories that must be met (100% increase in our baseline scenario). The first constraint simply states that the minimization problem is constrained to meet the required calorie increase in aggregate, based on existing crop production. The second set of constraints defines P_{xy} as the current proportion cultivated for each grid-cell. Thus, for each x-y pair, we constrain our optimization to only consider cells that have at least 0.05 proportion of the grid-cell under cultivation and to never increase cultivation beyond 0.95 proportion cultivated. The final set of constraints defines how large of an increase in proportion cultivated each grid-cell can achieve. This is achieved by defining assumptions on how much the proportion cultivated can increase when the resulting proportion is less than 0.95 (defined by the unconstrained growth factor g_u) and how much it can increase if unconstrained growth would increase the proportion past 0.95 (defined by the constrained growth factor g_c). Together, these constraints attempt to limit the potential extensification choices we consider in our optimization to be limited to ones where per-crop data is sufficiently numerous, where land is not approaching complete cultivation, and also, to limit the increase by 2050 to be smaller than completely closing the yield-gap present in each grid-cell.

The business-as-usual (BAU) is defined as increasing the share cultivated of each grid-cell by the percent necessary to meet the caloric goal. BAU increases are subject to the same feasibility constraints and limits a in the optimal simulation but ignore carbon storage.

The following steps describe how we processed our data and solved the maximization.

Step 1: Construct Caloric Yield (CY) per Ha and per Grid-Cell for Each Crop

From the EarthStat dataset, we have yield per hectare for 175 crops (Fig. A1.1 illustrates this for Maize), along with the proportion of the grid-cell harvested for each crop (Fig. A1.2). Multiplying A1.1 with A1.2 and adjusting for the number of hectares present gives us the tons of Maize present in each grid-cell (Fig. A1.3). We combine the data in A1.3 for each crop with data from FAOSTAT on caloric content of Maize (Fig. A1.4), adjusted so that both units are in dry-tons. Summation of 175 crop-specific versions of map A1.4 creates Fig. A1.5, which is caloric yield, CY_{xy}, the numerator of our crop advantage indicator.

Step 2: Calculate Marginal Effect of Extensification on Carbon Storage (ΔC)

We calculate the marginal change in carbon storage from equation (4). Fig. A1.6 shows the change in crop carbon, CC_{xy}. Fig. A1.7 shows the carbon storage in potential natural vegetation, $PNVC_{xy}$. Fig. A1.8 shows the carbon storage change attributable to soil carbon SC_{xy}. Using equation (4), we derive ΔC_{xy}, shown in Fig. A1.9.

Step 3: Calculate the Crop Advantage Indicator (CA)

Using equation (1), we use the maps calculated in steps 1 and 2 to create the map of Crop Advantage, shown in Fig. A1.10. This represents the marginal value of extensifying in each grid-cell.

Step 4: Find Optimal Extensification Choices Step-wise

To find the set of extensification choices that maximize the amount of carbon storage present while still meeting calorie demand, we identify which cell has the highest crop-advantage and then extensify cultivation on that cell according to the growth factor (subject to feasibility constraints). We chose to increase extensification in proportion to observed extensification rather than to a biophysical maximum to reflect the lack of data and sub-grid-cell heterogeneity on soil, topography or other conditions at finer resolutions that may prevent extensification. We repeat this procedure on each grid-cell in order of decreasing crop-advantage until we meet the global calorie production goal. The output of this algorithm, shown in Fig. A1.11, identifies where new calories should be produced according to the optimal simulation.

We implement the above algorithm in a software application, Geospatial Economic Modeler (GEM), which is a visualization and analysis program written by the lead author of this paper. GEM is written in Python and is ready for integration into other Python based analysis programs, such as InVEST from the Natural Capital Project.

Step 5: Find BAU Extensification Choices

Using the extensification rules defined above and a similar iterative procedure, we proportionately increase cultivation until we meet the increased calorie demand. We assume that the future BAU scenario has the same proportion of each crop as existed in 2000 for each grid-cell. Fig. A1.12 shows where new calories are produced under BAU.

Step 6: Identify Difference between BAU and Optimal Extensification Choices

Figs. A1.11 and A1.12 describe dramatically different agriculture systems, despite the fact that at global resolutions the differences are difficult to see. Because of this difficulty, it is more useful to analyze the difference between the simulations. We define a Net Land Protected map (shown in Fig. A1.13) as the difference between the proportion cultivated in BAU and the proportion cultivated in the optimized simulation. Positive values (shown in green) of this number indicate the proportion of a grid-cell that is spared from extensification in the optimized simulation but is not spared in BAU. Negative values (shown in red) indicate where extensification must be higher in the optimal simulation than in BAU to continue reaching the caloric demand.

Step 7: Calculate Difference between Simulations in Carbon Storage

Note that in red cells in Fig. A1.13, there is a net *loss* of carbon storage in the optimal scenario. This is because the optimal simulation produces as many calories as BAU, and thus must sacrifice some lands to reach the calorie goal. However, grid-cells sacrificed have systematically lower carbon storage values (a direct result of optimizing), and thus we see that total carbon is increased in the optimal simulation. To calculate how much carbon is gained or lost in each grid-cell, we multiply the values in the Net Land Protected map (Fig. A1.13) with the values in Fig. A1.9, Marginal Effect of Extensification on Carbon Storage (ΔC). We present the net change of carbon storage in Fig. A1.14. Summation of all grid-cells in A1.14 gives us the global amount of carbon saved. In the scenario illustrated here (more pessimistic than the baseline scenario in the text), map A1.14 sums to 12.08 billion metric tons of carbon.

Illustration of Optimization Routine

In this section, we include global maps that illustrate each of the steps described above. In general, the differences between maps appear slight at the global scale, but become pronounced at regional scales. To illustrate the range of analyses possible, the maps presented in this section are from a more pessimistic scenario in which only 50% of the increase is met with intensification.

Figure A1.1: Maize Proportion of Grid-Cell Harvested.

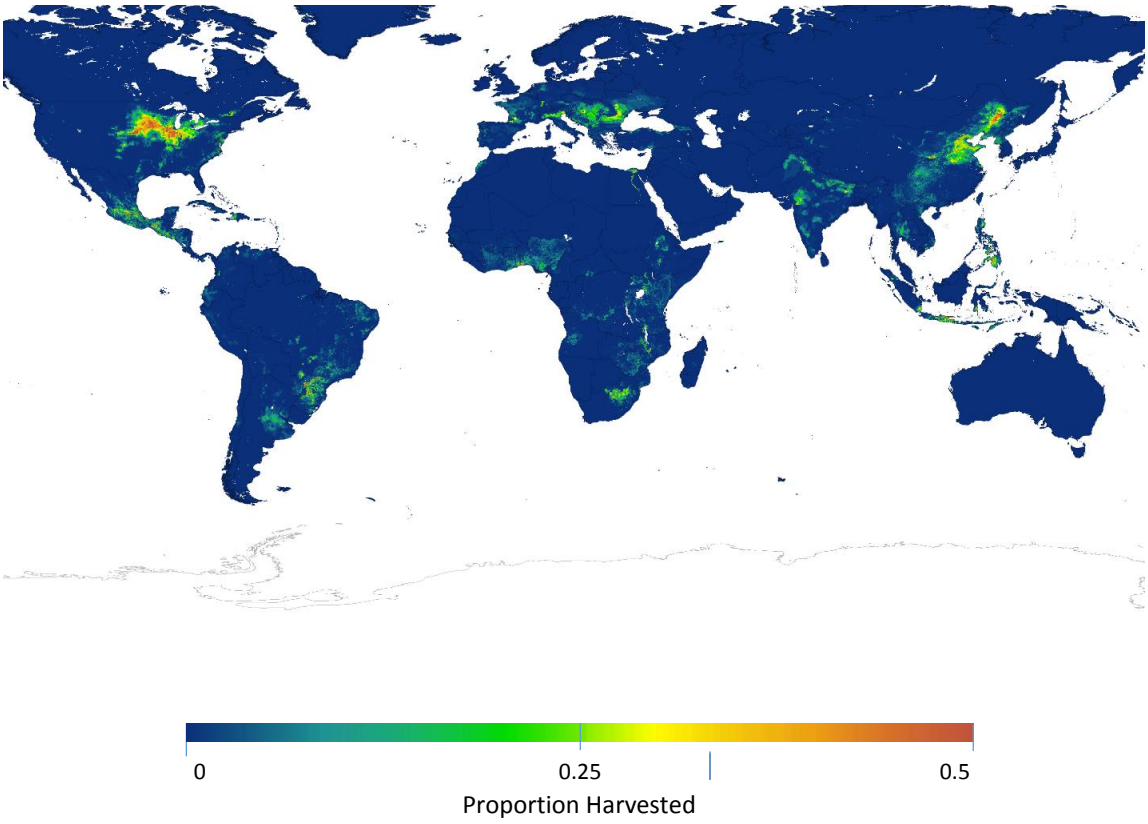


Figure A1.2: Maize per Ha Yield (tons).

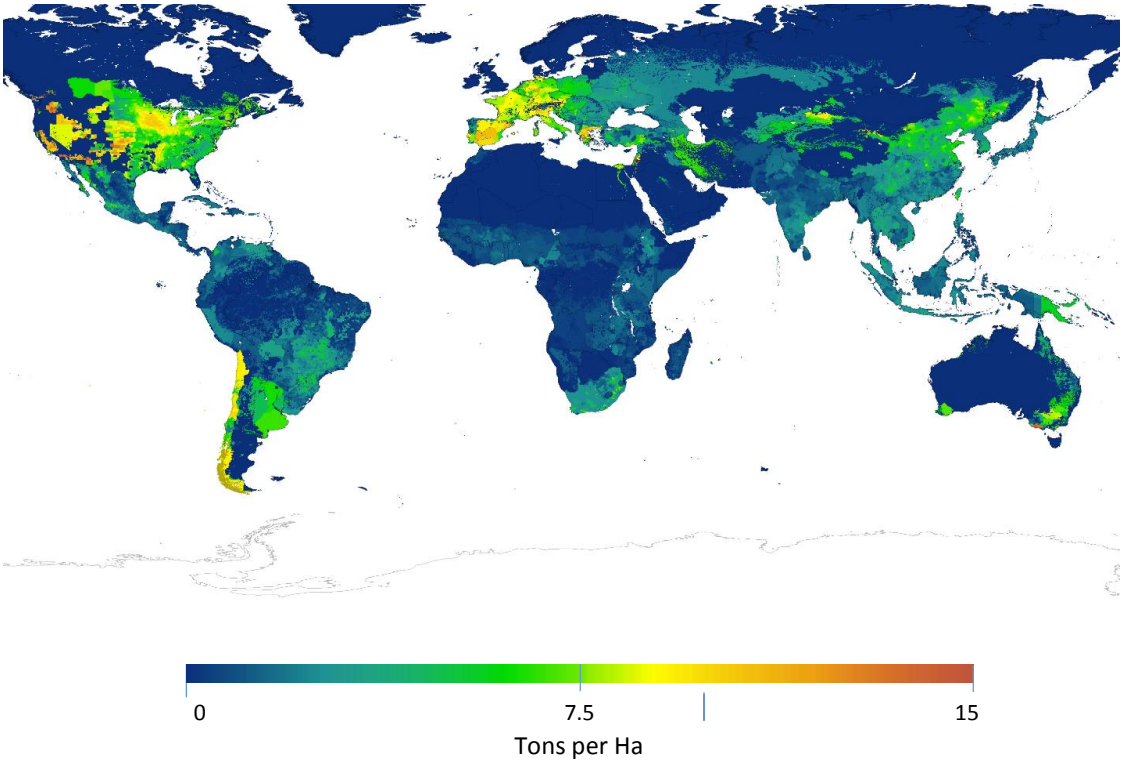


Figure A1.3: Maize per Grid-Cell Yield (tons dry weight).

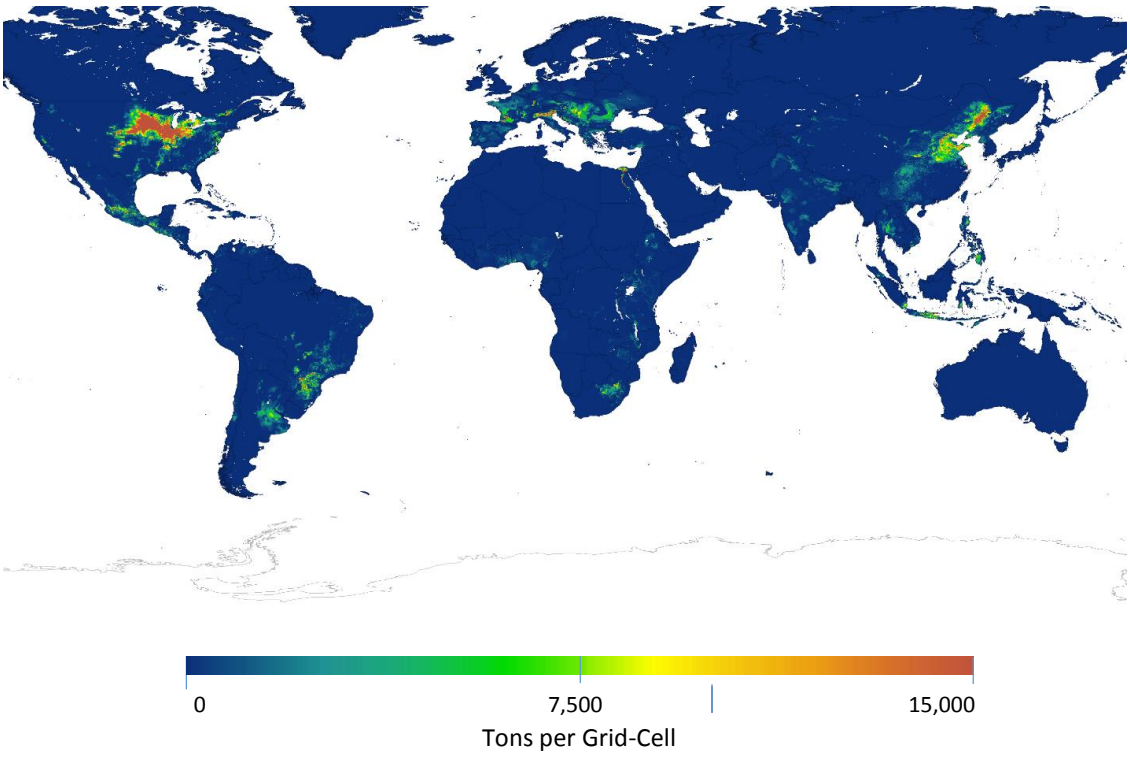


Figure A1.4: Maize per Grid-Cell Caloric Yield (kcal).

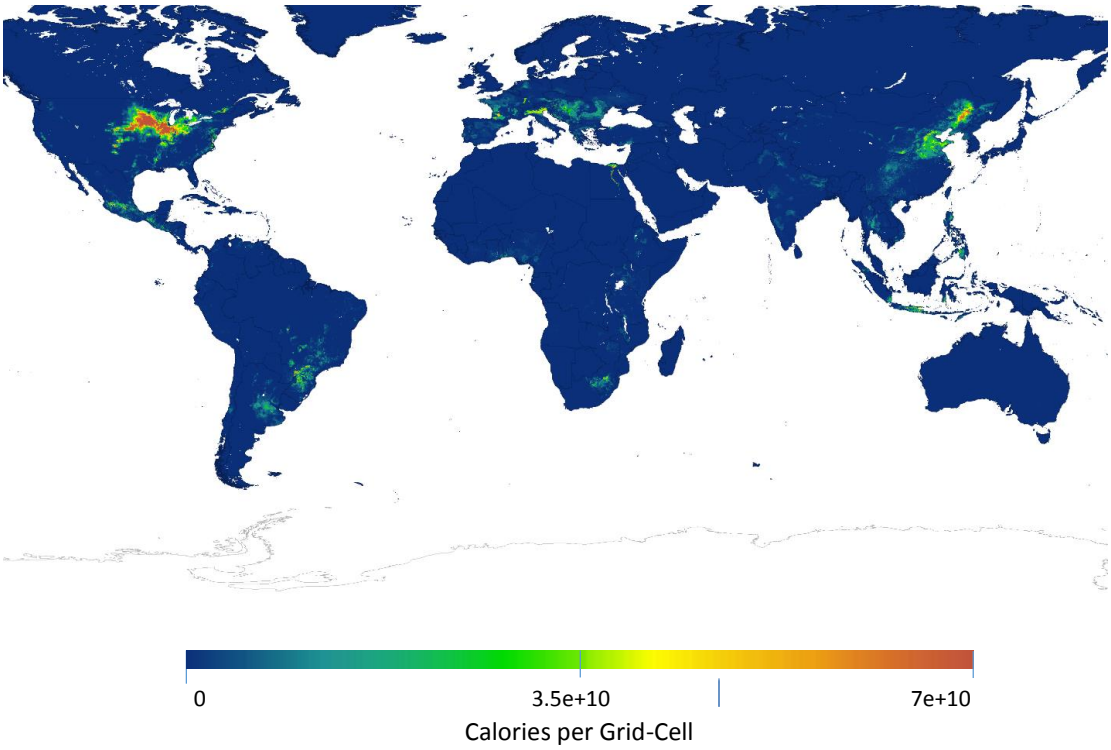


Figure A1.5: All Crop per Grid-Cell Caloric Yield (kcal).

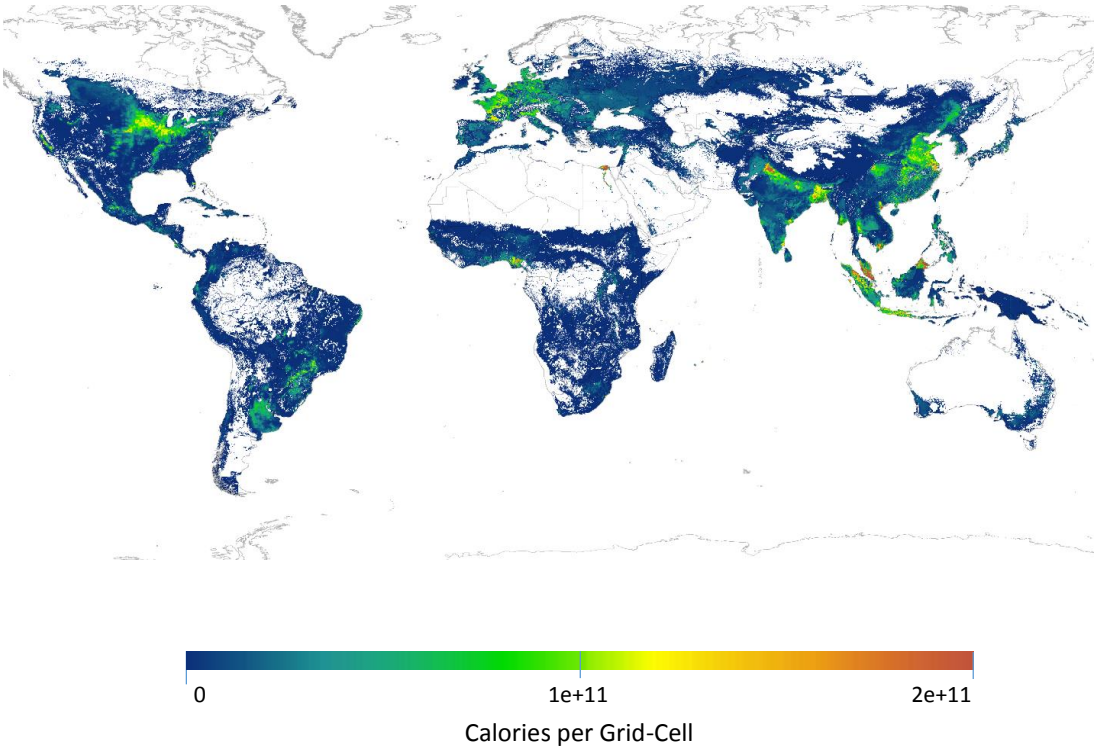


Figure A1.6: Carbon Storage in Crops, Assuming Complete Extensification of Grid-Cell at Current Crop Proportions (tons).

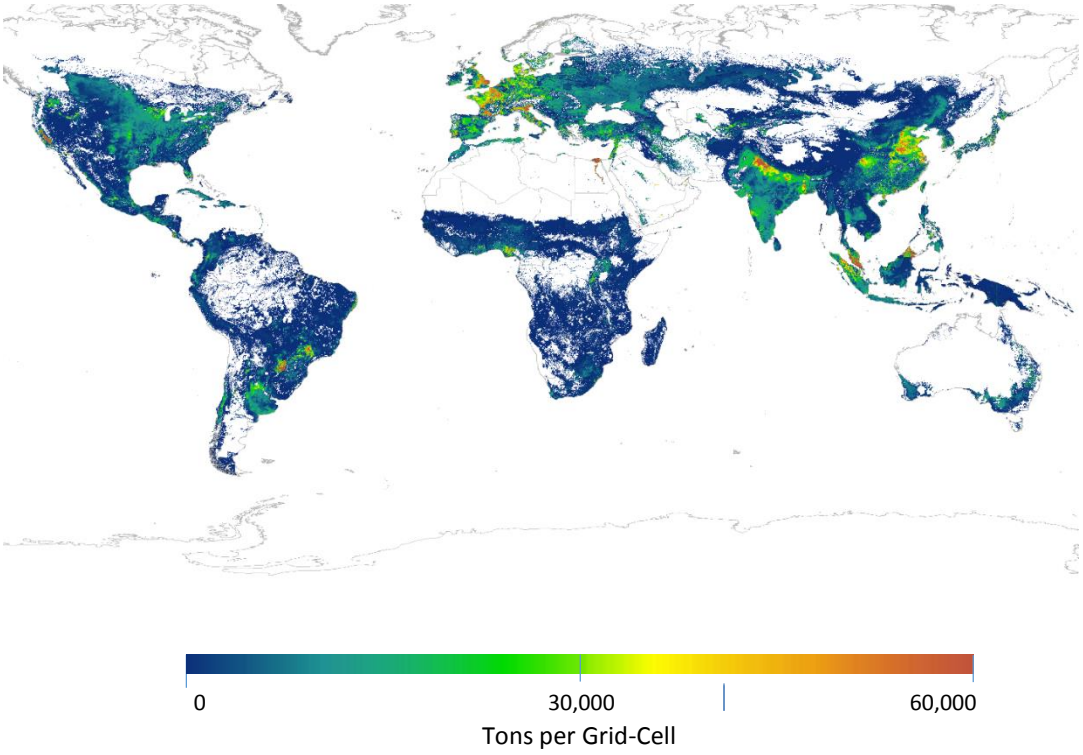


Figure A1.7: Carbon Storage in Potential Natural Vegetation (tons)

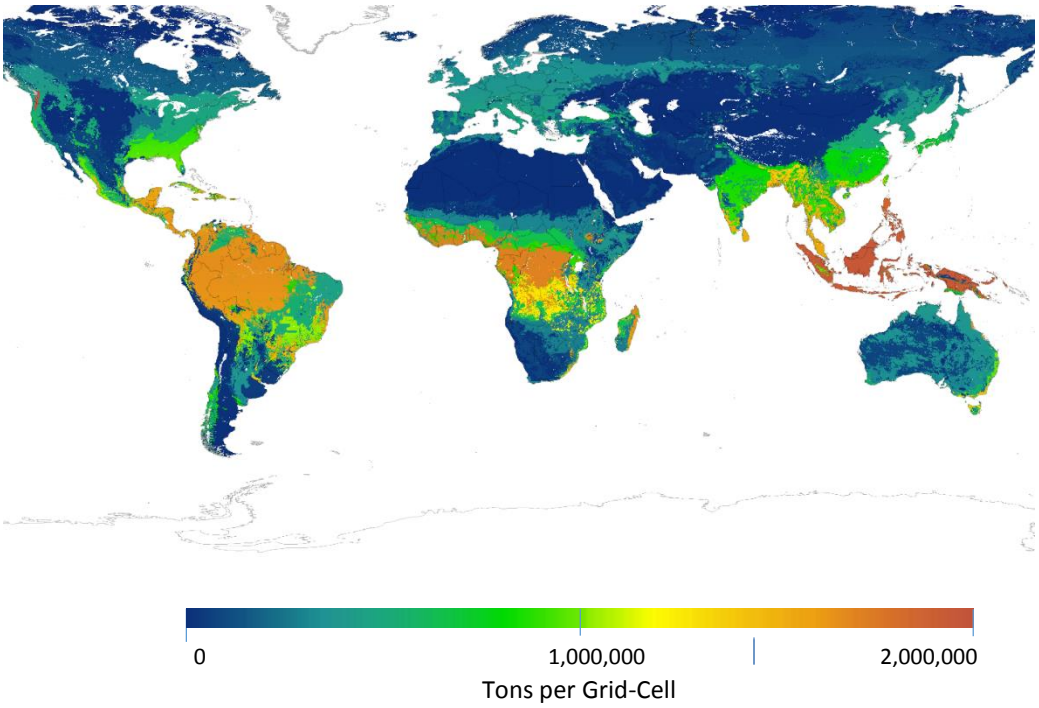


Figure A1.8: Carbon Storage in Soils (tons per Ha)

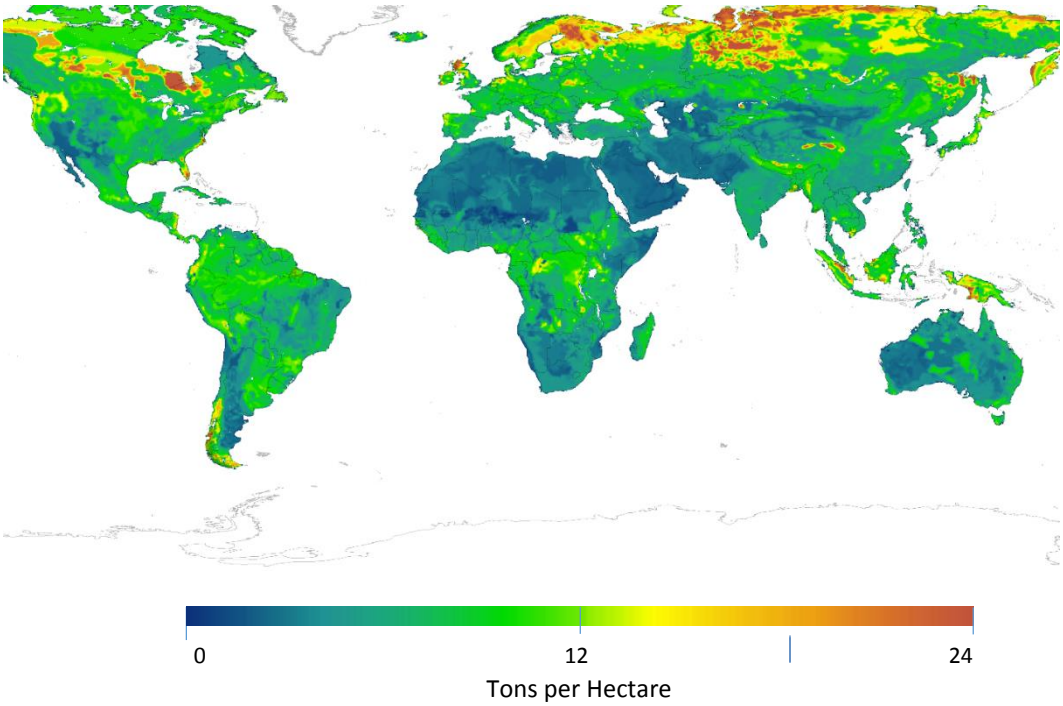


Figure A1.9: Marginal Effect of Extensification on Carbon Storage (tons C per grid-cell).

