Land Use Change and Economic Opportunity in Amazonia: An Agent-based Model

by

Arthur Raymond Cabrera

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Economic changes such as rising açaí prices and the availability of off-farm employment are transforming the landscape of the Amazonian várzea, subject to decision-making at the farming household level. Land use change results from complex human-environment interactions which can be addressed by an agent-based model. An agent-based model is a simulation model composed of autonomous interacting entities known as agents, built from the bottom-up. Coupled with cellular automata, which forms the agents' environment, agent-based models are becoming an important tool of land use science, complementing traditional methods of induction and deduction. The decision-making methods employed by agent-based models in recent years have included optimization, imitation, heuristics, classifier systems and genetic algorithms, among others, but multiple methods have rarely been comparatively analyzed. A modular agent-based model is designed to allow the researcher to substitute alternative decision-making methods. For a smallholder farming community in Marajó Island near Ponta de Pedras, Pará, Brazil, 21 households are simulated over a 40-year period. In three major scenarios of increasing complexity, these households first face an environment where goods sell at a constant price throughout the simulated period and there are no outside employment opportunities. This is followed by a scenario of variable prices based on empirical data. The third scenario combines variable prices with limited employment opportunities, creating multi-sited households as members emigrate. In each scenario, populations of optimizing agents and heuristic agents are analyzed in parallel. While optimizing agents allocate land cells to maximize revenue using linear programming, fast and frugal heuristic agents use decision trees to quickly pare down feasible solutions and probabilistically select between alternatives weighted by expected revenue. Using distributed computing, the model is run through several parameter sweeps and results are recorded to a central database. Land use trajectories and sensitivity analyses highlight the relative biases of each decision-making method and illustrate cases where alternative methods lead to significantly divergent outcomes. A hybrid approach is recommended, employing alternative decision-making methods in parallel to illustrate inefficiencies exogenous and endogenous to the decision-maker, or allowing agents to select among multiple methods to mitigate bias and best represent their real-world analogues.

Acknowledgements

This thesis would not have been complete without the input of several individuals, to whom I express my thanks.

I would like to first express my gratitude to Prof. Peter Deadman, my advisor during my master's degree and my employer in an earlier co-op position, for introducing me to agent-based modelling and to the field of geography. My experiences in geography have been a pleasure. Thank you for the opportunities to apply myself to a number of fields such as land use change, landslide modelling and interactive learning. Most of all, thank you for your advice and support.

I thank Prof. Alexander Brenning for such opportunities and for introducing me to spatial analysis and other analytical methods, which have been invaluable in the preparation of this thesis. Thank you for challenging me on numerous occasions, from spatial statistics assignments, our collaboration on Carla, to your invitation to guest lecture a spatial methods course.

I also thank my committee, Prof. Doug Dudycha, and the readers of my thesis, Prof. Bob Sharpe and Prof. Rob Feick, for their time and feedback. Their input and discussion has led me to explore new insights within the field of simulation modelling and its ability to complement other analytical methods.

Thanks to Kristina Lüüs for assisting in the initial development of the environmental model and for our collaboration in the early stages of this thesis.

I thank Eduardo Brondízio, Andréa Siqueira, Nathan Vogt, Scott Hetrick, Miguel Piñedo-Vasquez, Christine Padoch and Robin Sears for providing me with the knowledge and data related to land use change in the Amazonian várzea. I also thank the National Science Foundation (Human and Social Dynamics) and the Department of Geography and Environmental Management for their funding and support.

Dedication

To my parents, Arturo and Aida Cabrera,
and Anna, Louis, Matteo and Kai.

And, to Carol.

Thank you all for your unwavering love, encouragement and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Agent-based modelling has recently become an important tool of land use science. Building upon the scientific methods of induction and deduction, simulation has become a "third science", allowing researchers to codify assumptions and test them in a virtual environment. An agent-based model is a specific type of simulation, built from individual interacting entities known as *agents*. Agents have the ability to sense aspects of their environment, but are not necessarily perfect observers. They possess some cognitive ability, ranging in complexity from reactive to deliberative.

Agent-based models have been applied to the study of land use/cover change (LUCC) in the last decade (Parker et al., 2003). ABM/LUCC are often coupled with one or more layers of cellular models to represent the agents' environment. Cellular transition rules simulate natural transitions and socioeconomic institutions such as land markets. Examples of such models include Fearlus (Gotts et al., 2003; Polhill et al., 2001), Sypria (Manson, 2004, 2006a) and Lucita (Deadman et al., 2004; Lim, 2000; Robinson, 2003). Early versions of Fearlus, such as the version described by Polhill et al., emulated no specific study area but served to explore the land use actions of land managers in a spatially and temporally heterogeneous environment. Other models were developed as case-specific models, based on field data and remotely sensed images. Sypria, for instance, was used to explore the Southern Yucatán Penninsula and the area of Transamazon Highway west of Altamira, Brazil was simulated by Lucita.

With the same basic design as these agent-based models, a new agent-based model, Maria (Multi-Agent Reasoning in Amazonia), is designed to explore alternative decision-making strategies in the context of certain drivers of land use change in the Amazonian floodplain, focusing on the community of Paricatuba near Ponta de Pedras, Brazil. Like Sypria and Lucita, agents in this model are capable of agricultural activities, in addition to other economic activities. While Lucita used a classifier system and, later, decision-tree heuristics, and Sypria used genetic algorithms, neither of these models were used to evaluate significantly differing alternative models side by side. The choice of decision-making method is a matter of debate (Schreinemachers and Berger, 2006), but to date, a case-specific agent-based model has not been presented with alternative decision-making methods.

1.2 Economic opportunities and land use change in the Amazonian várzea

Throughout the history of the Amazonian *várzea* (floodplain forest), local households extracted, cultivated and consumed açaí palm fruit (*Euterpe oleracea*) and manioc flour as a significant part of their diet (Brondízio et al., 1994; Murrieta et al., 1999; Wallace, 1853). The last few decades have seen a boom in the açaí market, which can be characterized in stages. Prior to the 1970s, açaí was consumed and extracted from the local forest and made available in markets in the rural estuary. As early as the 19th century, noted by Wallace, açaí was routinely prepared as a sweet beverage when mixed with water and sugar. In the past, açaí harvesting practices were characterized as extractivist, but have recently been recognized as agroforestry management (Brondízio and Siqueira, 1997).

In the 1970s, the migration of low-income earners from rural to urban areas has led to a trend of "ruralization". This term does not describe emigration toward rural areas, but refers to an increasing influence of rural preferences in urban centres. Specifically, ruralization has led to an increased demand of açaí in urban markets, transforming it from a strictly rural and indigenous staple to an urban staple food by the 1990s (Brondízio, 2004). Açaí soon became available in urban and regional markets, in addition to those found in the rural estuary.

More recently, ease of transportation, better açaí preservation technology, and the trend of globalization has opened up markets for açaí on a world stage. Beginning in the mid 1990s, demand for açaí in goods as a healthy *fashion food* has led to a dramatic increase in the price of açaí (Brondízio et al., 1994). Açaí in a refined form, derived from açaí pulp and powder, can now be found in grocery stores worldwide within products such as health drinks and yogurt.

These trends have had a net positive effect on the price of açaí, leading to an increase in intensive açaí management, from smallholder households to commercial farms (Brondízio, 2004). In addition, the difficulty in transporting açaí has led to additional economic opportunities such as açaí trading (*marretagem*).

1.3 Motivation for research

In the Amazonian *varzea*, land use change is resulting from shifting economic opportunities. Recent economic changes include rising selling prices of goods and the availability of alternative sources of employment. Traditional farming practices in this region include extensive cultivation, such as intensive açaí management and the farming of annuals and biannuals, and extraction of açaí fruit and timber. This, in turn, has led to a further intensification in açaí management in an effort to produce more yield. However, the degree to which this has occurred is difficult to evaluate.

The observability of land use change is made difficult by the challenge of differentiating unmanaged floodplain forest from intensively and intermediately-managed açaí agroforestry in remotely-sensed imagery. Field observations in this area were recorded for a very limited number of sites in the period of 1991–1994 by Brondízio (2008). This includes the yields of intensively and intermediately-managed açaízals (açaí stands) in a few experimental sites, as well as local demographics. This data is supplemented by classified Landsat and IKONOS images, which roughly estimate the degree of intensive açaí production and other land cover such as savannas. However, the resolution of these images limits their usefulness as housegardens are typically too small and sparse to be shown. An agent-based model, a bottom-up system designed based on assumptions of social institutions, economic opportunities and constraints, will aid in the exploration and explanation of drivers of land use and economic change. Often facilitated by a coupled cellular model, an agent-based model can also produce spatial output. A model of incentives and constraints is developed, presenting utility-maximizing agents with an environment of agricultural and economic activities. Specific variables to be examined include the prices of goods, labour requirements of agricultural practices and employment requiring temporary or permanent emigration from the farm.

With respect to the decision-making capacity and strategy taken by agents, the use of rational versus boundedly rational decision-making methods is subject to much debate (Gigerenzer and Todd, 1999; Schreinemachers and Berger, 2006). Agent-based models in studies of LUCC are typically implemented with only one major type of decision-making method at a time. Fearlus is one exception, a version of which was tested among many imitative, optimizing and randomizing algorithms (Polhill et al., 2001). A model by Jager et al. (2000) developed two types of agents, *Homo economicus* and *Homo psychologicus*, differed by their aspiration levels and perceived uncertainty, and in turn, the decision-making method employed (among deliberation, repetition, imitation and social comparison). However, neither of these models was applied to the study of a real-world study area. Overall, boundedly rational agents are the most common in ABM/LUCC, as they characterize local decision-making without assuming infinite cognitive capacity on the part of the decision-maker.

This thesis attempts to evaluate the utility of rational versus boundedly rational agents in the context of land use change in the Amazonian *várzea*. This comparison will be made in a *case-specific* model, unlike the general, theoretical environments used in Fearlus, a model by Jager et al. (2000), and Axelrod's Prisoner's Dilemma tournaments (Axelrod, 1984), where such comparisons have already been made.

1.3.1 Problem definition

Given the rapid economic changes in this area and the difficulty in observing land use change, an alternative approach to facilitate scientific exploration is desired. Agent-based models are beginning to be proven as a scientific approach, using bottom-up design to test assumptions. The interactions of agents in these models often produce emergent effects which may be difficult to predict—based on the design of individual agents—before the model is tested. That is, by designing individual rules from the bottom-up, complex phenomena can be produced from a comparatively simple model.

The decision-making method employed by agents has been identified as a research issue (Schreinemachers and Berger, 2006), but to date, a comparison has not been made for a real-world study area. Between two broad classes of decision-making, optimization and bounded rationality, which is more suitable for a case-specific agent-based model of land use/cover change? Using heuristics as an example of bounded rationality, this thesis will explore this question in the context of shifting economic opportunities within the Amazonian *varzea*.

1.4 Goals and objectives

Using an agent-based model, this thesis sets out to explore decision-making methods employed by agents in the Amazonian floodplain within an environment of rising selling prices and other economic opportunities related to intensive forest management and agriculture. The role of multi-sited households are introduced, in the current incarnation of the model, as resulting from migration to pursue employment, but other economic linkages and reciprocities based on property encroachment, fishing and trade exist in the area. This thesis will include migration-related linkages only, explicitly modelling the movement of labour between farming and non-farming practices.

Furthermore, the question of the utility of rational agents will be explored in comparison to fast and frugal heuristics. In separate simulations, two populations of agents will be formed, each utilizing one decision-making method between rationality and heuristics. These populations will be compared in terms of land use trajectories and relative economic success. Fast and frugal heuristics are quick and require little information on the part of the decision-maker, but require that case-specific beliefs, desires and decision-making methods must be modelled explicitly. In contrast, rational agents require the definition of an objective function, codifying the needs and desires of an agent in terms of an expression to be maximized or minimized, subject to constraints which are expressed as equations. The rational decision-making process is treated as a black box, producing the optimal solution without emulating case-specific decision-making methods. Positive and negative arguments exist for both approaches, but an explicit comparison between the two methods in a case-specific ABM is yet to be made.

These two primary goals will be realized using an agent-based model named Maria (Multi-agent reasoning in Amazonia). Some components of Maria are influenced by an earlier model, Lucita, which was developed as a pioneering, deforestation model of the Transamazon Highway near Altamira, Brazil. However, the biophysical and socioeconomic attributes of the study areas differ greatly, especially as Lucita was developed for an upland area far from the floodplain study area modelled by Maria. This necessitates the creation of a new model, since most assumptions from one model cannot be said to hold in the other. Maria is developed as a modular system, allowing for the addition and replacement of agents and methods. This facilitates rapid evaluation of alternative decision-making models under a variety of scenarios, including price variation and urbanization. As this model is being created in parallel with ongoing research, this system should also accept new data and better models (such as demographic or soil models) as they become available. The objectives of this thesis are:

1. Develop a modular architecture and implementation of Maria, integrating assumptions derived from Brondízio (2008) and other sources while considering future uses of the model and expected data.
2. Evaluate and compare the suitability of optimizing and heuristic algorithms for ABM/LUCC using scenarios based on theoretical and empirical data, where available.
3. Qualitatively evaluate spatial land use allocation, as it reflects the transition from shifting agriculture to intensive management in the model.
4. Explore the drivers of land use change, considering market prices, labour requirements and multi-sited households.

1.5 Structure of this thesis

This chapter presented an introduction to agent-based models and a few examples of applications in land use science. Some of the drivers of land use change in the Amazonian estuary were discussed, as were some of the issues concerning decision-making in agent-based models. Chapter 2 is comprised of a brief history and description of the study area within Marajó Island near Ponta de Pedras, Pará, Brazil, focusing on smallholder farming characterized by the small rural community of Paricatuba. Chapter 3 presents the current state of agent-based models and decision-making methods, discussing the distinction between rational and boundedly rational models. The design, implementation and analysis methodology of Maria is discussed in Chapter 4, followed by results in Chapter 5. The final chapter discusses the limitations of the model as well as potential future uses of Maria.

Appendix A describes the distributed computing strategy used to perform large parameter sweeps and Monte Carlo simulation across an ad hoc cluster of computing nodes. Appendix B contains the list of parameters used to generate the model results presented in Chapter 5.

Chapter 2

Land Use and Economic Opportunity in Marajó Island, Brazil

2.1 Introduction

The purpose of this chapter is to provide the reader with some context of the forces affecting land use change and decision-making in the riverine Brazilian Amazon. The content of this chapter is the basis of the design of the human and environmental models forming MARIA. In particular, MARIA will develop selected components of farming and other local activities, such as açaí cultivation and emigration leading to multi-sited households. Other characteristics of the study area, though not implemented in the first version of the model, will be provided for through its design, which will provide a framework to implement these in the future. This chapter first presents a background and a short history of Marajó Island, while the subsequent sections discuss the present demographic, social and economic characteristics of three communities within Marajó Island near Ponta de Pedras. The chapter concludes with a discussion of social changes and potential implications on land use change.

2.2 Geography

Ponta de Pedras is a municipality located in the lower Amazonian estuary at the southeast part of Marajó Island, in the state of Pará, Brazil (Brondízio et al., 1994). The Pará River flows to the east and south of the island. The annual mean temperature of Ponta de Pedras is 27 °C (Murrieta et al., 1999). There are two main seasons, rainy and dry. Average monthly rainfall ranges from approximately 500–800 mm between December and April, dropping from 400 to nearly 0 mm between May and November. Total rainfall is approximately 3000 mm/year.

Ponta de Pedras has a population of 25 743 and is located just west of Belém, the capital of Pará, a city of 1 408 847 (IBGE, 2007). Near Ponta de Pedras, there are three small farming communities, each characteristic of a distinct type of farming arrangement to be discussed later in this chapter (Section 2.4): Paricatuba (smallholder), Marajó-Açu (sharecropper), and Praia Grande (co-operative). These communities are shown in Figure 2.1 (Environmental Systems Research Institute, 1992; MDA Federal, 2004). In 1994, these three communities had populations of 144, 371, and 117, respectively. Paricatuba is a small riverine community located south of Ponta de Pedras, along the Paricatuba river. Marajó-Açu is located northwest of Paricatuba on the north side of Rio Marajó-Açu. It, too, is a riverine community. Upland, to the east of both communities, north of Rio Marajó-Açu, is the community of Praia Grande.

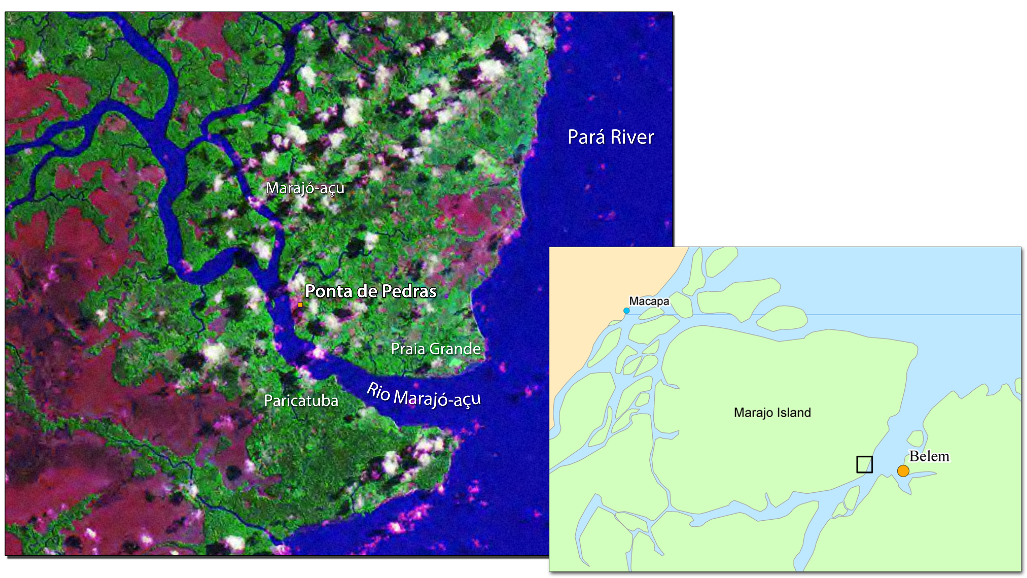


Figure 2.1: Communities near Ponta de Pedras, on a Landsat 4/5 TM image.

Marajó Island contains regions ranging from upland forest, floodplain forest and grassland savannas. Transitional forests are present between areas of forest and savannas. The community of Paricatuba, representing the smallholder focus of this thesis, is located in the floodplain region. This area consists primarily of dense floodplain forest (várzea), with several species of palm as well as dominant families of leguminosae and arecaceae (Brondízio et al., 1994). These species are highly adapted to frequent tidal flooding. Unlike many other areas of the Amazon, characterized by small numbers of many species of vegetation, the várzea is dominated by relatively few species. As such, this area of forest can be classified as an oligarchic forest (Peters et al., 1989). In terms of individual numbers, açaí (*Euterpe oleracea*) is one such dominant species in this area. Açaí is a slender palm tree, about 60–80 ft. high, providing fruit in the form of berries (Wallace, 1853). The fruit of the açaí palm is a local staple, comprising as much as 30% of the diet of the local population (Murrieta et al., 1999). Although açaí is the dominant palm species in the area, other palm trees are economically significant, providing goods such as lumber and medicinal oil. Açaí may also be used for lumber, but is often regarded as too valuable for this purpose due to its fruit. Overall, the floodplain forest is spatially heterogeneous, with areas regarded as unsuitable for intensive management due to existing land cover or topography (Brondízio, 2008).

The built environment in the community of Paricatuba consists primarily of wooden homes and elevated walkways. Paths near the household are elevated above the tide. In contrast, pathways in the upland and transitional forested areas are created through the annual burning of dense grass.

Transportation is facilitated through the many streams and rivers in the area. The nearby state capital city of Belém is located across the Pará River and is accessible from Ponta de Pedras by a five-hour ferry ride offered twice daily. Another mode of transporting goods to market is through the employment of middlemen, though some smallholders have invested in motorboats to deliver goods to market directly (Brondízio, 2008).

2.3 A brief history of Marajó Island

Marajó Island was home to some of the earliest settlements in the Amazon. The Caboclo populations are the largest non-tribal native population of the Brazilian Amazon, a racial mixture of Amerindians, Europeans and Africans (Pace, 1997). In a non-racial context, the term Caboclos is often used to refer to the poor peasantry in the Amazonian estuary. Since the term "Caboclo" is also a stigma, implying degrading racial connotations, it should be noted that the term is used by Brondízio (2008) and Siqueira (1997) as a social category in reference to a portion of the Amazonian peasantry, uniquely identifying this population to contrast them from more recent immigrants. Much debate has been ensued over the use of this term (Cleary, 1993; Nugent, 1993; Pace, 1997), in terms of racial overtones and its adequacy for describing the indigenous population, but that will not be addressed in any further detail here. Another term used in the literature includes *ribeñeros* (riverine people), but this term applies to a much wider range of people and has been used to describe similar peoples in Peru (Brondízio et al., 1994). Meanwhile, the Caboclo identifier applies to urban residents with roots in this community. As described by Siqueira, the term Caboclos is used in an academic context to refer to the traditional, rural occupants of the area to contrast them from recent immigrants. For lack of a suitable alternative term (Brondízio, 2009), the Caboclo identifier is used here. Encouragingly, however, one community has been recorded as identifying itself as Caboclo, in a positive sense, to reflect its identity as native to the area with deep knowledge of the forest (Silva-Fosberg, 1996, cited by Siqueira, 1997). Ideally, the use of the term Caboclo with a positive meaning can help wear away the social stigmas of the past and recognize this population as uniquely knowledgeable.

The Caboclos cultivated and managed the Marajó Island region for the last 150 years. Traditional activities of the Caboclos include fishing, shrimping, swidden agriculture, agroforestry (management and extraction of forest products) and gathering (Murrieta et al., 1999). These activities are a result of the assimilation of Portuguese and other European immigrants whose culture combined with that of the native Amerindians (Siqueira, 1997). European colonization also involved the subjugation of the Caboclos, resulting in uprisings in the mid-19th century with little long-term effect. The rubber boom of the late 19th century and early 20th century resulted in further immigration and assimilation into Caboclos society.

Demand for rubber derived from the local plant *Hevea brasiliensis* resulted in an economic boom in the late 19th century, followed by a bust in the early 20th century. Accordingly, most available labour was allocated to rubber production, resulting in the abandonment of much of the food plantations in the estuarine region. Caboclos labour and knowledge were both highly useful in harvesting naturally-growing rubber in this area. However, as Caboclos typically did not own land, they would extract rubber for landowners. Under a patron-client relationship (Siqueira, 1997), they would be supplied all rubber-tapping equipment exclusively from the landowner at above-market prices and forced to sell all rubber to the landowner at the landowner's price, keeping the worker indebted to the landowner. This kept the rubber boom from improving the economic welfare of the Caboclo people. Eventually, the rubber economy in the Amazon declined as seeds were smuggled to create large plantations in

Malaysia. Uncontrolled tapping also led to decreasing yields, year by year. A smaller rubber boom followed during World War II, since the Allies did not have access to Malaysian rubber, but this boom was followed by a bust at the end of the war.

Economic development programs have since been sponsored by the government, attempting to integrate and develop Caboclos society. Many of these programs have been regarded as failures, due to their inability to improve their economic welfare. Instead, programs often focus on replacing traditional practices with technology to improve export. Investments have been said to "manufacture invisibility" of the Caboclos, keeping them without new infrastructure or better living conditions (Brondízio, 2009). Initially, government programs favoured small landownership, but later abandoned this in order to encourage large-scale production by the mid-1970s. In combination with the development of highway infrastructure, the latter large-scale programs resulted in significant deforestation and the displacement of landless farmers westward. Many of the Caboclos lost their land during this process. By the mid-1980s, local knowledge and tradition was seen as a potential economic boon, in addition to providing environmental sustainability (Padoch et al., 1985).

A boom in açaí has been present since the 1970s. Urbanization of nearby centres such Belém has led to an increase in "rural" preferences in these areas, including diet. This, in combination with economic programs intended to increase the export of goods (including açaí) from the Amazon, resulted in its availability in wider areas. The price of açaí increased fourfold in the approximate period of 1984–1994 (Brondízio et al., 1994) as it became an urban staple. A worldwide trend of fashion food has increased demand for açaí further, as it is valued for its antioxidants and its environmental sustainability (Brondízio, 2004). This açaí boom has translated in a shift in the production of açaí, evolving from indigenous extraction to intensive management, and later, corporate farms. Small-scale production remains viable for both smallholders and sharecroppers, as evidenced by continuing production trends in the community of Marajó-Açu and, to a lesser extent, Paricatuba. Marajó-Açu has moved largely to intensive açaí management, while Paricatuba maintains a significant amount of traditional forest-fallow agroforestry.

2.4 Farming households and activities

Farming household arrangements in the Amazonian estuary can be classified as sharecroppers, co-operative communities and small landowners (Brondízio, 2008).

Sharecroppers are farmers who reside on and cultivate land for a resident or absentee landowner, sharing their harvest with them based on an informal agreement. For instance, such an arrangement may assign half of açaí production to the owner. More informally, this arrangement may preclude the granting of prized game to the landlord, while the owner will provide needed medicine to the sharecropping family. Sharecroppers are assigned a clear designation of açaí stands, which they reside near, while areas for extraction may be shared among many sharecroppers (though only within land owned by the same landlord). In other cases, such as crop production, areas and land uses are explicitly delineated by or encouraged through prior arrangements.

Medium landowners (*proprietários*), which include both resident and absentee landlords employing a small number of sharecroppers, own 50–200 ha of land. In contrast, large landowners (*fazendeiros*) own over 200 ha of land or more. Typically, these large landowners are non-resident, dividing their land among many sharecroppers. As many as dozens of sharecroppers occupy a property, each residing near their assigned açaí groves. Smallholder houses in Paricatuba are separated by a distance of 20–300 m, varying to a range of 10–500 m apart in the mainly sharecropping community of Marajó-Açu (Brondízio, 2008; Siqueira et al., 2000).

Açaí production practices are often characterized as extractivist, despite the intensive management practices which are followed in many cases (Brondízio and Siqueira, 1997). The management of açaí can be divided into steps: selective thinning of undesirable species, pruning, planting of palm seeds and annual weeding. Undesirable species include other forest species, understory vegetation and vines.