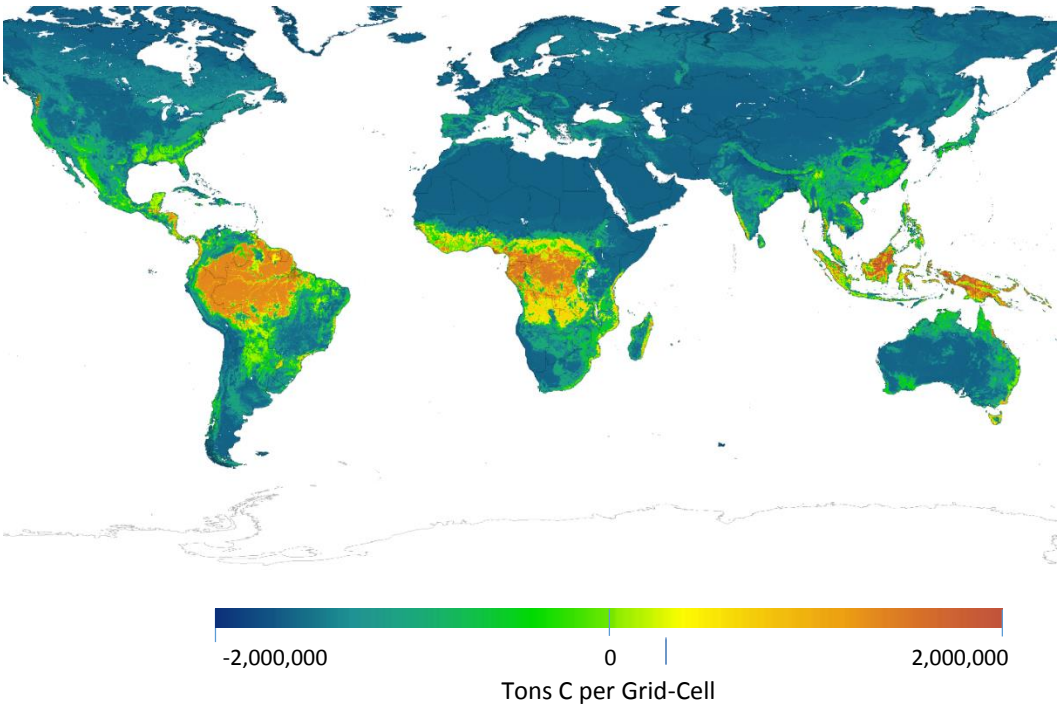


Figure A1.10: Crop Advantage (caloric yield per ton carbon storage loss).

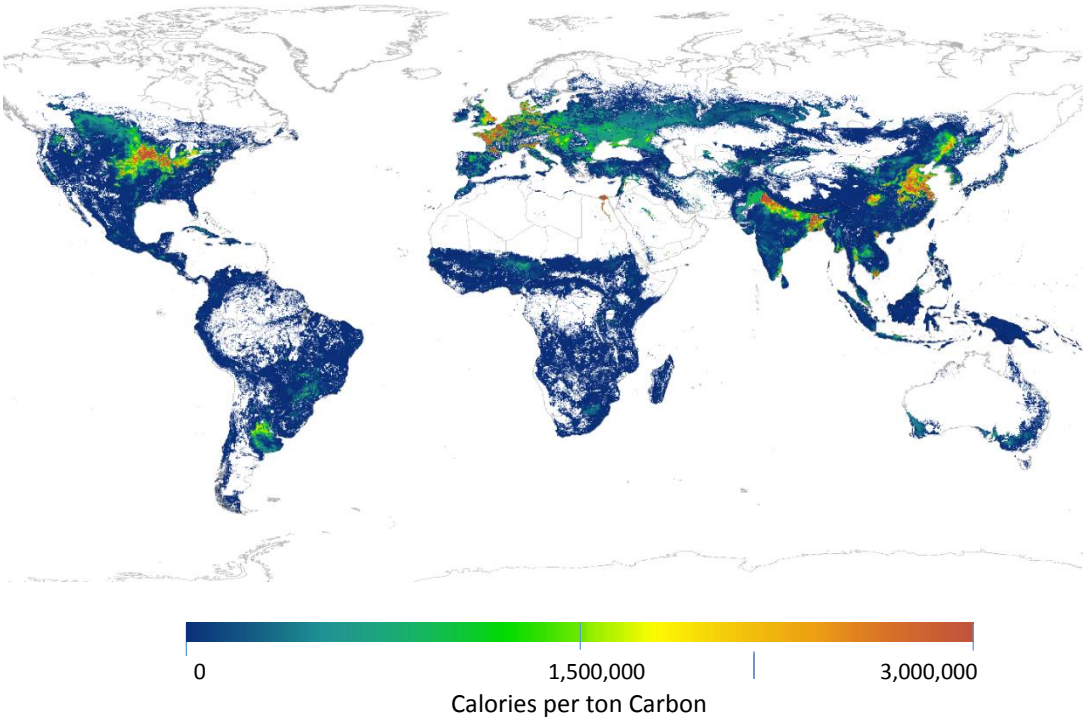


Figure A1.11: Change in Caloric Yield (post Optimization Simulation).

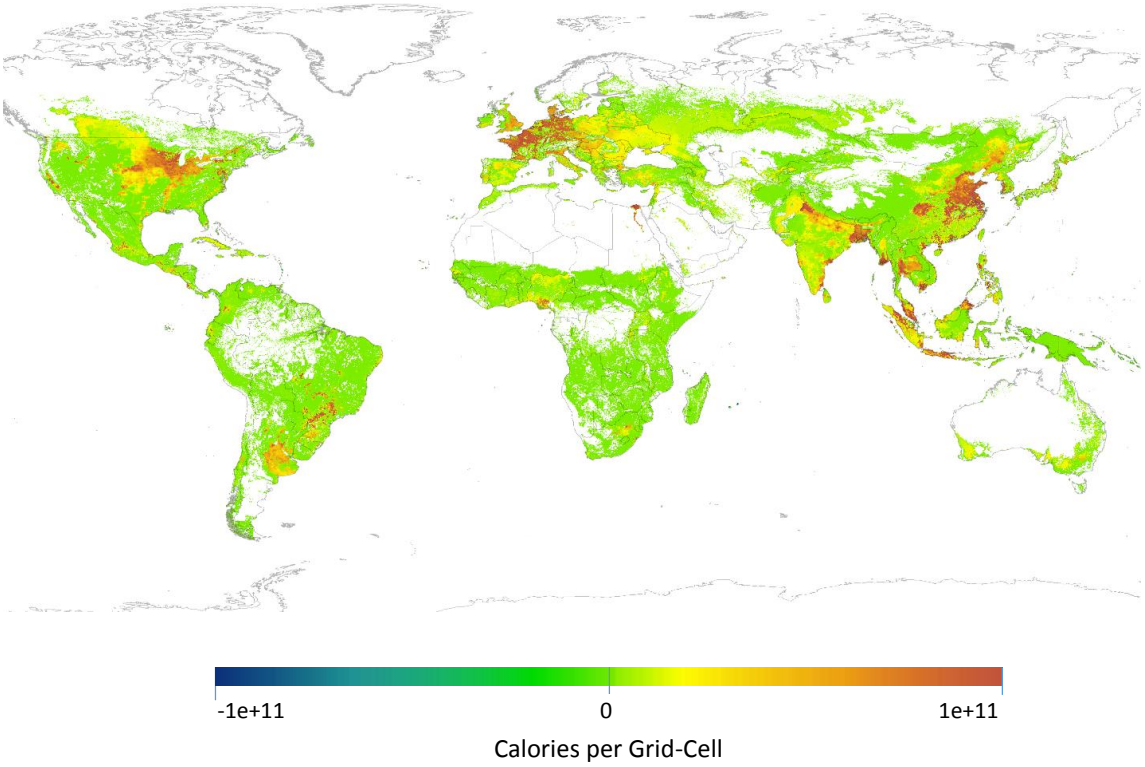


Figure A1.12: Change in Caloric Yield (post BAU Simulation).

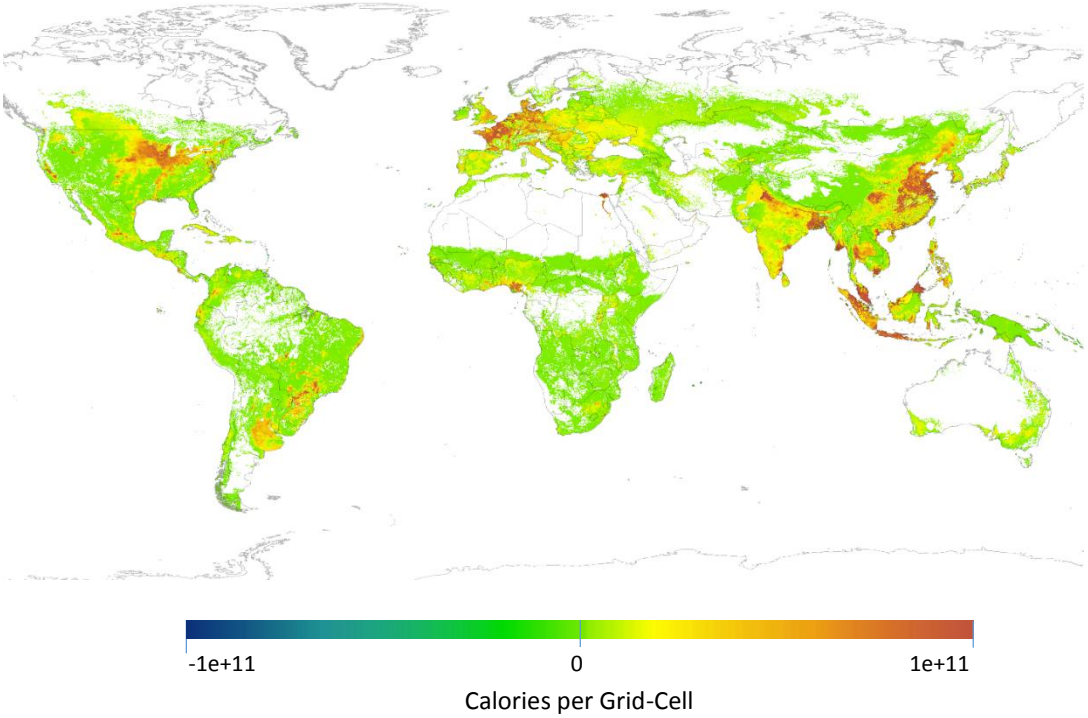


Figure A1.13: Net Land Protected (Proportion Harvested in BAU less Proportion Harvested in the Optimal Simulation).

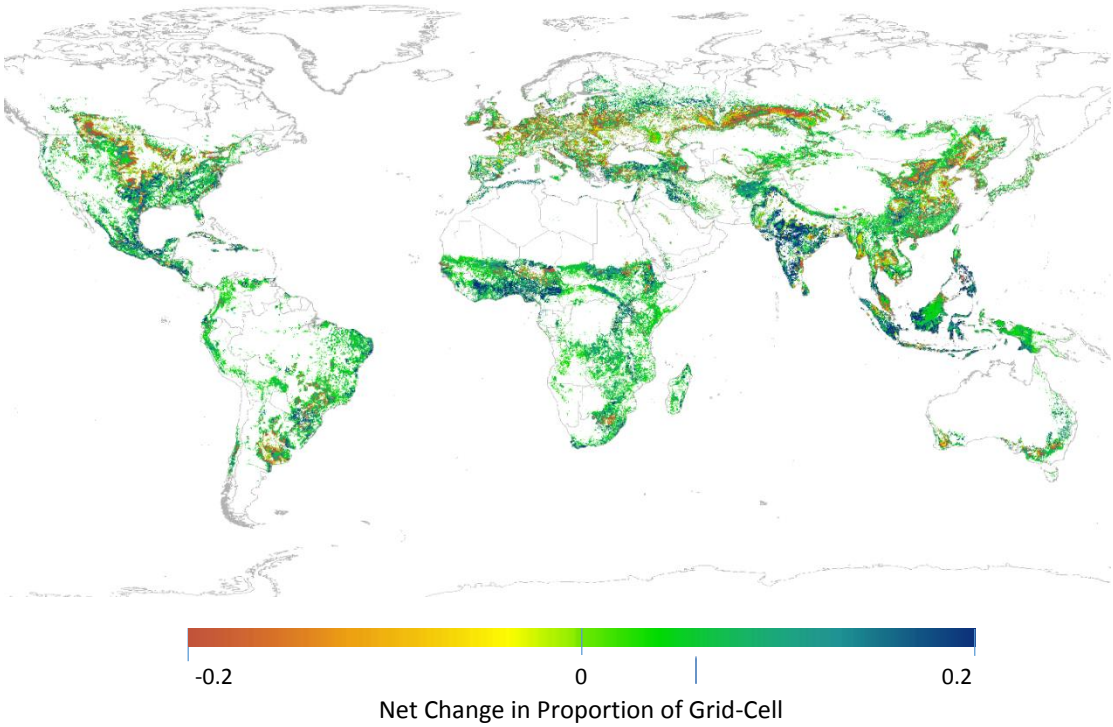
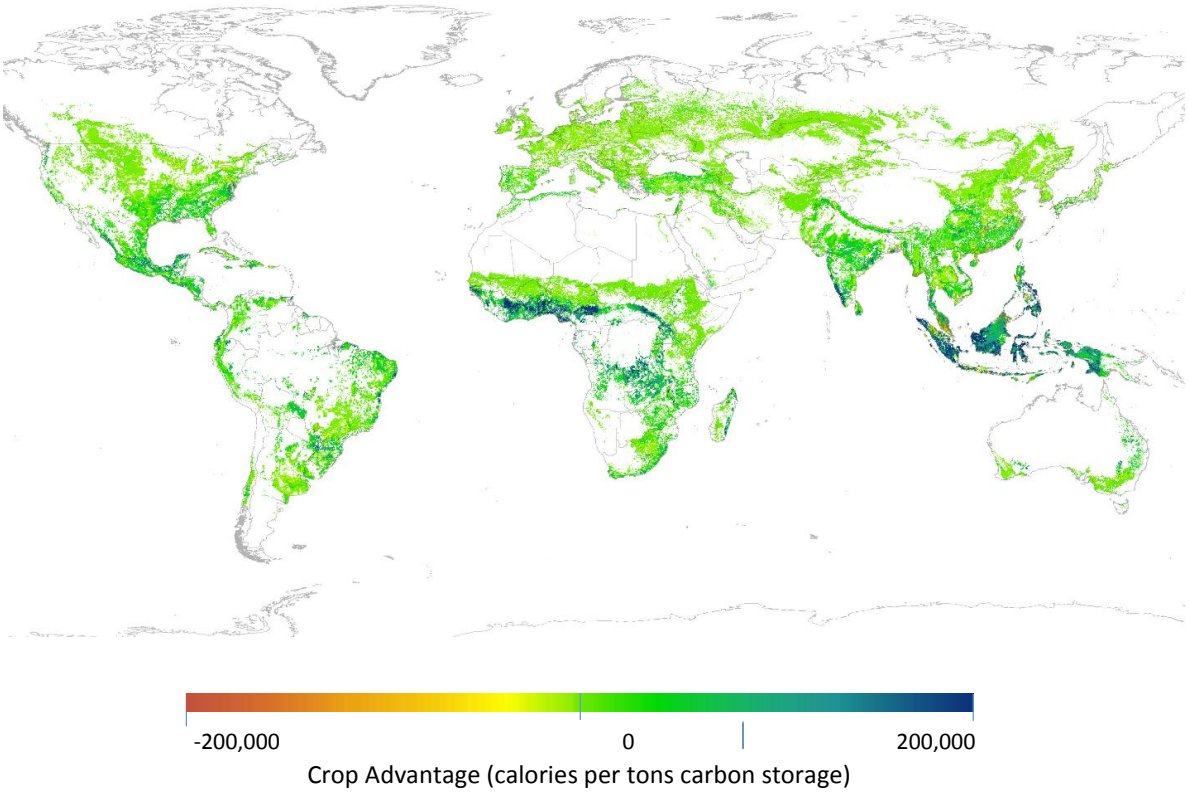


Figure A1.14: Net Carbon Storage Change (tons per grid-cell).



Sensitivity Analysis and Additional Results

Our base scenario, defined as producing 100% more calories with 75% of the increase coming from intensification, may underestimate the challenges we face, especially with regard to lower yield increases in the future. We discuss results from a second, more pessimistic, scenario than in the text. Detailed results for this scenario, in which only 50% of the increase is met with intensification, are presented here. This section also provides more detail on the optimization technique with extended documentation on each step of the calculation of the pessimistic scenario.

Our analysis depends on six parameter assumptions, discussed in the text and presented as column labels in table A1.1. The 20 sensitivity analyses presented in this table are full optimization runs with different parameters. Scenarios 1 and 2 correspond to the base

scenario and pessimistic scenario discussed in the text, while scenarios 3 – 20 are representative modulations of the input parameters. Many additional runs have been completed to test sensitivity but are excluded here. The rightmost column presents the carbon saved by optimizing under each of these assumptions sets. None of the sensitivity analyses depart substantially from the base scenario and nearly all of the variance can be explained by different assumptions on how much future intensification will contribute to meeting calorie demand.

Table A1.1: Sensitivity Analysis Results

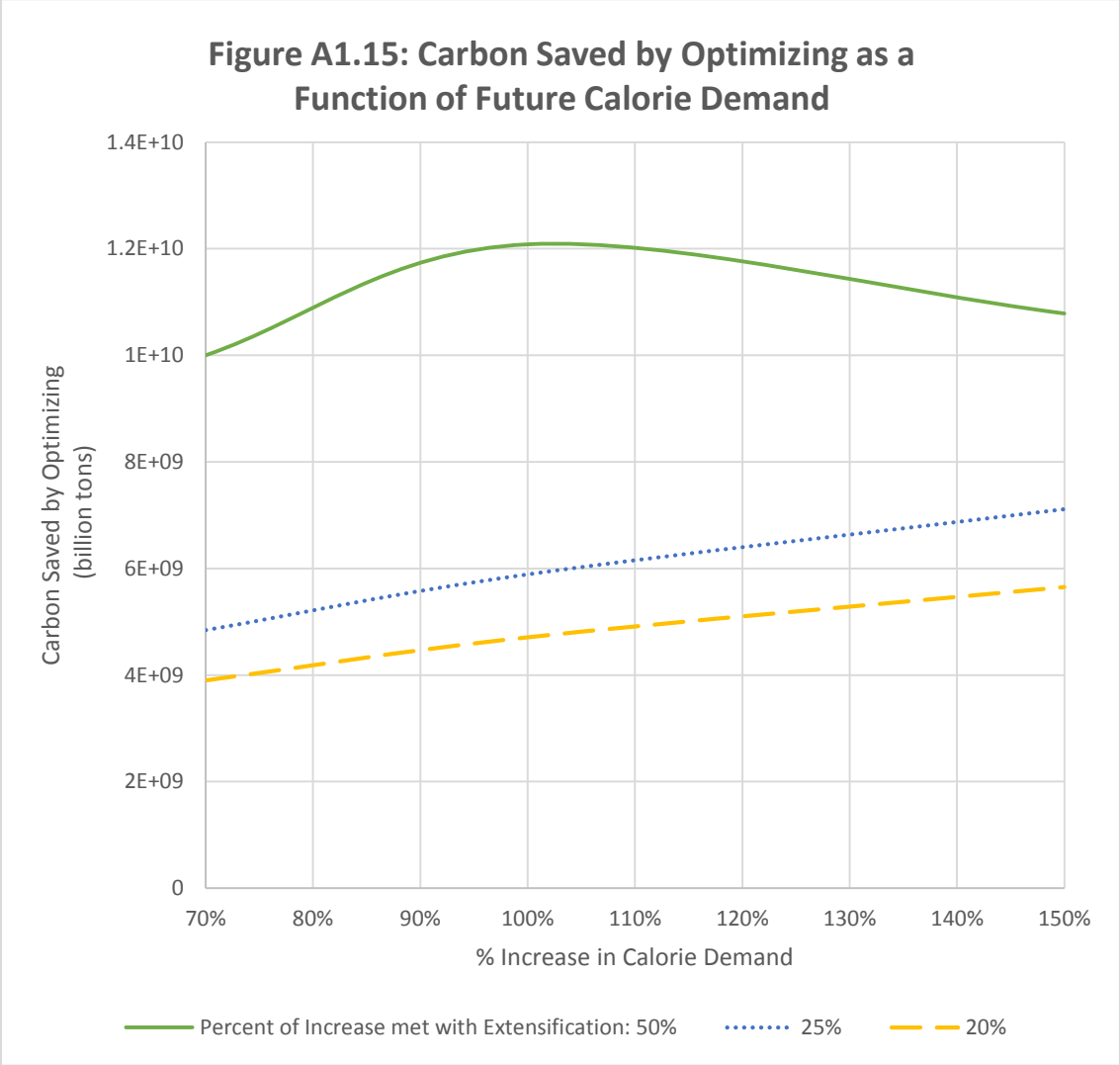
Sensitivity Analysis Scenario	*Percent Increase in Calories*	*Percent met with Intensification*	*Min Extensification Constraint*	*Max Extensification Constraint*	*Constrained Growth Factor*	*Unconstrained Growth Factor*	*Carbon Saved by Optimizing (billion tons)*
1	100%	75%	5%	95%	75%	4	5.89
2	100%	50%	5%	95%	75%	4	12.08
3	100%	80%	5%	95%	75%	4	4.71
4	150%	75%	5%	95%	75%	4	7.11
*5**	150%	50%	5%	95%	75%	4	10.79
6	150%	80%	5%	95%	75%	4	5.65
7	70%	75%	5%	95%	75%	4	4.84
8	70%	50%	5%	95%	75%	4	10.00
9	70%	80%	5%	95%	75%	4	3.90
10	100%	75%	5%	95%	50%	4	5.61
11	100%	75%	5%	95%	65%	4	5.82
12	100%	75%	5%	95%	85%	4	5.93
13	100%	75%	5%	95%	100%	4	5.97
14	100%	75%	1%	99%	75%	4	6.04
15	100%	75%	0%	100%	75%	4	6.05
16	100%	75%	5%	95%	75%	3	5.82
17	100%	75%	5%	95%	75%	5	5.91
*18**	100%	75%	5%	95%	75%	2	5.04
19	100%	75%	5%	95%	75%	6	5.92
20	100%	75%	5%	95%	75%	9999**	5.94

* Indicates a scenario in which the increased calorie demand could not be met.