These and other agricultural activities performed by the sharecroppers may have longer-term implications on their land tenure rights. Sharecroppers do not own their own land, nor do they acquire their land over time. However, they may claim compensation if the land has increased in value during their residence. For this reason, landlords often restrict the practice of infrastructure building and slash and burn agriculture, since these actions may increase the value of the land and thus the sharecroppers' rights to land tenure or compensation. In contrast, other practices which increase return on land without implicitly providing additional land tenure rights may be encouraged. Examples of such practices include intensive açaí agroforestry and swine husbandry. In one experimental area discussed by Brondízio (2008), a sharecropper was restricted to thinning and pruning and was unable to perform more intensive management of açaí. Due to these restrictions, a household may not be able to grow subsistence crops. However, households may be able to purchase these and other desired goods, such as manioc flour, through profits from açaí production (Siqueira, 1997).

In a sharecropping arrangement, landlords control when the harvest is reaped. Landlords may schedule harvesting during September to November, when prices are lower, forcing sharecroppers to harvest unripe fruit which would be better suited for harvesting in later months for a higher profit. Local interviews conducted by Brondízio (2008) indicate that an unlimited amount of açaí can be consumed by the sharecropping household, but their studies have indicated that landlords may reserve açaí as it becomes scarce or profitable.

The length of sharecropper tenure in this area varies from under one year to as many as three generations. However, among newer sharecroppers, there is a great deal of turnover: More often than not, sharecroppers do not stay long due to mutual mistrust. Between 1990 and 1994, more than 50 % of sharecroppers emigrated and were replaced by new sharecroppers (Brondízio, 2008). Absentee landlords believe sharecroppers are withholding açaí production, while sharecroppers feel exploited by their landlords.

There is no evidence of sharecroppers who later become owners of their assigned land. Instead, the tradition of sharecropping has often been broken through external intervention. For instance, the Roman Catholic diocese in the area has purchased land from large landowners to establish co-operatives, such as COPIUPPE in Praia Grande (Siqueira, 1997).

A number of small landowners have been able to inherit or purchase land from a larger land owner. Small landowners, or smallholders, are farmers who cultivate their own land, owning as little as 1 ha of land to as much as 50 ha. Another community near Mazagão is held by smallholder farmers, who do not hold official tenure to the land, but treat it as such (Menzies, 2007).

Smallholder farmers are free to exercise actions based on their own interests, but are still subject to constraints of land, labour and capital. Small landowners perform intensive and intermediately-managed açaí agroforestry, like sharecroppers. However, without restrictions on agricultural activities, small landowners are able to practice intensive management as well as swidden agriculture on both floodplain and upland forests (Brondízio et al., 1994). This includes the production of *roçado de várzea*, a type of floodplain garden involving a mixture of annuals and biannuals, followed by the planting of açaízals (Brondízio et al., 2002).

Households in the mainly smallholder community of Paricatuba practice shifting agriculture, shrimping, fishing, hunting, forest cultivation and extraction of other forest products (Siqueira et al., 1993, cited by Siqueira, 2009). Paricatuba is representative of traditional farming, in the sense that they practice diverse agricultural activities. Siqueira et al. (2000) note that in this community, unlike Marajó-Açu and Praia Grande, economic activities do not compromise any subsistence activities.

Across these three study areas, where swidden agriculture is practiced, the fallow period is about 5 years. This period depends on available labour, available land and quality of soil (Siqueira et al., 2000). In other riverine study areas, the fallow period is closer to 10 years. However, this *swidden-fallow* agriculture does not refer to the complete abandonment of the fallowed plot, but to the mixture of annual crops and perennials with natural forest regrowth (Dufour, 1990). In this manner, farmers are able to extract resources from the area throughout the fallow period.

2.5 Demographics of local communities

Three populations near Ponta de Pedras were studied by a multi-disciplinary, multi-institutional team from 1989 onward (Siqueira, 1997). These populations are characteristic of three types of farming household arrangements in the area: small landownership, sharecropping and co-operative. Paricatuba, located south of Ponta de Pedras, comprises 21 households, 26 families and 144 individuals, according to demographic data collected in 1994. Paricatuba consists primarily of small landowners, but includes two absentee medium owners (owning approximately 50–200 ha. of land each) employing four sharecroppers in total.

Marajó-Açu is a community of 43 households, 46 families and 371 individuals. More than 65 % of households in Marajó-Açu are sharecroppers, working the land of three large land owners. Most of these sharecroppers inherited the land through family, though many have arrived recently, replacing other sharecroppers who have emigrated.

Praia Grande, a co-operative community of 19 households, 21 families and 117 individuals, is located upland along the shoreline of Marajó Bay. The local Roman Catholic diocese purchased a tract of land in order to form this community. Collectively, the community maintains land ownership, as the entire community belongs to COPIUPPE (*Cooperativa Mista Agropecuária Irmãos Unidos de Ponta de Pedras*), an agropastoral co-operative. Through the development processes establishing this co-operative community, residents have moved from floodplain forest to upland terrain along a dirt road (Siqueira, 1997). This community practices mechanized agriculture, producing beans, corn, coconuts, and rice, as well as the mechanized preparation of land for cattle and pasture. Praia Grande is the only community out of these three which practices mechanized agriculture and cattle ranching. This community is supported by subsidies from the church and has eliminated the practice of swidden cultivation (Brondízio, 2009).

Households in all of these communities may also include *agregados*, extended ("aggregated") household members not necessarily linked by kinship. While these household members share no biological ties of kinship, they may sleep under the same roof and contribute labour. Other "household members" do not necessarily reside at the same location, but contribute labour, food or money (Siqueira, 2009). Such households can be regarded as multi-sited and often result from circular or impermanent migration to nearby urban centres. Unlike a single-sited household, a multi-sited household has distributed labour and capital resources, presenting a challenge in estimating labour and capital constraints.

2.6 Other economic activities

Employment in nearby urban centres has allowed individuals to pursue work other than farming. However, the limited availability of urban employment has led to impermanent or circular migration. Migration patterns differ among women and men (Siqueira, 1997): Women often leave the rural household as teenagers in order to work as maids for landowners living in Belém, returning later in life to marry local men. Men, on the other hand, practice more circular migration, moving between urban centres and their rural origin throughout life. Men pursue economic enterprises, either by commercializing their household's farm products or trading the goods of others as middlemen or brokers.

Fishing and shrimping are also important activities in this area. Though açaí and manioc flour comprise over 60 % of local diet in Paricatuba (Murrieta et al., 1999), fishing is the most important source of protein, followed by pork. The long stretch of land required to implement shrimp traps is often a source of conflict between neighbouring properties, as the catching area of shrimp often extends beyond the property line.

Outside of the local açaí growth season, males engage in açaí trading as middlemen, especially in Marajó-Açu. This trading is locally known as *marretagem* (Murrieta et al., 1989, cited by Siqueira, 2009). Açaí is available elsewhere in the region, but unavailable locally. They travel up to 2–4 days by boat to purchase açaí, then resell it in Belém, culminating in a trip as long as 15 days. In the region of Marajó-Açu, *marretagem* is one of the main economic activities, along with açaí production and shrimp fishing (Siqueira, 2009).

Swine husbandry is encouraged by landlords in the area, since it is highly compatible with açaí production and offers high economic return for the land. However, swine requires the implementation of some infrastructure such as fencing, as loose pigs can damage manioc gardens. Fencing and other measures do not offer complete protection, so farmers face a choice between pork husbandry and some loss of manioc. Nevertheless, pork husbandry is an attractive option: It requires only a small amount of labour (other than infrastructure if desired) and provides a high return, since pork meat is undersupplied in the area.

Deforestation, while prevalent in much of the Amazon from the mid-1970s to the mid-1980s, was not experienced to the same degree in the floodplain regions.

Other activities in the region include the preparation of açaí baskets and shrimp traps, made from the wood of açaí palm. Açaí baskets are made to hold a standard amount, approximately 12 kg, and are thus marketable in the region.

2.7 Markets

Brondízio (2008, Ch. 8) discusses the price dynamics of açaí in three scales: daily, seasonal and decadal. On a daily scale, the market prices vary greatly, according to the supply of açaí unloaded from boats, as well as the quality of the berries. The originating location of açaí is important, since açaí spoils quickly. Açaí is only fit for consumption during the three days following harvest (Brondízio et al., 2002). Açaí originating further inland, such as from Maranhão, is less desirable than that cultivated from Ponta de Pedras or the islands near Belém such as Ilha das Onças. Açaí producers consider the quality of their goods, sorting goods into barrels of similar quality, presenting the most perfect berries on top.

There are three general roles taken at the açaí market: producers, middlemen and brokers. Producers may sell to brokers directly, but they must arrange their own transport. Alternatively, middlemen may purchase goods from producers at a lower rate to resell to brokers at the market price, handling the transportation from the farm to the market then selling to brokers (Muñiz-Miret et al., 1996; Siqueira, 1997).

There are three price models taken between sellers (producers and middlemen) and brokers. At the peak of the açaí harvesting season, when supply is greater than demand, an average pricing model is often followed. Through this pricing model, the seller is paid by the purchaser at the end of the market day, either by the average of the opening price and the closing price or at the closing price. This protects the broker from buying high and selling cheap, instead making transactions at an average price. An alternative price model is the hourly price, in which the current price of açaí is paid to the seller on delivery. When demand is greater than supply, sellers may select purchasers based on their bids. Larger producers, such as corporate farms, may also use a contracted price model, in which a price is prearranged for the duration of each season. This arrangement buffers the prices when harvesting times cannot be controlled. Otherwise, açaí is best harvested when prices are highest to provide the most profit.

The seasonal scale highlights price manipulation on the part of larger producers. Large landowners create scarcity by controlling sharecropper production. These large landowners control a supply sufficient to influence price. Brondízio (2008) refers to a group of 5 large producers which produce 7 000–29 000 baskets per season. This price manipulation affects the staple food industry more than that of the fashion food. While fashion foods, locally and internationally, use processed and preserved açaí pulp and powder, açaí as a staple fruit or pulp is required to be fresh, making it more sensitive to changes in price. Preservation technology is beginning to dampen seasonal price variation outside of the peak season through stock control.

On an annual scale, year-over-year growth in açaí prices becomes readily apparent. As it has been previously discussed, açaí prices rose as demand for açaí as a fashion food grew. This rise in açaí prices is also a result of inflation, as indicated by its comparison with IPA-PARA, the Agricultural and Husbandry Price Index for the state of Pará. Both the API (Açaí Price Index) and IPA-PARA showed significant growth throughout the rampant inflationary period, with açaí showing growth beyond inflation. However, açaí sells for a significantly higher price during the second half of its harvesting season, as much as 2–2.5 times IPA-PARA (Brondízio, 2008). Ultimately, açaí producers achieve a better than average return over other agroforestry and husbandry goods.

2.8 Chapter summary

This chapter introduced the Caboclos society in the floodplains of Marajó Island, near Ponta de Pedras, Pará, Brazil. The Marajó Island area has seen rapid economic changes as a result of market demands for rubber and, more recently, açaí. As a result, Caboclo farmers have been exploited for their labour, but have recently been recognized as having local knowledge of sustainable farming practices. In the community of Paricatuba, there are approximately 21 smallholder households who practice swidden-fallow cultivation and intensive açaí management, in addition to other economic activities such as employment as middlemen or in urban areas. Urban employment has led to long-term or short-term circular migration, as an individual may perform work off-farm and return during the açaí season. Such processes have led to multi-sited households, in which members living outside the physical home may contribute labour or money. The most recent market shift has been a boom in açaí, as it has evolved from a rural staple to an international fashion food.

Research in this area has identified açaí prices and multi-sited households as drivers of change. Broadly speaking, to what degree have these drivers influenced land use change and household decision-making? Recent research (Siqueira, 2009) has begun to address internal household decision-making as a research issue. In turn, how does decision-making affect land use change and economic welfare?

Chapter 3

Agent-based models of land use change

3.1 Introduction

Agent-based modelling (ABM) is a technique coming to popularity in recent years, a branch of multi-agent systems (MAS) of distributed systems research and artificial life (Bousquet and Le Page, 2004; Matthews et al., 2007; Robinson et al., 2007). It is undergoing increasing popularity in the land use science community, especially in the last decade. Beginning with theoretical models such as SugarScape (Epstein and Axtell, 1996), ABMs have since evolved into modern tools of virtual experimentation based on empirical data (Deadman et al., 2004; Manson, 2006a), policy analysis (Berger, 2001), and scientific or participatory collaboration (Castella et al., 2005; Pignotti et al., 2004). However, agent-based models have not become operational decision support systems, lacking numerical accuracy and relevance to end users, among other issues (Matthews et al., 2007). Nevertheless, such models can be useful for illustrating the dynamics of a complex system, explaining causative factors of some phenomenon or as a teaching tool (Axelrod, 2003; Epstein, 2008).

An agent-based model is a model based on autonomous software entities, agents, which sense and act upon their simulated environment. They are typically designed with a bottom-up approach, the modeller having codified the attributes and behaviour of individual agents and their environment into software form. Stochastic methods compensate for uncertainties in model parameters. Running the model once or several times in a Monte Carlo simulation, the interactions between agents and their environment result in emergent macro-properties that cannot necessarily be predicted from the behaviours of the individual.

The simulation of interacting individuals in a common environment leads to the concept of emergence. ABM/LUCC often illustrates emergence due to its bottom-up design. The concept of emergence may be explained, quite simplistically, as "the whole is greater than the sum of its parts" (Parker et al., 2003). In other words, micro-interaction of individuals leads to observable higher-order macro-patterns. Given the uncertainties involved in modelling a real-world system, agent-based models use probability distributions as a proxy for uncertain outcomes. Due to the stochastic nature of ABM and the uncertainties involved in modelling a real-world system, an ABM should only be expected to model *possibilities*, not to predict outcomes with absolute certainty. Each run of an ABM, associated with some random seed, produces one possibility. Running an ABM several times produces a set of possibilities, enhancing its use as an exploratory tool.

Agent-based modelling has been combined with cellular automata models to create ABMs of land use and land cover change (ABM/LUCC) (Parker et al., 2003). This chapter will discuss agent-based modelling as a tool for studying LUCC. It begins with a discussion on a rationale of modelling as a viable tool. This is followed by a discussion on the history and state of cellular automata and agent-based modelling as it relates to LUCC. A brief discussion of modelling tools will be provided. Given the variety of agent-based models developed for land use/cover change and the broad backgrounds of model developers, many fundamental components are implemented in novel ways. The bulk of this chapter consists of a review of the decision-making models and spatial methods employed by ABM/LUCC. Finally, the chapter concludes with a discussion on the feasibility of verification and validation on an ABM/LUCC.

3.2 Why model?

Explicit models, such as agent-based models, simplify a system into a form of codified assumptions and data which can be analyzed to highlight theoretical properties and outcomes which cannot be observed in the real-world system. Supplementing field observations and laboratory experimentation, an agent-based model allows a researcher to perform experimentation on a virtual population which would otherwise be infeasible or unethical. Such experiments may include price manipulation, for instance.

A significant challenge faced by land use scientists is posed by the complexity of human-environment interactions. Land use change is a complex process, resulting from feedbacks in biophysical and socioeconomic systems (Aspinall, 2008). It is argued that such a process may not be evident when studied in the frame of one system alone (Liu et al., 2007): Systems such as climate, ecology, demographics, landscape, economics and culture may define or influence local behaviour and land use change. Modelling can integrate these systems into an experimental frame, in which the assumptions of each of these systems can be codified. An experimental frame, as defined by Ziegler (1976), is a limited set of circumstances observed in the real system. A modeller may not necessarily have expertise in each system, but expert input can help shape the model and its assumptions. In this case, modelling can be used as an experimental tool and, among multiple experts, a collaborative tool. Ultimately, a model provides, at the very least, a definition of a system as a set of rules which can be analyzed in terms of its design and the data it generates. Furthermore, by the nature of its bottom-up design, agent-based modelling often results in the discovery of emergence in the system. Emergence refers to macro-scale properties which are not easily predictable based on the individuals the model is built upon (Verburg et al., 2004). Emergence occurs as a result of interactions among individuals and their environment.

Epstein (2008) argues in support of modelling, first contrasting between implicit and explicit models. Implicit models, he states, include those created in the human imagination. By defining modelling this way, he argues that all people are modellers when constructing their understanding of a system based on their observations or data. This obviates the modelling argument. However, the question remains of whether *explicit* modelling is useful. The remainder of this discussion will discuss explicit models, where assumptions are recorded and codified. Core to Epstein's main argument is that in an explicit model, assumptions are defined and can be tested. In an agent-based model, these assumptions are defined in terms of parameters and computer code. By modifying the parameters and aspects of the code, assumptions can be tested. Best-available data and expert knowledge related to the problem domain can be integrated to gain a fuller understanding of the system under a variety of scenarios.

Box and Draper (1987) famously stated that, "Essentially, all models are wrong, but some are useful." Epstein makes an important related note, distinguishing prediction from explanation. While a model cannot and should not be expected to predict future events with absolute certainty, the model may serve to explain certain aspects of the system. At best, model outputs may provide bounds on future events or estimates on likely outcomes. Depending on the quality of data, likely outcomes may be predicted with reasonable confidence. In the case of a more theoretical model, prediction is not necessarily the goal. Instead, trends in the simulation outcome may aid scientific explanation or discovery (Axelrod, 2003). With this in mind, the purpose of modelling is not merely the study the end state of a system: Instead of regarding the end state as the only model output, the researcher can study the functionality of the system, the relationship of its components and the trends experienced throughout the simulation. In this manner, the model may be used as a virtual laboratory, allowing the researcher to manipulate a hypothetical system and study causality on a deeper level than can be observed in a real-world system, especially when the system is unobservable or complex.

However, there are arguments against the use of agent-based modelling. Beyond simple models like SugarScape (Epstein and Axtell, 1996) and early versions of Fearlus (Gotts et al., 2003; Polhill et al., 2001), significant amounts of data are required to capture site-specific intricacies. An agent-based model cannot include all of the complexity of the real-world system: Some simplifying assumptions must be made to ensure it is feasible to implement and test. As more complexity is introduced into a model, it becomes much more difficult to attribute outcomes to causative factors. Couclelis (2001) briefly highlights some challenges faced by the modelling paradigm and questions the "considerable" effort placed into adding complexity into ABM/LUCC models and whether the benefits of complexity outweigh its costs. Just as Couclelis notes the failure of modelling to find its place as a decision support tool in planning and policy-making, ABM has not realized the same level of applicability with end-users as it has within academia (Matthews et al., 2007). This is related to the current limitations and data requirements of ABM, as well as the poor usability of ABM software as a decision-support tool. The latter issue can be addressed through end user training and improved software packaging targeted toward decision-support end users, or as Matthews et al. suggest, end user participation throughout the model design process.

3.3 History of agent-based modelling

3.3.1 The software agent

The software "agent" has evolved from research of artificial intelligence (AI), specifically distributed artificial intelligence (DAI). In the field of DAI, the research objective is not to emulate or provide the knowledge and reasoning of a single intelligent agent, but to study the knowledge, reasoning and co-ordination of multiple heterogeneous agents (Bousquet and Le Page, 2004). DAI can be classified into two main areas of research, one of which is the area of multi-agent systems (MAS) (Moulin and Chaib-draa, 1996). MAS use autonomous, possibly heterogeneous, software entities—agents—in co-ordination to solve problems. An agent can be loosely defined as an autonomous software entity which is capable of sensing its environment and acting upon it (Russell and Norvig, 2002). Finer definitions and classifications of agents, particularly those described by Russell and Norvig (2002) and Moulin and Chaib-draa (1996), are described in Section 3.6.1.

3.3.2 Cellular automata

Cellular automata are uniformly shaped cells arranged in a discrete lattice, most often in a two-dimensional rectangular arrangements for geographic applications (Torrens and O'Sullivan, 2000), though alternative lattices such as 2D hexagonal and cubic are possible. For each cell comprising the lattice, a set of rules is applied once per simulation step, which modifies the cell's state based on its previous state and that of its neighbours. In terms of a rectangular lattice, a cell's neighbours are those it is adjacent to, either vertically or horizontally, in the case of von Neumann neighbourhoods. Moore neighbourhoods include cells in the von Neumann neighbourhood, in addition to diagonally-adjacent neighbours. Beyond these neighbourhood-based rules, generalized cellular automata (GCA) may use rules which are not limited to adjacent neighbours (Takeyama and Couclelis, 1997).

One of the first cellular automata models was developed by John Conway in 1967 (Gardner, 1970). Using a physical checkerboard and flat counters with two colours, representing two states, "alive" and "dead", Conway devised Life as a "simulation game" meant to represent the rise and fall of several generations of organisms. He achieved this by creating rules which satisfied certain criteria, namely preventing (provably) unbounded population growth while giving the ability to create patterns which oscillate or stabilize into a steady state. Cell transitions in Conway's Game of Life, as it is now popularly known, are based on states of cells in the Moore neighbourhood. A dead cell comes to life if it has exactly 3 live neighbours. A live cell dies from overcrowding if it has four or more neighbours or dies from isolation if it has less than two neighbours. In other cases, the cell retains its previous state. Running complex cases on physical counters proved infeasible, so a computer program was developed for the PDP-7, allowing Conway and other researchers to discover more complex patterns (Poundstone, 1985). This "game" stands as a precursor to more modern agent-based models.

Cellular automata formed the basis for simulated virtual agent worlds, intended to illustrate social phenomena with a simple set of rules. Schelling (1971) developed a model illustrating the emergent effects of individual preferences for like-neighbours toward a more global trend of segregation. In this model, an agent (coloured black or white) seeks an aspiration level of a certain number or percentage of like-neighbours and will move if this aspiration level is not met. A very segregated neighbourhood results even with a common aspiration level of as little as 30 %. More advanced virtual worlds have been developed through the field of agent-based computational economics (ACE) (Tesfatsion, 2006).

Approaching the application of models to land use/cover change, successive models of cellular automata were used to simulate patterns of urban growth. Described as a "computer movie", Tobler (1970) used a deterministic model of cellular automata to simulate urban growth patterns in Detroit from 1910 through 2000. Tobler observed that "everything is related to everything else, but near things are more related than distant things", naming this the First Law of Geography. Applying this Law to a computer model, with the spatial population distribution of Detroit expressed as attributes assigned to cells in a rectangular lattice, Tobler created a demographic model based on a linear relationship between neighbouring cells across time. The cellular automata rule governing the next population of a cell is an equation stating that the population is based on a weighted sum of cells within a Chebyshev distance of 2, with nearer cells weighted more than farther cells. Through these rules, the model takes into account net migration (the difference between immigration and emigration), birth rates, death rates and population migration to adjacent cells (spread). Given these rules, a computer movie is produced with a temporal resolution of 0.05 or 0.5 years/step. The computer movies were developed for educational and illustrative purposes rather than for population prediction.

Moving closer to a land use model, rather than a strict population model, Batty (1997) created a CA-based model of urban growth. Cells in this model carry one of two states, developed or undeveloped. A cell becomes developed with probability ρ if one adjacent cell is developed. If the cell is not developed on its first attempt (with probability $1 - \rho$), then its subsequent probability of development is ρ^2. This pattern is continued, with the probability of development on the n^{th} attempt equal to ρ^n, approaching zero as $n \to \infty$. This model was applied to the area surrounding Niagara

Falls, where initial seeds placed in the centers of the municipalities of Niagara Falls, Buffalo and St. Catharine's in a combined cellular landscape produced approximate shapes of these cities resembling their real-life counterparts. A more complex model, the Sleuth model, also uses self-modifying cellular automata to model urban growth. For a variety of cities worldwide, Sleuth is calibrated with input maps of slope, land use and other factors, to determine parameters of diffusion (spontaneous growth), breed (growth of urban centers), road gravity and other factors (Silva and Clarke, 2002). With such flexibility, Sleuth is developed as a "first vision" of a universal urban growth model, subject to appropriate calibration.

3.3.3 Agent-based modelling for land use/cover change

Agent-based models of land use/cover change have come to popular use in the last decade (see Bousquet and Le Page (2004); Matthews et al. (2007); Parker et al. (2001, 2003)). These models range from the theoretical to the empirical, with theoretical models leaning toward simple, generalizable concepts and empirical models requiring more complexity and case-specific data (Berger et al., 2001; Robinson et al., 2007). There have been a few attempts to organize or describe agent-based models of land use/cover change under a continua (Berger et al., 2001), taxonomy (Hare and Deadman, 2008), ontological framework (Polhill et al., 2008), or conceptual design pattern (Parker et al., 2008). Each of these discussions includes a review of the state of the art in ABM/LUCC.

Cellular models have been coupled with agent-based models, using the cellular model as the agent's environment. An example of a theoretical (non-LUCC) model is SugarScape, a very simple environment which Epstein and Axtell (1996) used to model trade, combat, disease transmission and a myriad of other social issues. This arrangement has become popular for land use/cover change applications (Parker et al., 2001). In an ABM/LUCC, agents typically reside on a cellular grid, owning a set of cells constituting a parcel of land (for examples see Deadman et al. (2004); Manson (2004); Polhill et al. (2001)). While the agents themselves are not necessarily cellular, the land they manipulate is represented by cellular automata. The modelling of agents upon a cellular automata allows researchers to model spatial processes and agent-environment interaction.

An example of an ABM/LUCC coupled with cellular automata is Lucita ("Land Use Change in the Amazon"). Lucita was developed to study deforestation and land use change in a rural area along the Transamazon Highway west of Altamira, Brazil

(Deadman et al., 2001, 2004). Lucita is a model of smallholder farming households in the Amazon who produce cash and subsistence crops, subject to constraints of land, labour and capital. In a recent version (Deadman et al., 2004), households in Lucita select land use activities with price-weighted probability, until capital, labour or land resources are extinguished. An older version of Lucita (Lim, 2000) used genetic algorithms and a classifier system, shared among all agents. Through a bitstring, an agent specifies its available resources to the classifier system. The agent also provides a list of possible land use strategies and the past effectiveness of each one. The classifier system then performs rule matching and determines the set of rules which should be implemented.

Between both versions of the model and a more recent unpublished one, the landscape is divided into cells of 1 ha with similar environmental models. The equations of Fearnside (1986) were used to model soil conditions at each cell. These equations govern changes in soil variables (N, P, Al, pH, C), which in turn affect crop yield. Though cultivation reduces soil nutrients, as practitioners of swidden cultivation, farmers can conduct burns to return nutrients to the soil. Lucita's environmental model uses cellular transitions which are based on the previous state of the cell and the agent's inputs into the cell. Other models take a similar approach, codifying environmental transition rules within cellular automata, including Sypria (Manson, 2006a).