** Chosen as arbitrarily high value to test maximal unconstrained growth.

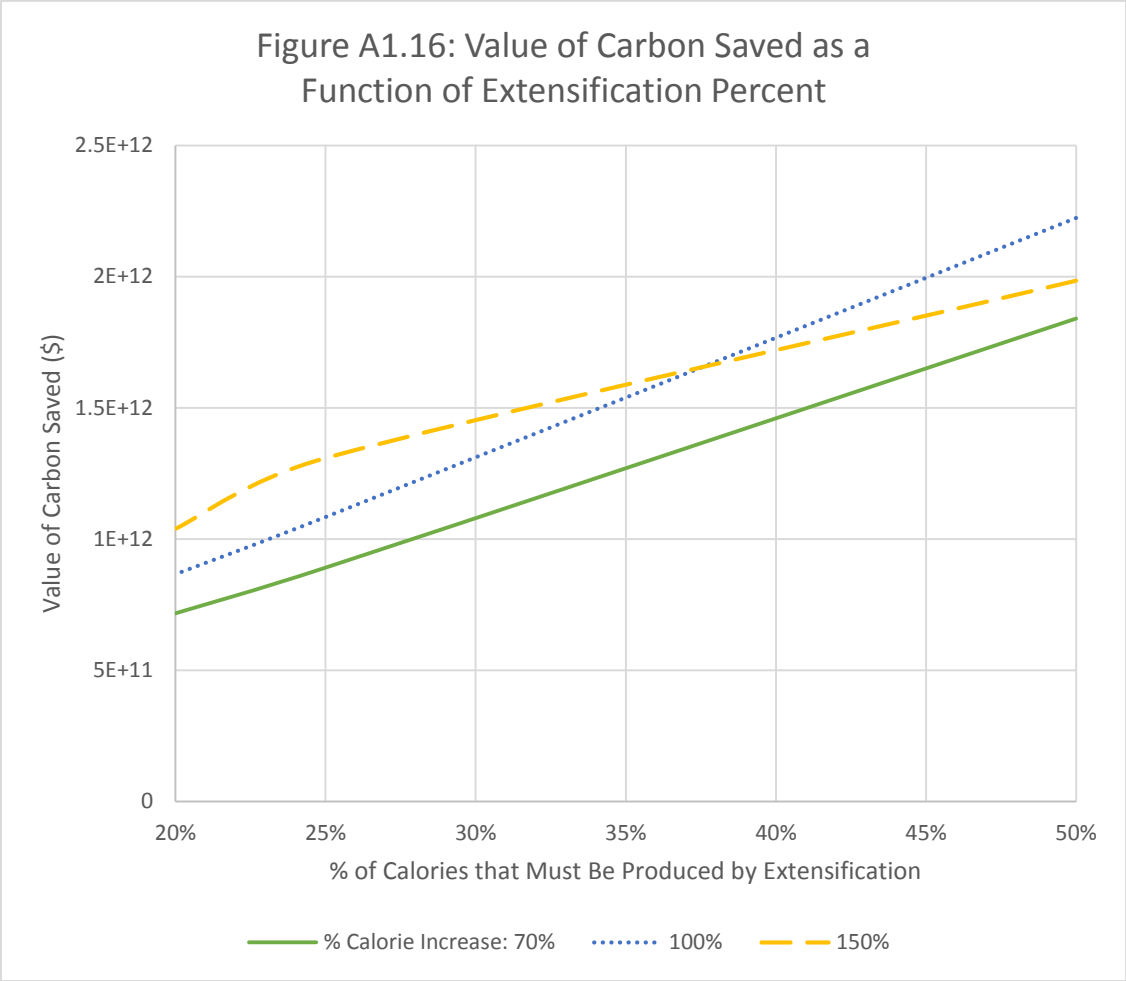
Additional Analysis and Discussion

The amount of carbon storage saved by optimizing depends primarily on how many more calories we must produce and how much of that increase must be met through extensification. Figs. A1.15 and A1.16 show these two effects, respectively. In Fig. A1.15, we solved for global carbon storage saved under the optimal solution versus BAU for a range of future calorie demands (70% more to 150% more). In most cases, higher calorie demand implied more carbon could be saved by optimizing. In the most pessimistic cases, such as meeting more than a 100% increase with only 50% coming from intensification, we see that the value of optimizing declines. This happens because in the most pessimistic scenarios, meeting future food demand becomes so challenging (and impossible in some cases) that even the optimal simulation must extensify on nearly every grid-cell. At the limit of feasible scenarios, the optimal solution converges to the BAU solution because there are no remaining alternate choices.



In Fig. A1.16, we show how the value of carbon saved depends on the amount of extensification required (using our central estimate for the social cost of carbon, $181/ton). Although we centered our analysis around the assumption of requiring 25% of calories to come from extensification, recent reductions in the rate at which yields are increasing suggests future agriculture may not be able to depend as heavily on yield

gains. This increases the importance of optimizing our extensification choices, as shown in A1.2.



Our estimates of how much carbon is lost in different scenarios considers only carbon lost from extensification-related land use change. Other sources of carbon storage loss and other drivers of land use change are excluded from our analysis. These exclusions partially explain why the values of carbon saved in our estimates are two orders of magnitude less than total carbon storage. In efficiency terms (calories produces per carbon storage lost) BAU is 15.5 times worse than the optimal scenario (see table A1.2 for supporting statistics). This suggests that making informed extensification decisions is important.

Our analysis depends on the accuracy of the EarthStat data set. There are considerable challenges in properly determining grid-cell cultivation statistics (particularly yield) and limitations arising from data interpolation. These concerns are offset, we believe, by the additional insights that can be gained from the higher resolution of the data.

Other values of natural ecosystems, such as the protection of biodiversity, are excluded from our estimates. Thus, the valuation we place on optimizing our extensification choices is likely a dramatic underestimate of the true value lost. Nonetheless, our optimization approach provides evidence on the importance of protecting tropical ecosystems. Even if our specific values are underestimates, the policy precepts and advice support the general conclusion of protecting areas rich in all types of ecosystem services.

Table A1.2: Descriptive and Model Statistics

Global Carbon Storage, Current (tons, aggregated IPCC Tier 1 estimates by grid-cell)	5.008e+11
Global Carbon Storage in Potential Vegetation (tons)	7.389e+11
Total Carbon Storage Change under BAU (tons)	-6.295e+09
Total Carbon Change under Optimizing (tons)	-4.051e+08
Number of cultivated cells in dataset	9.209e+05
How much more carbon is lost under BAU than optimal? (tons)	5.890e+9
How many times more efficient (in terms of carbon lost per calorie produced) is the optimal scenario compared to BAU?	15.54

Heterogeneity within Grid-Cells

Although our data are higher-resolution than national-, state- or county-level data, there still exists considerable heterogeneity within each grid-cell. To make global analysis computationally possible, each variable is represented with a single number for each grid-cell. This masks the heterogeneity that exists at the sub-grid-cell level. Fig. A1.17 (right) illustrates this point by comparing satellite imagery with a plot of CA at varying altitudes.

The image on the right, for example, shows one particular grid cell, which includes areas with very different extensification possibilities. Location one, for instance, may be quite suitable for extensification due to the relatively large extent of the undeveloped land. Area two is likely more expensive to bring into cultivation due to the difficulties of cultivating along the banks of a stream. Area three would be extremely expensive to cultivate because of the high-value of land in different uses (commercial and residential in this case). This type of sub-grid-cell heterogeneity will affect how much extensification is possible within the grid-cell. Due to computational difficulties of analyzing this higher-resolution, our model implicitly assumes that each grid-cell is homogeneous, but we apply constraints on how much extensification can happen within a grid-cell to reflect the heterogeneity. The assumptions are discussed in more depth in the text, but the most important assumption is that the maximum extensification within a grid-cell is a function of how much is currently cultivated. This captures, albeit in an ad hoc way, the possibility that grid-cells are physically unable to extensify past a certain point and that this effect is partially captured by current rates of extensification. Future work will address this point more directly by utilizing downscaled data applied to smaller regions of the globe.

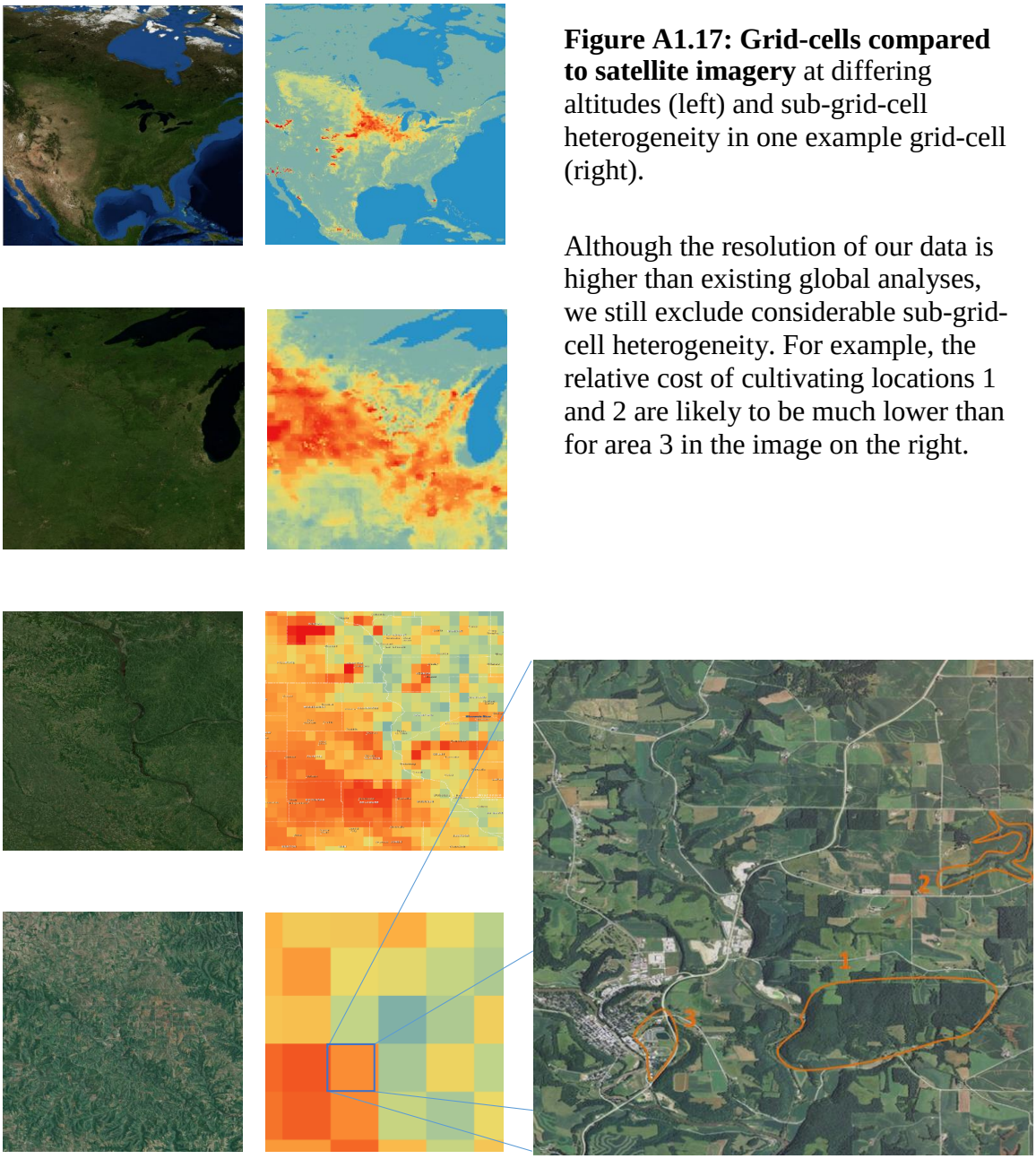


Figure A1.17: Grid-cells compared to satellite imagery at differing altitudes (left) and sub-grid-cell heterogeneity in one example grid-cell (right).

Although the resolution of our data is higher than existing global analyses, we still exclude considerable sub-grid-cell heterogeneity. For example, the relative cost of cultivating locations 1 and 2 are likely to be much lower than for area 3 in the image on the right.

Appendix 2: Supplemental Information for Chapter 2

This appendix provides detail on the model, results and validation of Chapter 2. The first section describes the notation I used to define the agent-based simulation, both in grid-cell and non-grid-cell network arrangements. The notation is then used to restate the model described in chapter 2 with more detail. Next, I provide a set of results parallel to those presented in the chapter focused on 250 representative agents rather than 10 million. This example is useful because many simulations may be done at the village level rather than the individual level with approximately this many villages as the agents. Moreover, this section presents additional methodological concerns that arise when using representative villages, such as how to identify the optimal zone definition or the optimal number of agents. Third, I provide more details on the specific algorithms and methods I used in computing the simulation in chapter 2, including a full discussion of the nested hierarchy of route-finding algorithms I used to balance accuracy with calculation speed. Finally, I provide additional information on how I attempted to validate this model.

Notation and Model Details

Describing a model that combines agent-based simulation with economics in a spatially heterogeneous environment requires new notation and data concepts. This section presents that notation and applies it to firewood foraging behavior. The result of this section is a comprehensive framework that can be applied to other geospatially specific economic phenomena. In a departure from traditional economic techniques, I do not use the concepts of equilibrium or optimization to define how agents act. Instead, I derive behavior rules *from* traditional economics and use those rules in a simulation to iteratively reach an outcome. This section is admittedly dense and rather long. However, I err towards comprehensive inclusion of notation so to allow for greater generalization. It is important to keep in mind that the technical details here are critically important to ensure clarity but once clarified can be safely deemphasized in favor of the broader and more intuitive results that emerge from this logical construct (see the results section for an example that shows the intuitiveness of results despite the complexity of notation defined here).

The notation I define will include:

1. Cell Network
2. Normalization Method
3. The Agent

4. Behavior Rules
5. Time and Simulation Steps

I begin this methods section by describing the cell network and a specific instantiation of it useful to spatial topics, the geospatial grid-cell network.

The Cell Network

The cell network, and many of the subsequent data types I define, require introducing three "primitive" data types I use: points, zones, and networks. These types are not interesting in themselves, but by carefully defining their usage, subsequent notation can be shortened.

Point and Correspondence

Points are zero-dimensional references defined in continuous space. All points (and subsequent data types defined below) are defined either with a Cartesian scheme tied to latitude and longitude, which I denote with the subscripts x and y, or in matrix notation, which I define with row, column notation with subscripts r and c respectively. For notational simplicity, these coordinates may be omitted if the primitive is indexed to another primitive that does have coordinates defined. In this event, the primitive must be in a defined correspondence list (in which I will index each element i). In practical terms, this allows us to connect a primitive to an agent-indexed variable, (such as labor expenditures, L_i^g in future sections) via an agent-point correspondence between the i-th agent and the x, y or r, c location: $i = (r, c)$ or $i = (x, y)$. Though seemingly trivial, the advantage of this framing includes that full precision may be maintained even when a geospatial transform is applied (such as gridded interpolation), vector-based mathematics and algorithms may be used,[9] points are a required to define full-precision shapes at higher dimensions, and points are the simplest way to combine non-spatially defined agent data (such as survey-based cross-sectional demographic data) with spatially defined agent locations.

[9] The important part of this aspect is that it maintains access to the wide array of geometric algorithms documented in computer science. This is a technical hurdle necessary for maintaining computational feasibility at the scope and resolution necessary to analyze firewood collection in Tanzania.

Zone

Zones are ordered sets of points where each point forms an edge with its two neighbors within the wrapped set. For instance, the zone $Z = \{p_1, p_2, p_3, p_4\}$ implies p_1 forms edges with p_4 and p_2 while p_2 forms edges with p_1 and p_3. Each zone must also have a defined reference point, p^r, that is used to pair the zone with other data constructs (similar to why point correspondences were used above). The reference point, if left undefined, is assumed to be the geographic centroid of the zone. An example of a reference point that might be better defined explicitly is representing the population of the state of Minnesota. If our interest in defining Minnesota is to understand how it acts as an agent in a simulation, it is nonsensical to assume the geographic center of the state is the best representative point of the state's overall economic activity. Rather it should be defined somewhere near the population centers of Minneapolis and St. Paul rather. If processing power was unlimited, the better approach would be to define smaller political units each with population figures, but in many applications it is important to aggregate.

Network

A network is a set of nodes n and corresponding link lists. The node is defined as a zone and reference point pair while the link list contains the set of other nodes to which the given node is connected and the attributes of the link. I define a network with the following notation:

$$N = \{(n_{1,1}, l_{1,1}), (n_{2,1}, l_{2,1}), \dots, (n_{R,C}, l_{R,C})\}$$

where

$$n_{r,c} = Z_{r,c}, p^r_{r,c}$$

and

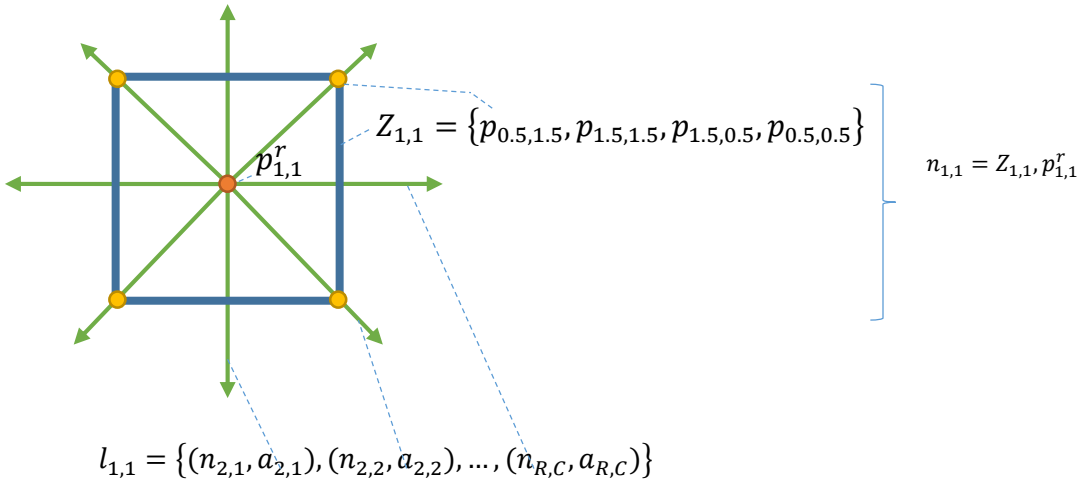
$$l_{1,1} = \{(n_{2,1}, a_{2,1}), (n_{2,2}, a_{2,2}), \dots, (n_{R,C}, a_{R,C})\}$$

In this definition, a link list is composed of all the other nodes to which the current node is linked paired with the attribute value of that link, $n_{2,1}$ and $a_{2,1}$ respectively. Networks allow for description of interactions between nodes. For instance, a network may describe bilateral trade relationships, travel time between locations, social networks or supply

chains. The advantage of using full network notation, as described above, is that it provides a comprehensive data frame that can contain all elements that comprise the network, rather than case-specific representations more frequently used (such as adjacency matrices, trade partner lists, or flow maps, which are case-specific subsets of a full network description).

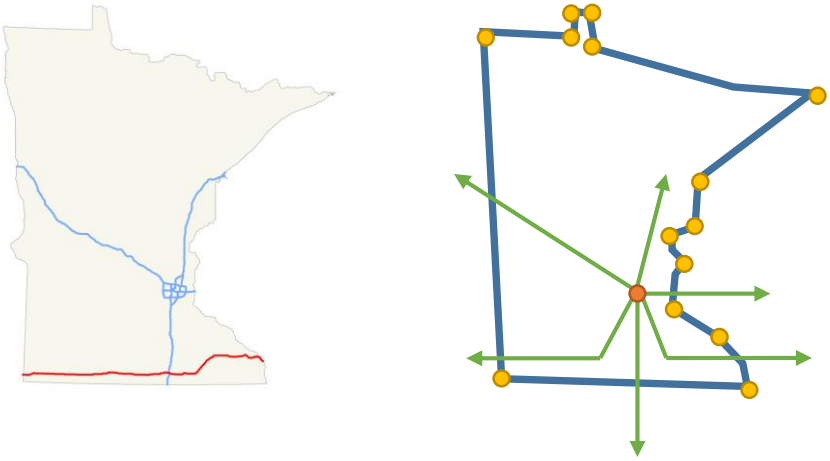
Combined, the point, zone, network data primitives create the first basic unit of analysis I use for geospatial economic simulations, the cell, and is shown in Fig. 2.1 with a summary of the notation used to describe the cell.

Figure A2.1: Example Cell with Notation



The cell above is only a simple instance of this basic unit type. The example in Fig. 4.2 below shows the same framework applied to the State of Minnesota, using population density to define the reference point and the interstate highway system to define the network.

Figure A2.2: Example of an Agent Applied to Minnesota with Interstate Highway Network Links

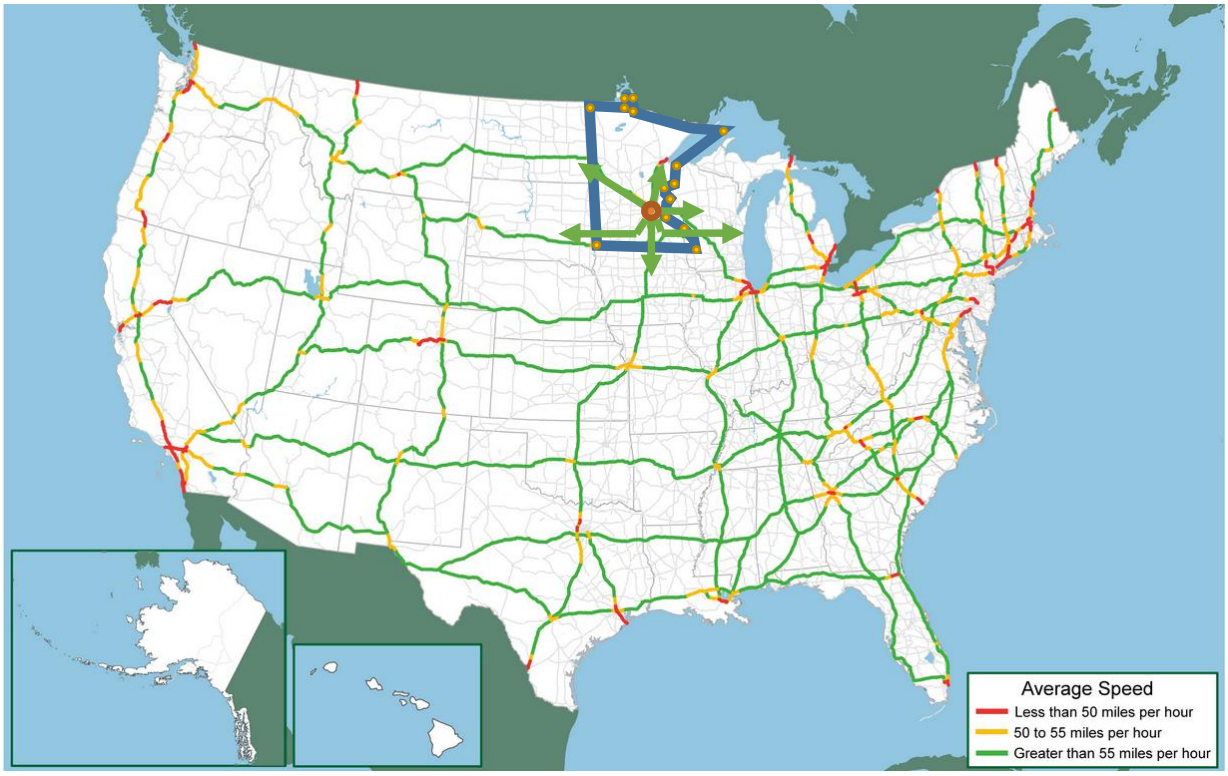


I further define a set of such cells as a geospatial cell network if it satisfies the following assumptions:

1. Each cell zone is polygon defined by the non-reference points in the zone definition.
2. No cell zones overlap
3. No area in the convex hull of all cell zone points is not part of one cell zone

Intuitively, a geospatial cell network is a set of cells that are correctly linked and correctly cover a geographic space. Fig. A2.3 shows an example of a geospatial cell network by placing the cell for Minnesota define above on a map of the United States along with cells for the other 47 contiguous states. Choosing how to define the grid-cells is still arbitrary, but is chosen to be relevant to a research question. In this example, the each network link has an attribute for average truck speed, and thus, the geospatial cell network may be particularly relevant to research questions involving inter-state transport of goods.