Sypria (Southern Yucatán Penninsular Region Integrated Assessment), another model of land use change in the tropics, focuses on the spatial allocation of land use activities. The cellular environment is based on cellular automata. The environmental model is implemented using cellular transitions, which are based on the state of adjacent cells. Agent decision-making processes have a spatial focus, taking into account relative theories of space and absolute theories of space. Relative theories of space infer that relative distances are strongly factored into decision-making, while absolute theories of space postulate that spatial heterogeneity is key. Using genetic algorithms, agents in this model consider factors such as soil quality (spatial heterogeneity) and distances to market when making land use decisions.

Current agent-based models of land use/cover change have integrated a wide range of environments and human decision-making strategies. Environments have ranged from simple bitfields (Polhill et al., 2001) to transitions of soil characteristics based on mathematical models (Lim, 2000; Matthews, 2006). Decision-making methods have included optimization, random selection, imitation, heuristics, classifier systems and genetic algorithms. Current challenges of models of LUCC which have been addressed are verification and validation (Berger et al., 2001; Ormerod and Rosewell, 2009; Pontius and Schneider, 2001; Qudrat-Ullah, 2005; Xiang et al., 2005), spatial representation (Huigen et al., 2006; Manson, 2006a), and data collection methodology (Huigen, 2004). Decision-making has been identified as a research issue (Schreinemachers and Berger, 2006), but has only been tested in theoretical models lacking empirical data (Jager et al., 2000; Polhill et al., 2001).

3.4 Software

There are a variety of open source and commercial software packages for creating agent-based models. Currently-maintained software platforms include NetLogo (Wilensky, 1999), Repast Simphony (North et al., 2007) and Swarm (Minar et al., 1996), among others. These software platforms vary in their ease of use, performance and potential for complexity, so models have often been ported across platforms to satisfy particular needs (Millington et al., 2008; Parker et al., 2008).

Based on Logo, a programming language featuring a "turtle" who responds to movement commands, NetLogo and StarLogo each offer a high-level programming language and environment for creating agent-based models. NetLogo is developed at Northwestern University on a Java and Scala platform. Scala is a functional programming language built upon the Java Virtual Machine. Although NetLogo programs are written in a high level interpreted language, intended for easy accessibility for non-programmers at the expense of performance, code is partially compiled to Java bytecode for performance improvement (Wilensky, 1999).

StarLogo is developed at the MIT Media Lab and has been used to implement agent-based model models, including an urban growth model developed by Batty (1997). Inspired by an early version of StarLogo, NetLogo was developed as an alternative with additional features and has since been applied to several land use models (Gilbert et al., 2008; Millington et al., 2008; Parker et al., 2008). Often, models are prototyped in NetLogo for its simple syntax and ease of use and ported to Swarm, Repast or outside a software framework in a language such as C++.

While NetLogo and StarLogo are high-level simulation platforms, which provide a closed environment and a custom high-level language, Swarm and Repast are *framework and library* systems (Railsback et al., 2006). In such as system, the model is developed using a lower-level language such as Java in a loose conceptual framework made to organize model development. Software libraries provide common tools, such as a scheduler or a pseudorandom number generator. High-level platforms like NetLogo are easier to use, but have less potential than framework and library systems such as Swarm (Bousquet and Le Page, 2004).

Swarm is a simulation package allowing models to be written in Objective-C or Java and has been used to develop land use models such as Fearlus (Gotts et al., 2003). Repast was originally derived from Swarm as a Java-based simulation package. It has recently evolved into Repast Simphony, allowing ABMs to be written using flowcharts, Java or Groovy, a dynamic language built onto the Java Virtual Machine. Flowcharts greatly simplify the modelling process for simple models, but greater complexity requires the use of Groovy or Java. Prior to Repast Simphony, earlier versions of Repast (North et al., 2006) have been applied to land use models such as MameLuke (Huigen, 2004), Sludge (Parker and Meretsky, 2004) and Lucita (Deadman et al., 2004).

All of the aforementioned platforms are free or open source software. In contrast, AnyLogic is an example of commercial software. It has been used to implement land use models such as the farming structural change model by Albisser and Lehmann (2007). While AnyLogic provides agent-based modelling capabilities in a Java environment like Repast, AnyLogic provides support for system dynamics and discrete-event modelling. Repast Simphony is beginning to include these features as well.

A comparison of several simulation software packages is provided by Nikolai and Madey (2009) for a broad audience. LUCC and pattern-oriented directed comparisons between a much smaller set of software packages have been prepared by Berger et al. (2001) and Railsback et al. (2006), respectively. Railsback et al. compared NetLogo 2.1, Mason (another Java-based simulation framework) version 10, Repast 3.1 and Swarm 2.2—versions current as of September 2005—comparing ease of development and performance, to a certain degree. As Railsback et al. noted, their comparison was outdated even at the time of publication, as new features and platforms were developed in the meantime. Having implemented one model across all platforms, their comparison was meant to be more qualitative than quantitative, but found that their model was fastest on Mason. Repast was found to perform almost as quickly, with a negligible difference for complex models. Swarm was, by far, the slowest platform for complex models, but fastest for the simplest models.

In terms of the ease of development on the various platforms, documentation for Mason, Swarm and Repast was found to be lacking. NetLogo was found to be the easiest to use, due to its simplified programming model. As models become more complex, the simplicity of the programming language may become a limitation. Furthermore, all model code is to be included in one file, which may be difficult to manage in a large model. The development of complex models was identified as a potential difficulty in Mason, due to the platform's restrictive scheduling framework and confusing terminology.

3.5 Scheduling

Discrete-time simulations can be implemented with either fixed-time or discrete-event scheduling. The distinction between these is that fixed-time scheduled models execute actions at predetermined intervals, while discrete-event models execute events on a dynamic schedule. Cellular automata models are examples of fixed-time scheduled models: *All* cells execute their actions at *every* simulation step.

The length of a simulation step is specific to the model. In some cases, a step or tick represents nothing particularly analagous to a real-world amount of time, especially in general, theoretical models. Models of real-world systems may assign a fixed length of time to be represented by a simulation step: For instance, the Sypria (Manson, 2005) and Lucita (Deadman et al., 2004) models use a step measuring one year. Tobler's computer movie of urban growth in Detroit (1970) uses a simulation step measuring either 5 % or 50 % of one year, with a finer temporal resolution resulting in a smoother computer movie.

In contrast to a fixed-time scheduled model, discrete-event models are reactive, executing or scheduling actions in response to events. In a discrete-event model, the schedule is initialized by placing one or more events onto the schedule. A schedule can be expressed as a table of events and their execution times, sorted in chronological order (most immediate first). To execute the next event, the simulation clock increments to the time of the most immediate event—the topmost row. This event is executed, removed from the schedule and may explicitly schedule future events. This process is repeated until there are no more events to be executed.

A discrete-event schedule allows for finer temporal resolution without necessitating the execution of every event at every time step, since the model only executes scheduled events. In between scheduled events, the simulation clock immediately increments to the time of the next event, regardless of the time interval between events. In contrast, an increase in the temporal resolution of fixed-time scheduled models would schedule every agent's actions more often. However, running a discrete-event model requires support on the part of the scheduler: While Repast and Swarm provide facilities to create a discrete-event model, NetLogo abstracts the scheduler, preventing the explicit scheduling of future events (Railsback et al., 2006).

In the case of many ABMs, a fixed-time scheduler is as appropriate as a discrete-event scheduler, if all agents execute methods at the same temporal resolution. While fixed-time systems execute all actions at every step, discrete-event systems require the explicit scheduling of future actions, requiring an increase in code complexity which may be unnecessary.

If multiple actions are scheduled at the same step, they will execute in priority order. Actions executing with the same priority and at the same time should execute in random order. While this rule is not followed for all models, this mitigates any consistent advantage or disadvantage resulting from execution order, especially when multiple runs are analyzed through Monte Carlo simulation.

3.6 Reasoning and decision-making methods

3.6.1 Agent categories

While agents are typically described in ABM literature as autonomous or interdependent software entities, Russell and Norvig (2002) define an agent as one who senses its environment through *sensors* and acts upon its environment through *actuators*. This latter definition necessitates a differentiation between *human agents* and *artificial agents*, such as software agents. Couclelis (2001) discusses *designed* vs. *analyzed* agents, where designed agents refer to software or hardware (robot) agents and analyzed agents refer to natural subjects such as human or animal. In discussing the model, this thesis restricts the discussion of agents to software agents, though they may be representative of human or other physical counterparts in the real world.

Agents can be classified by their problem-solving capabilities: Moulin and Chaib-draa (1996) classify agents as reactive, intentional and social agents, based on classifications made by Demazeau and Müller (1991). Similarly, Russell and Norvig (2002) classify agents as simple-reflex, model-reflex, goal-based and utility-based agents, in order of increasing cognition.

A reactive or reflex agent simply reacts to environmental changes or received messages without the ability to “reason” its own intentions. Another term for this type of agent is a rule-based agent, as a reactive agent’s behaviour is based on the execution of pre-defined rules. The reactive agent classification can be sub-divided into Russell and Norvig's simple-reflex and model-reflex categories: The simple-reflex agents has no concept of history, while the model-reflex agent has a concept of its internal state. The model-reflex agent updates its internal state to keep track of what it can no longer observe. However, the model-reflex agent uses the same rule-based reasoning as the simple-reflex agent, but has knowledge from its internal state to draw from.

Agents more advanced than reflex agents can be classified as intentional agents. Intentional agents have the capability of reasoning, whether in order to find a goal or resolve conflicts. (Moulin and Chaib-draa (1996) add a third classification, a social agent, to the classifications of Demazeau and Müller (1991). A social agent is said to contain models of other agents, containing the beliefs, goals and plans of these agents, so that it can plan and act with respect to the behaviours and actions of other agents.)

As a subset of intentional agents, goal-based agents have the capacity for some level of reasoning: They have an idea of some desired binary or discrete state, such as "happy". In addition, agents can estimate or have knowledge of potential results of possible actions to be taken. Instead of pure reactive rules, goal-based agents utilize their notion of how their environment is changing, the consequences of their actions, and a set of goals to decide on a particular plan of action. These agents use reasoning such as searching and planning.

Another subset of intentional agents, more complex than the goal-based agent, is the utility-based agent. This type of agent is distinguished from the goal-based agent: Utility is not a binary or discrete state, but a function mapped onto the set of real numbers. The utility function is a mapping of a state to a real number, which expresses the "goodness" of the state. Russell and Norvig (2002) argue that an agent is only rational if it acts "as if it possesses a utility function whose expected value it tries to maximize". Given the uncertainty of a partially observable environment, an agent cannot calculate the exact utility at any point in the future, but can determine the best course of action by calculating the expected utility of the action.

A classification scheme employed by Schreinemachers and Berger (2006) categorizes agents by their decision-making methods and application. Agents are first classified as either optimizing or heuristic agents. Optimizing agents are further classified by their application, those used for normative purposes and those used for positive purposes. Normative agents are used to discover new, optimal solutions within resource constraints, while positive agents are used in empirical models to represent some real-world analogue. As this research is focused on the positive exploration of land use dynamics of rural Marajó Island, rather than, say, a search for more optimal farming practices, this thesis will focus on positive agents.

3.6.2 Unbounded rationality

Traditional economic models assume *rational behaviour*. A rational agent, *homo economicus*, will always act on the optimum solution, having clear preferences and given all available information. Given the uncertainties in the future, the rational strategy may not make the most optimal choice *in hindsight*, but chooses the most optimal alternative given all information known at the time. Such unbounded rationality can be regarded as unrealistic, given that unlimited time is necessary to enumerate the outcome of all possible solutions. Nevertheless, proponents of unbounded rationality argue that humans act as if they are rational, so they consider an unboundedly rational model to be a suitable model for human decision-making.

Russell and Norvig (2002) assert that a rational agent must act as if it is optimizing some utility function. Mapping the expected value of each available action to a numerical value, rational agents are able to objectively rank the expected value of each action in order to select the most optimal course.

Linear programming

In order to make an optimal, purely rational decision, a solver must find the most desirable action or set of actions. Given that every possible solution can be mapped to a numerical value of desirability, the calculation of this value for *every* available action is computationally infeasible, given the potentially infinite number of possible actions or action paths. Therefore, it is desirable to find the optimal solution in an efficient manner by evaluating a minimal number of solutions.

Optimization may be implemented using mathematical models known as linear programs, introduced into widespread use in 1947 by George Dantzig. Linear programs were not invented by Dantzig, as they were described by Fourier and de la Vallée Poussin in 1823 and 1911 respectively. However, it was Dantzig's novel method of solving these linear programs as well as its applicability for solving military-related problems which allowed linear programming to achieve popularity. A more detailed history of linear programs is recalled by Dantzig (2002) as part of a 50th anniversary issue of the journal, *Operations Research*.

A linear program expresses an optimization problem as a linear objective function combined with a set of linear constraints expressed as equalities or inequalities. The solution to a linear program is the set of variables which maximizes or minimizes the value of this objective function (utility function). (Historically, the term "program" in this case refers to military plans, such as logistics or schedules, which were early applications of this mathematical model.) The maximum (or minimum, if appropriate) value of the objective function corresponds to the optimal solution. Linear programming can be used to find an optimal solution provided that the decision-maker's objective and constraints can be expressed as linear inequalities and that at least one feasible solution exists within the constraints. As a decision-method within a (non-ABM) model of LUCC, Chuvieco (1993) used linear programming as a land allocation method.

A sample problem to illustrate optimization: Suppose an agent wishes to develop 5 plots of land and may choose from land uses A and B. Land use A requires 1 units of material and 9 units of labour, while B requires 2 units of material and 4 units of labour. Profits from land uses A and B will be \$2 and \$3 respectively. The agent has 8 units of material and 36 units of labour. Wishing to maximize profit, a linear program can be developed, where a and b refer to the number of plots of A and B, respectively.

$$
\begin{aligned}
&\text{maximize } 2a + 3b && \text{(profit)} \\
&\text{subject to} && \\
&\qquad a + 2b \leq 8 && \text{(material)} \\
&\qquad 9a + 4b \leq 36 && \text{(labour)} \\
&\qquad a + b \leq 5 && \text{(land)} \\
&\qquad a, b \geq 0 &&
\end{aligned}
$$

Since it operates in only 2 dimensions, a and b, this linear program can also be expressed graphically, as shown in Figure 3.1.

The first equation of this linear program specifies the objective function, in this case, a maximization of profit. The inequalities specify constraints of material, labour and number of available plots (assuming that up to a single land use can be used to develop each plot). The final inequality is a non-negativity constraint, used to prevent invalid solutions, such as a negative number of plots of either land use. The intersection of these inequalities is the feasible region of solutions.

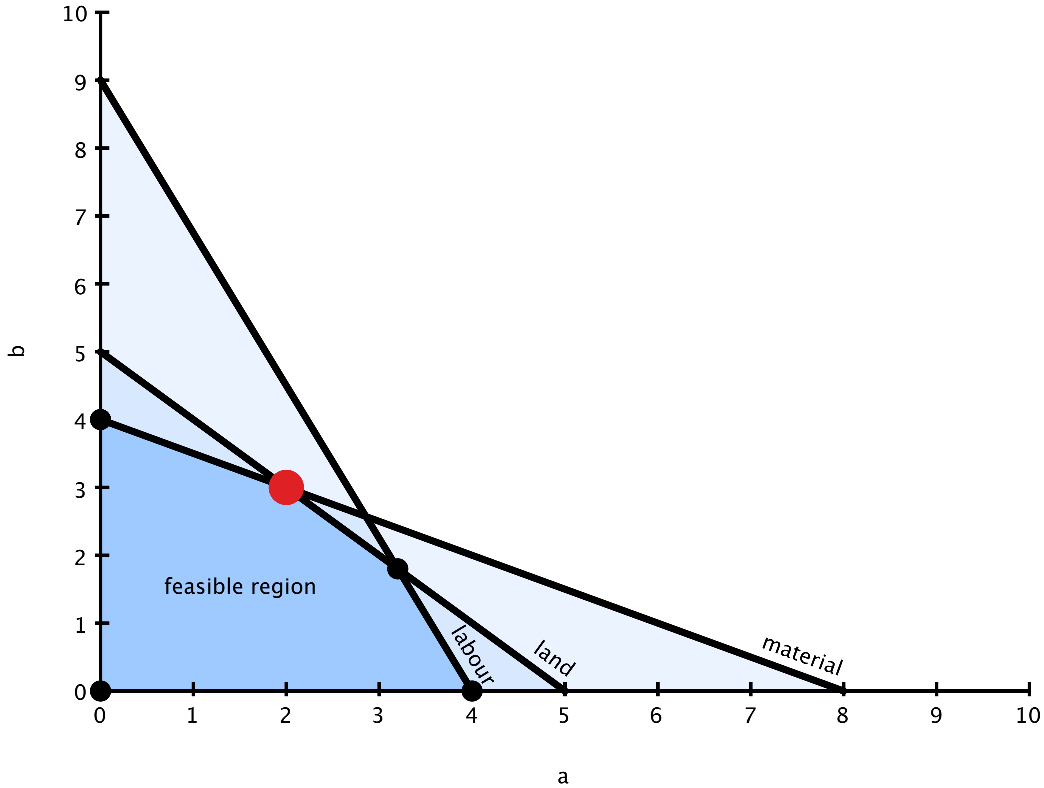


Figure 3.1: Graphical representation of a linear program

The optimal solution is found by exploring the space of all feasible solutions while traversing only a small set of nodes. Solution algorithms, such as the Simplex method developed by Dantzig (Dantzig, 2002; Nash, 2000), may employ admissible heuristics to perform an efficient enumeration of feasible solutions. The Simplex algorithm traverses only a few nodes at the edge of the feasible region, shown by the circles. The solution which maximizes the objective function is shown by the largest circle, in red.

3.6.3 Alternative models of rationality

Models of rationality can be divided into two major categories, shown in Figure 3.2. Jager and Janssen (2003) argue for the use of behaviourally realistic agents. A purely optimal approach to rationality can be regarded as unrealistic, given that it also assumes that the search cost is zero or, in other words, that the *homo economicus* decision-maker possesses "demonic powers of reasoning" (Gigerenzer and Todd, 1999). There is an alternative model of rationality, *bounded rationality*, in which the decision-maker has limited computational ability, time or cognitive capacity (Simon, 1957). This notion of rationality has been explored in LUCC literature (Manson, 2004, 2006b). While pure rationality assumes access to perfect information or infinite ability, bounded rationality accounts for the fact that this may be impossible. Bounded rationality also accounts for the need to learn or adapt.

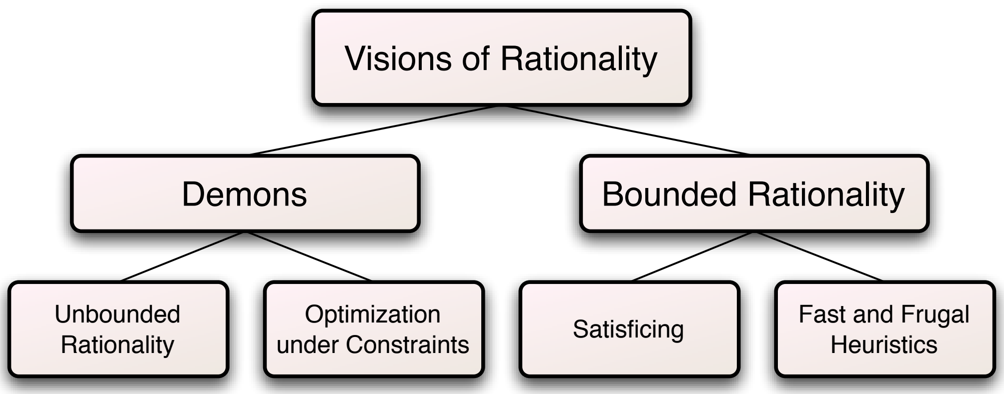


Figure 3.2: Models of rationality, from Gigerenzer and Todd (1999)

Similar to bounded rationality, in recognition of limited cognitive capacity, there is an alternative model of rationality termed *optimization under constraints* (Gigerenzer, 2006; Gigerenzer and Todd, 1999). The optimization under constraints model recognizes a non-zero search cost. As solutions are considered, a *stopping criterion* determines when the search for a more optimal solution should be stopped. This stopping criterion stops the search at the point in time when the search cost matches or exceeds the expected benefit from a continued search. Optimization under constraints is *not* a model of bounded rationality: The optimization under constraints model requires more information and computation than an unboundedly rational model, since the expected remaining search time must be recalculated when considering each subsequent alternative to determine the optimal stopping time. In contrast, bounded rationality uses fast and frugal heuristics to determine stopping time.

Other than limits in time and cognitive ability, sources of irrational behaviour are said to be related to cognitive biases, such as availability, anchoring and loss aversion (Kahneman and Tversky, 1979; Tversky and Kahneman, 1973, 1974). These biases are said to influence decision-making toward the irrational. In facing a choice between alternatives where the expected value of each choice could be calculated, rationality would prescribe that the alternative with the highest expected value be chosen. These cognitive biases suggest other factors be taken into account. For instance, a loss averse decision-maker would underweigh a choice with a high probability of failure (loss), even if the net expected value was positive. In contrast, a risk seeking agent would overweigh low probabilities of large success. A bias in either direction can lead to an irrational decision.

Gigerenzer (2006) argues against the use of unbounded optimization, optimization under constraints and cognitive biases to model human behaviour. He dismisses unbounded rationality and optimization under constraints since they present a decision-making model with unlimited computational resources. Cognitive biases proposed by Tversky and Kahneman, he argues, are untestable. Instead, Gigerenzer opts for *ecological rationality*. Ecological rationality is Gigerenzer's term for Simon's original definition of bounded rationality, which was originally postulated with two components (Simon, 1956): The first component refers to the limitation in cognitive capacity and time. The second component, which is often ignored in the literature (Gigerenzer and Todd, 1999), refers to the decision-maker's environment and social norms. Ecological rationality stresses that the decision-maker's rationality is closely tied with the decision-maker's environment. (Simon (1956) uses the term "environment" to refer to the needs, goals and drives of the decision-maker.) Depending on this environment, a decision-maker may not employ utility functions or a full enumeration of all alternatives.

Without simulating the psychology of agents too deeply, current ecological-economic literature suggests that decision-makers employ simple heuristics to make decisions, such as social comparison, imitation and repetition (or *autoimitation*) (Jager et al., 2000; Polhill et al., 2001). Heuristics are rules which are used to govern decision-making (Schreinemachers and Berger, 2006). Gigerenzer and Todd (1999) argues that humans employ *fast and frugal heuristics*. Heuristics are fast if they can be computed in little time and frugal if they can be computed with little information.

Rational vs. heuristic decision making

It has been said that models using optimal solutions tend to look for inefficiencies exogenous to the agents' internal cognition, while *satisficing* methods model inefficiencies in the decision-making process itself (Schreinemachers and Berger, 2006). "Satisficing" is a term coined by Herbert Simon, combining the words "satisfying" and "sufficing" (Simon, 1976). Satisficing agents (whether human or artificial), unlike optimizing decision-makers, search for a solution until one is found that is "good enough". Alternatively, satisficing may be described as a recognition that there are non-zero search costs involved in solution-finding, whereas an optimizing calculation is assumed to be costless (Schreinemachers and Berger, 2006). Therefore, a decision-maker may prefer to stop enumerating further options rather than search for a definitively optimal—or better—solution.

Schreinemachers and Berger (2006) argue that both optimizing and non-optimizing methods are useful as tools in empirical models of land use/cover change. Optimizing algorithms are regarded as unrealistic, since they assume perfect information and can consider an infinitesimal amount of possible actions. However, they suggest that optimizing algorithms can be combined with heuristic algorithms to produce a more realistic result: Heuristics are used to limit the solution space to a set of *perceived* available options. An optimizing algorithm can be used to determine the most desirable solution from this subset of feasible solutions.

Satisficing

Pettit's method of satisficing (1984) involves choosing an aspiration level, at or above which a solution is deemed "good enough". Then, each solution is evaluated, one at a time in some unspecified order, until a solution is found which satisfies the aspiration level. A version of Fearlus, a MAS/LUCC, utilizes this decision-making method explicitly (Gotts et al., 2003). For this method to be feasible, the solutions must be commensurate and comparable with the aspiration level. In a computational model, the aspiration level is expressed as a scalar value, so the goodness or desirability of each solution must be quantifiable as a comparable scalar value. Byron (1998) states that the aspiration level need not be chosen in advance, since the decision to satisfice (instead of optimize) may be taken while enumerating solutions.

This method can be contrasted with an optimizing one, in which the decision-maker evaluates (or appears to evaluate) every solution. If all solutions are evaluated, then the most optimal solution can be chosen. However, a satisficing algorithm does not evaluate all solutions, but evaluates solutions until one satisfies some level of desirability. In what order are potential solutions evaluated? Alternatively, how are solutions selected for evaluation or disregarded? The following sections discuss decision-making methods, each of which choose a solution from the same solution space. Many of these methods do not have stated aspiration levels, but instead select an option based on some heuristic.

Decision trees

Decision trees represent hierarchical sets of condition-action (if-then) rules and are an example of fast and frugal heuristics. Decision trees are often used for decision analysis and support, but can also be applied to classifier systems (Friedl and Brodley, 1997) and as a decision-making method in ABMs (Robinson, 2003). In fact, most ABM/LUCC use a condition-action rule system to model behaviour (Schreinemachers and Berger, 2006). Each non-leaf node represents a choice or an uncertain condition (chance). The leaf nodes of a decision-tree represent alternative solutions which are feasible if their ancestor choices and conditions are satisfied.

Choice nodes are often expressed as squares. An example used in the design of Maria is illustrated by Figure 4.9 on page 65. The outbound edges from a choice node represent the feasible choices. Chance nodes, not shown in the figure, are represented as circles, often posing the unknown as a question or variable. Each outbound edge represents a particular circumstance, specified as an answer to the question or an exact value or range of the variable. When the decision tree is constructed for decision analysis, the outcomes of chance nodes are also assigned expected probabilities so that the expected value of each alternative can be calculated (Peterman and Anderson, 1999). However, for a decision tree in an agent-based model, chance nodes are unused since the agent uses the decision tree as a fast and frugal heuristic without calculating expected values. Therefore, chance nodes are unnecessary (assuming the agent itself does not perform decision analysis as part of its cognitive process).