Figure A2.3: Example Geospatial Cell Network using States as Zones and Interstate Highways as Links. Source: U.S Department of Transportation, Federal Highway Administration, Office of Freight Management and Operation, Freight Performance



Normalization Method

The network defined above in Fig. A2.3 defines a geographic space on which agents can operate and uses non-approximated vector based definitions. Simulating agent actions on precise geographic spaces is possible when there are fewer than 1,000 cells, but becomes computationally impossible given current technology with many more cells. In this section, I define a method of normalizing the cell structure defined above that allows for computation of many more agents, including as many as 10 billion agents, without losing

theoretical generality.[10] The simulation methodology I use requires having a valid normalization method, M, so that all input networks and agent sets may be converted to a data representation that works across types while maintaining computational speed, transparency and replicability.

To normalize a cell network, I convert the vector-based cell described above to approximated grid-cells. This requires choosing a tesselatable regular shape (almost always a square in 2 dimensions or a cube in 3 dimensions, but is best done with hexagons or triangles in a small set of situations), a uniform size of this shape (for instance, the edge length of the square) and a row, column schema for defining the location of all grid-cells. Finally, the schema requires a geographic extent defined so that we know what rectangle of space on the globe our matrix represents (see Chang 2010 for a description of how geographic information systems, GIS, approach this problem).

Assuming we chose squares as our shape, we can define this schema as a *grid-cell matrix*, where each element in the matrix is a single valued number that describes a particular attribute of the cell network and where the row, column index of the matrix is sufficient to identify the location and extent of each grid-cell, both in geographic[11] space and network space. Grid-cell matrix, as defined here, also go by the aliases raster, GeoTIFF,[12] or map, however each of these terms connotes a slightly different data type an application, so I use the term matrix throughout this text.

Four different types of matrices emerge from the process of gridding the cell network, described below. These types are able to describe the cell network in matrix form, and thus enable highly efficient computation methods (such as use of NumPy and SciPy; see Olophant 2006).

1. A **zone identity matrix** stores the cell reference identity (from the $(i = r, c)$ correspondence described above) in the grid-cell matrix of whichever cell covers the most of the grid-cell in question.
2. A **point identity matrix** is similar to a zone identity matrix, but only stores the cell reference if the zone's reference point is in the grid-cell. If there is no

[10] 10 billion is a useful number because it is near projections of maximum human populations in the 22nd century.
[11] Throughout this text, I have used the word geographic to refer to any location-designation system that corresponds to 2- or 3-dimensional Euclidean space.
[12] GeoTIFF is the name of the file type used in many GIS applications to represent gridded data.

reference point within the cell, the grid-cell is valued at zero. Thus, if a zone covers more than one grid-cell, the identity reference will not be repeated in each zone cell as it was for the zone matrix.

3. An **attribute matrix** defines the value that each grid-cell takes for a particular variable. Because the row, column index of each attribute corresponds to the row, column index of zone and point identities in the same location, we implicitly know to which cell the attribute applies without the need to directly specify it.

A **network matrix** stores the information on network links between each cell. Because the links may overlap, skip cells, and are in the intersection *between* grid-cells, it requires a level of abstraction to describe a network in matrix form. To do this, I use the zone matrix created in step one rather than the unnormalized cell network. Then, I subdivide any link that spans multiple grid-cells into multiple links, each describing a route with start and end grid-cells that are adjacent (or later, diagonal).[13] This means that to travel three grid-cells, one will traverse three link. Finally, I split the cost-of-traversing each grid-cell into the network matrix so that the traversal cost of each link is defined by its starting and ending grid-cells: $t_c = \frac{start\_value_{rc}}{2} + \frac{end\_value_{rc}}{2}$. If the link is diagonal, we must account for the longer distance across the diagonal of a square: $t_c^{diagonal} = \sqrt{2}\left(\frac{start\_value_{rc}}{2} + \frac{end\_value_{rc}}{2}\right)$. The result of this process is a single-valued matrix of the same dimensions as the zone matrix and point matrix that describes the cost of traversing from any location to any other location on the zone and point matrices.

Fig. A2.4 illustrates the process for defining each of the data types described above. The input data for this example is shown in Fig. A2.4.a, which is satellite imagery of approximately 4 square miles of farmland in southeastern Minnesota. Fig. A2.4.b defines ten cell zones that roughly correspond to land use and/or land ownership. This step can be done by using zoning data from government entities, automatic classification algorithms or manual input according to classification rules (these method are well documented in the remote sensing literature, such as Singh 1989 or Marpu et al. 2013).

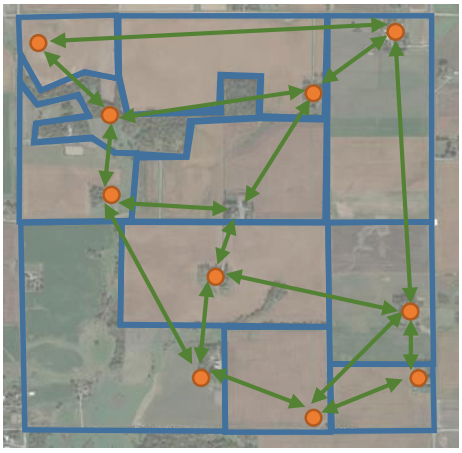
[13] I have decided to focus my exposition on the subset of networks described by square grids. This approach can be criticized for incorrectly taking the square as somehow fundamental to spatial phenomena, but it turns out this may not be a bad assumption given the broad proclivity of humans to arrange things in a gridded fashion. The reasons for the frequency of grid-schemes in human systems is complex and out of the scope of this thesis, but a likely reason is that a square is the only regular shape that can be tessellated without gaps, rotation or off-setting of the shape. The square *is* fundamental.

The zones defined in A2.4.b are not full definitions of cells yet, as they lack reference points and network links. Fig. A2.4.b adds this information based on where within the zone the farm buildings are located and a rough approximation of travel to other locations via nearby roads. Note that although the network links are depicted as straight lines, this is just an abstraction. Each link can be constructed with a travel time attribute on each network link that reflects the shortest *road* route.

Fig. A2.4.c shows reference points for each zone and defines the set of network links between zones. With zones, points and network links in place, this defines a full cell network as described above.

I then normalize the cell network, which results in matrices shown below. Figs. A2.4.d shows the normalized point matrix, A2.4.e shows the normalized zone matrix and A2.4.f shows a network matrix that roughly constructed from the cost of traversing from adjacent cells based on the location of the reference point within the cell and the presence of roads.

Figure A2.4: Process for Defining Cell Network and Normalized Matrices



1	4	4	8
2	5	5	8
3	6	6	9
3	3	7	10

1		4	8
2	5		
		6	9
	3	7	10

1	2	2	1
1	3	2	2
3	2	3	1
2	1	1	1

The process of normalizing the cell network shown above maintains all of the detail in original network but now allows for much faster computation. Choosing the correct zones (step 2 in the figure above) depends on what data are being analyzed and what type of behavior is being simulated. A further simplification of the geospatial cell network described above allows us to bypass the arbitrary choices of how to define zones.

To summarize, I have applied the notation defined above to a network construct and a normalization method, such that for Fig 2.5.d, in which we have 16 cells defined, we have reference points $p^r_{1,1}, p^r_{2,1}, \dots, p^r_{4,4}$, zones $Z_{1,1} = \{p_1, p_2, p_3, p_4\}, Z_{2,1} = \{p_1, p_2, p_3, p_4\}, \dots, Z_{4,4} = \{p_1, p_2, p_3, p_4\}$ and a cell network:

$$N = \left\{\left(p^r_{1,1}, l_{1,1}\right), \left(p^r_{2,1}, l_{2,1}\right), \dots, \left(p^r_{4,4}, l_{4,4}\right)\right\}$$

$$l_{r,c} = \left\{\left(p^r_{r-1,c}, a_{r-1,c}\right), \left(p^r_{r,c+1,}, a_{r,c+1}\right), \left(p^r_{r+1,c,}, a_{r+1,c}\right), \left(p^r_{r,c-1,}, a_{r,c-1}\right) \mid p^r_{r,c} \neq \{\emptyset\}\right\}$$

$$\forall\, 0 \leq r \leq R, 0 \leq c \leq C$$

Gridded data are frequently used by physical scientists and a broad literature exists on how to create matrices to represent spatial data. Scientists less frequently use gridded data to understand human agents and their behavior, which limits the ways that physical science can make policy-relevant recommendations. Economists, conversely, rarely use gridded data, and instead use tabular data to represent cross-sectional or time-series data on individuals or aggregated agents and their behavior. When gridded data are used in economics, it is often just as an input to be converted to tabular data. By shedding spatially precise information, tabular data solves the computational problems that arise when 1,000 or more agents are considered, but this is at the expense of losing geospatial detail. The method I present here defines a data type that can define both physical phenomena and agent behavior in a unified framework along with a normalization method that allows for broad computational feasibility without shedding either agent or spatial information.

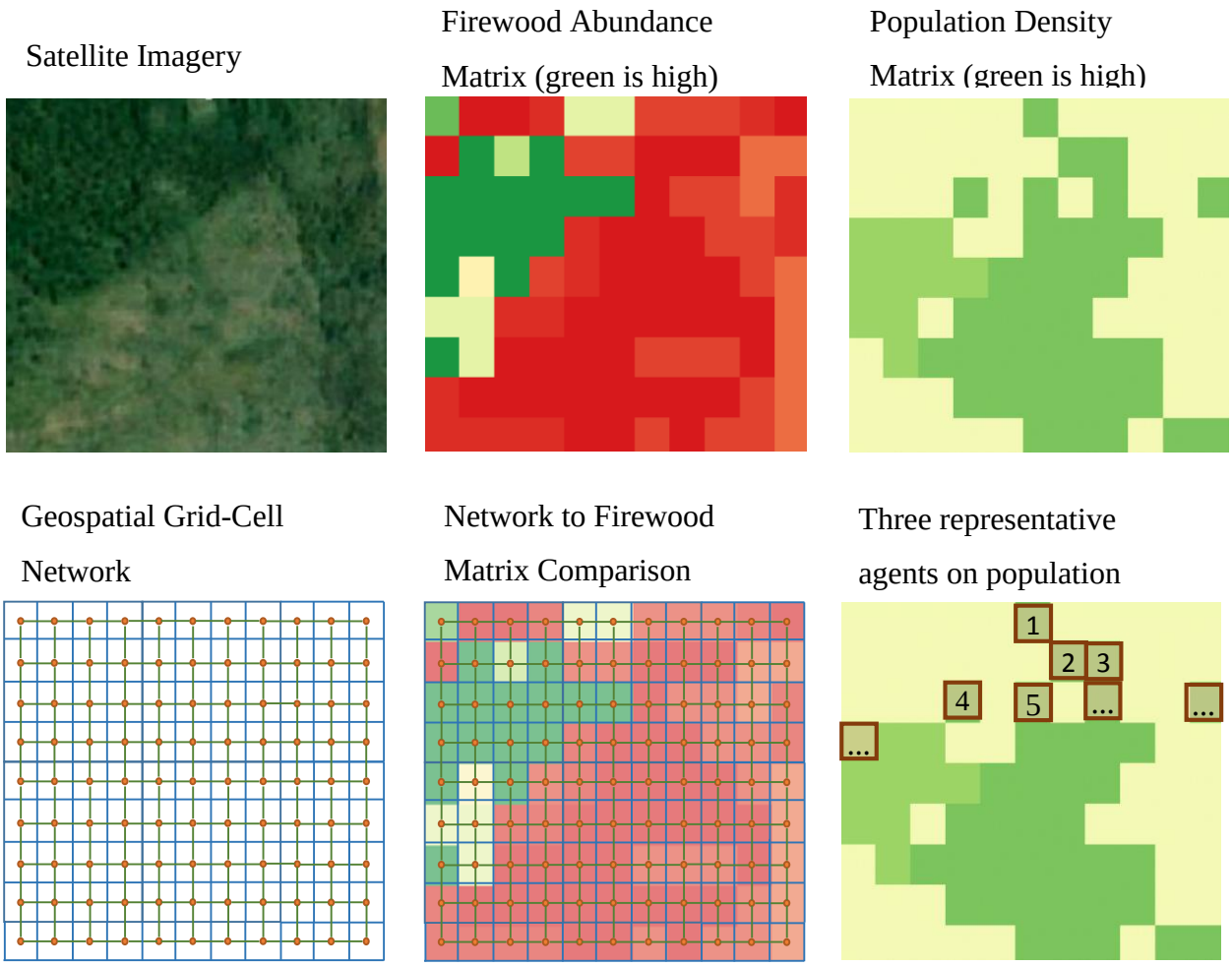
The cell as defined here differs from the other basic entity in my modeling approach, the agent, in how the entities interact over time. Thus before defining the agent, I first use define with more specificity how my modeling approach handles time.

Additional attributes can be added to define an agent specific to a particular simulation task. Figure 4.7 below shows the process for defining our cell network and agents based on a small forest plot in Eastern Tanzania (for now I will exclude stating variable

definitions, units and colorbars to focus on the method used. See the Data section for details on the data used). Fig. A2.7.a shows satellite imagery of the location we will use, which is a mixture of secondary forest and mosaic cropland with shrubland. Fig. A2.5.b is a matrix depicting how much firewood is available to be foraged on each grid-cell based on the underlying land cover. Fig. A2.5.c is the number of persons whose homes are on each grid-cell. Fig. A2.5.d shows a 9 row, 11 column spatial grid-cell network. This network was constructed so that there is exactly one grid-cell defined for each matrix element and so that the area of a grid-cell is equal to the area implied by the matrix. Fig. A2.5.e shows how the grid-cell network corresponds to the firewood abundance input matrix to illustrate how the network is constructed to reflect the input data. Each cell in the network has a cell attribute that corresponds to the value in the firewood abundance matrix. Finally, Fig. A2.5.f shows three agents chosen to represent the population per grid-cell input matrix, shown as grid-cell reference points, along with the grid-cell zones. Summing the total population of the input matrix laying in each of the grid-cell zones defines a "population" attribute for each agent. Because the agent represents multiple persons, I refer to it as a representative agent. The process for identifying how many agents is optimal depends on the specific research question being asked and on the size of the input data (which determines computational difficulty). As I will discuss in the data section, the model I ultimately construct is different than the example shown here. Instead of using agents that are representative of data from multiple elements in the input matrix, I construct many more agents so that there is exactly one agent per grid-cell.

Using the agents defined in 4.7.f, we construct an attribute table similar to the previously shown table. One category of attributes that will be included in the table will parameterize utility functions and production functions for each agent. The next section introduces how utility and production are used to define agent behavior in the simulation.

Figure A2.5: Process for Defining Geospatial Grid-Cell Network and Agents based on Tanzanian Data



Defining the Simulation

Using the elements above, I define the simulation as composed of a valid geospatial grid-cell network N and corresponding matrix normalization M, a set of valid agents A defined with attributes in an agent table, a set of behavior rules (max-marginal-gain iterative approach to utility maximization) R, and finally, a structure and order to time periods and iteration steps, T and S respectively.

$$Simulation = \{N, M, A, R, T, S\}$$

Given this definition of simulation, I solve the simulation by iterating through S and T for each agent in A acting on network N, given the normalization M.

Data additional comments

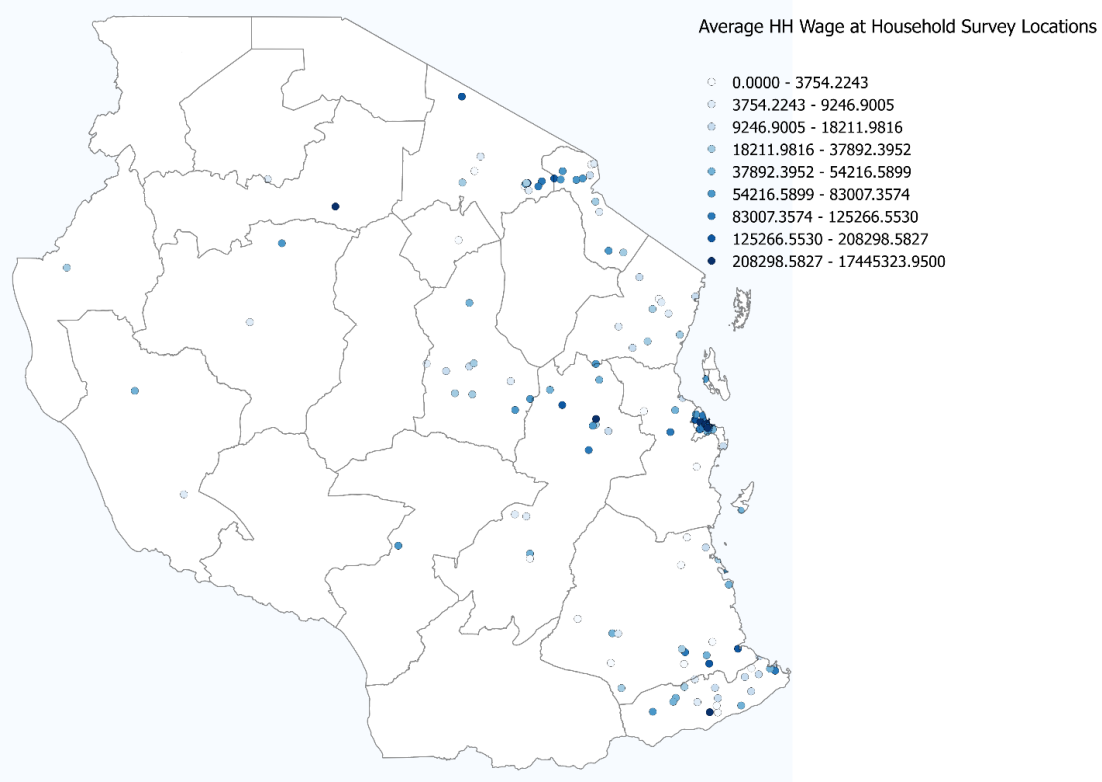
Part of the contribution of this paper is to present methods for systematically and transparently dealing with these data issues and creating a methodology (implemented in a software package) that combines survey data with geospatial data. Replicability is a key motivation here, especially given the many ways that errors and biases can be introduced when spatial and non-spatial data are combined. Existing literature in economics often does not report how aggregation issues such as these are treated, and thus replication is impossible. In this analysis, the main household survey information that was used was from questions on wage generating activities and firewood collection time. Table 4.3 describes these variables and a collection of additional variables that can be used for validation and future extension. To illustrate the spatial aspect of the NPS, Fig. A2.6 presents the per household wage income at each of the enumeration areas using in the NPS.

Table A2.1: Questions from NPS Used in Simulation

E15: How long does it take NAME to get to work from here? [TIME ONE WAY ONLY] (codes = Hour, Day, Week, Fortnight, Month, Quarter, Half year, Year)
E18: How much was NAMES last payment? IF RESPONDENT HAS NOT YET BEEN PAID, ASK: What payment to do you expect? What period of time did this payment cover?
E19: How many hours did NAME work last week?
E21: What is the value of those payments? Over what time interval? (same intervals as E18)
E32: What is the total value of your physical capital stock, including all tools, equipment, buildings, land, vehicles for the business?
E33: What is the total value of your current stock of inputs or supplies?
E34: What is the total value of your current stock of finished merchandise (goods for sale)?
E35 What gross income/takings did you get from your business of businesses last week/month? (Week = 1, month = 2)
E36 What was your net income (profit) from your business or businesses last week/month?
E38: What is/was NAME's total expenditure on wages last month?
E39 What was NAME's total expenditure on raw materials last month?
E42: What was your AVERAGE net income (profit) during the months you operated this business (TSH)?
E47: "How many hours did you spend yesterday collecting firewood (or other fuel materials)?"
Data note: 2 columns, first is for hours, second is for minutes.

J18 Major fuel used for lighting? (codes = Firewood, Paraffin, Electricity, Gas, Charcoal, Animal residual, Gas (biogas), other)
J8 Which is the household main source of cash income? (codes = Sale of food crops, Sale of livestock, Sale of livestock products, Sale of cash crops, Business income, Wages or salaries in cash, Other casual cash earnings, Cash remittances, Fishing, Other)

Figure A2.6: Household Survey Reported Average Wages (TSH)



One difficulty in using these data is that they are answered by one representative in the household for each member [NAME]. Thus, each household may range in the number of

workers from zero to many. This presents a problem when aggregating because opportunity cost must be per individual, not per household. Another challenge with these data is that each enumerate area (EA) comprises many households, but with this release of the data we only have the geolocation for the EA, not the household. Thus, there will be many observations stacked on a single point. As above, this presents interpretation challenges when understanding totaled values vs. per capita values.

The household survey data discussed here have considerable shortcomings when used on spatial data, particularly with respect to lack of coverage in certain locations and insufficient quantity of enumeration areas. These shortcoming will be discussed in more depth in the model validation section.

Representative Agent

I present two different implementations of the firewood foraging simulation model. These models represent two approaches to choosing how to summarize the complexity of the situation in a way that allows for feasible computation. In this section, I will describe an implementation that identifies the 250 agents that optimally (as defined below) summarize the full set of grid-cells in Tanzania. This implementation is useful in describing the generalized agent-based simulation approach and identifies key methodological solutions such as identifying which agents are best to analyze and which geographic zones we should ascribe to each agent. I will end discussion of this implementation with a discussion of where it was insufficient to model the hyper-localized phenomena of firewood foraging

Defining Representative Agents using Voronoi Specification

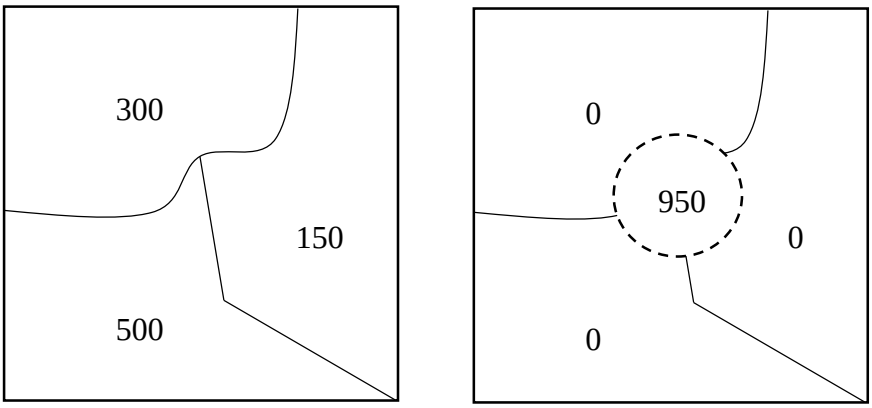
In many circumstances, it is necessary to identify representative agents for use in simulation to allow for feasible computation. However, model results may be very sensitive to how the agent was specified, so it is critical to correctly identify representative agents. In this section, I define what a representative agent is and describe a method to identify the optimal set of agents for an example landscape in Tanzania.

When village-level data are available, it can be best simply to use the village boundary as the definition of the agent zone. However, even when these data are available (and they often are not), they may still not be precise enough to model a phenomenon like firewood foraging that is sensitive to geographic scales much smaller than the village. Analyses that use single values for village zones imply, whether stated or not, that the population

value for the village zone shape is *uniformly* distributed. In many applications this assumption has no effect on the interpretation of results. In other cases, however, the zone value is compared to grid-cell input data that have higher resolutions (such as LULC data). In this context, the analyst must choose how to represent the zonal data in a grid-cell format. The uniform distribution assumption here means that every grid-cell used to represent the village zone has the same single value. This assumption becomes very problematic when local phenomena are assessed.

One method used to address the problems that arise with the uniform distribution problem is to attribute the value to single point, typically the village centroid, defined as the spatial gravity center of the boundary. In many cases, however, this location does not correspond to the actual village center and may instead identify a parcel of land that has few or no individuals present. Using the centroid raises additional problems and does not solve those raised by the uniformity assumption that can lead to estimation bias. Fig. A2.7 presents an example that shows these two types of problems that arise from using centroid-based or uniformly-spread data. The map on the right presents three hypothetical villages and their boundaries with the population of each village depicted by the number in the center of the village. Suppose, however, that the population of all three villages is actually concentrated in the center of the map, as shown in the map on the right. In this case, the assumption that all grid cells within a village zone have an identical share of the village's population is clearly flawed.

Figure A2.7: Example of Problems from Using Village Population Values



The method I present in this section is an alternative to using village centroid or a uniform distribution. I use high resolution population data to identify a grid-cell definition (both a point and a corresponding zone) that improves upon the methods discussed above. Because high resolution data are not always available, I also include a section below on how to use downscaling techniques to create high resolution data.

Optimal Agent Definition

Fig. 4.20 (upper left) shows the population density data described above. I apply an algorithm I created, which I call the N-Optimal Agent Identification Algorithm, to these data in order to identify the set of agent grid-cells that are the best n choices to represent the population. The algorithm minimizes the log-difference between the marginal reduction in distance between non-agent grid-cells and agent grid-cell and the number of agents chosen. Intuitively, the algorithm functions by choosing agents so that they are more likely to be represent a location with high population density, but because distance-to-nearest-agent also is in the heuristic, the algorithm also identifies some agents in remote, but potentially less dense, areas. The positive value obtained in this heuristic is then subtracted by the log of the number of agents identified. This second value approximates the computation time necessary to work with datasets of size n, and thus lowers the value of using too many agents. The approach balances the need to over-sample high population density areas with the need to assess the broader geographic extent in a project. For all of the agents chosen, whether in areas of relatively high or low population density, each agent represents a local maximum of population density, thus eliminating the problems of using naïve village centroids discussed above.

N-Optimal Agent Identification Algorithm

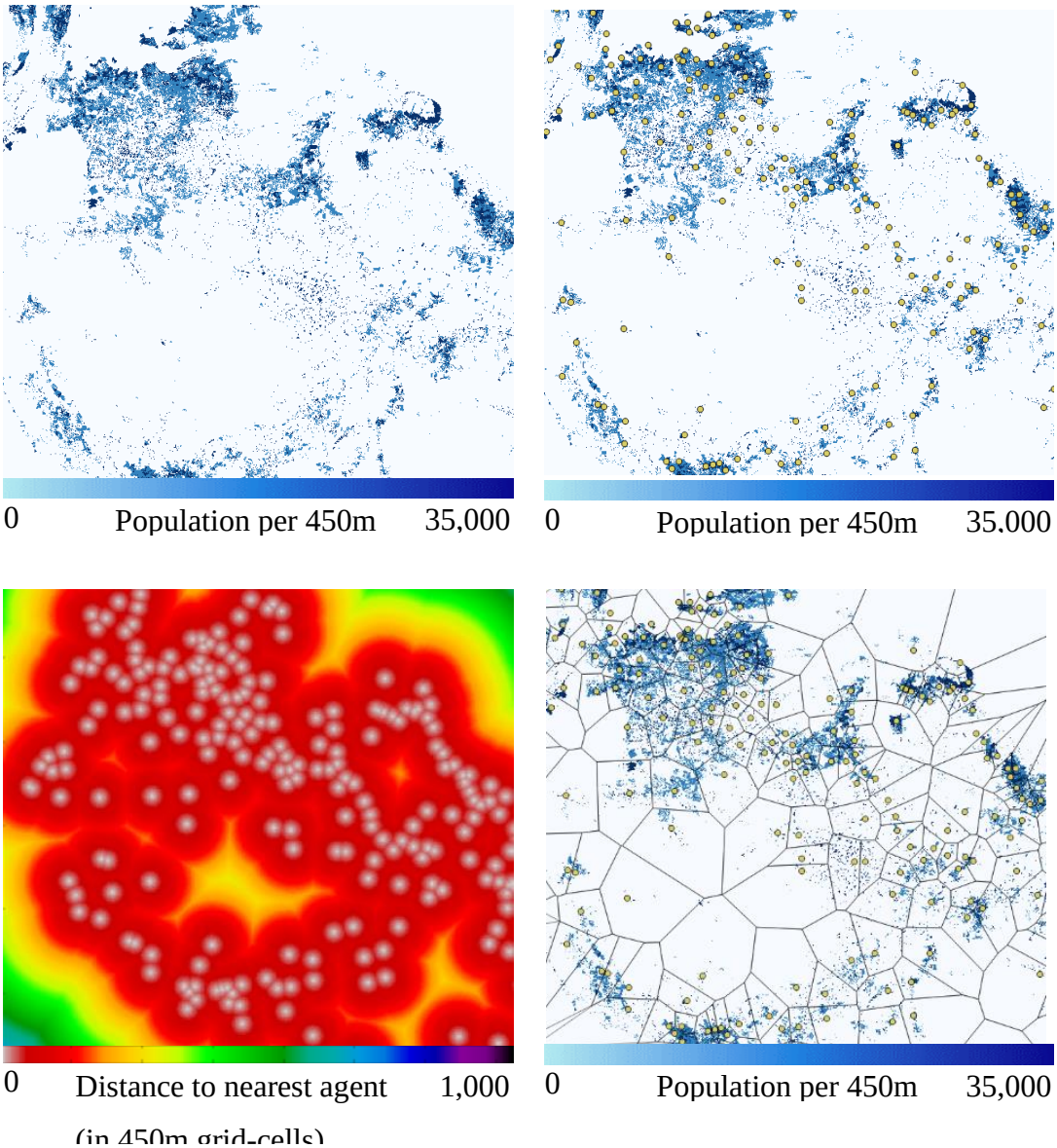
1. Identify the grid-cell with the highest population density in the grid-cell network. Starting with $n = 1$, define this grid-cell as agent n and add to the set of agent grid-cells.
2. Calculate the distance of all non-agent grid-cells to the nearest agent grid-cell.
3. Multiply the distance value(s) found in step 2 by the population density in the grid-cells. Define this value as heuristic h.
4. Identify the grid-cell with the highest heuristic value, define it as the n+1th agent and add it to the set of agent grid-cells.
5. Recalculate the heuristic and use it to calculate the quality of fit metric Q, where:

$$Q = \ln\left|\frac{\partial\left(\sum_x \sum_y h_{xy}\ \right)/\partial \mathrm{n}}{\mathrm{n}}\right| - \ln(n)$$

6. Repeat steps 2-5 until $Q = 0$. Define the set of agent-grid cells as the N-Optimal set of Agents.

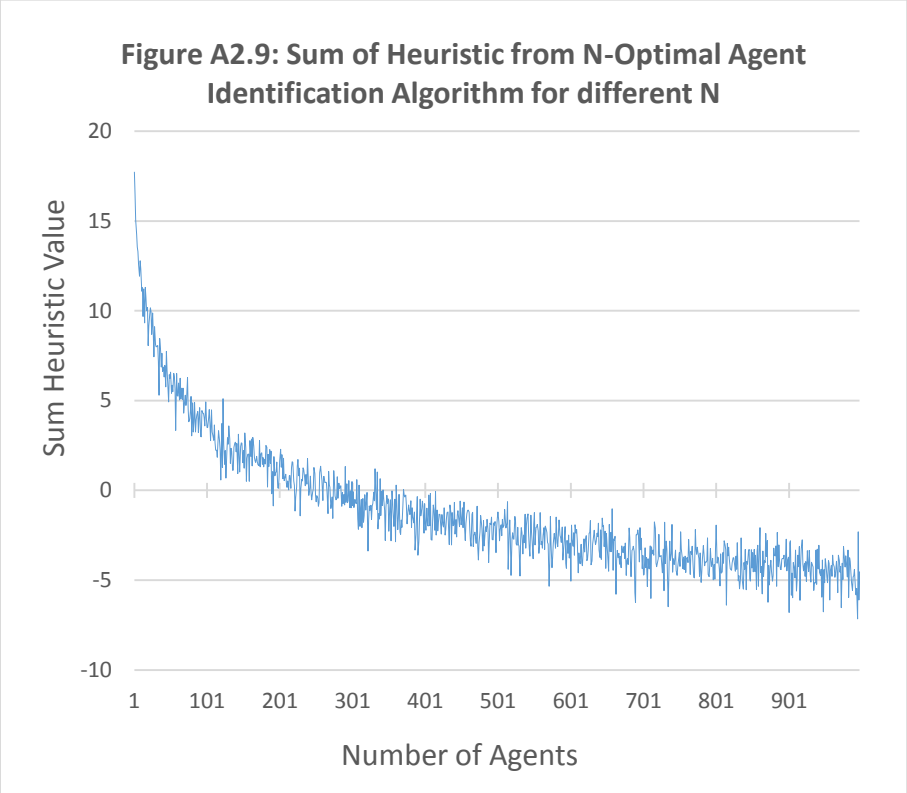
Fig. A2.8 below graphically illustrates the results of this algorithm applied to a sub-region within Tanzania. The map in the upper-left shows the population density input data while the map in the upper-right shows the 250-optimal agents the algorithm identifies. These agent-points were chosen iteratively using the distance-weighted heuristic. The image in the lower right shows the map of distance of non-agent grid-cells to their nearest agent grid-cell, which is the value multiplied by the population density to create the heuristic. Finally, the map in the lower right presents a corresponding set of agent zones for each of these agent point, thus identifying two parts of the cell network. The zones were identified by calculating the Voronoi zones for each point. This approach assigns all areas on the map to whichever agent point is closest.

Figure A2.8: N-Optimal Agent Algorithm Applied to Tanzania



Using this algorithm, one can choose the correct number of agents to balance computational feasibility with model accuracy. The algorithm provides a quality of fit metric for a subset region within Tanzania. Fig. A2.9 plots the quality of fit metric when

this algorithm is applied to a subset of Tanzania[14] with values of N ranging from 1 to 1,000. The algorithm value becomes negative at approximately 250 agents, suggesting that for this subset 250 is the optimal number of agents to define.



Application of Representative Agents to Tanzanian Data

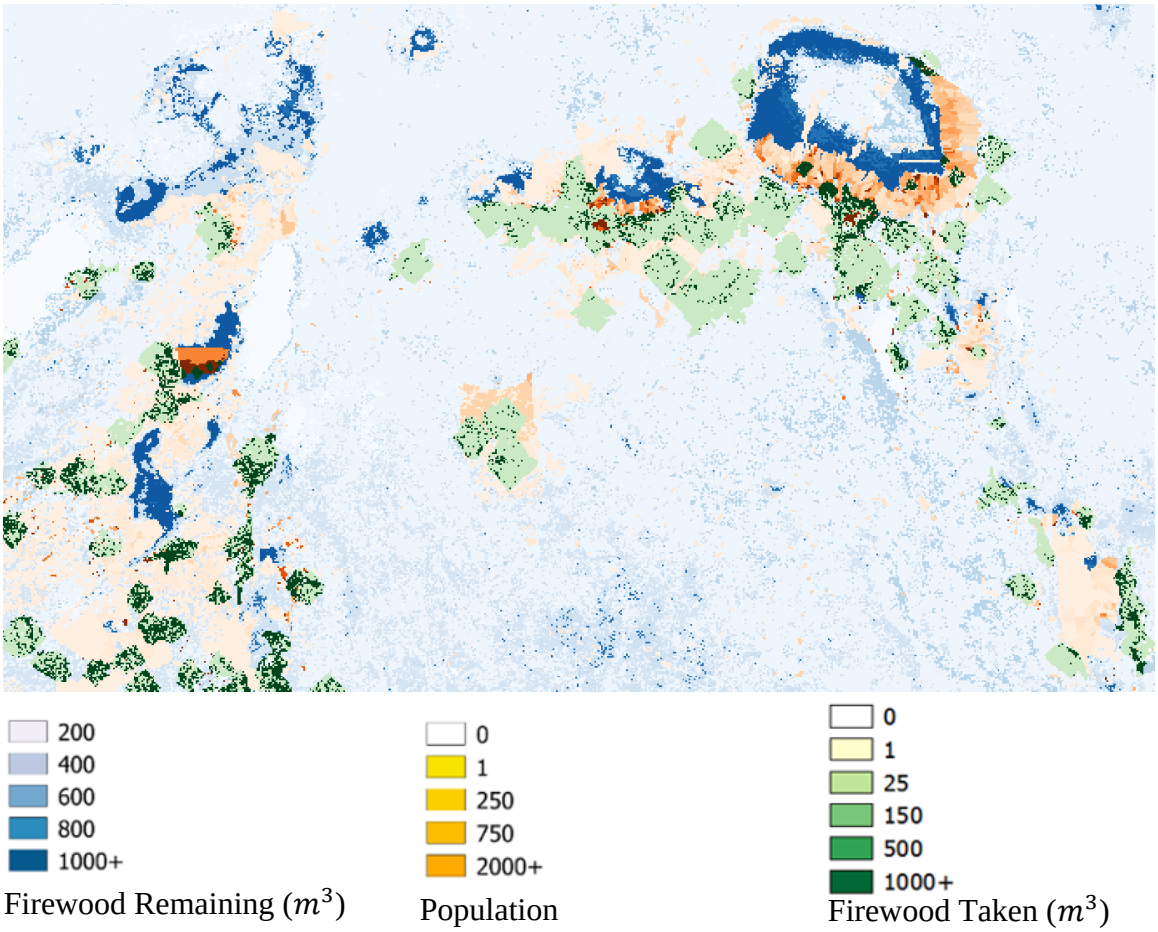
Using the 250 agents identified with the algorithm above, I present the first of two implementations of the agent simulation. I refer to this first implementation of the "representative agent" approach. The representative agent approach is inferior to the second approach discussed, which identifies every pixel as an agent rather than a zone of pixels (thus I will I spend relatively less time presenting details on this case). However, using representative agents has the positive attribute that the results are much easier to visually interpret. Moreover, many environmental end ecosystem service analyses focus

[14] Subset chosen to increase the visual clarity of the resulting zones.

on phenomena that are not as hyper-localized as firewood collection. In these cases, the value of using representative, zonally defined agents as in this first approach may be superior, such as when doing watershed-level analysis or farm-level analysis.

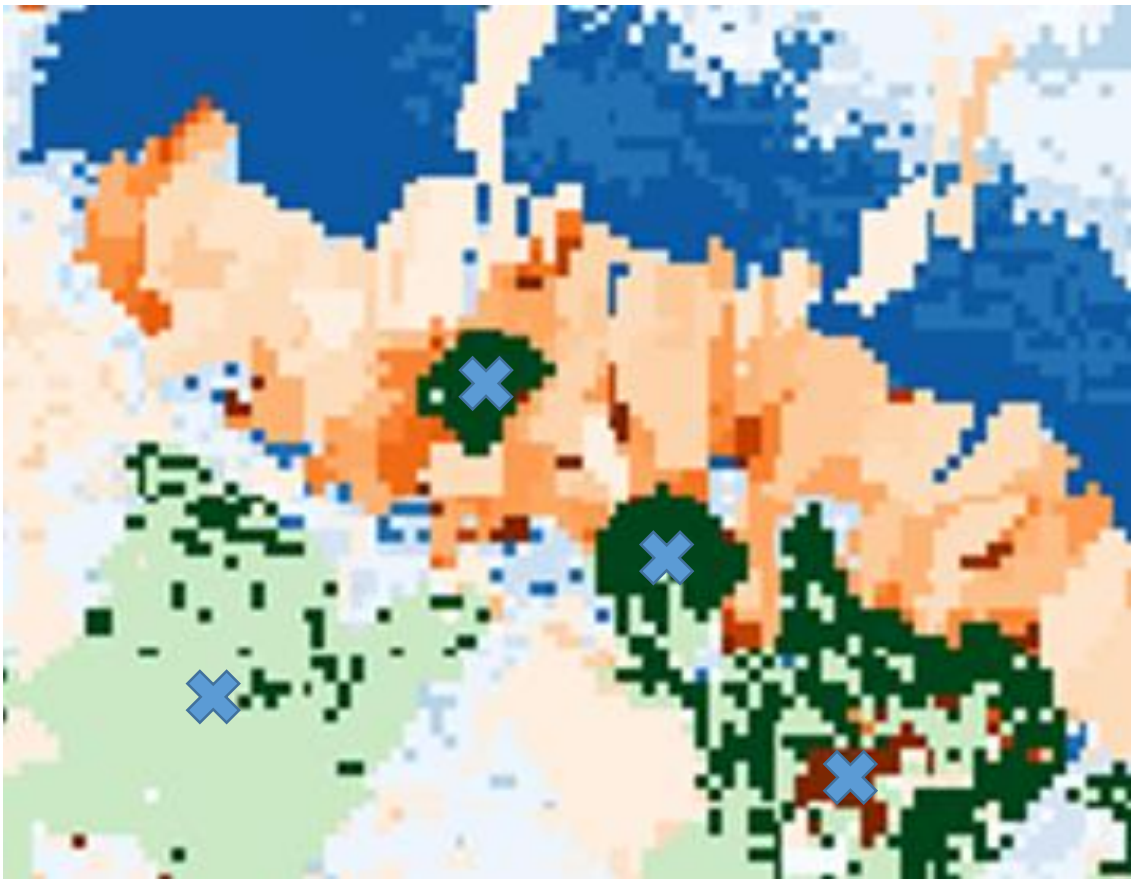
Fig. A2.9 shows results from the representative agent example. Three different variables are plotted, including the firewood remaining after the simulation has been run (blues), the initial population data that were used to identify the optimal set of 250 agents (oranges) and the amount of firewood that was gathered from each grid-cell (greens).

Figure A2.9: Results from Representative-Agent Simulation



In this example, we see the basic dynamic that arises from the simulation approach. We have specific agent locations (denoted in the close-up Fig. A2.10 below with an X mark) and specific cells from which they gather firewood. Rather than identifying aggregate behavior, this method identifies a full set of agent-home and agent-foraging locations, giving much more specific predictions about who benefits.

Figure A2.10: Results from Representative-Agent Simulation



Even with this many agents, however, we see some undesirable results. For instance, the population cluster near the center has some populated areas on the north side of the cluster *and* has available forest underneath, but it registers no foraging because that portion of the population node was summarized by an agent centered far enough to the south that these forests were not foraged. An additional problem shown above is the arrangement of foraged zones into diamonds. This is an artefact of the route-finding algorithm being based on grids, but more importantly shows how sensitive the results are to the edges of the zero profit zones that are implied by this travel algorithm.

For these reasons, I present an approach that extends the number of agents defined so that there is one agent per pixel of input data. I refer to this level of analysis as the "pixel-agent" approach. Before discussing the next approach, however, I first define a technique for creating the type of high-resolution data necessary for the pixel-level result to work.

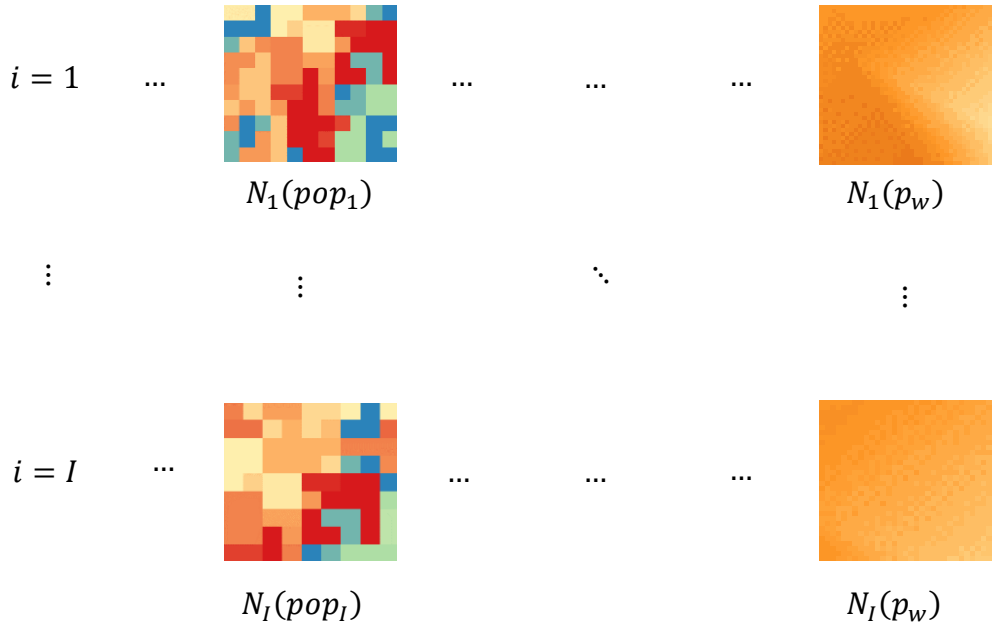
Additional Methods

The agent set defined here also has an attribute table and a corresponding set of attribute matrices (as shown in Fig. A2.11). The notation used in Table 4.5, for instance, $N_1(pop_1)$, references the normalized matrix for population as viewed from the $i = 1$ position in network N. Note that matrix references as used here are slightly different than standard matrices, represented in this case by the i-th index. This notation implies the data form is a "view" of a matrix such that it may be a subset of the full matrix defined by a reference point stored in the center element in the matrix and all the neighboring cells within square radius (defined by vertices calculated as $(r - radius, c - radius, r + radius, c + radius)$. The idea of using a matrix view allows for a dramatic increase in computational speed by eliminating data and calculations that may be considered unnecessary due to spatial remoteness from the point of interest. A matrix view with the largest possible radius is identical to the original matrix.