As a fast and frugal heuristic, a decision tree can be used in a divide and conquer strategy to quickly pare out infeasible or undesirable solutions from a large solution set (Quinlan, 1990). In this case, a traversal through a decision tree leads to a set of actions, rather than to a single solution. These actions represent the feasible alternatives available given a set of choices and circumstances. Among these alternatives, another decision-making method is used to select a solution. This is a useful arrangement when the other decision-making method is computationally intensive or if a satisficing solution is desired.

Another, possibly complementary, use of decision trees in agent-based modelling is to explicitly codify the decisions made by the agents' human or physical counterparts. Unlike many other black box decision-making methods, which have no clear real-world analogue, human knowledge and behaviour can be represented as condition-action rules. This is the approach taken by (Deadman et al., 2004) in creating the decision-making algorithm for farmers in Lucita. In this case, the non-leaf nodes are all chance nodes and all leaf nodes are actions or sets of actions. The agent makes a decision by traversing down chance nodes, selecting branches which satisfy the current circumstances. The agent arrives at a set of one or more actions, which can be reduced to one solution by adopting another decision-making algorithm. In the case of LUCITA, the decision tree is applied multiple times as many land use decisions are made before resources are exhausted for the current year.

Black box decision-making methods

Black box decision-making methods, like black box algorithms, hide their implementation from the decision-maker. Given a set of inputs, the black box produces a decision. These methods may either be invariant or adaptive over time. Genetic algorithms are an example of a black box decision-making method and have also been used as a proxy for bounded rationality (Manson, 2006b). Genetic algorithms produce novel solutions and can also be used as a toolbox for memory and learning (Manson, 2005). While this and other black box methods like neural networks have proven to be successful in ABM, they have been said to confound even their inventors (Gigerenzer and Todd, 1999).

3.7 Spatial methods in land use/cover change models

According to the formulation test (Berger et al., 2001), a model is spatially explicit if spatial concepts are present in behavioural rules. Agent behaviours in the ABM to be discussed require two spatial methods: one for settlement (land tenure allocation) and the other for land use allocation. Furthermore, there remains the question of how models deal with the initial conditions of the landscape, specifically its spatial variation across the landscape. In the absence of data, such as a DEM or classified land use image, a model must explicitly state the initial conditions of the landscape, whether spatially homogeneous or heterogeneous.

3.7.1 Settlement patterns

A spatial method for a settlement pattern involves the initial placement of agents onto the landscape. By far the simplest settlement pattern is random placement, adopted by such theoretical models as SugarScape (Epstein and Axtell, 1996).

The imitative, competitive FEARLUS model (Gotts et al., 2003) populates its entire landscape with land manager agents, each owning one cell. When an agent fails and vacates the simulation, its cell may be claimed by its adjacent (surviving) neighbours or a land manager new to the simulation. One claimant is selected among the candidates with equal probability.

Building upon the random placement strategy, SYPRIA (Manson, 2004) uses an exogenous population density map to probabilistically allocate immigrating agents into the landscape. Agents randomly select a cell, each weighted by its relative population density. This approach requires the availability of spatial population distribution data. The SLEUTH urban growth model is calibrated with external maps to determine cellular growth parameters.

LUCITA uses a simple boundedly rational approach for its settlement pattern, limiting land parcel search to exactly 3 parcels from the set of vacant parcels. Each parcel would be evaluated for its distance to the Transamazon Highway. After 3 parcels were evaluated, the household would select the land parcel which was closest to the highway. This simple algorithm was sufficient to produce a fish-bone deforestation pattern evident in this area, with relatively more deforestation in properties closer to the highway, due to longer settlement time.

Batty's (1997) cellular automata model of the settlement and urban sprawl of St. Catharine's, Niagara Falls and Buffalo used probabilities based on adjacency and land history to model settlement (development) patterns. In this model, cells take on a binary state (settled or unsettled). At each time step, a cell adjacent to a settled cell becomes settled with probability p (where $0 < p < 1$). However, if the cell remains unsettled, when reconsidered at the next (second) time step, it is settled with a lesser probability p^2, reflecting a diminished land value. The exponent increases by one at each subsequent time step until settled.

Each settlement method is appropriate to the study in question. Ultimately, an appropriate settlement pattern is chosen based on theory (Batty), available data (SYPRIA, SLEUTH) or heuristics (LUCITA). FEARLUS was developed to illustrate the relative strengths between algorithms, so a competitive land claim structure was appropriate.

3.7.2 Land Use Allocation Methods

LUCITA imposes an absolute ordering on plots of land, irrespective of the land use to be allocated onto a cell. Prior to the model run, each cell is ranked by its distance from the highway with respect to the other cells in the same property. Agents use this ranking to allocate land: The available cell closest to the highway is chosen. No effort was made to cluster similar land uses or model any specific land use patterns within a property. However, the LUCITA null model worked well to illustrate a clearcut fishbone deforestation pattern model in the binary case of forested vs. deforested cells. Since a household will always cut forest closest to the highway, this method would not model selective deforestation without necessitating some degree of increased complexity.

SYPRIA uses multicriteria analysis and multiobjective land allocation (Manson, 2004, 2006a). For a given production activity, agents in SYPRIA attempt to evaluate the suitability of a set of cells (S), expected as a weighted sum (with weights $W = w_1, w_2, \ldots, w_m$) of production factors ($V = v_1, v_2, \ldots, v_m$), subject to Boolean constraints ($B = b_1, b_2, \ldots, b_n$). Boolean constraints rule out infeasible actions, such as unavailable land due to land tenure arrangements. Agents determine weights, W, based on environmental and institutional factors, considering constraints B.

$$S = \sum_{i=1}^{m} w_i v_i \prod_{i=1}^{n} b_j$$

However, agents are boundedly rational and are unable to evaluate the optimal solution directly. Instead, agents approximate suitability using genetic algorithms.

3.7.3 Spatial variation

One test or criterion to determine if a model is spatially explicit is the spatial invariance test (Berger et al., 2001). If the agents in an ABM can be rearranged spatially without affecting the results, the model is not spatially invariant and—by this test—not spatially explicit. Existing ABMs treat the spatial variation of cells in different ways. FEARLUS, representing a theoretical landscape, used bitstrings to represent the characteristics of each cell (Gotts et al., 2003). This bitstring was derived by combining two strings, one representing spatial variance (biophysical characteristics) and the other, temporal variance (external conditions). The external conditions bitstring changes at every time step, but is common to all cells, while each cell's biophysical characteristics bitstring remains constant over time but is an attribute of each cell.

LUCITA represents a pioneering model, where the landscape was unsettled prior to 1970. It was therefore a reasonable assumption that the land was homogeneous prior to this time. In this model, spatial invariance resulted from the cultivation of the land. A similar approach can be taken for a non-pioneering model lacking data regarding initial conditions: The simulation's warm-up period, the time taken for the model to reach an initial steady state from some initial condition, can be used to model the period prior to the time in study. It can be assumed that the landscape was once virgin and homogeneous sometime in the past. To control the initialization bias resulting from this artificial initial condition, statistics and model output should only be analyzed from the run period following this warm-up period. Output truncation is one of the simplest and most common methods to mitigate initialization bias (Schruben, 1982). In this case, the warm-up period would begin from a homogeneous landscape, but agents would alter the landscape through cultivation prior to the run period.

3.8 Verification and validation

The modelling process occurs by defining and codifying assumptions, then observing emergent behaviour, unlike a data-driven approach which comes from the observation of a real-world system. One such methodology for this process is the Third Science methodology introduced by Platt (1964) and applied to ABM/LUCC by Robinson (2003). Simulation is a "third" science in the sense that it is contrasted with traditional scientific methods of induction, the discovery of patterns in empirical data, and deduction, the formulation and scrutiny of hypotheses based on real-world observation (Axelrod, 2003). Simulation, as Axelrod describes, "aids intuition" by allowing the researcher to analyze data generated from rigorous sets of rules rather than proving hypotheses from real-world data.

While verification and validation both refer to the assessment of a model, they should be distinguished. The verification of a model refers to the assessment that the model has been transferred from another model (such as a conceptual model) with sufficient accuracy (Banks, 1998; Xiang et al., 2005). In other words, verification ensures that the model has been programmed as intended. Verification is performed using the process of debugging and sensitivity analysis (Manson, 2001; Parker et al., 2003). By sweeping model parameters across a wide spectrum of values, shortfalls in the model and its parameter limitations can be identified. Since uncertainties in the model are often described by stochastic processes, Monte Carlo simulation is useful for analyzing the distribution of output variables. Several samples of outcomes can be obtained using Monte Carlo simulation, where a model is run several times with different random seeds. A Monte Carlo simulation provides a set of both likely and unlikely outcomes. While an outcome may not be predicted with absolute certainty, the probability or likelihood of certain outcomes can be estimated.

Validation, on the other hand, refers to the assessment that the model correctly represents the real world. Ormerod and Rosewell (2009) describe verification as "the process of determining that the equations are solved correctly" and validation is "the process of determining that we are using the correct equations". Furthermore, validation can be subdivided into structural validation or outcome validation. Similar to verification, structural validation is the assessment that the software model represents the conceptual model correctly (Manson, 2001). Outcome validation involves comparing the model results with empirical data. Structural validation is especially useful in bringing confidence in the model despite poor representation of empirical outcomes (Qudrat-Ullah, 2005).

There are significant challenges in the validation of agent-based models of land use change, particularly outcome validation, due to the complexity of the real-world environment and the relative simplicity of the model. While a model may be validated, it can only be regarded as valid in a certain experimental frame with respect to certain criteria (Ziegler, 1976). In the case of the land use change model under discussion, while the model may consider changes in market prices and urban employment, it ignores pension programs and middlemen. The model may be considered valid in the context of prices and employment, but would not be considered valid if the scope of observations is widened or shifted to include pension programs or middlemen.

3.9 Chapter summary

Agent-based models have been introduced in this chapter, beginning with an discussion on the usefulness of modelling for scientific exploration. While agent-based models have found a niche in academic and scientific exploration, significant inroads as decision support tools have not been made. However, agent-based models have evolved from theoretical exploration, through models such as SugarScape, to site-specific analysis in land use science. The role of coupled cellular models has also been discussed, as it provides an environment for the agents, often in the form of land cover and soil characteristics. ABM software has been discussed, highlighting the evolution from Logo to NetLogo and more recent software packages such as Repast Simphony.

Spatial algorithms for ABM/LUCC are discussed in terms of settlement patterns and land use allocation algorithms, ranging from random or pre-determined ordered placement to multicriteria analysis and multiobjective allocation. This chapter concludes with a brief discussion on verification and validation, highlighting the use of Monte Carlo simulation and sensitivity analyses.

The nature of human rationality has taken the form of two major types of "visions", unbounded rationality, common in economic models, and bounded rationality, which accounts for limitations in cognitive capacity. Bounded rationality has been interpreted in two ways. Satisficing accounts for the fact that humans seek a solution which is "good enough", but can be swayed by psychological influences such as anchoring and loss aversion. Alternatively, one school of thought proposes ecological rationality, in which fast and frugal heuristics—based on environmental factors such as social norms—are used to make decisions quickly and with little information. An example of a fast and frugal heuristic is a decision tree.

Decision-making methods, especially among rational and boundedly rational agents, have been identified as a research issue of agent-based models in general, where comparisons in case-specific models have not been made. The following chapter discusses the implementation of an agent-based model which addresses this issue by allowing alternative decision-making models to be compared in a common framework. Maria presents a comparison of rational and boundedly rational agents within the community of Paricatuba in Chapter 5.

Chapter 4

Methodology

4.1 Overview

This chapter presents the methodology undertaken in the design, implementation and analysis of an agent-based model, Maria (Multi-Agent Reasoning in Amazonia). Maria is being developed to study the role of decision-making, whether rational or boundedly rational, on land use change and economic welfare. First, the design of the model is presented, providing a broad overview of the scope and architecture of the model and its human and environmental sub-models. Data preparation is briefly discussed, covering the conversion of remotely sensed imagery into suitable data for the model. The bulk of the chapter concerns the implementation of the model, from the selection of its software platform to more detailed implementation decisions, made to allow the model to encompass multiple decision-making algorithms and future empirical data, should it become available. The implementation section also includes a discussion of data output, including 3D GIS visualization, sensitivity analyses and database design. The chapter concludes with a brief description of the runs and analyses to be presented in the next chapter.

Maria has been developed to evaluate alternative decision-making methods in the context of external markets and economic opportunity in the community of Paricatuba, Pará, Brazil. Paricatuba, a small community just south of Ponta de Pedras and west of Belém, was chosen for its relative simplicity in comparison to other study areas nearby. While Praia Grande is a unique co-operative community and Marajó-Açu is populated with many sharecroppers, Paricatuba is primarily composed of smallholder households (Brondízio, 2008). In the model, it is assumed that the decision-making of a smallholder family is made at the household level only, without the influence of an external landlord nor through the internal negotiation of its members.

To create this first version of the model, a framework is established in which classes can be injected into the code or substituted through polymorphism, a feature of object-oriented programming. Typically, polymorphism is implemented through inheritance: An object class may extend another object class, overriding its methods. The extending class can be treated with the same interface as the original class, though the extending class' methods will be called instead of the original class. Alternatively, an interface may be defined to be common among multiple classes, allowing these classes to be utilized through the same interface.

This framework recognizes a future need to replace simplified assumptions with more realistic realizations as more complex data becomes available or desired. Substituted classes can be alternative implementations of the same agent type, such as the `Household` class. In this thesis, two alternative implementations of household agents are used to compare and contrast optimal and non-optimal decision-making algorithms.

In addition, the scheduler is set up such that agents may schedule themselves without modifying code outside the agent. This allows for entirely new agents, such as employers in nearby urban centres, to be injected into the model.

The model description is separated into two sections, a broad overview of its design followed by a detailed description of its implementation. These model-related sections are separated by a discussion of data preparation methods. The remainder of this chapter discusses parameter sweeps and the analysis of the model results.

4.2 Design

4.2.1 Collaboration

The agent-based model fits into a broader study facilitated by a National Science Foundation grant in the area of Human and Social Dynamics (Behavioral and Cognitive Sciences). The grant focuses on the study of the effects of global economic change on local socioeconomic and biophysical dynamics in the Amazonian estuary, a broader study area than the one discussed in this thesis. Through this grant, research teams at the Anthropological Center for Training and Research on Global Environmental Change (ACT) at Indiana University-Bloomington and the Center for Environmental Research and Conservation (CERC) at Columbia University, collaborated with the author and the author's advisor, Peter Deadman (also a co-Principal Investigator). These other teams, consisting of anthropologists, a botanist, ecologists, and remote sensing experts, conducted field and studies of the area. The author's input was the design and development of an agent-based model which would integrate recent and historical data. Model development was iterative, providing illustrative prototypes to the other researchers at the meetings, first demonstrating the capabilities of agent-based models, and later, preliminary results. For brevity, this thesis discusses only the most recent iteration of the model.

4.2.2 Scope

Maria is a designed as an empirical model of smallholder households and land use change in rural Amazonia. It is an empirical model, as opposed to a theoretical one, as it integrates case-specific data for the purposes of extrapolating emergent data in hypothetical, yet realistic scenarios. The complexity of the model has been chosen to sufficiently include a certain scope of detail. Meetings with research teams from ACT and CERC have highlighted global markets (specifically that of açaí) and multi-sited households as two key components to include in the model, building upon the demographic and land use models explored in Lucita.

Market prices are modelled as exogenous factors, externally driven by forces far more global than the small set of household agents composing the majority of the model. The main scenario of market prices is that of steeply-rising prices of açaí, using the açaí price index derived by Brondízio (2008). The prices of other goods are derived from IPA-PARA (Agroforestry and Husbandry Price Index for the state of Pará) published by the Fundaçao Getúlio Vargas (FGV).

The definition of a multi-sited household was a source of contention in the meetings. For the purposes of model development, the term is defined to indicate households who maintain economic linkages and mutual interest, though these households are not necessarily linked through kinship. This definition is meant to include *agregados*, aggregated household members who are not related, but maintain a relationship with the household. In this first version of Maria, multi-sited households are generated by household members moving from the household to urban areas seeking employment. The new household maintains economic linkages with the old household, creating a multi-sited household. The kinship relationship of household members, whether in a single-sited or multi-sited household, is not defined: *Agregados* are treated the same as family members, as far as economic decisions are concerned.

MARIA is developed with the community of Paricatuba (Brondízio, 2008; Siqueira, 1997) in mind, but can be adapted between alternative study areas in the region, such as Marajó-Açu, Praia Grande and communities near Mazagão. These areas differ primarily in household behaviour, as far as the model is concerned. Sharecropping arrangements and community behaviours would need to be modelled explicitly, as they would affect the decision-making and constraints placed upon household agents. Local knowledge and preferences vary slightly. For example, intensive agricultural periods of 2 years—involving cultivation of floodplain gardens—are preferred by those near Ponta de Pedras while periods of 2.5 years are preferred by those near Mazagão. Furthermore, the locations of markets and waterways would need to be adapted to represent those in the study area. This can be done by replacing the underlying rasters used for calculation in the model. These assumptions would have to be modified between the two study areas, but the model architecture and decision-making processes as a whole would remain the same.

4.2.3 Model architecture

MARIA is developed using Repast Simphony, chosen due to familiarity with the Java programming language, in addition to its features of distributed computing and GIS integration. The model divided into two sub-models, or "contexts", as they are termed in Repast Simphony. As LUCITA is separated into soil, land cover and human sub-models (Deadman et al., 2004), MARIA is separated into sub-contexts, one for the natural environment and the other for human interaction and decision-making. These sub-models are not independent, as the human model will manipulate the environmental model through cultivation and feedbacks from the environmental model will affect the human model. The models, as implemented, begin with an initial state of homogeneous floodplain forest with heterogeneity resulting from cultivation.

The environmental context can be described in terms of layers. Like several land use models, such as LUCITA, Fearlus , and SYPRIA (Manson, 2006a), the environmental model is based on cellular automata arranged in rectangular grids. Cells in these grids have several attributes for soil, land cover and history. For the purposes of discussion, the cellular grid, and thus each cell, can be divided into two layers: soil and land cover. The land cover layer, modified through human interaction or natural, internal transitions, is affected by the soil layer in terms of yield. The soil layer's fertility—in a sense, its carrying capacity—is affected by land cover, increasing in period of fallow and natural succession and decreasing through intensive cultivation.

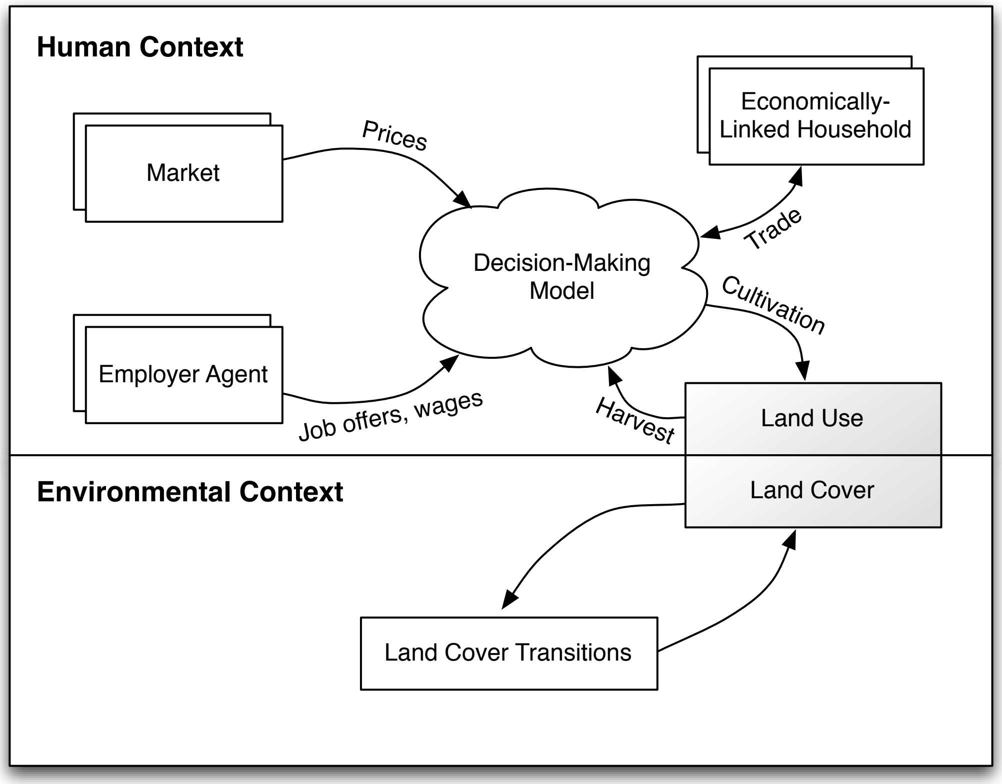


Figure 4.1: Human and environmental contexts

The human context introduces a layer above that of the environment. The human context includes farming households, who manipulate land cover for cultivation, as well as external markets and employers which do not interact directly with the land. External markets and employers do, however, influence the decision-making and resources of farming households. External markets influence selling prices, which in turn affect the desirability to cultivate certain goods. Employers may encourage farmers to take up non-farming vocations, potentially resulting in increased capital or decreased labour, which may increase or reduce cultivation.

The interaction of these two contexts is expected to result in a dynamic system in which market feedbacks and economic opportunities influence a changing landscape. By modifying the initial parameters of the model and by manipulating yearly events, experiments can be run to diagnose key influences of land use change.

4.2.4 Environmental context

The environmental context is designed as a cellular automata model, based on rules derived from research in the region (Brondízio, 2008). These rules constrain land cover transitions, as determined by typical regrowth periods and soil characteristics as well as current and historical land cover. While a detailed environmental model would include soil characteristics such as phosphorus, pH, and nitrogen, in Maria, these are abstracted as a fuzzy variable, fertility. The fertility variable allows the model to abstract biophysical changes in the soil without introducing unnecessary complexity into the model. Future models may implement a more detailed soil model, but the current focus of this initial version of Maria is the human model.

Cells are arranged in a rectangular 5×5 m grid. The source data for these cells (SRTM and Landsat images), have been interpolated from their original resolution using Kriging, as discussed in Section 4.3. Cells are classified into land and water cells. Water cells remain in the model, but are currently unused. Future uses for water cells may include shrimp farming and transportation. However, agents are aware of the distances from each land cell to the nearest water cell, as this data is stored as an attribute in each land cell.

Land cells contain constant attributes set before runtime, including distances, elevations and other terrain attributes, as well as soil and land cover variables. Soil conditions are aggregated as a single fertility variable. Land cells can support multiple land uses, provided they do not exceed the carrying capacity of the cell. A cell stores its land use composition and land use attributes as a set of fuzzy variables: age, density and health. Age and health are modified internally, depending on soil conditions and actions performed by the agent on the cell, while density is manipulated directly by the agent. Cell-level constraints prevent invalid states, such as an attempt to cultivate too much on one cell, or invalid variable values, such as fuzzy variables greater than one or less than zero. For density, fuzzy variables are used instead of crisp variables to allow agents to make land use decisions at a scale smaller than the cell size.

Land use and land cover transitions

Possible land cover in Maria are reflective of the most important land uses in the riverine study area: intensive açaí, gardens, forest and forest-fallow. The “gardens” land use type is used to represent both housegardens and *roçado de várzea* (floodplain gardens). Other land uses, which would be included in a more upland study area, include cattle fields and slash-and-burn fallow. Feasible land use transitions based on field studies in the floodplain region (Brondízio, 2008) are shown in Figure 4.2.

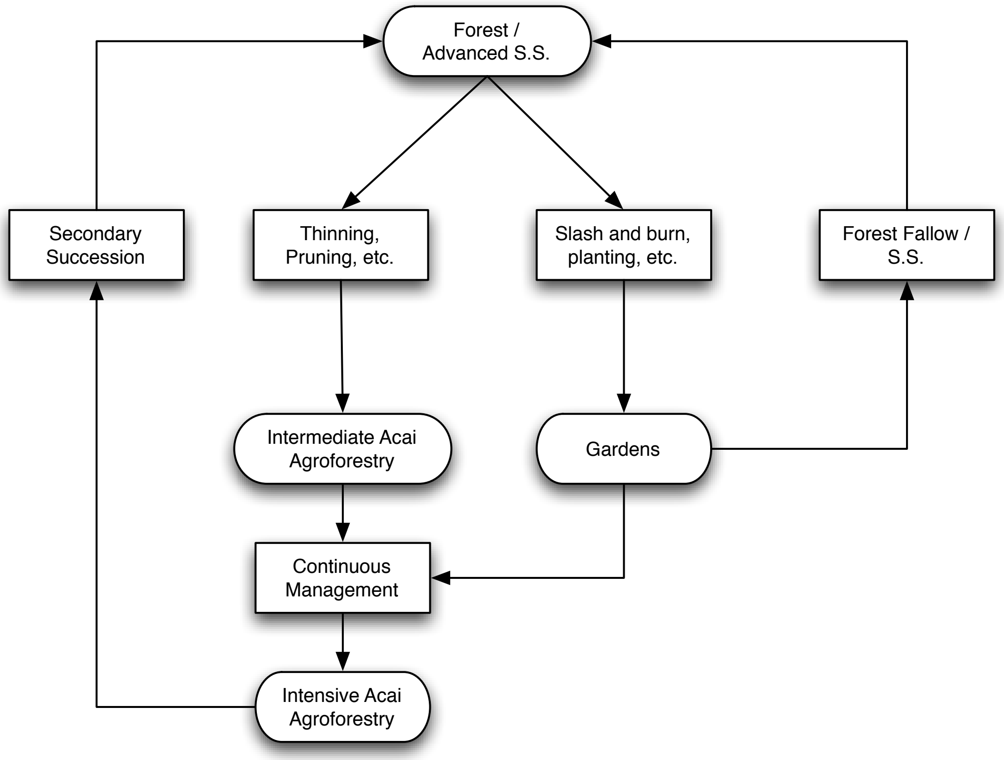


Figure 4.2: Land use transitions in Maria, adapted from Brondízio (2008)

Figure 4.2 shows cellular states and processes required or occurring between states. The processes (in rectangular containers) in the two center columns are related to human actions, whereas the outer processes are more natural. During secondary succession, households may continue to extract resources despite decreased management.