Table A2.2: Agent Attributes

Agent ID	Location Cell	Agent Population	α_i	β_i^c	β_i^{cl}	p_w	... other attributes
$\boldsymbol{i = 1}$	(r_1, c_1)	$N_1(pop_1)$	$N_1(\alpha_1)$	$N_1(\beta_1^c)$	$N_1(\beta_1^l)$	$N_1(\beta_1^l)$	…
$\boldsymbol{i = 2}$	(r_1, c_1)	$N_2(pop_2)$	$N_2(\alpha_2)$	$N_2(\beta_2^c)$	$N_2(\beta_2^l)$	$N_2(\beta_2^l)$	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\boldsymbol{i = I}$	(r_I, c_I)	$N_I(pop_I)$	$N_I(\alpha_I)$	$N_I(\beta_I^c)$	$N_I(\beta_I^l)$	$N_I(\beta_I^l)$	…

Figure A2.11: Normalized Matrices Corresponding to Elements in the Attribute Table



Calculating Travel Cost

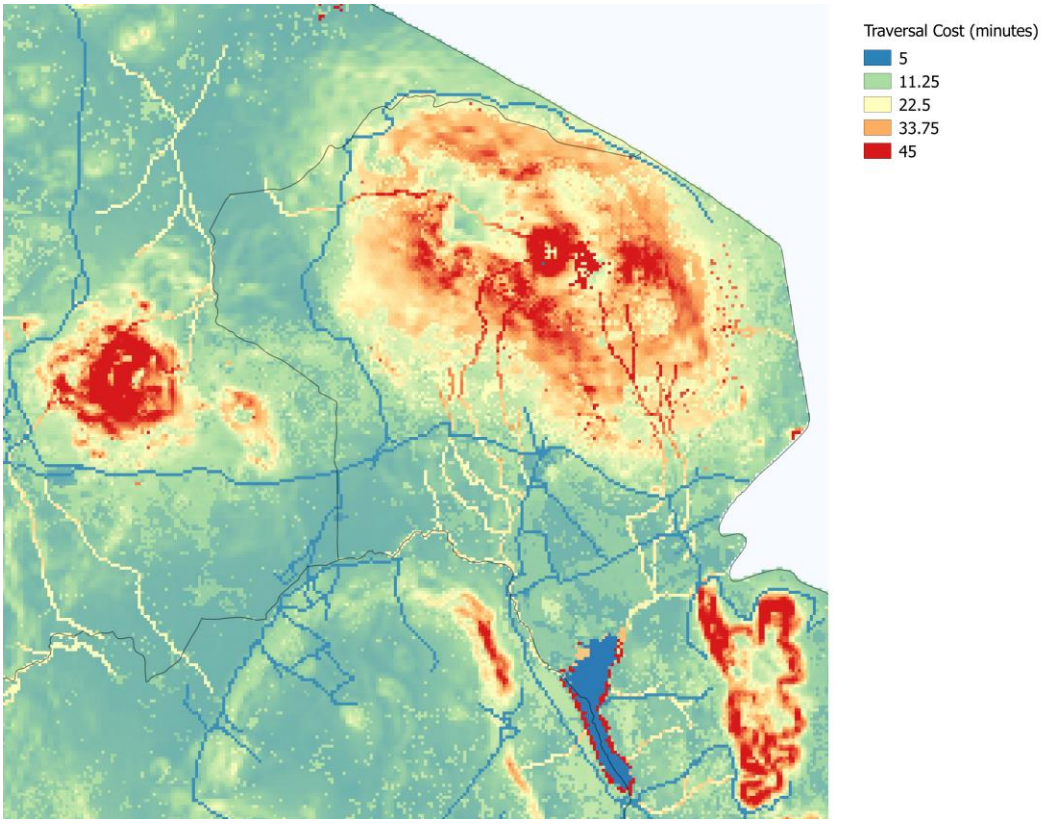
One of the most basic calculations necessary when simulating agent behavior on a landscape is to identify the cost of traveling from one location to another. Even when using matrix views as discussed in the previous section, finding the optimal route can become computationally infeasible, especially if the destination is remote from the origin. In this section, I identify a method that allows for route-finding even in large data sets. I do this by identifying four different route-finding algorithms and then create a composite algorithm that combines all four approaches hierarchically to apply the correct algorithm (defined as maximizing accuracy while minimizing computational intensity). This approach is able to provide a large computation speed increase because many route finding problems are quite simple (for instance, traversing a uniformly flat field) and the simpler algorithms identify the optimal route in a faction of the computational time necessary for more complex route-finding algorithms.[15] Before presenting these algorithms, however, I first discuss how the traversal cost grid-cell network was created.

Constructing the Traversal Cost and Arrival Cost Grid-Cell Networks

I distinguish between three types of travel values in this section that have unavoidably similar terms. For the first term, I use *traversal cost* to denote a network or matrix that stores the amount of time it takes an agent to cross from one side of the grid-cell to the other when traveling on one of the cardinal axes. Fig. A2.12 below shows the traversal cost matrix for Tanzania.

[15] In the course of this work, I also created an algorithm that works well even when extremely long routes must be calculated. This algorithm works by solving the hierarchical search algorithm at varying levels of resolution for the traversal cost data and then combining the results to calculate the full route. I have omitted it from this analysis due to time constraints.

Figure A2.12: Traversal Cost Grid-Cell Network



The second term is the *traversal step cost*. When an agent travels from one cell to another, the agent incurs a travel cost equal to the sum of the traversal cost in the origin cell and the destination cell, divided by two to reflect that the agent is presumed to start from the middle of the origin cell and thus only travels half the distance of the cell. I refer to this value as the traversal step cost because it is accounts for traversal cost values in both of the adjacent cells.

$$TraversalStepCost(o_{rc}, d_{rc}) = \frac{1}{2}\left(TraversalCost(o_{rc}) + TraversalCost(d_{rc})\right)$$

Here, and in this rest of this section, o_{rc} and d_{rc} refer to the origin and destination points with row indexed by r and column indexed by c. When calculating traversal step cost, the origin and destination must be either adjacent or diagonal. The above traversal step cost function is only valid for adjacent cells but in many applications it is useful to allow

travel to diagonal cells. In this case, a slight modification of the function above is required to account for the greater distance from the center of a square to one of its corner vertices:

$$TraversalStepCost(o_{rc}, d_{rc}) = \frac{d}{2}\left(TraversalCost(o_{rc}) + TraversalCost(d_{rc})\right)$$

where $d = \sqrt{2}\ if\ (r_o - r_d)(c_o - r_o) = 0\ or\ d = 1\ otherwise.$

Finally, the third term I use is *arrival cost*, which denotes the sum of traversal step costs in the set of moves necessary to reach a destination cell not adjacent to the origin cell (i.e., the arrival cost sums the traversal step cost of all grid cells the agent must cross). The following section will define the algorithms used to identify the cells that must be traversed in the optimal route. For now, while delaying definition of how the route cells were determined, denote the arrival cost as:

$$ArrivalCost(o_{rc}, d_{rc}) = \sum_{i=0}^{\#steps} TraversalStepCost(RouteCell_i, RouteCell_{i+1})$$

In applications where an agent is considering what action to do based on multiple potential locations, it is useful to define an arrival cost geospatial grid-cell network and corresponding matrix. This matrix (or matrix view) requires that the origin cell be exactly located at the center-element of the matrix and that the matrix is square with an odd-valued dimension so that the center point of the matrix corresponds with the centroid of a cell. Thus, all non-central elements in the arrival cost matrix represent the arrival cost calculated with the center-cell of the matrix as the origin of the route and the non-center cell as the destination of the route. It will become clear when constructing net profit why this data structure is useful, but for now note that it is a way to represent the full set of transport costs an agent must consider when searching multiple grid-cells for their next action. Figures are included in the subsequent sections giving examples of arrival cost matrices.

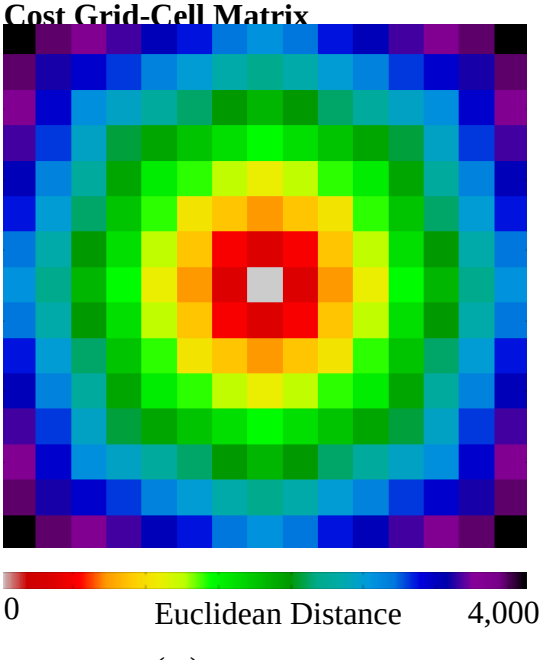
Euclidean Route

The first route-finding algorithm is quite simple, but ignores any data besides the origin and destination coordinates:

$$Euclidean\ Route\ Cost = (((o_r - d_r)^2 + (o_c - d_c)^2)^{.5}) * (gridcell\ size) * (travel\ cost)$$

The route identified by this algorithm is simply the set of grid-cells intersected by the line connecting the origin to the destination. The set of cells contained in this route will be exactly the same as the next algorithm, but the Euclidean route ignores the traversal cost in each cell and instead calculates the route cost blind to the cells it is traversing. Figure 4.30 applies the Euclidean route to create an arrival cost grid-cell matrix. As discussed above, all arrival cost maps are defined such that the value of each cell is the cumulative travel time it takes an agent to travel from the center of the map to the cell in question.

Figure A2.13: Euclidean Distance Arrival Cost Grid-Cell Matrix



Bulldozer Route

The bulldozer route-finding algorithm identifies exactly the same set of grid-cells as the Euclidean route, defined as all grid-cells intersected by the line connecting the origin grid-cell and the destination grid-cell. However, the Bulldozer route determines the cost of arrival by summing the traversal costs of all cells through which the agent passes. Thus, the use of this algorithm (and all subsequently defined algorithms) requires

constructing a traversal cost grid-cell network as described above. The name “bulldozer” was chosen because the route travels through all cells that are in a straight line between the origin and destination, regardless of traversal cost. Calculation time for the bulldozer route is approximately n-times slower than the Euclidean route, where n is the number of cells that must be traversed.

Myopically Optimal Route

Consider how a person chooses a route to travel through a crowded room when they cannot see beyond the area immediately adjacent to their current position because their view is blocked by other persons. In this circumstance, the person knows approximately the direction they want to go (to the bathroom, for instance) but cannot identify which movement (go straight, veer left, or veer right) is the optimal choice. Suppose that the person will proceed towards their goal by choosing whichever movement yields them the greatest advancement towards their goal per travel cost incurred by making that movement. This means that the person will prefer to go straight forward towards their goal if all choices’ traversal costs are the same, but will veer left or right if the direct route is expensive. In the bathroom finding example, this algorithm means one prefers to go straight towards the bathroom but will deviate from the straight line if there exists something that is particularly expensive to pass through, such as a thick pack of persons in a discussion. This process iterates until the person arrives at their goal. I refer to this as being myopically optimal because at each step the agent chooses the movement that is best if only the immediately adjacent cells are observed. Clearly this method may fail to find the optimal route, but it has the very important attribute that it only needs to consider the subset of adjacent neighbors. This lets the route be found approximately 100 times as fast as fully-optimal route-finding algorithms. When applied to the traversal cost matrix for Tanzania, the myopically optimal route algorithm performs almost identically to the fully-optimal route except in cases where there are complex road networks, tight mountain passes or few river crossings.

The Myopically Optimal Route-finding Algorithm proceeds as follows:

1. Draw a line between the current location and the destination. Define the first cell intersected by this line as the “go straight” choice. Define the cells adjacent to the current cell but to the left or right of the line as the “veer left” and “veer right” choices respectively.

2. Calculate the heuristic value for each of these three cells[16] as follows:

$$h(c) = \frac{d(c,d)}{d(s,d)} * p * t(c)$$

where $d()$ is a function to calculate the Euclidean distance between a potential choice $c \in \{veer\ left, go\ straight, veer\ rigt\}$ and the destination d while s represents traveling straight. The traversal cost of the choice is drawn from the traversal cost matrix and is represented as $t(c)$ above. Finally, p represents the agent's preference for traveling straightness. I will solve for the value of p that optimizes this heuristic below.

3. Choose whichever choice has the lowest heuristic value and define that cell as the current location. Repeat steps 1-3 from the new current location until the agent arrives at the destination cell.

The preference for directness is a calibration parameter that can be determined by testing the algorithm on different landscapes. I chose the value for p by solving the algorithm for multiple values of p and choosing the value that achieves the minimum travel time. Fig. A2.14 shows the results of this process, which identified $p = 32$ as the optimal choice.

[16] I also used, but excluded from this exposition, versions of the algorithm that considered five choices (90 degrees left, 45 degrees left, straight, 45 degrees right, 90 degrees right). This version showed promising results in situations of complex terrain and thus I will explore it in future research.

Figure A2.14: Identification of the Optimal Preference for Directness in the Myopically Optimal Route-Finding Algorithm

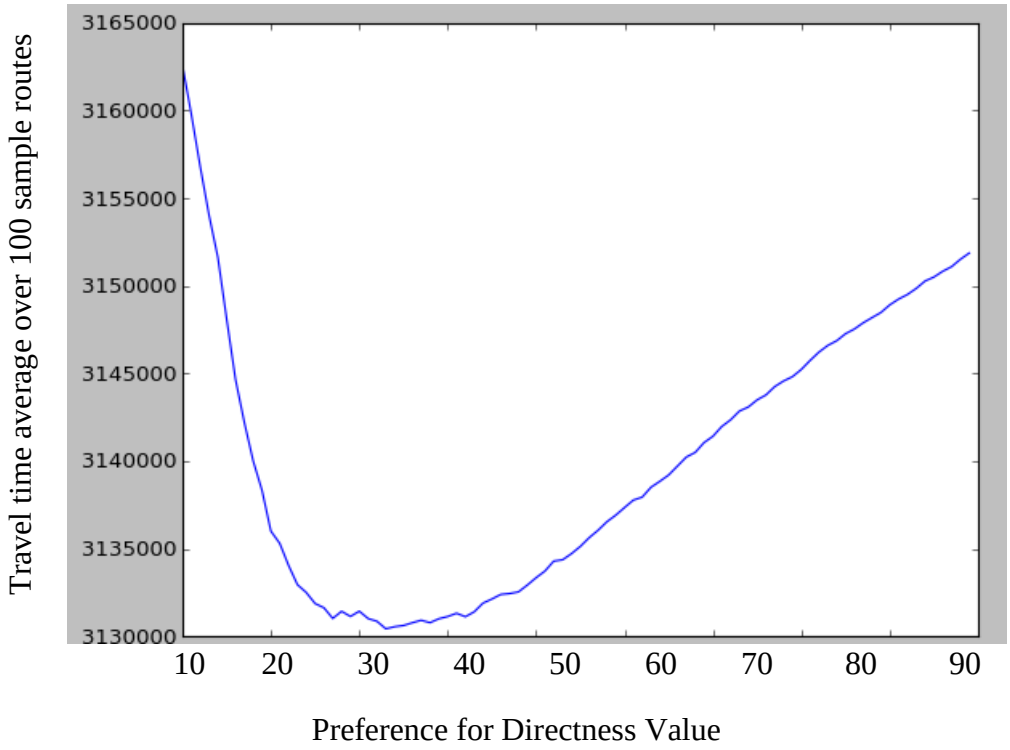
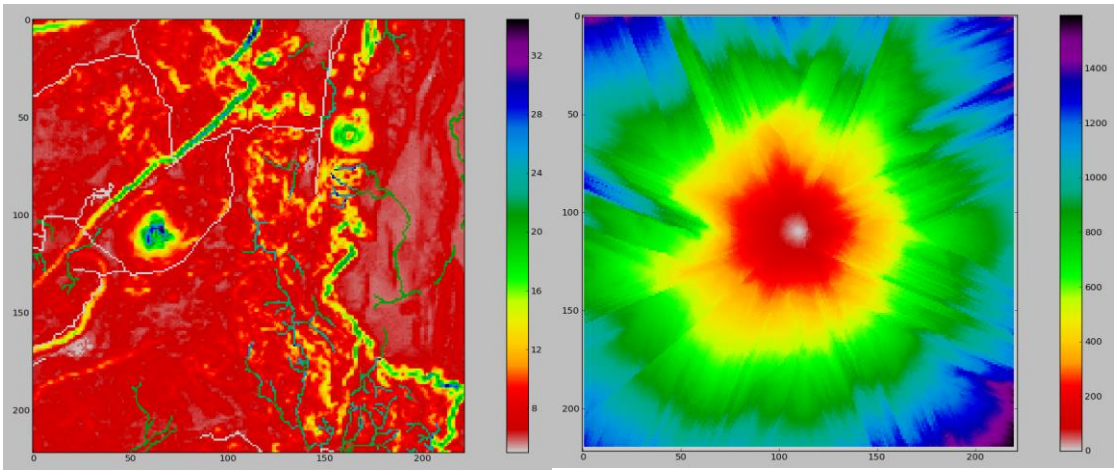


Fig. A2.15 shows the arrival cost grid-cell matrix for the myopically optimal route side-by-side with the traversal cost map used to generate the routes. Notice that elements like the mountain in the southwest or the road running through the center have very noticeable impacts on the shape of the arrival cost surface.

Figure A2.15: Identification of the Optimal Preference for Directness in the Myopically Optimal Route-Finding Algorithm



Traversal Cost (minutes)

Myopically Optimal Arrival Cost (minutes)

Optimal Route (A-Star)

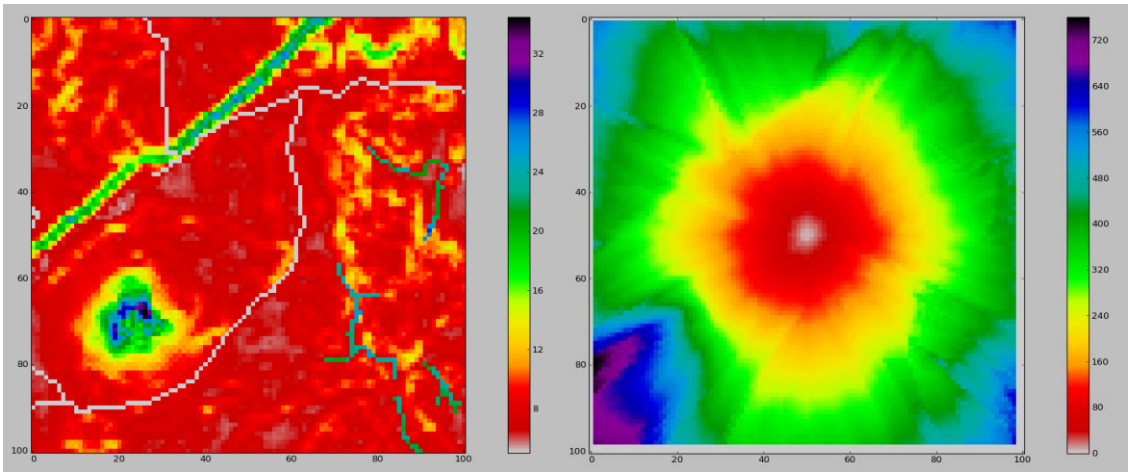
There exists a vast literature in computer science and transportation studies that addresses how to find the optimal route through a network (see, for instance, Hall 1986). I will not compare the many approaches to finding the optimal route, but instead, will compare one of the most common and efficient algorithms, the A-Star algorithm, to the other algorithms already discussed in this section. The class of fully optimal route-finding algorithms all share the feature that they are too slow on their own for the route-finding tasks in this paper. However, they do have the desirable trait of being fully optimal under broad assumptions, and so should be used when the other approaches identify a suboptimal result.

The A-Star algorithm uses a graph tree and a sorted priority queue of alternate paths to find the least cost path between two nodes. Fig. A2.16 shows a pair of traversal cost and arrival cost maps generated with the A-Star algorithm. The conditions under which the A-Star algorithm identifies the optimal route are well studied in computer science. The main requirement relevant to the work here is that the heuristic used is admissible, defined as never overestimating the minimal cost of reaching the goal.

Combined Route Finding

The method I use for calculating travel costs overall is to combine all four approaches into a decision hierarchy that chooses the algorithm that maximizes speed subject to achieving a threshold of accuracy. For every travel route, I use the Euclidean distance to estimate a first-pass estimate on traversal cost. In the (relatively uncommon) event that all of the traversal costs are uniform (such as traveling over a flat field), choosing the route that minimizes Euclidean travel cost identifies the fully optimal route. In the more common cast that it does not identify the optimal route, I instead use it as the heuristic in the next level algorithm. So, if Euclidean distance does not identify the optimal route, use it as a heuristic to enable computation of the bulldozer route. If the bulldozer route yields a different value than the Euclidean route, we know that the Euclidean route is not accurate and can disregard it. Next, I calculate the myopically optimal route. If there is no or little difference between the myopically optimal route and the bulldozer route, the algorithm ends here and uses the myopically optimal route to determine arrival cost. If the myopically optimal route is considerably different than the bulldozer route, I proceed to using the A-Star algorithm using the myopically optimal route as a candidate solution to speed up calculation. Fig. A2.16 shows the output of this process using the same area as in the myopically optimal algorithm (but with a smaller geographic extent due to calculation constraints). The reason this process yields a dramatic speed increase is because many routes a villager must travel are actually quite simple and thus applying a complicated routing algorithm is unnecessary.

Figure A2.16: Combination of Myopically Optimal Arrival Cost Calculation with A-Star Algorithm



Traversal Cost (minutes)

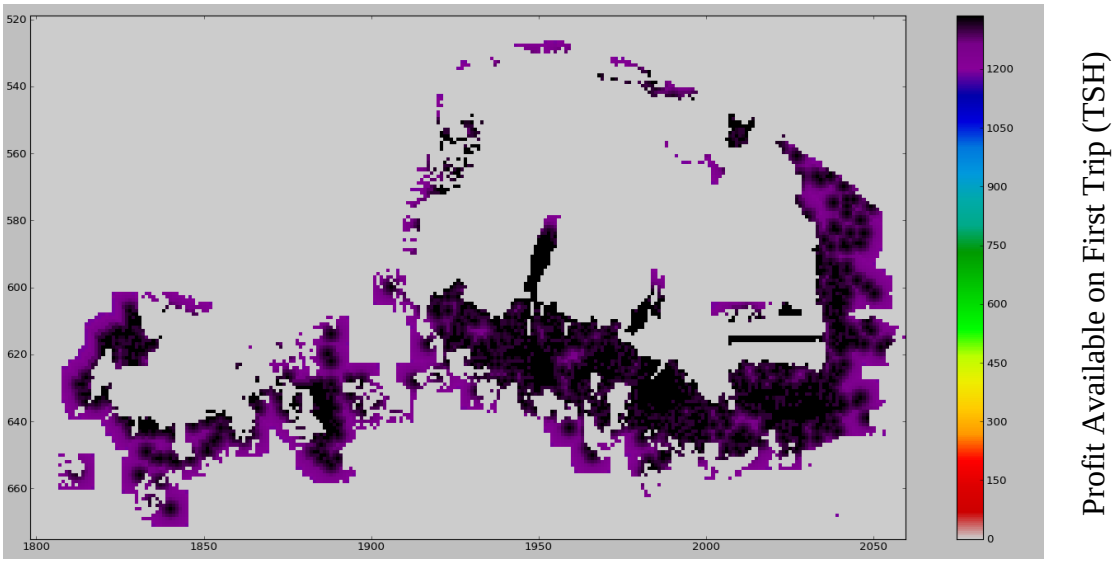
Myopically Optimal Arrival Cost (minutes)

This combined algorithm is then applied to all of Tanzania to identify the arrival cost of traveling from every grid-cell to every other grid-cell that might profit greater than zero. The result of this process is rather difficult to express in words or on paper, but is a very important mathematical construct that is useful in spatial analyses that define behavior happening over space during time. The algorithm described above creates a 2-dimensional array of arrival costs reflecting the optimal routes between the center cell and all other cells. This 2-dimensional array is created for every grid-cell in Tanzania, thus creating a 4-dimensional array that combines traversal cost with the arrival cost maps on each pixel of the traversal cost map. Fig. A2.17 illustrates a portion of this array with two of the arrival cost maps shown for two example grid-cells.

Process for Defining Action Order

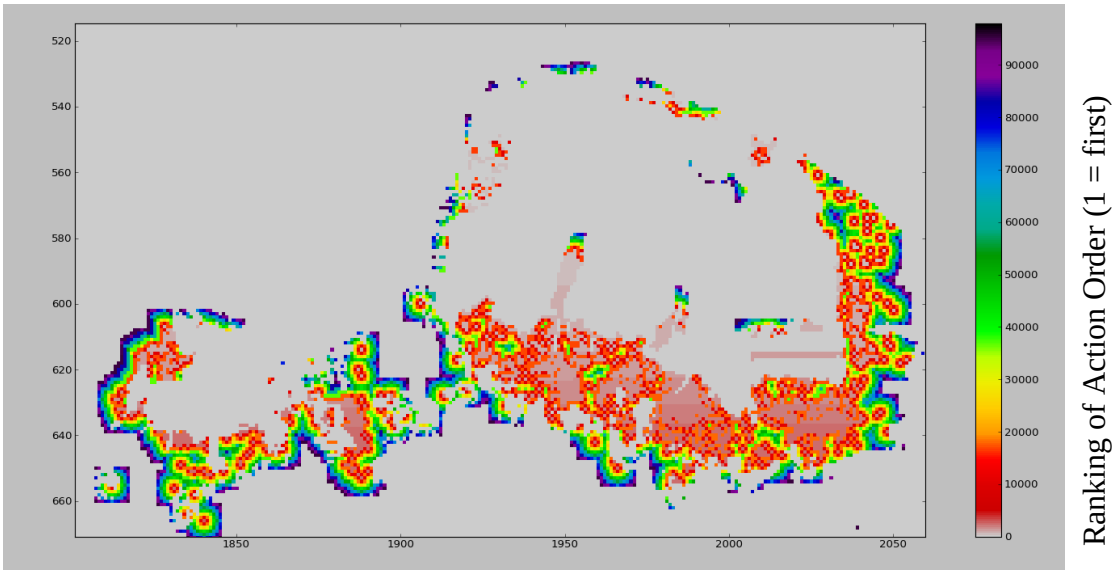
Step 1: Identify the order in which agents will forage. For this problem, I define the order so that the agent with the highest initial marginal profit available goes first, followed in descending order by all other agents and ruling out of consideration any agent that doesn't have any positive marginal profit within a search depth. Fig. A2.17 shows a mapping of agents by their available marginal profit. Currently, this only uses Euclidean distance as a heuristic for least-cost distance.