Within the model, the suitability of a particular cell for cultivation is based on the cell's internal land cover state, which is in turn based on its land use history. An attempt to cultivate an unsuitable cell will result in a poor yield. The yield of a cell is determined by multiplying the potential yield of the cultivated good by the cell's fuzzy fertility variable. Since the domain of a fuzzy variable is between zero and one, or 0 % to 100 %, fertility can be regarded as the percentage of potential yield which can be harvested from the cell. However, farming agents do not know the value of the fertility variable and must rely on their internal knowledge base to know where and where not to plant. This abstraction forces some amount of boundedness on the agents' rationality.

Intensively-cultivated cells only produce ideal yield during a certain cultivation period, which is determined by land use history (as a proxy for modelling soil characteristics). Cells will not produce ideal yield if cultivated within a period of time *after* an intensive cultivation. For floodplain gardens, which can be intensively cultivated for 2.5 years, the recovery time is 5 years.

After the cultivation period, the fertility variable drops to nearly zero quickly. During the recovery time, which begins after deintensification, the fertility variable rises to its maximum value of 1 linearly. Vegetation on a cell with low fertility will produce little yield. However, a fallowed cell continues to produce yield (Hedden-Dunkhorst et al., 2003).

Unlike gardens, açaí grows naturally in the area and can be extracted immediately. However, to reach full potential, intensive management strategies are adopted, involving pruning, weeding and the selective cultivation of açaízals (açaí stands) (Brondízio, 2008). Intensively managed açaízals may continue to bear ideal yield for many years, as long as continuous management is practiced. Through secondary succession, an abandoned açaí plot returns to a more natural forest state or advanced secondary succession within 5–10 years.

4.2.5 Human context

The human context is an agent-based model, wherein households are represented as the primary agents of the model. While the members of the households are also implemented as agents, they play a lesser role, providing some amount of labour to the household without any personal capacity for decision-making. Households aggregate the contribution of their members to determine available resources, such as capital and labour, similar to LUCITA (Robinson, 2003).

Markets

The rising price of açaí is implemented in the variable price scenarios as a list of prices determined *a priori*, but revealed to agents year by year. Market prices are revealed by Market agents, who send messages in each step containing the year's prices before the household decision-making stage. Future models may incorporate market prices which react to supply of rural households and the increasing demand of urban households, reflecting as the recent trend of ruralization of urban areas (Padoch et al., 2008). However, this would require an economic sub-model. The current model treats market prices as externally driven.

In particular, rising prices of açaí are derived from studies conducted by Brondízio (2008) and are defined in terms of an inflation and currency-adjusted price index, since the Brazilian currency changed five times during the period under study (1984–1999). Açaí prices are assigned the values of the açaí price index (Api) shown in Figure 4.3, normalized to an index price of 100 in the year 1994.

Prices of other agricultural activities are published in the monthly journal *Conjuntura Econômica* by Fundaçao Getúlio Vargas (FGV). IPA-PARA (Agricultural and Husbandry Price Index for the state of Pará) is an index for 24 farming products and 7 husbandry products, but does not include açaí, so it is used as the price of "other" goods produced by the general "gardens" land use type, for lack of an index that excludes husbandry products. Brondizio adjusted IPA-PARA to the same timeframe and index as the API. Prices for açaí and other goods will be derived from the API and IPA-PARA. Since these are in the same units, they will be scaled by the same multiplier during model calibration. (Model calibration will involve the sweeping of many input parameters in order to create a balanced steady-state system.)

Since available price data for both indices are only available during the period of 1987–1999, the years of 1970–1984 and 2000–2008 have been extrapolated. The acai price during early years is assumed to have been very low and has plateaued in recent years. As discussed by Brondizio, IPA-PARA, also indicates that prices of other goods have remained relatively steady in recent years.

This scenario is intended to illustrate agent behaviour in 3 phases: The first, in which agents cultivate açaí only as a subsistence good, followed by a second phase when açaí surpasses other goods in value. Finally, the price of açaí plateaus at a steady state, along with the prices of other goods.

Households

A household is a family unit of one or more members living in one settlement. Extended households spanning more than one settlement are implemented as multiple households linked together. Each household has some amount of capital, which is increased or decreased through revenue and expenses from farming and transportation. Farming households own land, while urban households do not as

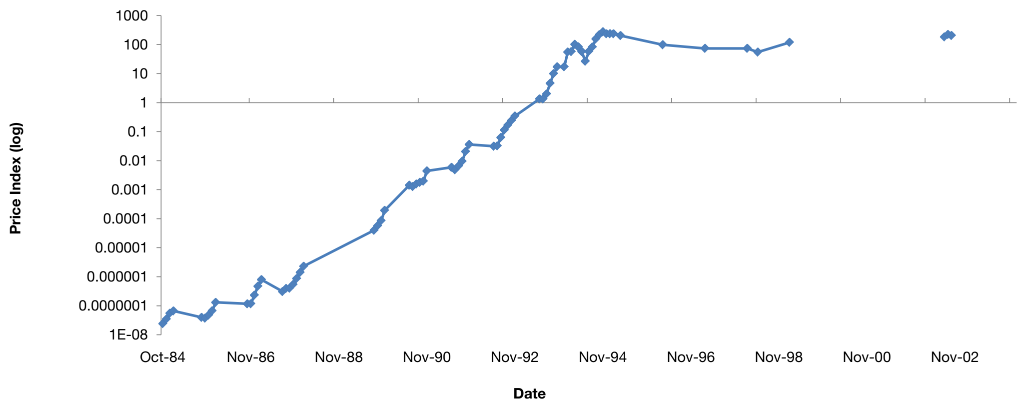


Figure 4.3: Açaí price index (API) in the period of 1984–1999, adapted from Brondízio (2008)

far as the model is concerned. (Land tenure, within the simulation, is restricted to include only land for forest, agroforestry or agriculture and not urban property.)

Decisions are made at the household level. Individuals do not make their own decisions nor do they share or negotiate decisions among each other. Rather, the household, as a whole, makes decisions as a single unit.

The first decision-making model implemented in Maria is the similar to that used in the latest version of Lucita, a heuristic model. Heuristics can be described as cognitive shortcuts which reduce complex tasks to simpler operations (Tversky and Kahneman, 1974). In selecting a new land use to expand the number of managed cells by one, heuristic agents select the land use by sampling from a set of preferred land uses, weighted by expected revenue. This is similar to the approach used by Lucita version 2 and newer (Robinson, 2003). However, Lucita used crop prices (per kg) and not expected revenue, since there was evidence that this was common household practice in the Altamira region (Moran, 1981). New cells are managed until (any of) labour, capital or land are extinguished. The second model is an optimizing one, using linear programming to maximize revenue under constraints of labour, capital and land. These are discussed in detail in Section 4.4.4.

These two mutually exclusive decision-making models are implemented as alternative types of households. When a household enters the simulation, it acquires one of the decision-making models and cannot change during the model run. For an imitative analysis (Polhill et al., 2001), which may be performed in the future, runtime substitution may be implemented.

The decision-making of a household is separated into stages of planning, execution and retrospection. The decision-making agents' time step begins with the planning stage, in which knowledge, belief, goals and constraints are integrated to produce a schedule of actions for the year. These actions take form in cultivation, extraction, harvest and other employment. Finally, an agent's year ends with a retrospective stage, through which the year's actions and outcomes are evaluated.

Household networks

A network can be described as a graph: a set of nodes and edges. The multi-sited household network consists of households as nodes and edges representing economic linkages. Initially, all households are unlinked nodes. Household members taking off-site employment spawn new households, which are linked to their parent households through the multi-sited household network. Each of these connected households makes decisions independently, but negotiations can be performed through message passing.

In the current version of Maria, there are no capital-sharing negotiations, since the focus of the model is on a comparison between multi-sited households and independent households and not on a thorough exploration of negotiation methods. Instead, networked agents send excess capital back to the parent household. This allows the household network's capital to be measured at a single location, simplifying the results collection process. If identified as a research issue, a more detailed model in the future may include capital-sharing negotiations, allowing more complex processes such as reciprocity.

Employment

In certain scenarios, agents from nearby towns will offer off-farm employment to household agents. Offers include a stated annual income, which the agents may use to evaluate choice of employment. These offers are limited in number and not available to all agents. If a member of a household takes a job offer, the member creates a new household in the town of the employer. The new household is economically-linked to the old household, facilitating trade of capital.

4.2.6 Timeline of a simulation step

Putting the environmental and human models together, events are organized into the order shown in Figure 4.4. Actions within each flowchart step are executed in random order. The implementation of this scheduling is detailed in Section 4.4.2.

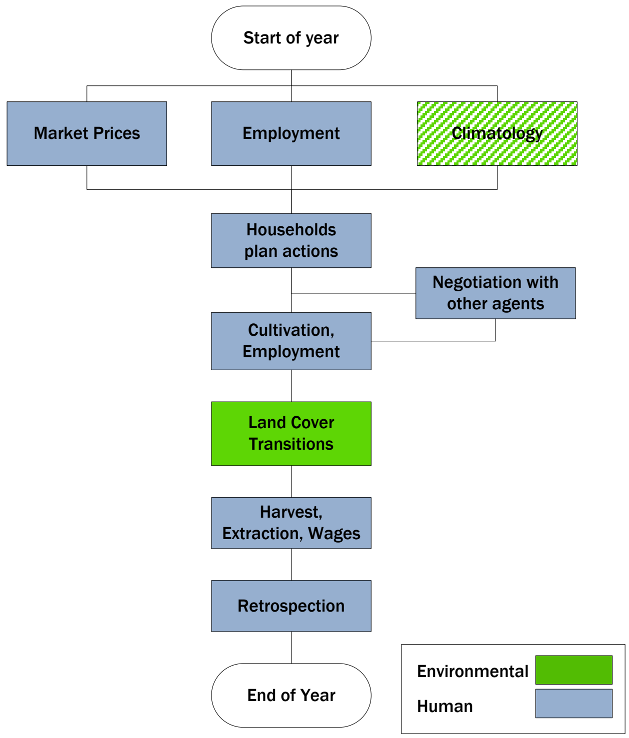


Figure 4.4: Timeline of a simulation step

4.3 Data preparation

4.3.1 Raster processing

Landsat 5 Thematic Mapper (TM) images from 1992 and 2006, as well as a 90m SRTM DEM, were used to prepare rasters for the environmental cellular automata model. Images from 1992 and 2006, in particular, show Paricatuba with only a small amount of cloud cover. To illustrate the Landsat image, bands 2, 3 and 4 (green, red and near infrared, respectively) were used to generate the false-colour image shown in Figure 4.5.

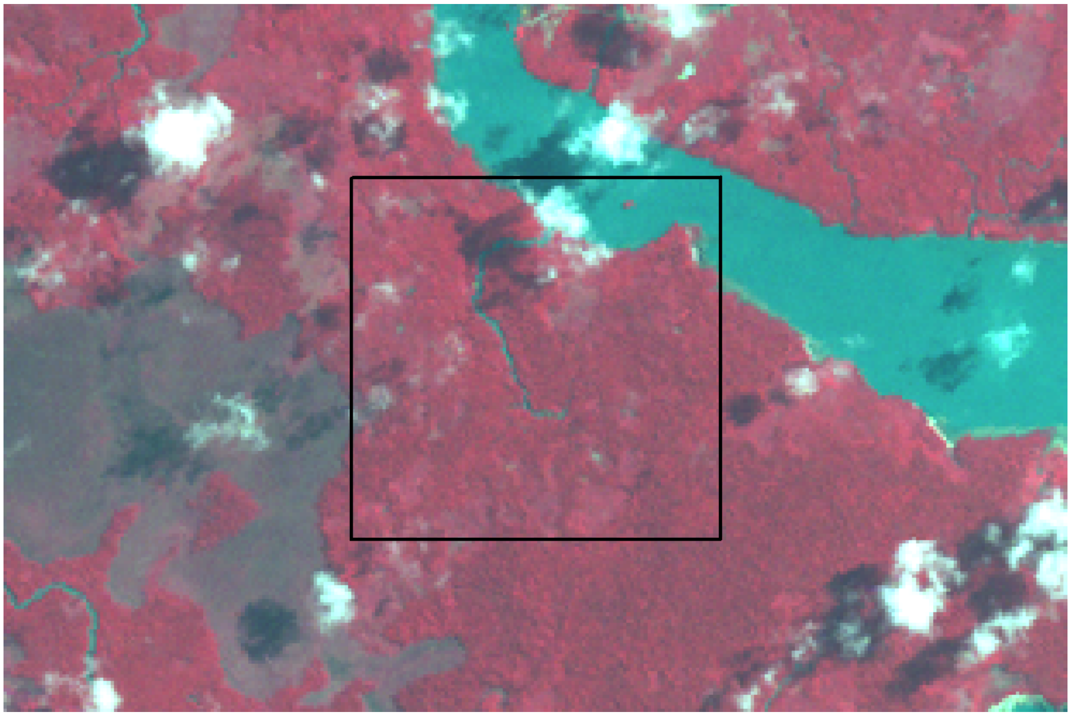


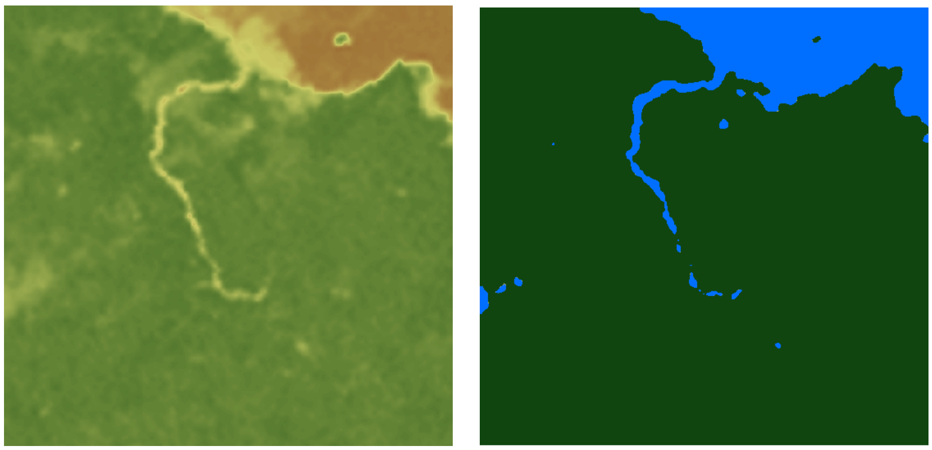
Figure 4.5: False-colour image of Paricatuba with model study area extents

For efficiency, the model assumes that all of its rasters are provided with the same extents and resolution. The model runs at the resolution (cell size) provided by its images. All images were interpolated using Kriging to 5 m resolution and snapped to the same extents before processing so that the processed images would be output at the same 5 m resolution, representing each of the same cells. Interpolation was done on the raw images, rather than processed rasters (e.g. land/water classification), to mitigate jagged edges resulting from the interpolation of crisp rasters. This was performed only on a rectangular area representing the area of interest, effectively clipping each source image to a 612×600 raster at 5 m resolution. The extents of this clipped area are outlined in Figure 4.5.

Using the normalized difference vegetation index (NDVI), raster cells were classified into land/water to restrict agriculture to land cells. NDVI can be used as an index of the health of vegetation, but in this case, it serves well to separate water from the land, which is assumed to be covered in vegetation. NDVI is calculated as:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{RED})}{(\mathrm{NIR} + \mathrm{RED})}$$

While vegetation has a high positive NDVI, water has either a low positive or a negative index, as shown in Figure 4.6(a). A threshold value was chosen such that most of the *igarapé* would be classified as water without producing too many false positives inland. A hard threshold of 0.25 produces the classified image in Figure 4.6(b).



(a) Calculated NDVI raster

(b) Classified land/water raster, threshold of NDVI = 0.25

Figure 4.6: Use of NDVI for land/water classification

Constant variables, such as Euclidean distances from water, are calculated using ArcMap tools. These variables are input as rasters into the model, which can be looked up by agents. This saves agents from recalculating variables during runtime. Each raster is converted to cellular variables during model setup.

4.4 Implementation

4.4.1 Platform

Maria is implemented in the Java language, using the latest version of the Sun Java Development Kit (currently Java 6u12), and version 1.2 of the Repast Simphony agent-based model development framework (North and Macal, 2007; North et al., 2007). Non-spatial results are written to a local or remote MySQL database for later analysis. R is used for data reporting and analysis, using either the RMySQL or RJDBC package to connect to the database. Spatial results are written to PNG files, displaying a classified land use map for each simulation tick.

4.4.2 Scheduling

MARIA is designed using a fixed-time scheduler, since decision-making and land use/cover dynamics occur at the same temporal resolution.

Evolving from the approach taken by LUCITA, MARIA arranges events within a simulation into some order. LUCITA used `preStep()`, `step()` and `postStep()` methods as the 3 stages of a simulation tick. Generally, `preStep()` was used for agent initialization code, `step()` was used for the bulk of the simulation, while `postStep()` was used for agent termination (when appropriate) and reporting. All agents were processed in the same non-random order between steps—not true Monte Carlo simulation.

MARIA utilizes Repast Simphony's support for scheduled method priorities to order events into stages while randomizing agent execution within each stage. Like LUCITA, MARIA orders initialization code at the beginning of a step and reporting code at the end of a step, but also orders decision-making separate from the execution of those decisions. This ensures that agents make decisions given the same information. Due to the random agent execution order within each stage, minor advantages from earlier or later execution will be acquired randomly, but steps have been taken to mitigate these relative advantages.

All stages and related method priorities are listed in Table 4.1. Methods are executed in descending order of priority, though the exact priority values are otherwise arbitrary. (However, negative priority values were used so that priorities would be executed in increasing order of magnitude. This was merely a decision based on aesthetic preference.) Methods with the same priority, such as methods shared between agents, are executed in a random order according to the simulation run's random seed.

Using these priority values, an agent's method can be scheduled into the simulation without modifying any code outside the agent class. The method is scheduled using the `@ScheduledMethod` annotation, as shown below in 4.1. The first line of code, the annotation indicated by the @ symbol, declares that the immediately-following method (`harvest()`) should be scheduled from the first method on, at every step (an interval of 1 step), with the `HARVEST` priority. By convention, most methods are named for the stage in which they are executed. Exceptions to this rule are many methods in the `ENVIRONMENTAL` stage, which are better described by their specific function, such as `transition()` and `offerJob()`.

Context	Stage	Priority
Environmental	ENVIRONMENTAL	∞
Human	MESSAGE_PASSING_1	0
Human	PLANNING	-1000
Human	MESSAGE_PASSING_2	-1100
Human	INTERMEDIATE	-1500
Human	MESSAGE_PASSING_3	-1900
Human	ACTION	-2000
Human	MARKETS	-10000
Human	MESSAGE_PASSING_4	-13000
Environmental	BIOPHYSICAL	-20000
Human	HARVEST	-50000
Human	RETROSPECT	-10^6
All	DATA_PREPARATION	-10^7
All	REPORT	-10^8
All	FINAL_REPORT	-10^9
All	CLEANUP	$-\infty$

Table 4.1: Scheduled method priorities

```
@ScheduledMethod(start = 1, interval = 1, priority = MariaPriorities.
    HARVEST)
public void harvest() {
  // ... harvest code ...
}
```

Listing 4.1: Implementation of goodness function for spatial land allocation

According to the priority schedule, all climatological models and exogenous actions execute first. These include the determination of market prices and employment offers and may, in the future, include flooding models if appropriate. This is followed by stages of human decision-making, communication and action stages. Within each human, non-messaging stage, household agents deliberate or act internally. At each message-passing stage, agents are able to send messages, but cannot process them until the next deliberative or action stage.

The effect of this arrangement is that all non-message passing stages are strictly internal to each agent and involve no information sharing between agents. The message-passing system implements a delayed update mechanism. This is necessary since Repast Simphony's agents update asynchronously: An agent's attributes update immediately upon modification. The message-passing system effectively synchronizes agent updates through communication and mitigates the relative advantages and disadvantages of execution order.

The scheduling of the most important stages and messages are shown with a UML sequence diagram in Figure 4.7, excluding the data recording and reporting stages. Each of these stages will be discussed in the context of their agents in the following sections. In general, land cells execute during the priorities marked "Environmental" in Table 4.1 and household agents execute in "Human" stages. Market agents and employer agents execute during the initial "Environmental" stage since they model exogenous factors. Finally, any and all agents execute the data preparation, reporting and cleanup code when appropriate.

4.4.3 Land cells

Each land cell contains several variables indicating the condition of the soil and its land cover. The soil condition is described by a fertility variable, an artificial counter variable abstracting a more detailed nutrient model. Land cover is described by attributes of land cover type, age, density (intensity), and health. Age is an integer variable in units equal to the length of a simulation tick (one year). The density/intensity variable is a fuzzy variable and can be described as percentage cover. While most cellular automata models strictly assign one land use per cell, Maria uses fuzzy variables within the environmental context to allow multiple land uses in one cell, subject to constraints: At the cellular level, rules ensure that the land cover composition of a cell is valid. Validity rules ensure that cells do not support more than their capacity.

The health variable is closely tied to the fertility variable, but a distinction is made in case land cover health is dependent on other variables. For example, later models may include upland areas and land uses better suited to *terra firme*. If upland vegetation is cultivated on floodplain, the health of the upland vegetation would remain low while the fertility of the soil is high. The current version of Maria ties land cover health to fertility, in the absence of other factors, such as abandonment (characterized by a lack of maintenance).

Internal cell state transitions determine yield. State transitions are Markovian, based on land cover history and current state variables. At each simulation step, the `transition` method is called. This method first increments the age of each land cover on the cell, then transitions 10-year-old secondary succession to mature forest.

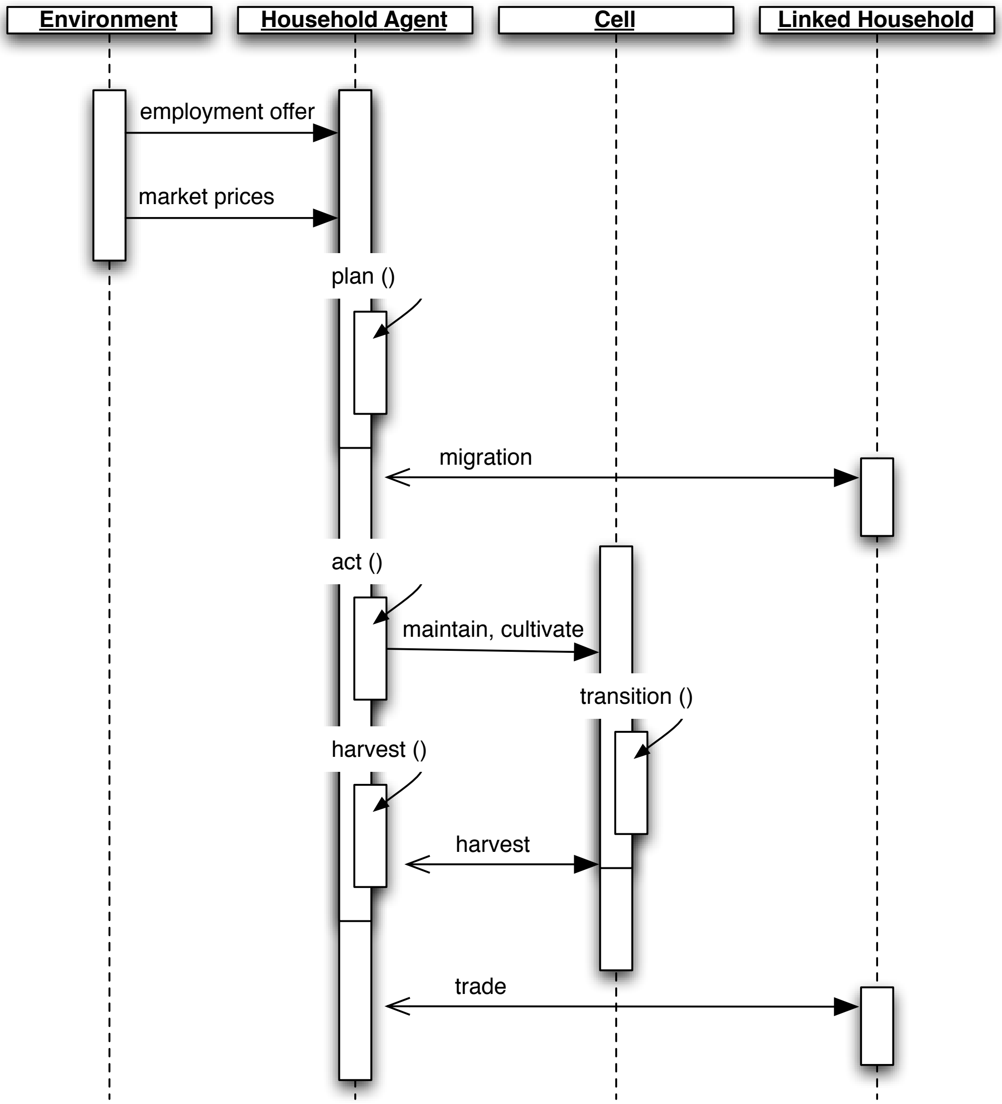


Figure 4.7: UML sequence for a simulation step

The fertility variable counts down or up as nutrients are removed or returned to the soil, respectively. If the land is being cultivated, the fertility variable is decreased by a small amount based on the land cover's cultivation period (Section 4.2.4). The fertility variable is decreased by an amount such that it will reach zero at the end of the maximum cultivation period for the current land use, assuming the fertility variable begins to decrease from a value of one. If the cell's land cover density or intensity is less than 100 %, the rate of decrease is scaled by the density.

$$\text{fertility} = \text{fertility} - \sum_{c} \frac{1}{p(c)} \times d(c)$$

However, if the land cell is undergoing secondary succession or is covered by mature forest, the refertility rate is calculated based on the last land use, l, based on the recovery period ($r(l)$). In this case, fertility is adjusted as follows:

$$\text{fertility} = \text{fertility} + \sum_{l} \frac{1}{r(l)} \times d(l)$$

After fertility calculations have been completed, health variables of land cover are adjusted if fertility is too low. Currently, the model decreases land cover health by 50 % is fertility is lower than 5 %. Health is similar to the fertility variable, though it refers to biomass health rather than soil conditions. That is, health reflects the yield of the land use/land cover, while the fertility variable acts as an abstraction of soil resources. Only when the fertility variable is too low, the health of the land cover begins to decay.