Figure A2.17: Initial Profit Available



Step 2: Create a matrix that represents the order in which agents will act by assigning an incremental counter to each cell, ranked by descending marginal profit, shown in Fig. A2.18.

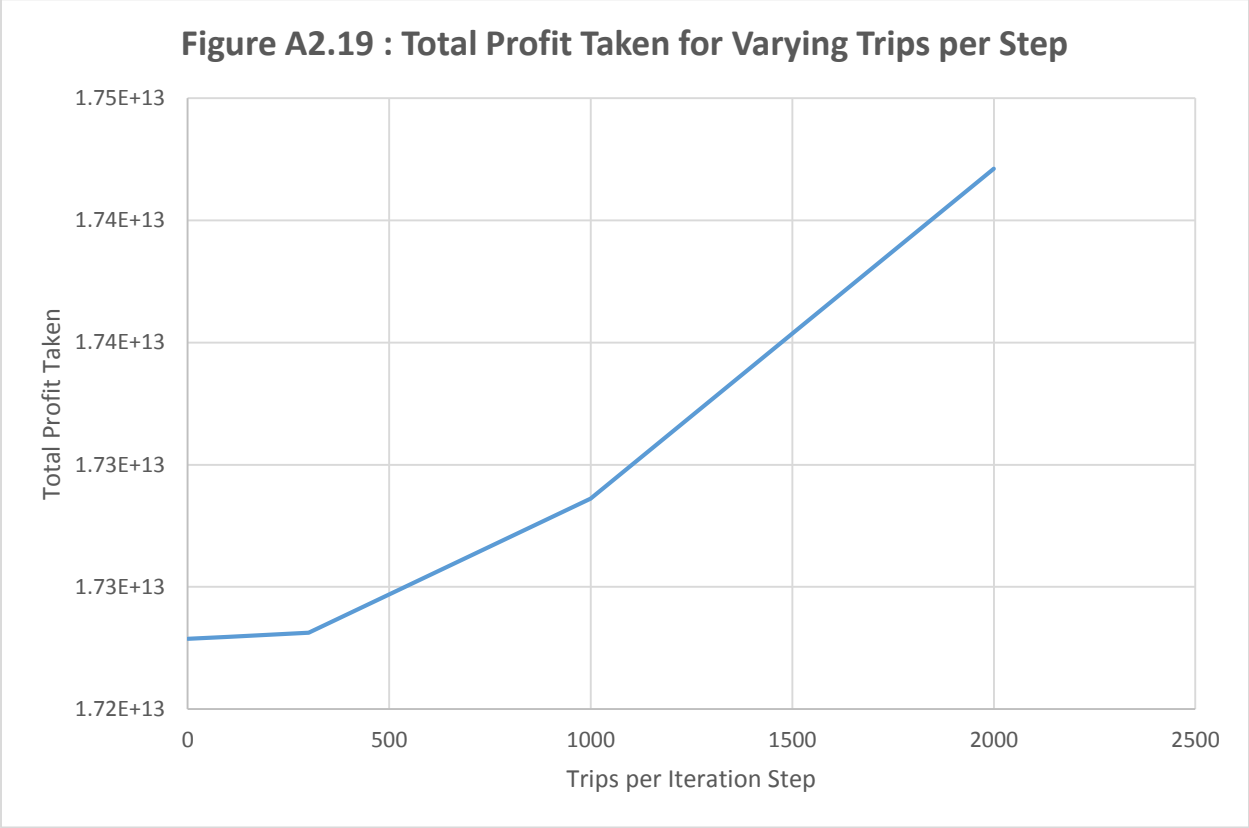
Figure A2.18: Action Order of First 100,000 Agents



Sensitivity Analysis on Number of Trips

The results from this simulation are not sensitive to order. I tested this by trying many different orders and saw that there was nearly no change. Testing with step sizes as low as the amount of cargo one actual individual could carry (28 kg for an adult female), and then comparing it to higher step sizes, I found very little difference. Additionally, I tested how sensitive the results were to the number of trips each pixel-agent can take in one step.

Starting with one trip per iteration step and increasing to 2000 trips per step, I found that the results changed very little. The maximum difference in the outcomes was found by comparing the one-trip example to the 2000-trip example, which yielded only a 1.1% difference in total profit collected. When comparing the one-trip example to the 300-trip example, there was only a 0.013% change in total profit collected. These results are shown below in A2.19. Additionally, I compared the same example runs to see if specific agents had different results (but with a similar total profit). I did not find evidence for this and the deviation was almost exactly the same per agent as it was in aggregate.



The action order together with the original firewood abundance map constitute the initial conditions for the model to begin. Once these conditions are defined, running the model is relatively straightforward. The model iterates through every iteration step and every time period while letting each agent identify and then do their max-marginal-gain choice. This is done in sequence and updates the 4-dimensional profit map at each step so that all other agents are making profit decisions based on the depleted landscape. Given this process, we obtain our results, which primarily lie in observing and reporting how this profit array changes over time and how it is affected by different policies. The next section presents a subset of results from this process, focusing on showing profit-maps and post-depletion abundance maps after the simulation has reached a resting state.

Defining the Action Order

One tricky assumption that becomes more important in pixel-based approaches is defining the order by which agents get to forage. A naïve algorithm that chooses the top-left pixel and iterate down and then across, though computationally fast, can introduce

non-trivial bias. The reason why this matters more for the pixel approach is that many agents are immediate neighbors while in the more aggregated approach, agents have a large set of non-contested objective cells. In other words, the sensitivity to competition effects increases as the number of agents increases, and thus order of play becomes more important.

The ordering approach I use prioritizes agents who have the highest initial profit, defined by ranking all agent cells by their max available profit at the start of the simulation, and then letting them play in that order. This method has the benefit that agents adjacent in the order of play are not frequently adjacent spatially. This minimizes ordering-sensitive bias. While still as arbitrary as nearly any other ordering scheme, this method has the positive attribute that it lets those act first who are most motivated to act – an assumption that may be close to reality. Nonetheless, recall that the bias is also dealt with by choosing very small iteration steps. As the amount of firewood each agent can gather each step goes to zero, the bias also goes to zero. I show sensitivity results for this in the next section. First, however, I define the algorithm I use to identify the action order.

#See Appendix 2 for more details on how the action order was established as well as a sensitivity analysis that shows changing the action order has extremely minimal effects on the results.

Validation

Validating an agent based model is critically important because the results are sensitive to assumptions on a variety of variables. This section discusses validation results and potential future methods for validation. There are two basic levels of validation that I discuss. First, I validate at a regional level. Specifically, I test to see if the model predicts aggregated values correctly, such as for the amount of firewood consumed. Secondly, I discuss validation of grid-cell level predictions.

Aggregate-Level Validation

The focus of this chapter has been on the modeling framework and methodology I used and has not been on the specific values and policy recommendations that result from my model. However, the calibration parameters chosen (described in the data section of this chapter) were drawn from peer reviewed sources and should be generally correct. For this reason, we can validate the model on an aggregated level to see if it makes predictions consistent with observations in Tanzania. Although aggregate validation like this is much

less specific than the agent-level validation discussed in the next section, it matches the general scope that policy makers often want, focusing on general conclusions such as what is the overall magnitude of the non-timber forest product valuation, what is the general distribution of benefits, what regions are particularly harmed, and other such questions.

National Firewood Consumption

Given that demand for firewood is not given as an input parameter but instead is calculated endogenously as a function of the landscape and the agents' behavior rules, the first value for calibration is to see how accurately the modeled demand matches literature estimates on firewood demand. Values from existing studies suggest that demand is between 25.8 and 55.5 million cubic meters. The baseline scenario predicts 66.7 million cubic meters of firewood will be gathered. This is above, but within the order of magnitude reported in the literature. The overestimated value here is very likely a result of wrong assumptions on mean harvestable volume per hectare. The range of values I drew from the literature (14 – 117 $m^3 ha^{-1} yr^{-1}$) is based on a relatively small sample of land which was primarily covered in woodlands and newer estimates suggest the value should be closer to the minimum value of this distribution than to the mean (which is what I used). Additionally, given that the majority of Tanzania is scrublands and savannah, which may have lower densities of firewood, it is not surprising that using the woodlands estimates would result in overstating the national effect. For this reason, the sensitivity analyses (presented in Table 4.7 in the results section) may be especially important because they show cases in which land is assumed to have less firewood abundance. For instance, the "half-forest, quarter-shrubs" scenario predicts approximately 41.4 million cubic meters, which is comfortably within the range of estimated consumption figures.

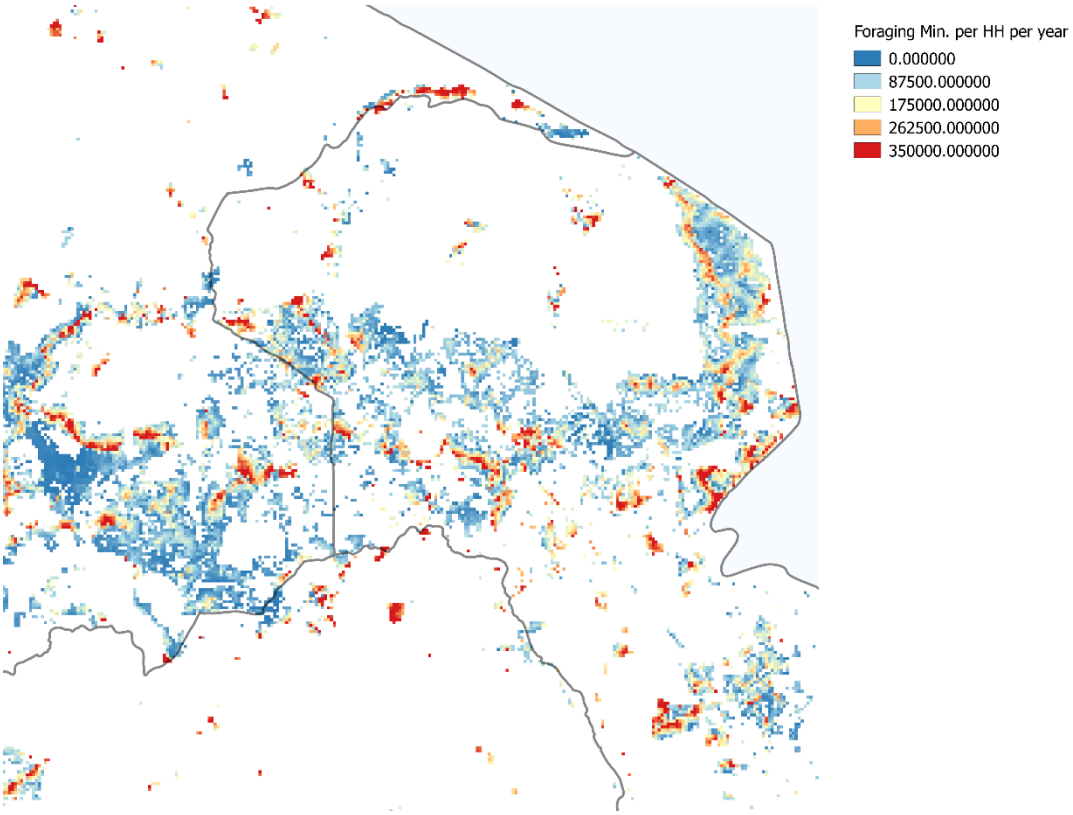
Future work will be able to devote more time to calibration that focuses on original work in the project area. For instance, I have already secured research-time with partners in Cambodia to make a new data-set of field observations on firewood abundance and gathering behavior. This study will be able to identify individual foragers' home location and foraging location. Pairs of home and foraging location pairs will allow for more precise validation and testing.

Agent-Level Validation

The basic framework for validation I use on the grid-cell level is to run the model and make predictions about, for instance, how many hours each agent will forage for

firewood, and then compare these predictions to household survey data on firewood collection. Figure 4.61 shows the travel spent for agents near Mount Kilimanjaro. Note that there is extreme variation at small scales. This is an expected outcome of the model given that firewood is not transported very far and that it has high transportation costs.

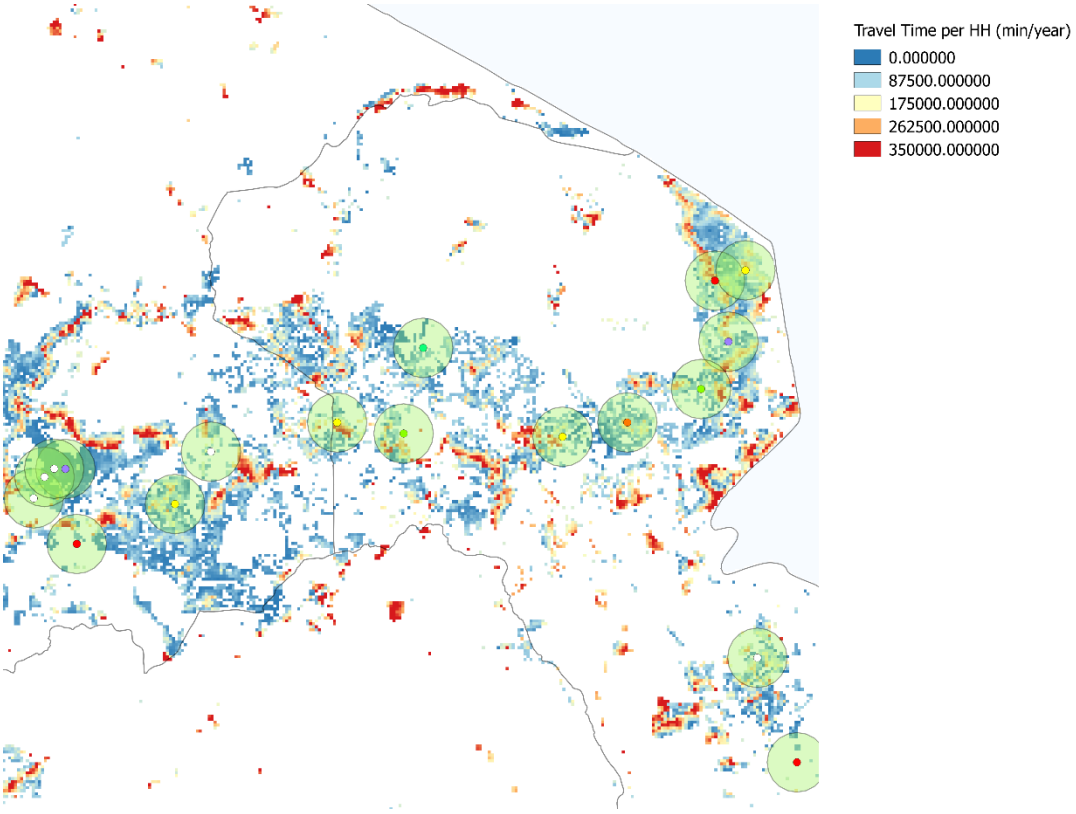
Figure A2.20: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)



Unfortunately, validation at the grid-cell level was not possible with the National Panel Survey because the household coordinates were randomly offset by 10 kilometers for privacy reasons. In many applications, this offset does not prevent geospatial analysis, but with a hyper-local production decision, as with firewood collection, the offset

presents insurmountable problems. Figure 4.62 expresses the difficulty with this by showing the where the enumeration areas are located along with a 10 kilometer radius circle overlaid on the point to show where the enumeration area may actually be located. The circle illustrates the size of the region in which the household is actually located. As Fig. A2.21 shows, there is a very high degree of variation in what my model predicts will be the traveled hours. This results in dramatically higher standard errors of any statistical comparison of the predicted results with the NPS observations.

Figure A2.21: Travel Time Spent in Minutes per Household per year (near Mt. Kilimanjaro)



The NPS data are insufficient for several other reasons. First, population in Tanzania are spatially located such that most areas have very low population while a handful of locations have extremely high population (cities and villages). When plotted on a two

dimensional geospatial map, the results resembles tall spikes on an otherwise flat surface. The households surveyed in the NPS are very likely located in one of these spikes, but because they are offset for privacy reasons, the location stated in the NPS is very likely located in a non-spike location. Thus, the results I presented here must be taken as correct on an aggregated level but not yet validated on a local level. This is not surprising given that I have specified a structural model with a resolution thousands or millions of times more detailed than existing structural models. For this reason, the value of the results lies in the methodological advances that are consistent with microeconomic theory but applied to a high resolution environment.

Towards overcoming the validation challenges discussed here, I present two methods for making validation succeed at the agent-level. First, as is illustrated in Fig. A2.22 and A2.23, one can construct a contour plot that identifies the continuous two dimensional surface that is consistent with the observed point data and assumes that the value changes in a linear fashion between all pairs of points. The background color of this figure is this contour surface derived from the survey points (similar to how the wage surface was created, as described in the data section). Also plotted in the model-predicted travel time spent. Although I do not provide statistical proof, it does seem evident that the contour plot is consistent with the modeled results (the regression however has an r-squared of only 0.33 due to the standard error problem discussed above).

Figure A2.22: Validation using Contour Surfaces (National)

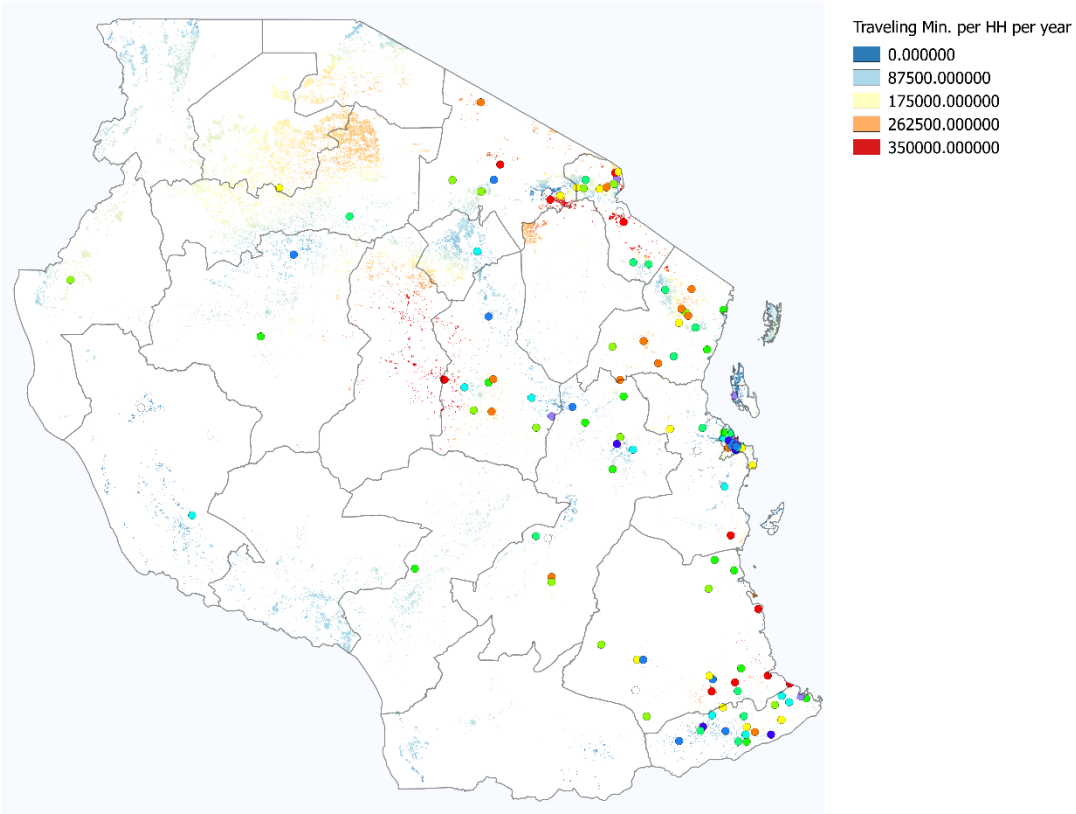
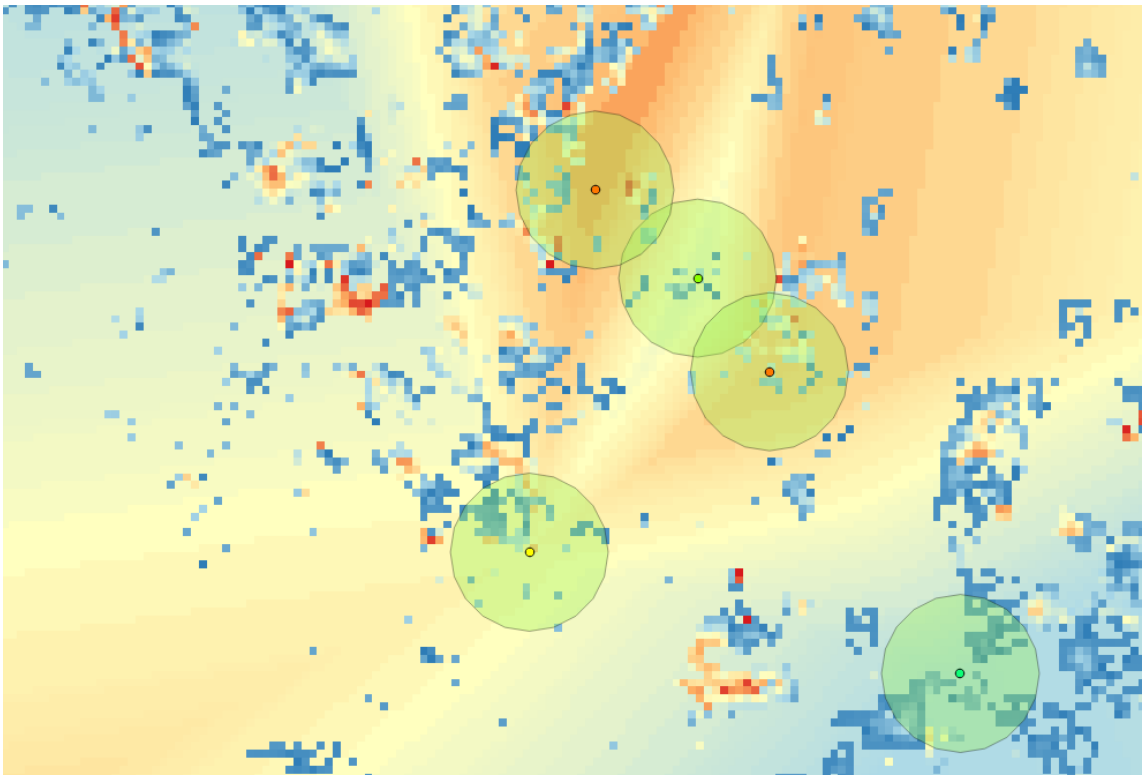
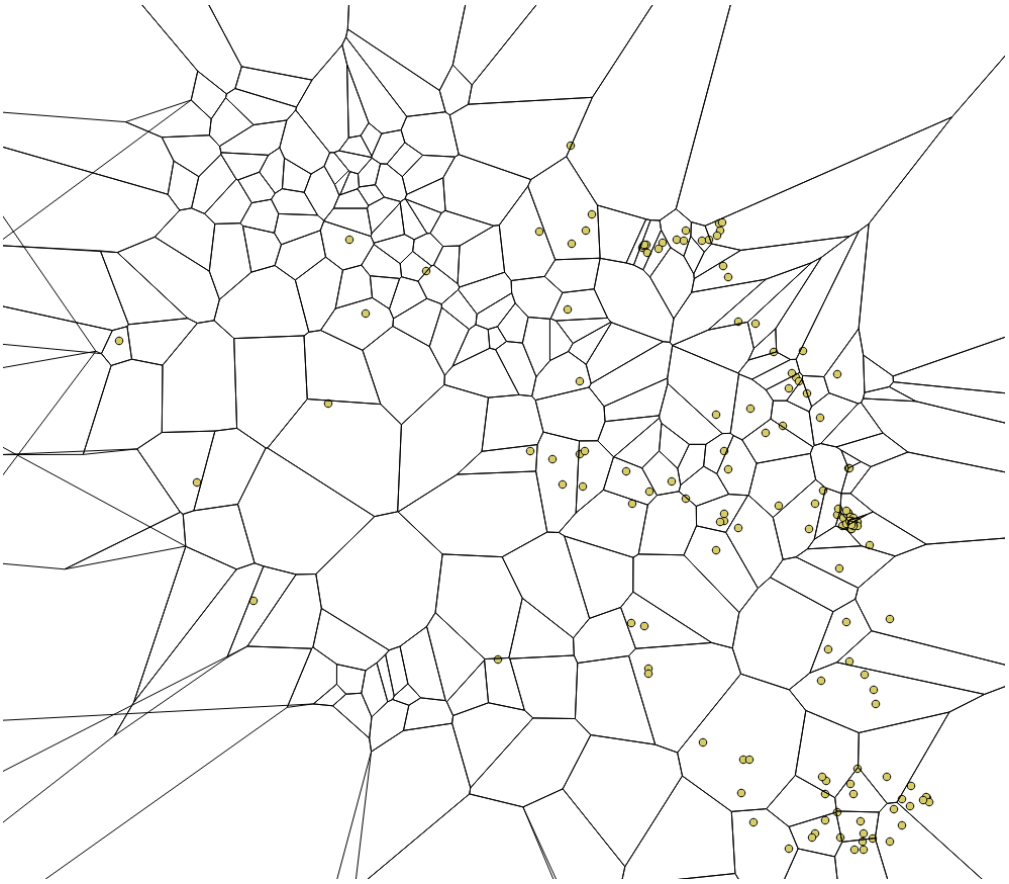


Figure A2.23: Validation using Contour Surfaces (near Mt. Kilimanjaro)



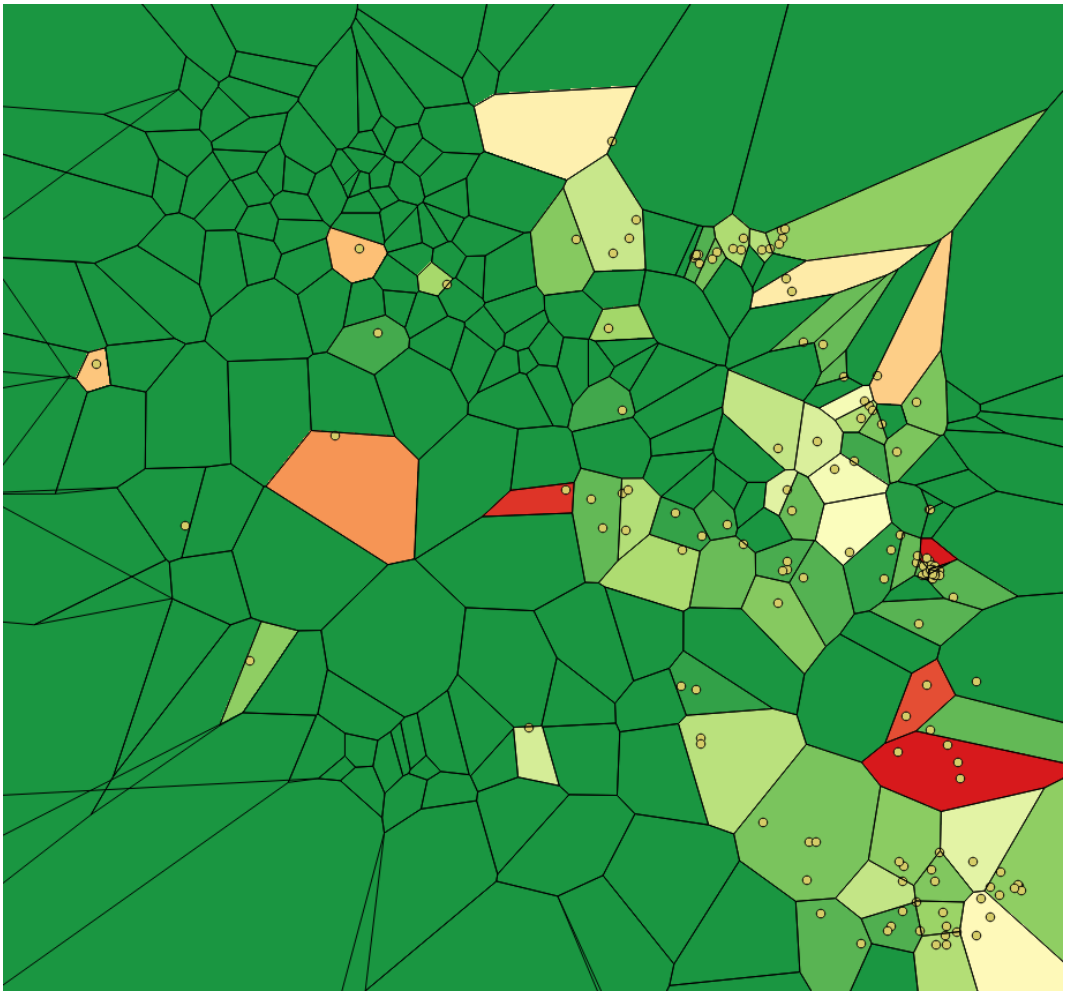
The second method that can be employed to validate data of this type is to create zones and fill them with an algorithm I describe below. For zone definitions, I used the 250 optimal agent zones discussed in the methods section, presented again for convenience in Fig. A2.24.

Figure A2.24: 250 Optimal Agent Zone-Point Pairs



To fill spatial zones with sampled survey data, it is critical to correctly deal with missing value and zones for which we have no survey observations. Fig. A2.25 shows a naive approach to this, in which I assign the mean value from survey responses to each zone. I extend this approach in Fig. A2.26 by filling the adjacent zones according to rules.

Figure A2.25: Naïve Approach to Zone Assignment: Mean or Zero



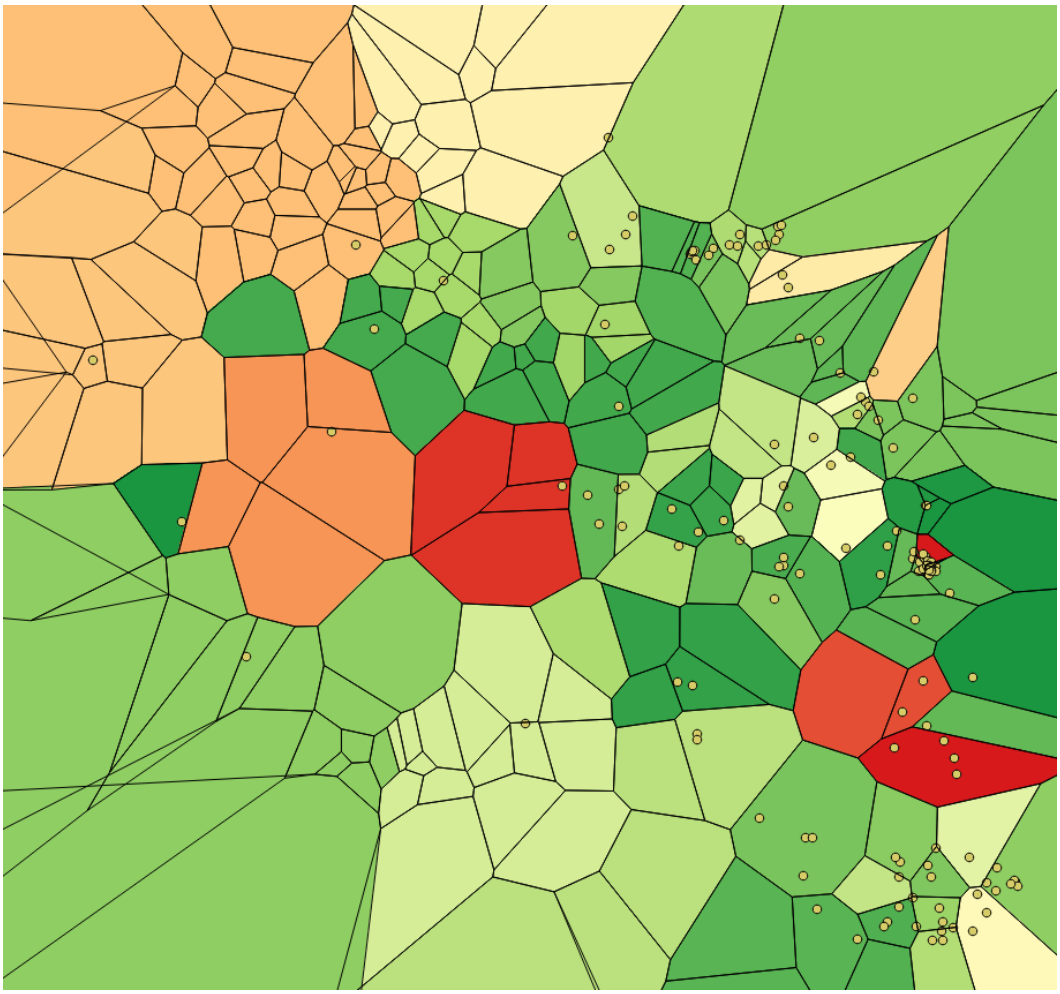
The approach above is flawed because it assumes that the value of all locations where a household survey was not conducted must be zero. Instead of assuming zero values for the empty zones, I instead apply the following algorithm to fill all of the zones. If an agentshed has 1 survey data point, use it as the average for each individual within the agent.

1. If an agentshed has more than 1 survey data point, use the average

2. If an agentshed has no survey data points, save until all other possible zones have been defined using steps 1 and 2, and then for the remaining unassigned zones, use my m-direction neighbor search algorithm:
    a. M-direction neighbor search algorithm, assuming m=4 (ideal for rasters):
    b. Starting from the agentpoint, define 4 directional vectors $v = \{v_!, v_2, v_3, v_4\}$ where 1 = up, 2 = right, 3 = down, 4 = left. For a set of vectors not equal to four, assign starting at 12-oclock with each vector $\frac{360}{m}$ degrees separated.
    c. Travel 1 grid-cell away from the agentpoint grid cell following each of the 4 vectors and check the value in those 4 cells.
    d. If any of these cells have a non-zero value, record it in the corresponding position in the set $n = \{n_1, n_2, n_3, n_4\}$ and ignore that vector for the rest of the algorithm.
    e. For the remaining vectors, continue stepwise until each has found a non-zero-valued grid-cell.
    f. The value for the target agentshed is the average if the vector's results: $\sum_{i=1}^{m} \frac{n_1}{m}$.

A first attempt at improving the "Mean or Zero" assignment rule is shown in Fig. A2.26. Here, I use the above algorithm to fill all of the zones.

Figure A2.26: Algorithmic Approach to Zone Filling



The approach shown in Fig. A2.26 is an improvement over the approach in Fig. 4.66 in situations where assuming zero for non-observations would introduce bias. The approach in Fig. A2.26, however, assumes that an empty zone takes on the value of its nearest valued neighbor no matter how distant that neighbor is. In many cases, it may be more reasonable to assume that the predictive accuracy of a neighbor smoothly decreases with distance. This is the assumption used in many interpolation methods, but typical interpolation methods do not work well in our case with defined, irregular zones and non-uniformly sampled data points.

Thus, I extend the algorithm to calculate the value as the spatial average using the distance weighted average of the z-nearest neighbors. Distance weighting for zone x ($spatial\ average_x$) is done by calculating:

$$spatial\ average_x = \frac{\frac{v_1 d_{min}}{d_1} + \frac{v_2 d_{min}}{d_2} + \cdots + \frac{v_z d_{min}}{d_z}}{\frac{d_{min}}{d_1} + \frac{d_{min}}{d_2} + \cdots + \frac{d_{min}}{d_z}}$$

Where v_i is the value in the i-th nearest cell to cell x and d_i is the distance to cell x of the i-th nearest cell for $i \in \{1,2, \dots, z\}$. This equation considers the value of each of the z nearest neighbors to the target cell but considers the nearest cells more important. d_{min} is the distance to cell x of the nearest cell.

Appendix 3: Supplemental Information for Chapter 3

The traditional commons problem that exhibits strong free-riding can be expressed as

$$\max_{c_{it},\, a_{it}} \tilde{u}_{it}(c_{it},\, X_t(X_{t-1}, a_{it}, a_{\sim it}))$$

$$Subject\ to\ p_c c_{it} + p_a a_{it} \leq w_{it}$$

in which we define player i's payoff at time t as $\tilde{u}_{it}(c_{it},\, X_t(X_{t-1}, a_{it}, a_{\sim it}))$, where c_{it} is consumption, a_{it} is abatement and $X_t(X_{t-1}, a_{it}, a_{\sim it})$ is the total GHG emissions present at time t, which is a function of the stock in the previous period, X_{t-1}, and the abatement activities of all players. Given these payoff functions, together with prices of consumption and abatement, p_c and p_a, and wealth w_{it}. Normalizing prices to unity and assuming two identical players, i and j, we can re-express consumption in terms of wealth and abatement. Further, assume each player makes a single mitigation choice at the beginning of the game, which results in one time period. We now have the maximization problem:

$$\max_{a_i} \tilde{u}_i \left(a_i, X(a_i, a_j)\right)$$

Which is solved by taking first order conditions with respect to a_i: $\frac{\partial\, \tilde{u}_i(a_i, a_j)}{a_i}$ and defining best response functions for each player, $br_i(a_{\sim i})\ \forall i$. These can be solved as a system of equations to find the optimal abatement. Note that this solution is identical to our reciprocity model if we set $\beta = 0$.

The level of abatement in the strong free-rider's problem will be less than the socially optimal solution to the climate dilemma. To identify the socially optimal level, we consider a benevolent social planner that solves the following modified commons dilemma:

$$\max_{a_1, a_2, \ldots, a_n} \sum_{i=1}^{N} \left(\tilde{u}_i(a_1, a_2, \ldots, a_n) \right) - X(X_0, a_1, a_2, \ldots, a_n) \Big)$$

The solution to this model is found by calculating $\frac{\partial \tilde{u}_i(a_i,a_j)}{a_i}$ for each i and solving as a system of equations. The difference in this problem is that the social planner is maximizing aggregate utility and choosing all abatement levels. Our reciprocity model defines the same allocation as the social planner's problem so long as $\beta = \frac{n}{n+1}$. The level of utility in the social planner's problem is the maximum utility possible .

Together, the self-centered and socially optimal solutions define a spectrum of outcomes ranging from the worst-case scenario to the best. Our reciprocity model results in a solution necessarily between these two bounds.

Application to Forestry Management

To end this section, I will apply the commons reciprocity utility model to a different commons situation: sustainable forestry vs. clear-cut logging. Forest management results in commons dilemmas similar in type (though obviously not in scope) to climate change. This section applies the extended utility framework I have developed to understand how individuals operate in common-property forests by defining an agent based simulation (with the methods in Chapter 3). This section does not present results due to time constraints approaching the defense of this thesis (which is unfortunate because the results were almost done).

I define the forest management agent based simulation as

$$Forest\ Management\ Simulation = \{N, M, A, R, T, S\}$$

The time periods and iteration steps, T and S, are similar to chapter 3 but now emphasis will be on a more complex growth function for trees in between time periods. For simplicity, I assume a very small network, where N is 15 by 15 grid-cells and the normalization method M is identical to chapter 3 (gridded-normalization to matrices). The abundance matrix will be assumed to be uniform (arbitrarily defined so that each grid-cell has 3 cargo-loads worth of the objective) but the location of agents will be irregular. The main elements that differentiate this simulation from chapter three include the set of agents and behavior rules.

For the set of agents, I use the notation and decision structure of the commons reciprocity utility model. Thus defined, agents choose between a private consumption activity (clear cutting a forest grid-cell) or practicing sustainable harvest (which will accelerate the tree

regrowth rate). These actions are defined as c_i and a_i, but in addition to choosing the level of each variable, the agent must also specify in which grid-cell it does its action. Thus, while omitting the x, y subscript to represent the geospatial location on each variable and including a discount rate of r, the agent solves:

$$\max_{a_{it}, c_{it}} \int e^{-rt} \Big(u_i^D\big(a_{it}, c_{it};\ a_{jt}, N_{it}, R_{j\in\sim i}\big) + \beta_{ij} R_{ij}(\cdot)\, u_j^D\big(a_{jt}, c_{jt};\ a_{it}, N_{jt}, R_{\mathrm{i}\in\sim i}\big)\Big)\, dt \quad (17)$$

The above function will likely remain unsolvable in analytical form, due the complexity of the optimal control problem of choosing harvest rates on multiple grid-cells over continuous time with potential for rival consumption, but it can be approximated with an iterative algorithmic approach (as discussed in Chapter 3).

I define the response function between i and j as the ratio of potential utility

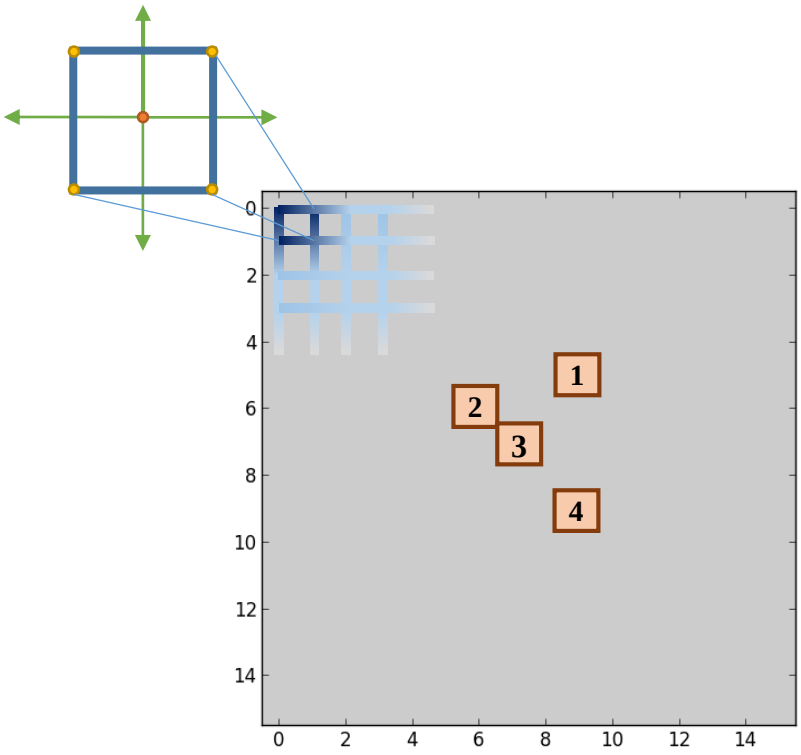
$$R_{ij}(\cdot) = \frac{L_i(A_{jt})}{L_j(A_{it})}$$

Where L_i represents a "potential lost" function that calculates the potential utility that could have been gained by agent i if agent j had not clear-cut the forest in a given cell, and where $A_{it} = \{a_{ik}\}_{k=0}^{t}$, the cumulative choices of agent i up until the current time t. This response function is much simpler in words than in math; agent i responds positively if j has made them lose less potential utility than they have made j lose. So, agent i will regard agent j positively if j has not caused them more harm than i has done themselves. Note that this example also incorporates time more than the climate change example whereby the response changes with the cumulative actions of the other player(s). It may be best to discount actions that happened in the more distant past when calculating the cumulative potential lost, but for now I assume that all losses are treated equally.

With a response function defined like this, agents will gain utility by seeing those agents who have helped them in the past do well and by seeing those agents who harmed them do poorly. This affects the agent's choice of c_{it} and a_{it} because each agent knows their actions can have negative impacts on neighbors, and thus the agent can reward cooperators by not clear-cutting near their location and can punish defectors by the opposite. I summarize the notation presented so far (excluding only behavior rules) in Fig. A3.1. In the upper left of this figure, there is one example grid-cell, but this is

defined for every location in the 15 by 15 landscape. I define 4 agents on this landscape, denoted by numbered squares.

Figure A3.1: Definition of Agent Based Simulation for Forestry Management

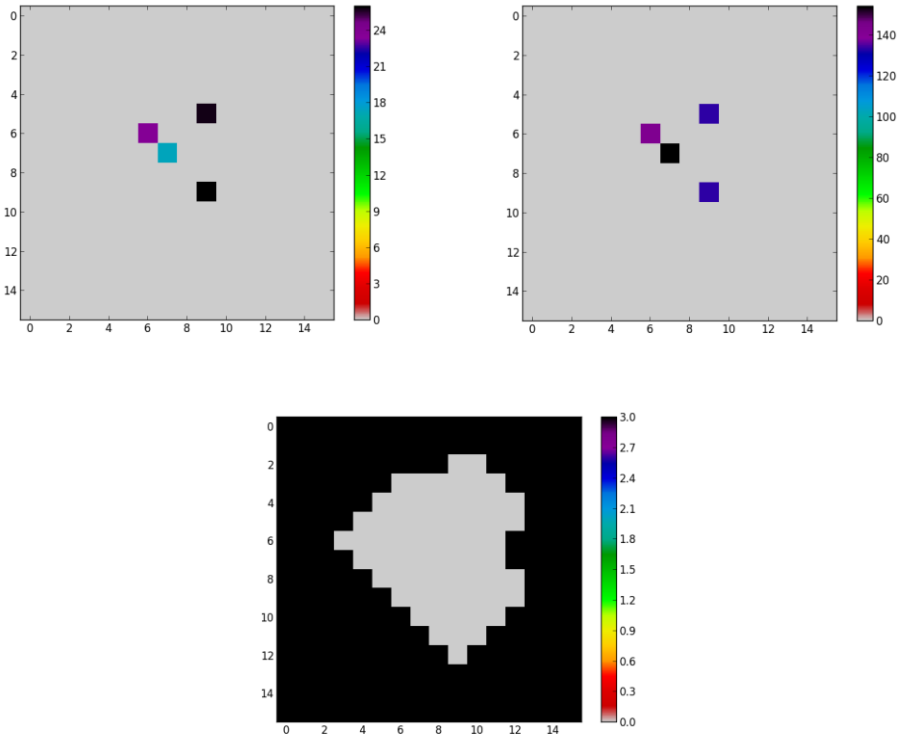


The second key extension from chapter 3 is deriving behavior rules from this relatively more complex maximization problem defined in the commons reciprocity utility model. The basic approach remains the same: agents will analyze every grid-cell within their zero-profit boundary an choose to do whichever action(s) yield the maximum marginal gain, as defined in equation 17. I make the simplifying assumption that agents respond only to past actions of other players and do not strategically anticipate future actions (however, they will consider the future value. This assumption means that all of the complexity that arises from including reciprocity is fully accounted for in equation 17. The only additional complexity we must consider is that the agent is now choosing

between two potential actions with a long horizon of regrowth. This is an example optimization that is ideal for the use of genetic algorithms to find the best decision rule (see Deb et al. 2002 for discussion of genetic algorithms and their application to difficult optimizations like this). I have not yet implemented these algorithms due to time constraints, so instead I present preliminary results from a simplified version of this model and discuss how it can be extended.

The main simplification I make is to assume clear-cutting is the only source of wood and that sustainable forestry is defined as just letting the forest regrow (thus reframing the maximization as a choice between cutting now or waiting to cut). I use a basic allometric regrowth function using parameters chosen arbitrarily from dipterocarp forests present on the island of Borneo (see Kohyama et al. 2003 for figures and functions). Finally, I simplify the model so that each agent only considers the harm from other agents in the previous time period and assume that agents are completely myopic to the dynamic optimization elements of this problem. Fig. A3.2 shows results from this simplified version of this model. The image in the upper left represents the utility gained by each agent from firewood cut (assuming a simple direct utility function similar to that used in the climate example). In the upper right I present the average travel time expended by each agent per time period. Finally, the image in the bottom row shows the supply of forest that remains uncut in equilibrium. These results are very preliminary and intended only to be illustrative of the possibility of combing the work in chapters 3 and 5. Nonetheless, the results seem quite intuitive insofar as the agent located in the middle travels farther and gains less utility than the other agents.

Figure A3.2: Application of Agent Based Simulation to Forest Management with Reciprocal Utility



Although these results are very preliminary, they suggest future progress can be made through pursuing this approach.