When a harvest is available, the health variable is used to model yield. The land use type's baseline potential yield, determined from field surveys, is multiplied by the health of the land use. Since health is a fuzzy variable, it is used as a percentage of potential yield.

$$\text{yield} = \text{health} \times \text{potential\_yield} \times 100\%$$

4.4.4 Households

Since alternative decision-making methods are to be compared between simulation runs, it is preferable to ease the replacement of one type of decision-making by another. Substitution is performed using polymorphism. Specifically, the Strategy design pattern (Gamma et al., 1995) is used to promote code reuse. A common `HouseholdAgent` abstract class defines a common interface to the rest of the model and provides common functions, as shown by the UML diagram in Figure 4.8. For instance, a data collector may need to retrieve labour and capital information, so common `getLabour()` and `getCapital()` methods need to be exposed and implemented by both types of households. New types of households are created by instantiating a concrete implementation of `HouseholdAgent`, such as `DecisionTreeHouseholdAgent`.

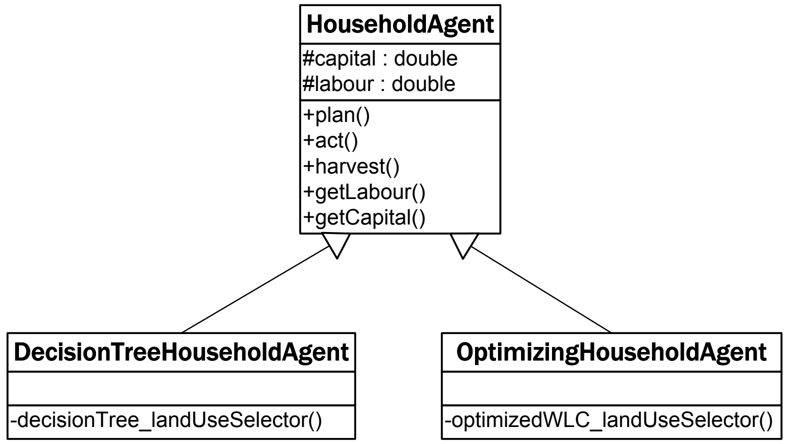


Figure 4.8: Household UML class diagram

With respect to the schedule detailed in Table 4.1, the primary stages executed by a household agent are the planning, action and harvest stages. When appropriate, the household will also participate in message passing. In general, a household collects an inventory of its land use assets and devise a plan in the planning stage. In the action stage, the household manages its land, performing maintenance and expanding cultivation if constraints allow. Finally, agents can harvest and sell goods in the harvest stage.

Decision tree-based household agents

As its name suggests, the `DecisionTreeHouseholdAgent` class provides decision tree-based implementations of household decision-making. In particular, its `plan()` and `act()` methods use decision trees to quickly limit the number of feasible solutions. Heuristics are used to select between the remaining alternatives. A household's decision tree is shown in Figure 4.9. Three of its terminators result in deterministic choices. The fourth choice recommends the expansion of managed agriculture, which requires further decision-making. The decision tree process is repeated until no feasible actions remain. This occurs when land, labour and capital resources run out.

In the planning stage, the household first considers any employment offers, which if taken would reduce agricultural contributing labour in favour of external income. Employment offers include a specific non-negotiable annual wage, which the agent can use to compare against past income. At the end of each simulation step, the agent calculates profit for the year, which can be divided by annual contributing labour in person-years to approximate the opportunity cost of labour lost in units of currency

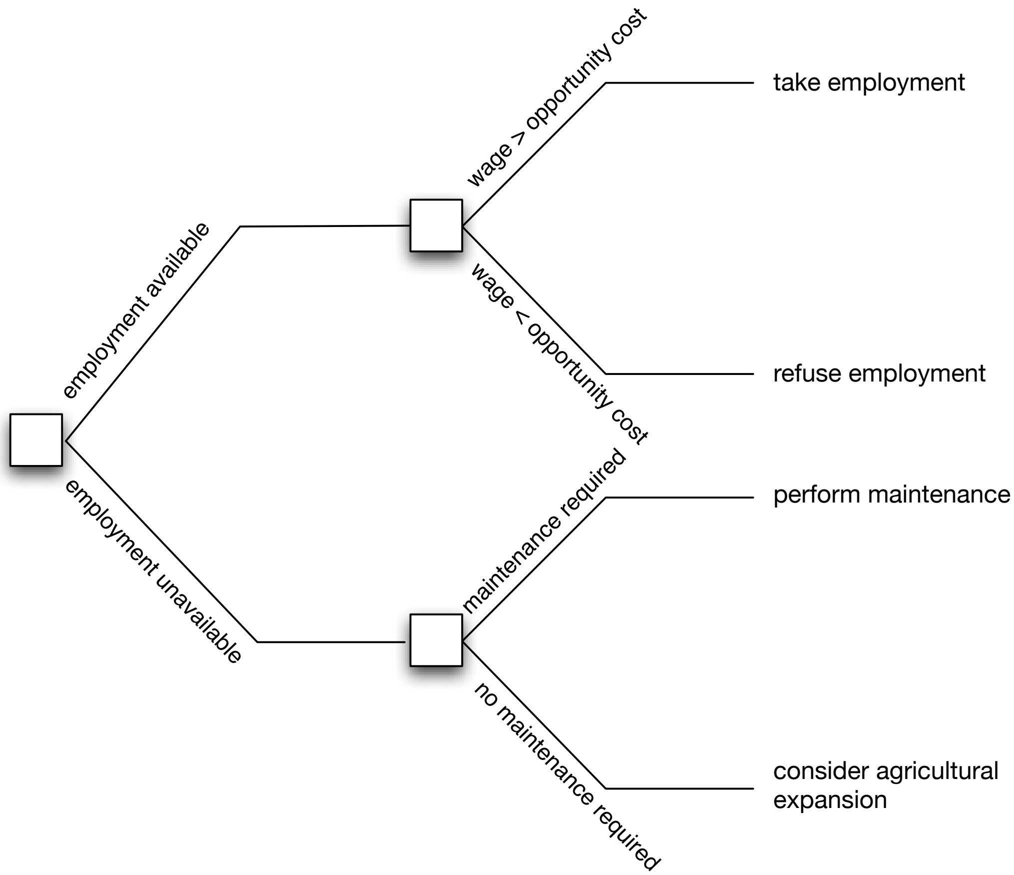


Figure 4.9: Household decision tree

per person-year. If a household member is available for employment, the household compares the annual wage with opportunity cost of foregone contributing labour. If the annual wage is higher than the opportunity cost, the household member leaves to pursue employment.

A household member is only available for employment if there exists an adult household member who is not the sole male or female adult in the household. This restriction roughly keeps family units intact and, in a very small way, accounts for the social cost of migration.

The household takes an inventory of its land cells and classifies land into cells available for development, cells requiring maintenance and cells of each land use. Cells are stored in reference lists, which will be used in later steps. At this point, nothing has been acted upon, but the household has collected an inventory of potential actions.

In the action stage, the decision tree household agent works on cells requiring maintenance first. Agents are assumed to be loss averse, not wanting to lose potential yield for lack of maintenance. After maintenance is complete, remaining labour and capital can be used to cultivate new land until either runs out. Cell by cell, the household selects a new land use randomly, weighted by the expected revenue of the new land use. A custom `WeightedSelector` class was built for this purpose. `WeightedSelector` has two important methods, `add(...)` and `sample()`. Land uses and expected revenues are added to an empty `WeightedSelector` using the `add(...)` method. The `sample()` method is used to select a land use. As long as there is available land, capital and labour, a household agent will select a land use and place it onto a cell. Details of how a cell is selected is discussed in Section 4.4.6.

Harvesting prioritization is performed in the same way. A new `WeightedSelector` is set up, containing all goods available for harvesting and weighted by revenue. A household samples a good from the `WeightedSelector`. The selected good is harvested from one cell and sold immediately. This process is repeated as long as capital and labour are available or until all cells have been harvested.

Finally, the `retrospect()` stage is used to calculate this year's profit. Combined with contributing labour, this information will be used to evaluate opportunity costs versus income from potential employment in later years.

Linear programming household agents

In this version of the model, optimizing household agents use linear programming to maximize revenue from scarce resources of capital, labour and land. The LP household agent uses two linear programs, one to allocate maintenance and new development and the other to allocate harvesting resources. The first linear program determines land use composition but not spatial land allocation, which is handled with a separate process. That is, the optimizing algorithm will determine the *number* of cells of each plot, but will not specify the locations of these cells. Land allocation is performed in the same manner as other household agents, as discussed in Section 4.4.6. When planning the cultivation of new plots of land, households first take an inventory of existing land uses. Using this inventory, the household determines the amount of land available for development and the amount of land under each land use. Job offers are also evaluated in this linear program. First, the youngest eligible members of the household are paired with available job offers, so the household knows the opportunity cost of lost agricultural labour. Household members working off the farm are also considered for recall. Their annual wage is known, as well as their contributing labour, allowing the linear program to optimize labour assignments between agricultural and non-agricultural work for maximum revenue.

The optimization problem to maximize revenue is set up, where n_{landuse} denotes the number of new cells of some land use and m_{landuse} represents the number of maintained cells of the land use. The selling price p is in units *per cell*, based on an average yield per cell known *a priori* multiplied by the current market price. Elements i are the members considered for emigration from the farm: e is a Boolean variable indicating whether or not the member is sent out (**`true`** if the member does emigrate) and w represents their annual wage. Similarly, elements j are members considered for recall back to the farm. r is a Boolean variable indicating whether or not the member is recalled, but for inclusion into a maximizing objective function, r is negated, so $\bar{r}$ indicates that the member is *not* recalled. l indicates the labour requirements of some land use activity or the contributing labour of an agent. c represents the capital requirements of a land use activity.

$$
\begin{aligned}
\text{max} \quad & p_{\text{acai}} n_{\text{acai}} + p_{\text{garden}} n_{\text{garden}} + p_{\text{acai}} m_{\text{acai}} + p_{\text{garden}} m_{\text{garden}} + \sum_i e_i w_i + \sum_j \bar{r}_j w_j \\
\text{s.t.} \quad & l_{\text{n\_acai}} n_{\text{acai}} + l_{\text{n\_garden}} n_{\text{garden}} + l_{\text{m\_acai}} m_{\text{acai}} + l_{\text{m\_garden}} m_{\text{garden}} \\
& \quad + \textstyle\sum_i e_i l_i + \sum_j \bar{r}_j l_j \leq \text{Labour} \\
& c_{\text{n\_acai}} n_{\text{acai}} + c_{\text{n\_garden}} n_{\text{garden}} + c_{\text{m\_acai}} m_{\text{acai}} + c_{\text{m\_garden}} m_{\text{garden}} \leq \text{Capital} \\
& n_{\text{acai}} + n_{\text{garden}} \leq \text{Land} \\
& n_{\text{acai}} \leq \text{Land}_{\text{acai}} \\
& n_{\text{garden}} \leq \text{Land}_{\text{garden}} \\
& \textit{... non-negativity constraints ...}
\end{aligned}
$$

The objective function states that the household wishes to maximize the expected revenue of all goods across all managed and unmanaged cells, calculated by multiplying current selling prices by the expected yield, in addition to the revenue gained from off-site employment. Values of n, m, e and r are selected to maximize the objective function while avoiding the violation of constraints. Off-site employment generates annual wages for both newly employed agents, i, and existing employed agents who are not recalled, j. The household does not attempt to forecast future prices of goods, but instead, uses current prices as an approximation of future prices. Non-negativity constraints are enforced on every variable, preventing the feasibility of negative solutions.

The optimization is performed using a linear programming solver, lp_solve, an open-source mixed integer linear programming solver (Buttrey, 2005) (currently available at `http://lpsolve.sourceforge.net/`). The lp_solve library uses the Simplex method (Nash, 2000) to find the optimal solution.

Note that the table of constraints does not include integer constraints to restrict the numbers of cells to integer solutions. Typically, a cellular automata model is constrained to strictly one land use per cell, which would necessitate integer constraints. However, since the cellular automata model is built upon fuzzy variables, there is support for multiple land uses. That is, agents can plant some proportion of both gardens and açaí on the same cell. However, the variables indicating employment or recall are Boolean variables, which do not allow partial employment or recall.

An underlying assumption of this simple linear program is that current selling prices will hold. More importantly, another assumption is that the future value of revenue is the same as the present value. Açaí, in particular, takes approximately 3 years to produce full yield. This simplified model serves to illustrate the differences between (nearly) optimal algorithms and those which are heuristic and non-optimal.

At harvest time, a similar linear program is constructed to determine which cells should be harvested. l and c represents the labour and capital requirements to harvest a cell. p is constructed by multiplying the current market price by the yield averaged over every cell of that land use type. Here, açaí extraction is separated from the harvest of intensively-managed açaí cells, since yields vary greatly between these types of cells. In this stage, agents are also able to recall working agents, at an opportunity cost of half their annual wage. It is assumed that the harvest season runs from September to February. Intensive management strategies have extended the harvest season, once from September to December, to the modern 6 month period (Brondízio, 2008). The linear program for harvesting optimization is as follows, with upper bounds and non-negativity constraints omitted for brevity:

$$\begin{aligned}
\max \quad & p_{\text{acai}} h_{\text{acai}} + p_{\text{managed_acai}} h_{\text{managed_acai}} + p_{\text{garden}} h_{\text{garden}} + \sum_j \frac{\bar{r}_i w_i}{2} \\
\text{s.t.} \quad & l_{\text{acai}} h_{\text{acai}} + l_{\text{managed_acai}} h_{\text{managed_acai}} + l_{\text{garden}} h_{\text{garden}} + \textstyle\sum_i e_i l_i + \sum_j \bar{r}_i l_i \leq \text{Labour} \\
& c_{\text{n_acai}} n_{\text{acai}} + c_{\text{n_garden}} n_{\text{garden}} + c_{\text{m_acai}} m_{\text{acai}} + c_{\text{m_garden}} m_{\text{garden}} \leq \text{Capital}
\end{aligned}$$

4.4.5 Household settlement

All households settle in the study area at the start of the simulation. To select a household settlement cell, each household takes a sample of 100 cells out of the 612×600-cell area. Of these 100 cells, the household chooses the land cell which is closest to the stream or river. The chosen cell is marked as the location of the agent's house, though it may be used for other land uses such as gardens or açaí. The location of the house becomes relevant for land use allocation, as agents may prefer to cultivate land close to home.

After the household cell is selected, the household claims 1–50 ha of land by claiming available plots of land. First, a property size of 1–50 ha is determined by sampling from a uniform distribution. The number of cells is then calculated using the random property size and the cell size of the model. The household claims that number of cells, if available, by claiming plots of land within an increasing distance from the initial settlement cell.

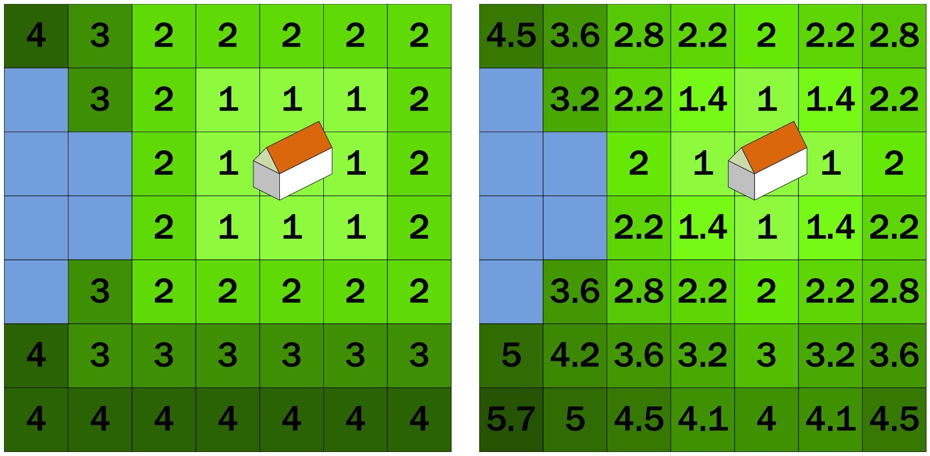
The distance from the house is calculated as the Chebyshev distance. The Chebyshev distance heuristic produces a square or near-square property. In contrast, a Euclidean distance heuristic would produce a more circular property. Without cadastral data, it is safe to assume that rectangular properties are more typical than circular ones, so a Chebyshev heuristic is preferred.

Figure 4.10 shows a comparison between Chebyshev and Euclidean distance heuristics. Land is claimed, cell by cell, in order of darkening cells. Cells of the same distance value may be claimed in any order. If a cell is a water cell, it cannot be claimed and no other adjustments will be made, since the household will continue to increase the distance window until the specified number of cells has been claimed. Previously-claimed land is also avoided.

Using the Chebyshev distance heuristic for 20 randomly placed households produces a settlement pattern such as the one illustrated in Figure 4.11. Settlements may coalesce, but land claims are granted on a first-come first-served basis. Each household settles and makes all land claims at once. Successive households settling nearby will not claim land from the first owner.

4.4.6 Spatial land allocation

In deciding where to cultivate land, agents first take an inventory of all used and available land. Counting and indexing each type of cell, agents can then rank cells



(a) Chebyshev distance (b) Euclidean distance

Figure 4.10: Chebyshev vs. Euclidean distance heuristics for land settlement

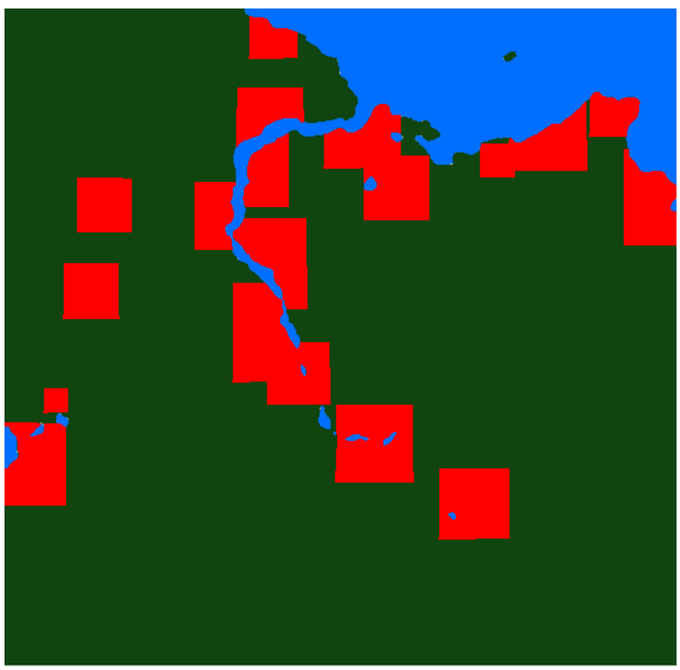


Figure 4.11: Land settlement pattern using the Chebyshev distance heuristic

based on preference. These preferences are implemented as heuristic functions which quantify the "goodness" of a cell as a single scalar value.

The heuristic function is wrapped in a `Comparator` class, which is used in conjunction with a `PriorityQueue` to sort all available cells by the heuristic "goodness" function. By storing all available cells in this `PriorityQueue`, cells are efficiently ranked by goodness as inventory is taken.

Note that the heuristic functions described do *not* produce optimal solutions, but merely provide methods for the agents to quickly evaluate the goodness of each cell. Since the land in this model is initially homogeneous (as discussed in Section 4.4.3), the chosen spatial land allocation method merely results in a particular visual representation and does not affect the economics of the model.

Effectively, land used within the recovery time period (Sections 4.2.4 and 4.4.3) will result in a sub-optimal yield. If these sub-optimal cells are removed from consideration, then all remaining cells will produce optimal yield. Alternatively, cells could be weighted by their fertility, but households do not have access to the fertility variable, nor are they aware that it increases and decreases linearly. Available land is not a constraining variable, since households have much more unmanaged land than managed land (Brondízio, 2008). The number of sub-optimal cells in a given year is comparatively small next to the amount of unmanaged, available land. The removal of these sub-optimal cells from consideration is expected to produce reasonable results.

Distance from house heuristic

One heuristic used by farming agents minimizes the distance from the house (settlement cell). This minimizes labour from walking to/from the house during cultivation. The heuristic function is simply:

$$f_h(c) = -\text{distanceFromHouse}(c)$$

The distance function is negated since the sign convention of all heuristics is taken such that positive values indicate goodness and negative values are unfavourable.

Like-land use adjacency heuristic

The like-land use adjacency heuristic attempts to cluster similar land uses. This heuristic counts the number of cells in the 3x3 Moore neighbourhood matching the candidate land use, l:

$$f_h(c) = \text{numNeighbours}(c, l)$$

Effectively, cells with more similar neighbours are weighted highly.

Composite heuristics

Multiple criteria can be combined into a single scalar value using a weighted linear combination of factors (Eastman et al., 1998). Taking the same goodness functions as above and assigning them weights, a scalar goodness value can be calculated as follows, where $f_h(c)$ is the heuristic function evaluated at cell c and $w(h)$ is the weight assigned to the heuristic h:

$$g = \sum_h w(h) \times f_h(c)$$

Since the heuristic functions, as previously discussed, vary significantly in range, the outputs of the heuristic functions must be normalized. For example, the distance from house function produces outputs ranging from 0 to values in the thousands, since it returns the distance in metres. Meanwhile, the like-land use adjacency function returns values from 0 to 8. Adding these heuristic functions together, unmodified, with the same weights would result in the distance to house being weighted significantly more than like-neighbours. Therefore, when adding heuristic functions together, care must be taken to normalize the heuristic functions. A simple way to normalize the distance function is to divide it by the maximum distance, the distance from the house to the furthest extent of the property, returning a distance value linearly scaled to the range of [0–1]. Similarly, the like-neighbour adjacency function can be divided by 8, as there are 8 neighbours in the Moore neighbourhood. Dividing the count by 8 will produce a heuristic output in the range of [0–1]. Non-linear scaling may also be performed in order to weight changes at one end of the range differently from changes at the other end, but such complexity will not be discussed here.

Implementation of heuristics into the decision-making process

Each land use can be assigned a different heuristic function, whether it uses a single or composite heuristic function. It is assumed that different land uses have different needs. For instance, housegardens are typically located close to the home, whereas açaí stands may not. However, açaí may be allocated closer to the nearest water body.

A code listing which illustrates the use of a goodness function inside a `Comparator` is shown in Listing 4.2. Note that the goodness values are negated within the `compare` method. This is done because the head of the `PriorityQueue` is the object with the *lowest* priority value. In other words, objects from a `PriorityQueue` are retrieved in order of ascending priority. This is in contrast to the scheduled method priorities described in Section 4.4.2, which were executed in descending order. To correct for this, goodness values are negated such that the highest goodness value, or magnitude, executes first.

```
public class AcaiGoodnessComparator implements Comparator<MyLandCell> {
  public int compare(MyLandCell o1, MyLandCell o2) {
    double w1 = -getGoodness(o1);
    double w2 = -getGoodness(o2);

    return Double.compare(w1, w2);
  }

  private double getGoodness(MyLandCell c) {
    return -c.getDistanceFromHouse()
      + c.getNeighbourLandUseCounts(LandUse.ACAI) * 100000000000d;
  }
}
```

Listing 4.2: Implementation of goodness function for spatial land allocation

4.4.7 Agent communication

Rather than using KQML (Finin et al., 1997) or FIPA-ACL (Foundation for Intelligent Physical Agents, 2000), both string-based messaging protocols which would require greater computational overhead for encoding and parsing, an application-specific object-oriented framework was developed for message passing. Interoperability with other agent systems is not a concern, so there is no need for a common protocol between systems. Instead, performance and maintainability are a priority.

All messages implement a common interface `Message`, exposing methods to determine the type of message as well as the contents of the message. This is shown by Figure 4.12, along with a few message types used by MARIA. The contents may be any Java object, while the message type is used to assist the receiver in determining the class of Java object. The receiver may execute an efficient **`switch`/`case`** code block instead of casting the message object with trial and error. If the message is tested with if/then statements and the **`instanceof`** operator, a message class must be checked against each possible message class until a match is found. Using the switch/case method with a message type integer enumeration, the message type is only checked once. If the receiver receives an unsupported message, it may throw an error or fail silently. The latter is preferable in a large batch run, while the former is useful during debugging to detect errant messages.

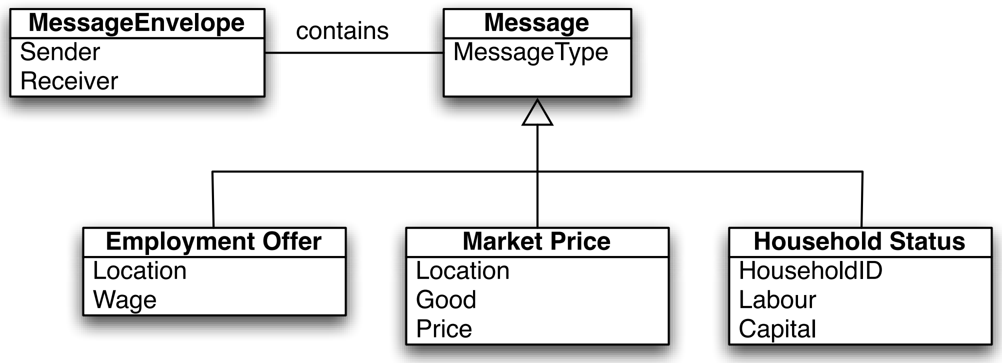


Figure 4.12: UML class diagram for a message

Each `Message` object is enclosed in a `MessageEnvelope` which adds sender and receiver attributes.

All messaging-capable agents are implementations of the `NetworkAgent` interface. The `NetworkAgent` interface exposes two methods, `store(MessageEnvelope env)` and `getAgentType()`. The `getAgentType()` method returns an enumeration encoding the agent's type, which is used to efficiently process incoming messages. The `store` method is used to send messages:

When an agent wishes to send information, it constructs a `Message` and obtains a `MessageEnvelope`. Sender and receiver attributes are attached to the envelope. Broadcast messages—messages sent to any and all agents—are not assigned specific receiver attributes. After the `MessageEnvelope` is fully constructed, the receiving agent's `store` method is called. This places a reference to the message envelope in the receiving agent's mailbox, which can be processed by the receiving agent at a later time. From the information stored in the `MessageEnvelope`, the receiving agent can determine the sender, the intended receiver (broadcast or unicast) and the message type. This allows the receiving agent to sort through messages quickly.

The message passing system is used by multi-sited households, market agents and employer agents. Multi-sited households share state messages containing available capital and labour, with the intent that trade of these resources can by negotiated between linked households. Market agents broadcast the year's current market prices to all agents, while employer agents broadcast job offers with proposed wages to a limited set of agents.

Note that the term "agent" is used quite loosely here, as market agents and employer agents do not "sense" their environment and, thus, cannot react to it. However, market and employer agents are implemented as agents, in case they may require this ability in the future. For instance, more advanced market agents may adjust their prices based on supply and demand, while employer agents may respond to declined offers with counteroffers. For now, market agents and employer agents are simply beacons of information, with the difference that employer agents' messages are not intended to reach all households.

Employer agents

A single employer agent, representing all sources of urban employment opportunities, is placed in Belém for illustrative purposes. This employer sends a number of job offer messages, which include annual wages and the employer's location. An accepting agent, an individual from a farming household, would emigrate to the location and collect annual wages until the agent migrates again (due to another economic opportunity).

Job offer events are modelled as a Poisson process, in which events occur independently of each other: The interarrival time between successive employment opportunities is generated from an exponential distribution. The exponential distribution's rate parameter, λ, is used to control employment availability.

Since the exponential distribution is continuous function and the model operates on a discrete-time schedule by year, fractional offer dates are truncated (rounded down) to integers in order to be scheduled. On model startup, a pseudorandom number generator based on the parameterized exponential distribution determines the dates of the employment offers for the duration of the model run. Each generated random number is an interarrival time, so an accumulating variable is used to date each offer. The accumulator's date is truncated and recorded as a scheduled offer. Multiple offers may be scheduled in one year. Households receiving each offer are selected at random with uniform probability and with replacement. New offers are not made if an offer is declined for any reason.

4.4.8 GIS integration and visualization

Since the cellular environment is derived from a remotely sensed, georeferenced image, agents living in this environment can be georeferenced as well. When an agent settles into a grid cell (Section 4.4.5), its UTM coordinates are derived from the cell's (x,y) coordinates and the grid's UTM coordinates. Outside of a grid, when an agent migrates to pursue employment, the employer agent's UTM coordinates are used. For each agent, random displacement is added to the employer's location to spread out the agents for visualization purposes.

The Landsat images use a UTM projection (zone 22S, WGS84 datum and ellipsoid). An earlier incarnation of Maria, calibrated for use in Mazagão, used classified remotely sensed images stored with a SAD69 projection. However, World Wind requires data to be provided in WGS84 with lat/lon coordinates. For visualization, all source data is transformed at runtime using the JTS Topology Suite (Vivid Solutions, 2003), a Java application programming interface for GIS.

Georeferenced agents and their social networks can be visualized on a 3D globe using NASA World Wind as shown in Figure 4.13. This allows for the visual exploration of agent settlement and social networks across a variety of scales. Spheres, in the distance, represent rural households while cubes show urban agents. A rural household's colour is determined by hashing its ID number into red, green and blue components using modulo arithmetic. Social networks and agents spawned from a household are assigned the same colour as the source household. The background image is derived from Landsat 7 images (bands 1, 2 and 3) donated to NASA by i-cubed for use within NASA World Wind.

Due to the number of cells in the landscape, rendering land uses as 3D sprites or polygons is computationally infeasible in real time, so 3D renderings are restricted to agents and networks. If desired, this can be addressed in the future by rendering land uses as a 2D image using a Plate Carrée projection. This image can be rendered as an overlay onto the globe (Boschetti et al., 2008). At the moment, land use and land cover maps are rendered as simple 2D grid images at a resolution of one pixel per cell and are not rendered on the globe.

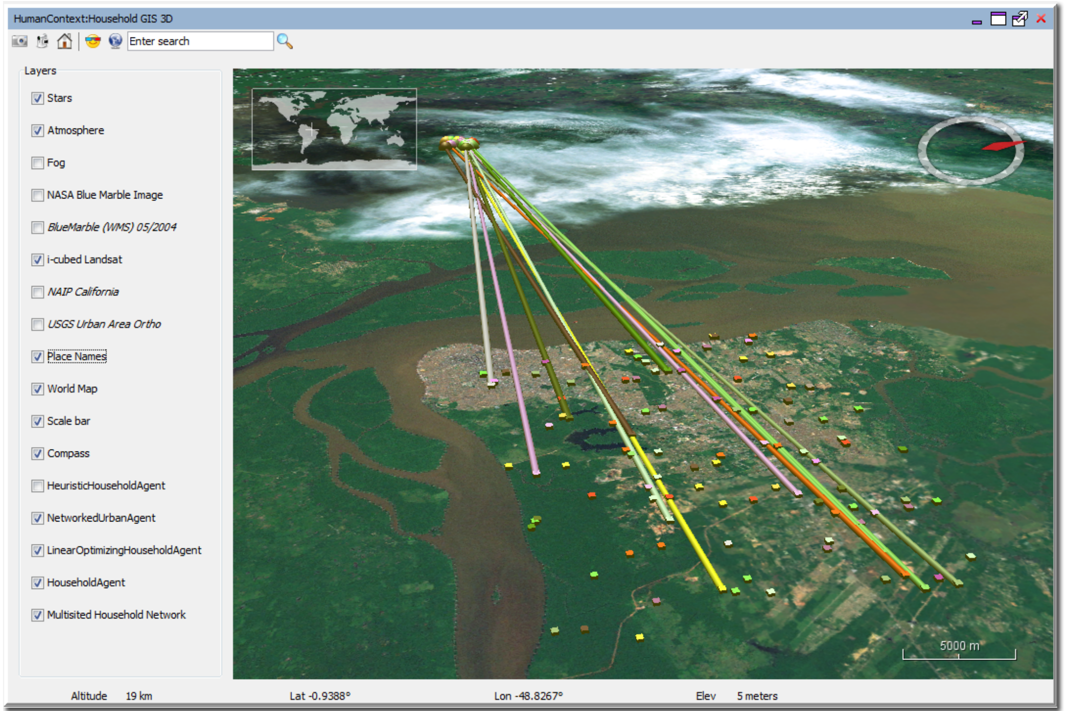


Figure 4.13: 3D GIS visualization of multi-sited households

4.4.9 Sensitivity analysis and parameter sweep implementation

Since the model is non-linear, it is infeasible to derive the model's sensitivity to a parameter variable mathematically, without running the model several times. Instead, a sensitivity analysis requires the model to be run several times, varying the parameter between each run. A batch of such runs is a parameter sweep. In Repast Simphony, a parameter sweep can be defined as an XML file. An example "nested" parameter sweep is shown below, in which for each value of the `acaiPrice` variable, all of the `capitalMultiplier` values are swept. In other words, the model is run with each possible combination of the listed pairs of the optimizing and heuristic household variables.

Using XML or other parameter sweep formats, such as Groovy scripts, arrangements other than lists and nesting are possible. However, these formats will not be discussed.

```xml
<?xml version="1.0"?>
<sweep runs="1">
  <parameter name="acaiPrice" type="list" value_type="double"
      values="1.0 2.0 3.0">
    <parameter name="percentOptimizingHouseholds"
        type="list" value_type="double" values="1.0 0.0" />
    <parameter name="percentHeuristicHouseholds"
        type="list" value_type="double" values="0.0 1.0" />
  </parameter>
</sweep>
```

Listing 4.3: Example list-based nested parameter sweep

The XML listing above produces a tree shown in Figure 4.14, with parameter names shortened. Each leaf node can represent a set of input parameters, which can be read by traversing from the leaf node to the root. When a leaf is visited, the input parameters of the model are set to the values of the node and its ancestors. Using Repast Simphony's internal `DefaultParameterSetter`, all leaves of the tree are traversed in left-to-right order.

Alternatively, if some desired model output is known, such as in a maximization problem, an optimized parameter sweeper may be used. An optimized parameter sweeper will only visit a subset of nodes such that "optimal" parameters can be found without searching the entire parameter space. Such optimizing algorithms include branch and bound algorithms (Lawler and Wood, 1966) and simulated annealing (Kirkpatrick et al., 1983). An optimized parameter sweeper is not appropriate for this thesis, since the model is not used for decision-support or with goal-finding in mind. Much of the analysis is performed by experimentation, a wide exploratory sweep of parameter values with no particular objective function to be optimized.

Details on the specific parameter sweeps used in analysis are discussed in Section 4.5. The strategy used to distribute jobs from one batch specification to a cluster of computing nodes is detailed in Appendix A.

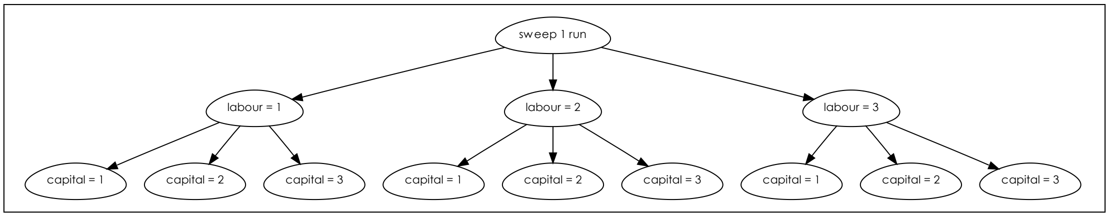


Figure 4.14: Parameter sweep tree generated from XML.

4.4.10 Reporting

Database design

Model results, such as input parameters, household state and networks, are logged to a central database, while spatial data is recorded as sets of classified images outside the database. Within the database, input parameters and a unique run ID are recorded at the start of a run. The random seed is included to allow results to be reproduced. In case a batch run is performed, the batch's name is included in the run's record to assist batch analysis, such as sensitivity analysis.

As each household is created, its unique ID is entered into the database's household table, linked to the current run by a foreign key reference. At the end of each simulation step, the household state is recorded, including land use composition and harvest amounts. Emigrant agents are recorded as "urban" (off-farm) agents, with states recorded at each time step as well. The stage attribute of the state tables makes reference to one of the stages in Table 4.1, recording exactly when—within a tick—the state is recorded. The database schema is shown in Figure 4.15, excluding foreign key attributes for brevity, which would otherwise be repeated from referenced tables. Outside the database, spatial images are organized into directories identified by the unique run ID used to generate the data. Spatial data may be integrated into a database in a future model, but this would greatly increase the amount of data transfer across the network in the case of distributed batch runs and detailed spatial analysis is not required at this time.

Each model run is treated as a single transaction, such that if the run fails to complete or is currently running, its partial results are not included in reports. Typical causes of run failure are memory errors, bugs in the code or simply that the process has been aborted. The latter is especially true in a clustered setting, where nodes or network connections may fail.

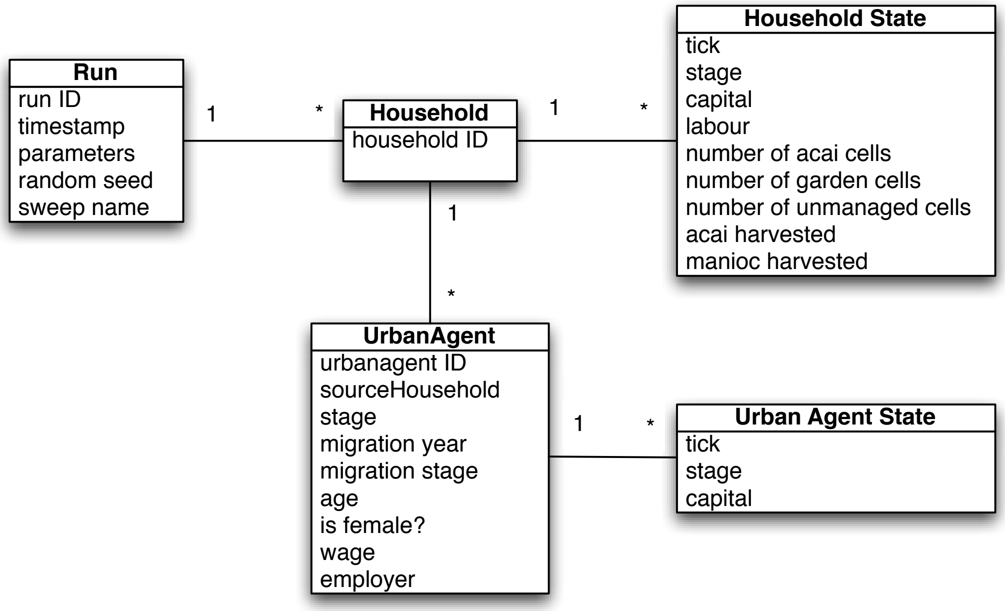


Figure 4.15: Database Schema

Non-spatial model results were logged using JDBC (Java Database Connectivity), an application programming interface used to facilitate communication between a client application and a database. JDBC provides a common programming interface for connection establishment between alternative database implementations, such as MySQL and Oracle. Through the JDBC connection, SQL (structured query language) statements allow the client to manipulate or retrieve data. However, there are minor differences in each database provider's implementation of SQL and in the features each database supports. Custom wrappers were written around the JDBC drivers to provide an even more common interface to the rest of the application. This simplified the development of the simulation by replacing implementation-specific SQL code with more friendly code such as `logNewHousehold(...)`. This also eased the transition between alternative database implementations.

The model application originally used the H2 database for its embedded database performance but migrated to MySQL later in development. MySQL was found to perform much better for post-run analysis especially due to its ability to aggregate data within the SQL query itself. H2 is optimized for use as an embedded database, where a Java-based client application has exclusive access to a database. H2 only has a limited implementation of SQL with no support for aggregate queries. Another database implementation, HSQLDB, was also evaluated but performed poorly in terms of speed and memory usage due to the size of the transactions and the limited amount of RAM available on each client workstation.

Spatial data handling

To limit the communication overhead from writing results to a database, spatial results in the form of classified images are written locally to a set of PNG (portable network graphic) files, one per time step. The PNG format, unlike JPEG, maintains the precision of discrete data, while losslessly compressing the image to a reasonably small size. JPEG images, in contrast, use lossy compression and are more suited for smooth images such as photographs, as artifacts are produced in areas of sharp transition. Although PNG files are small, one image is required for each step of each run. In a batch of thousands of runs, the bandwidth required to transmit spatial data is significant. Therefore, in a distributed batch run across a network of nodes (Appendix A), a compromise was made to minimize data transmission while maintaining some degree of data availability at a single site.

The arrangement of centralized non-spatial data and distributed spatial data is a compromise between performance and usability, reducing the network bandwidth usage at runtime while allowing for the quick aggregation of model results by non-spatial queries in a single, central database. After reducing the set of tens of thousands of model runs to a much smaller set of "interesting" runs, the spatial data for these runs can be reviewed. Otherwise, if all non-spatial and spatial data were centralized, the communication overhead would be much higher. Alternatively, if non-spatial data were distributed as well, several distributed databases would need to be queried during analysis. High availability and redundancy were not requirements, so database replication was not necessary.

4.5 Runs and analysis

Known parameters are integrated into Maria as described in their respective sections of Section 4.4 (summarized in Appendix B), using stochastic methods to compensate for uncertainties. Unknown parameters in Maria were calibrated by performing large parameter sweeps, made feasible through the use of distributed computing (Appendix A). Quantifiable seasonal or yearly labour requirements, for instance, are not available, so values for these requirements were determined by experimentation. Variables were calibrated by introducing them into a simplified model, one component at a time. Beginning with gardens, labour requirements, capital requirements and *constant* annual market prices for garden-produced goods were introduced into the model while other agricultural activities were made infeasible (through high labour requirements). Labour and capital requirements and constant market prices were experimentally determined such that the system reached a stable steady state for most households. That is, households did not fail due to lack of resources nor did they succeed with unbounded growth. The same process was followed to calibrate the açaí intensification activities.

In the variable price scenarios, açaí prices are recorded through the açaí price index (API), but relative only to açaí prices in a base year and not with respect to other variables. However, Brondízio (2008) rebased IPA-PARA, an index of agroforestry and husbandry products, to the same units as the API. However, capital costs in the model have not been appropriately rebased to API units. Therefore, instead of performing experiments to determine new capital costs, a price multiplier or scaling factor is used to rebase the price indices with respect to the other variables in the simulation. The price multiplier is swept until each household's output capital at year 40 is, on average, approximately equal to initial capital endowment at the start of the simulation. Outcomes from off-site employment are analyzed by performing sensitivity analyses in the same manner.

Maria is verified by analyzing output land use trajectories as well as capital and labour availability at each step. Sensitivity analyses are performed to determine issues such as the breaking point of decision methods.

Results are run for 21 households at a time. While this is a small amount of agents, it represents the entire population of Paricatuba. (While most ABM/LUCC are designed with hundreds or even thousands of agents (Berger and Parker, 2001), there are examples of models which have been run with fewer agents: The model by Jager et al. (2000) used 16 agents while one version of Fearlus (Gotts et al., 2003) was limited to 49 agents on a 7x7 toroidal grid.) Using the Monte Carlo method, thousands of runs are performed, obtaining a much larger sample of agents.

To compare decision-making models, each scenario is run with two alternative decision-making models in parallel. Linear programming is used as an example of rational, optimizing decision-making, while decision trees (with probabilistic selection among feasible alternatives) is used as a fast and frugal heuristic. By controlling the increase of complexity through each successive scenario, biases in each decision-making model can be identified before the model becomes too complex for such conclusions to be made.

Chapter 5

Results

5.1 Introduction

This chapter is a discussion of the model results, organized into three scenarios of increasing complexity. The three scenarios are selected to provide for "controlled complexity", in which sufficient complexity is included to analyze aspects of the decision-making method and outcomes. The chapter begins with a simple scenario of constant prices, where açaí and gardens can be cultivated and açaí can be extracted from the forest. The constant price is not a characteristic of real-world markets, but serves to illustrate nuances of the model, such as biases introduced by the decision methods. Successive scenarios introduce empirical or theoretical data, making for a more complete system. The next scenario presents variable prices based on the açaí price index (API) and the agroforestry and husbandry index, IPA-PARA. Both of these were calculated by Brondízio (2008). Introducing additional complexity into the model, multi-sited households are introduced in the third scenario by injecting an employer agent into the simulation, resulting in circular or impermanent migration.

For each scenario, two runs are performed, with one run entirely comprised of optimizing households, implemented using linear programming, and the other comprised of heuristic households, based on decision trees and choices based on weighted probability. These alternative decision-making methods will be presented in parallel for each scenario. The chapter concludes with a brief evaluation of spatial land allocation algorithms discussed in Section 4.4.6.

5.2 Scenario: Constant price scenario

The choice of decision-making method can lead to significantly different results, as the following constant price scenario illustrates. Figure 5.1 shows simulation results for two runs which differ only by the selection of decision-making method. Each run is comprised entirely of one type of decision-making household agent. One run uses linear programming (LP), an example of an optimizing decision-making method, while the other uses a decision tree and weighted selection: fast and frugal heuristics.

From this pair of runs, one household from each run is displayed, chosen by relative similarity in labour availability. The prices of açaí and manioc are set at constant values. These and other parameters were calibrated such that the capital resources of optimizing households would stabilize at a steady state. Descriptions of the parameters are listed in Appendix B and parameters for this specific run are detailed in Table B.2. In this scenario, the prices of açaí and other goods do not vary by the açaí price index or IPA-PARA but stay constant throughout the simulated period.

As shown by Figure 5.1, optimizing agents cultivate only gardens without any intensification of açaí, because this strategy provides the most income under the constraints of capital, labour and land. Since there is plenty of land and capital, and prices are unchanging, the optimizing agent simply increases its degree of garden cultivation until it can no longer expand. As expected, the optimizing agent is more successful than the decision tree agent in terms of profitability. The optimizing agent introduces a greater impact to its landscape in terms of area due to its exclusive use of shifting agriculture.

In comparison with the optimizing agent, the heuristic agent exhibits a slightly more complex pattern before reaching a steady state of intensive açaí management. Initially, the agent chooses a mixture of açaí and manioc gardens. However, intensively-managed açaí need not be fallowed or abandoned after time, as long as sufficient labour and capital are available to continue management. As long as a household possesses sufficient resources, there is no incentive to abandon açaízals. In contrast, gardens are required to be fallowed after 2–2.5 years. With land and other resources freed up to cultivate a new cell, the household makes a new selection weighted only by the expected revenue of each alternative land use, as in every case. Since the household does not take into account previous land uses, the land uses of açaí and gardens are chosen with similar probability. Thus, with existing açaízals remaining in perpetuity and a fair chance that the agent will choose to manage

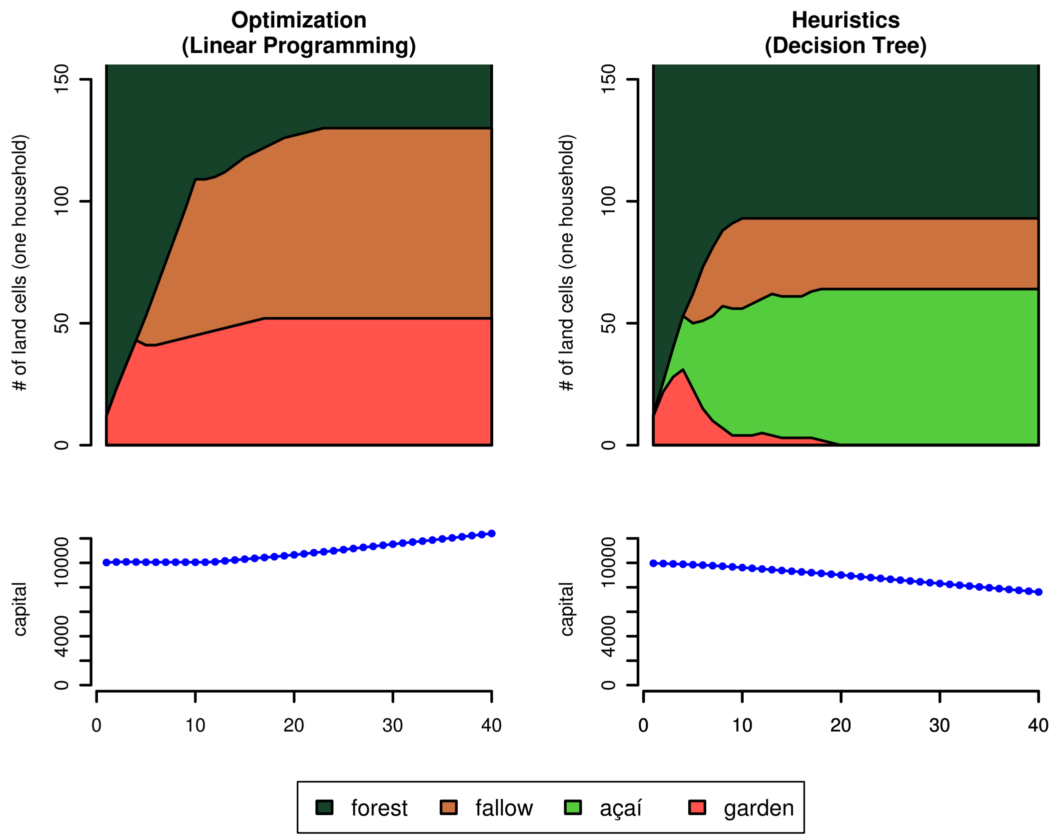


Figure 5.1: Single household trajectories under constant prices

new açaízals, there is a natural progression toward açaí even in a constant price environment. The steady state of açaí management as the single land use is a result of some household resource bound to the management and maintenance of açaí. Since açaí prices are similar to manioc prices, there is no significant incentive to pursue a change toward the latter, despite the year-after-year decline in capital resources. Furthermore, by nature of the order in which decisions are made (Figure 4.9 on page 65), these decision tree agents are loss averse, allocating resources to maintenance before expansion or change.

For the same run as above, the land use trajectories of all households are shown by Figure 5.2. again with gardens in red and manged açaí in green. The significant difference between optimizing agents and heuristic agents is evident in all households in this run. The price parameters of this run were selected to highlight the differences between optimizing and heuristic agents, as will become evident in the sensitivity analysis to follow.

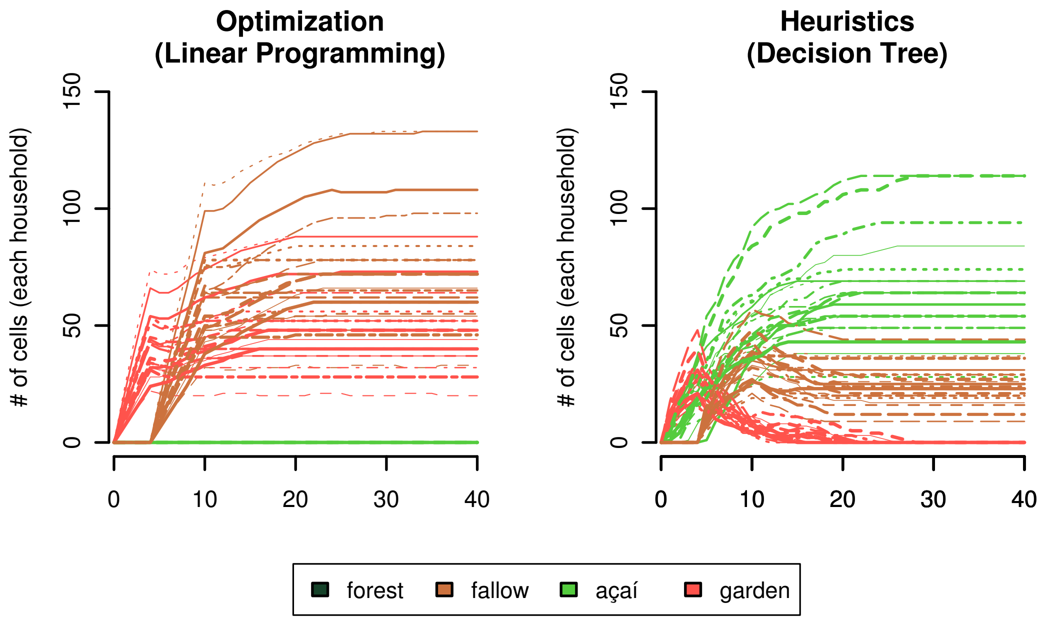


Figure 5.2: Multiple household land use trajectory under constant prices

5.2.1 Sensitivity analysis

Each constant price scenario illustrates one potential steady state outcome. By varying the price of one of the goods in a parameter sweep, a sensitivity analysis can be performed by Monte Carlo simulation. The model is run with açaí prices varying within 25% of a baseline price of 3.35×10^{-06}, with a difference in price of 1% between each simulation run. This produces a sample of 50 runs, each consisting of 21 households. (A Monte Carlo simulation was performed without modifying the baseline price and yielded expected results of a flat trend with some variability in the outliers.) The baseline price appears to be low since it is measured in terms of price units per kilogram, whereas other costs in the model are presented in price units per cell. Households perform the conversion to units per cell by multiplying the price by the cell's expected yield.

Household success is measured as the amount of capital at the end of the simulation run, year 40. Households begin with an initial capital endowment of 10000. Graphing household capital on a scatterplot by açaí price, household success can be expressed as a function of the price of goods. The scatterplots are smoothed by locally-weighted polynomial regression, using the LOWESS algorithm developed by Cleveland (1979, 1981). In this manner, two scatterplots are prepared (Figure 5.3) in order to compare optimizing agents with heuristic agents.

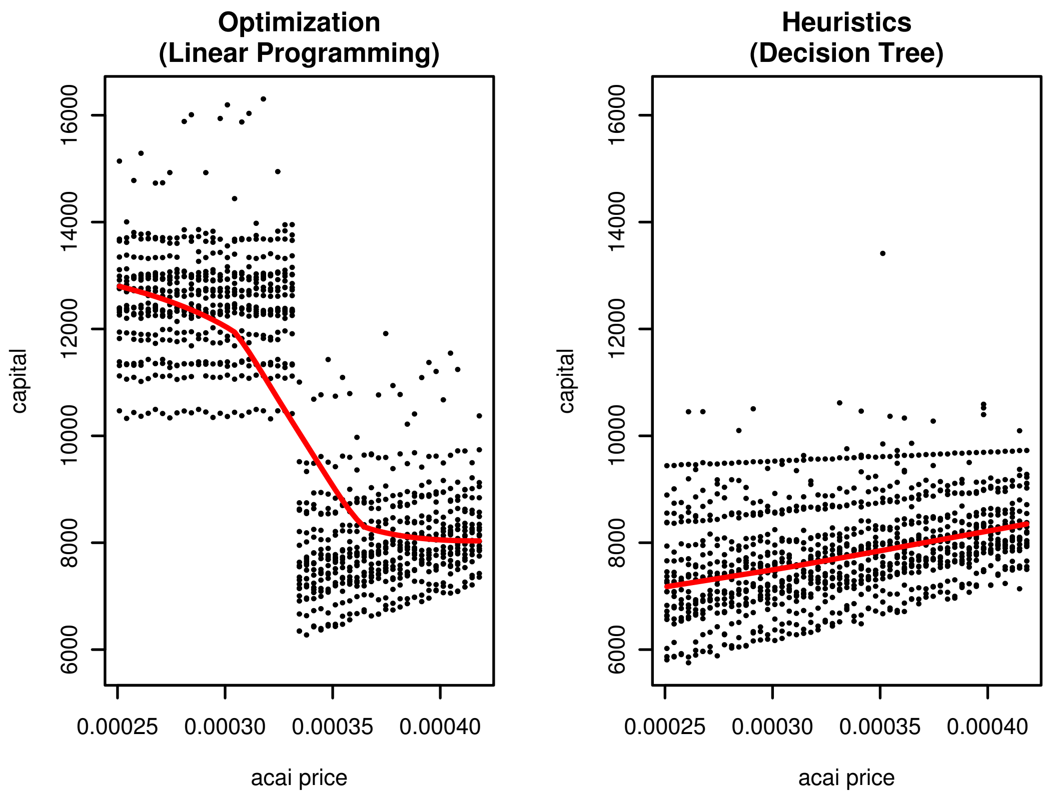


Figure 5.3: Sensitivity of constant price scenario to açaí price

A trivial hypothesis could be stated: Between two otherwise equal scenarios, a marginal increase in the price of a good should lead to a proportional increase in profit, at best, and at worst, there should be no loss. However, this is not necessarily the case for households which use linear programming, as the scatterplot shows. The baseline price of 2.8×10^{-06} was chosen as an approximate break-even point, where manioc plots are about as profitable as açaí plots, according to the understanding of the linear programming household. The linear programming agent's scatterplot shows that this is not the case, as the marginal increase in the selling price of açaí actually results in significantly reduced profit.

A significant limitation of the linear programming household is highlighted here, since it only optimizes the current year's activity without planning future years. Since the labour required to maintain intensive açaí management is higher than that required to maintain gardens, labour resources become more constrained as the farm becomes more invested in the management of açaízals. While the agent has infinite computational ability, it is limited in the factors it considers. A forward-thinking agent would realize that manioc is more profitable below an higher break-even price, taking future maintenance requirements into account. Below this break-even price, there should be no effect on economic behaviour. Since agents do not take future lags into account, by not discounting the future value of goods with respect to the present value, the relative price of açaí over other goods is weighted too highly. Effectively, the agent's break-even price is set too low, switching to açaí management even when it is a poor investment. However, above this break-even price, the optimizing agent only performs as poorly as the heuristic agent.

Interpreting the trends above the break-even price another way, the decision tree appears to produce a reasonable approximation of an optimizing agent in terms of capital success (Figure 5.4)). Like the optimizing agent, the heuristic agent does not consider future lags when making decisions.

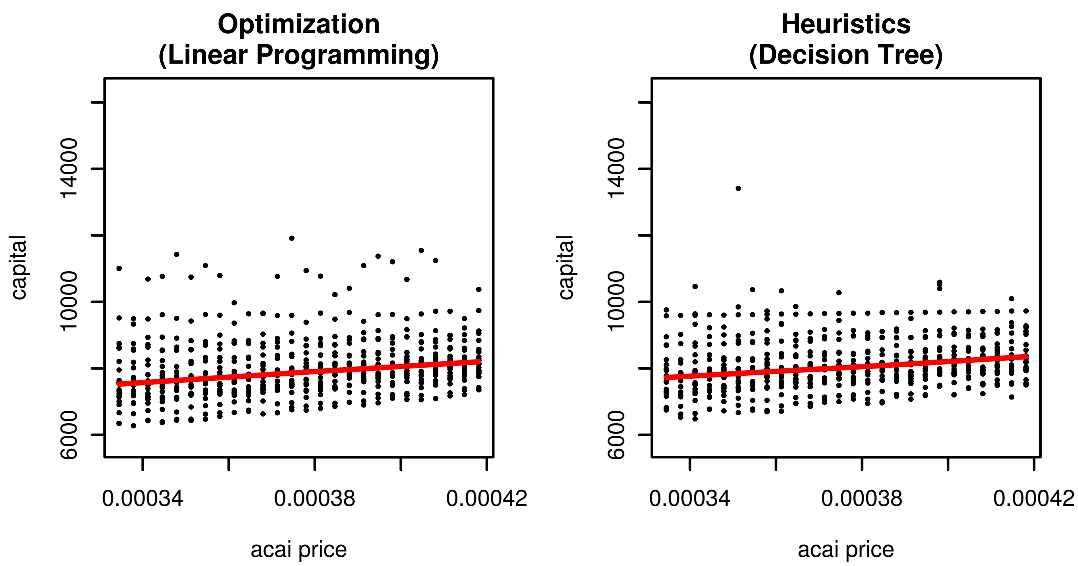


Figure 5.4: Sensitivity of agents to açaí price, above break-even point

The stochastic nature of the decision tree agent, using weighted selection among alternatives, results in a wider range of capital success and failure. While there are a few agents who are able to obtain slightly more than their initial endowment, there are also a number of agents who lose up to 50 % of this endowment. In comparison, optimizing agents are more homogeneous in their capital outcomes, exhibiting approximately the same distribution of outcomes as their heuristic counterparts, but without outliers. Optimizing agents are not stochastic in their actions.

Above the break-even price, where heuristic agents appear to approximate optimizing agents in terms of capital success, the land use composition produced by each type of decision-method highlights differences. One run, corresponding to an açaí price of 3.8×10^{-4}, is illustrated by Figure 5.5. Households in this plot are distinguished by varying line styles and widths, with land uses represented by the same colour palette used in other plots and maps in this thesis. This plot shows that unlike optimizing agents, heuristic households invest in a more diverse portfolio of land uses, but eventually drift toward açaí production. Optimizing agents, on the other hand, invest in only the most profitable land uses. Both types of decision-making method plateau at a steady state, indicating that the labour required to maintain existing plots (or maintain the same number of existing plots, if referring to shifting agriculture) leaves no additional labour for expansion or further land use change.

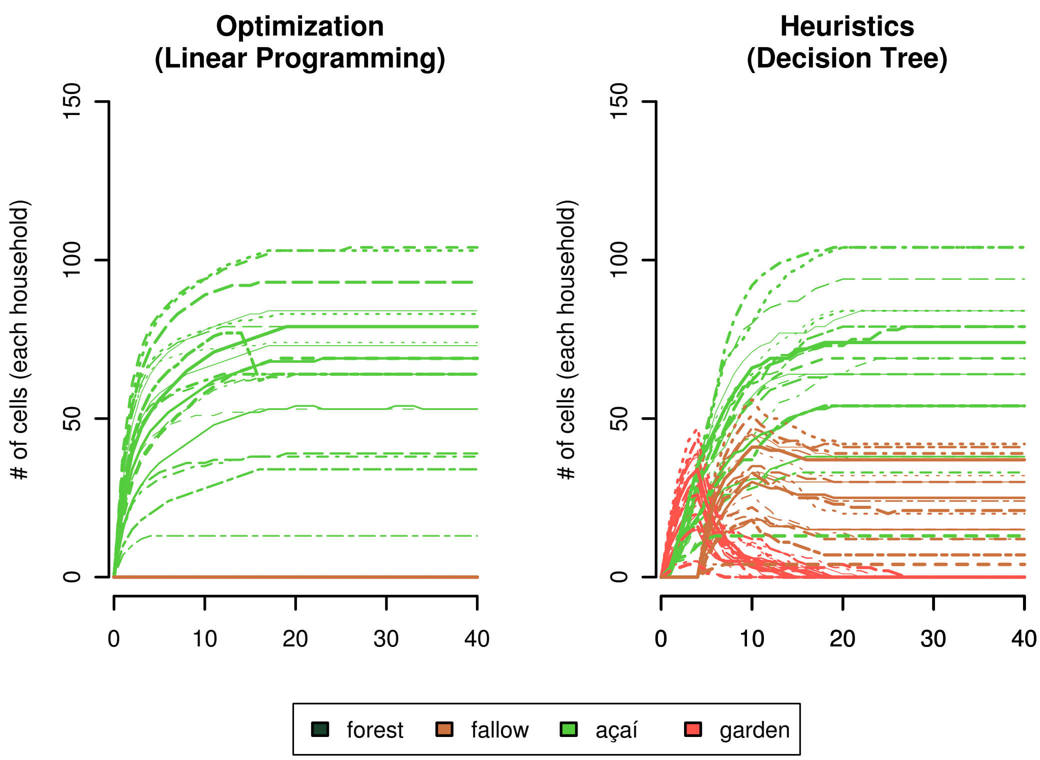


Figure 5.5: Land use composition of households above break-even açaí price

5.3 Scenario: Prices based on API and IPA-PARA

Performing a similar run as the previous section, but varying annual prices of açaí and other goods according to the açaí price index and IPA-PARA yields results shown in Figures 5.6 (single household comparison) and 5.7. In this scenario, like the constant price scenario, labour and capital costs of all activities are assumed to remain constant and all non-price parameters retain their previous values: All other things the same, the selling prices of goods vary over time based on real-world prices.

The land use and capital trajectories of one household are shown in Figure 5.6. This run corresponds to a price scaling factor of 9.0×10^{-6}, which is multiplied by the API and IPA-PARA to more closely match other costs in the model.

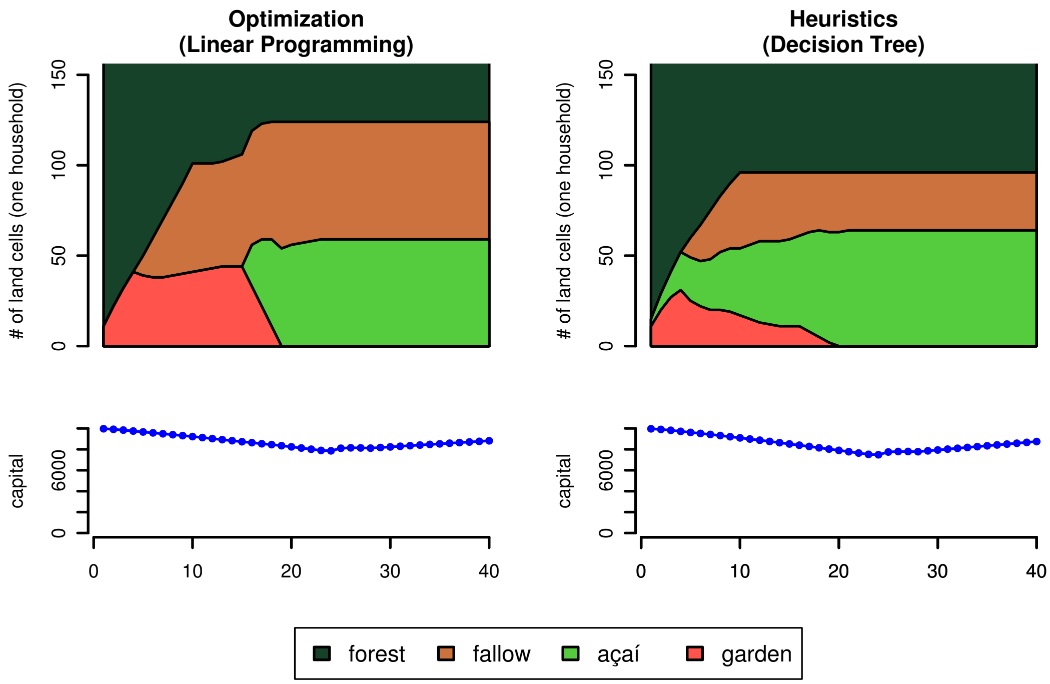


Figure 5.6: Single household trajectories under variable prices

In this case, the optimizing household exhibits extreme behaviour, allocating 100 % of its (managed) landscape to gardens, then shifting it entirely to açaí once açaí becomes more profitable. Additional land cannot be managed due to labour constraints. This is characteristic of all households in this scenario, as shown by Figure 5.7. This transition is briefly reversed just prior to year 20, as açaí becomes less profitable for a moment, but the trend toward açaí intensification resumes shortly afterward.

The heuristic agent shifts from gardens to açaí management earlier than the optimizing agent. Keeping in mind the natural drift toward açaí exhibited in the constant price scenario (Figure 5.2), the drift may be present here as well. The rapid transition to açaí earlier in time may indicate their affinity toward maintaining their açaízals, rather than their desire to cultivate the most profitable goods.

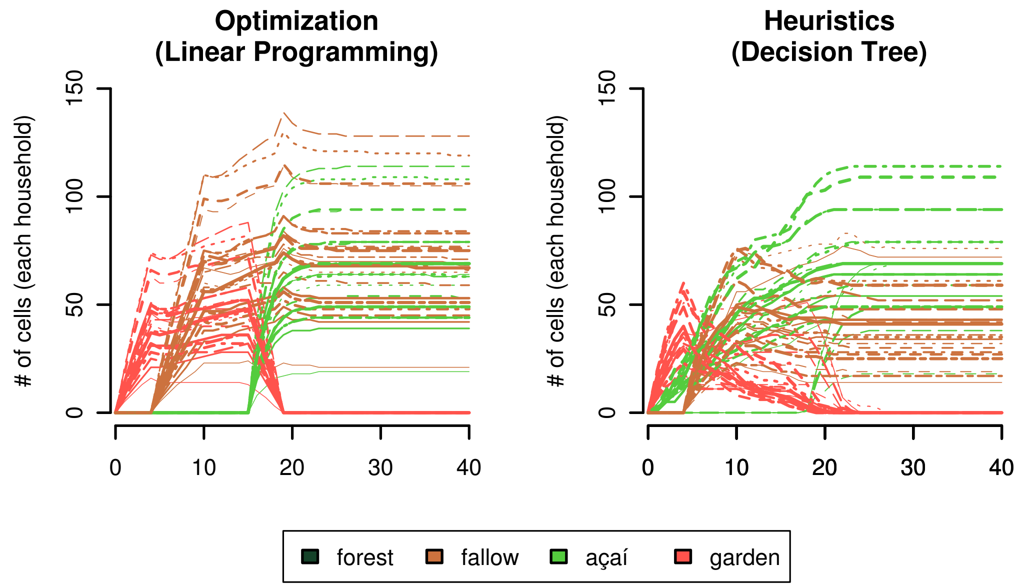


Figure 5.7: Multiple household land use trajectory under variable prices

The trend in the capital trajectory is a result of the use of constant costs and variable revenues from selling prices. For approximately the first half of the simulation, expenses outweigh income, resulting in year-over-year losses. As the selling prices of goods rise, goods become profitable. This issue can be addressed with additional data related to the costs required for agroforestry and agricultural duties as well as transportation.

5.3.1 Sensitivity analysis

The sensitivity of household capital success to the scaling factor of the price of açaí and other goods is plotted in Figure 5.8. The relative successes among households appears to converge at the price scaling factor of approximately 1.2×10^{-5}. Considering similar households one at a time, capital success varies linearly by the price scaling factor. This indicates that some types of households are much more sensitive to price than others.

Comparing two runs on either side of the convergence, it becomes apparent that the price scaling factor applied to both the prices of açaí and the prices of other goods does not affect household land use trajectories. However, the revenue gained from goods is insufficient throughout the simulation for households to gain profit.

From Figure 5.9, it is still difficult to relate capital success with household characteristics. In particular, the question lingers: Why are some households more

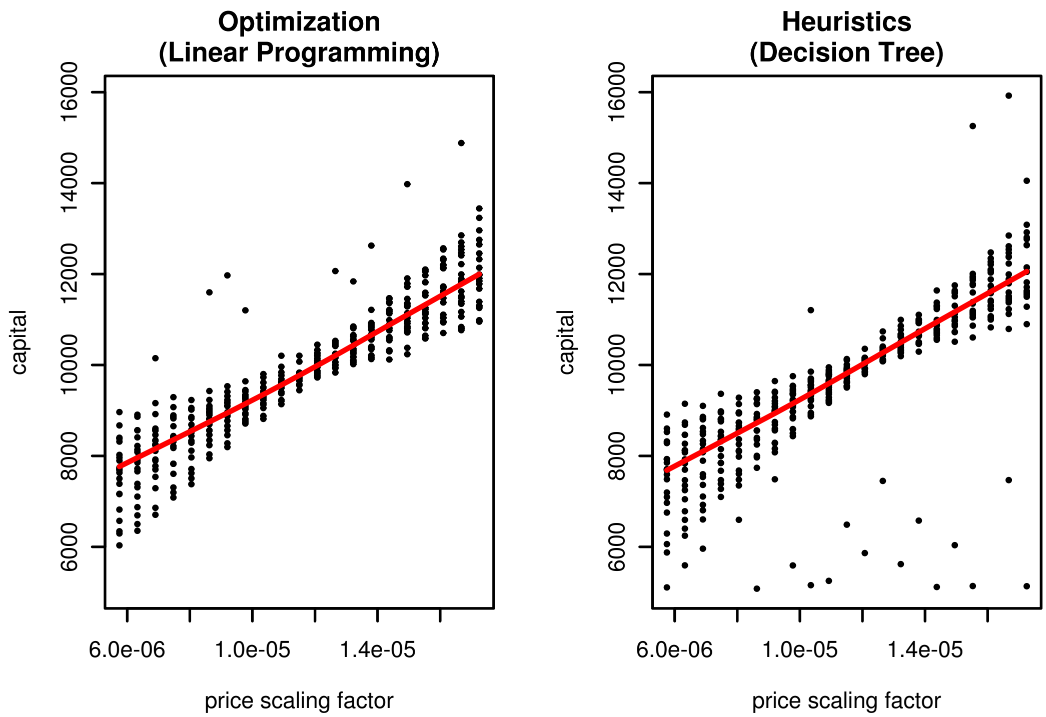


Figure 5.8: Sensitivity of variable price scenario to the scaling factor of prices

sensitive to price changes? In this scenario, households differ by two factors: available labour and land. All households are granted the same capital endowment at the start of the simulation and no households receive employment off-site. Since the land use trajectories indicate that households do not manage all of their land, they are constrained by available labour. Plotting capital success by labour for linear programming agents (Figure 5.10) shows a marked difference in household success. Below the break-even price, high labour resources are correlated with poor capital success. Above the point of convergence, labour correlates with capital success as expected. This is related not only to the lack of forethought in the decision-making process, but also to the nature of the objective function used. The linear programming agents maximize revenue, not profit. Households with larger labour pools are more sensitive to prices due to the size of their investment. Attempting to maximize revenue, a household will pursue the most desirable action within constraints. Since labour is the constraining factor of most households, larger labour pools are able to manage more land. Large areas of managed land translate to larger losses when revenues are low and larger gains when prices are high.

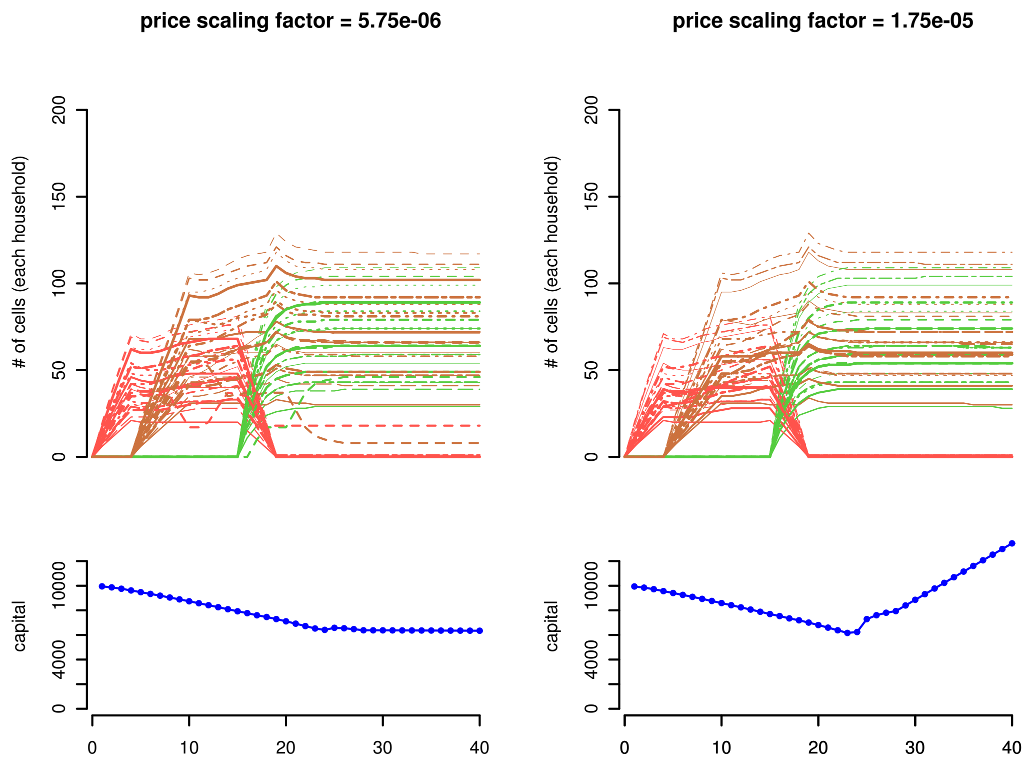


Figure 5.9: Land use trajectories of two LP runs

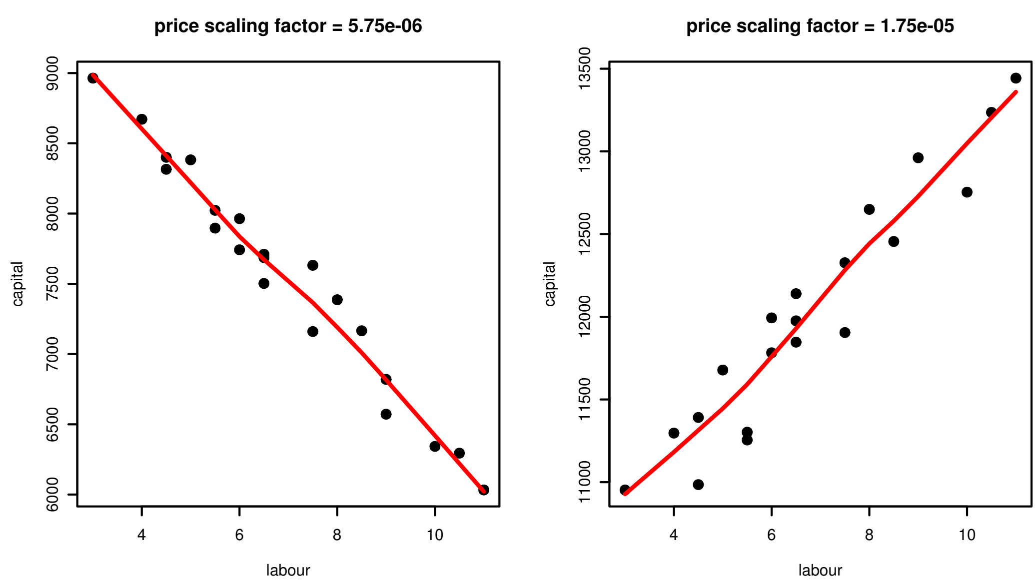


Figure 5.10: capital-labour plots for two LP runs

5.4 Scenario: Multi-sited households

Introducing off-site employment results in the creation of multi-sited households. With less labour available for farming duties, deforestation and intensive management trajectories are tempered. Two variables can be analyzed here: the availability of employment and the value of employment (wage). In a similar manner as the price variables, the sensitivity of the model to employment availability and wage are analyzed.

One employer agent is added to the variable price scenario to represent all sources of employment. The employer makes a random number of offers each year. The interarrival time between offers is based on an exponential distribution, with probability density function $p(x) = \lambda e^{-\lambda x}$. Employment availability is controlled using the exponential distribution's rate parameter, λ. λ is the constant average arrival rate: For example, $\lambda = 1.5$ indicates that an employment offer is made every 1.5 years. At the start of each run, employment offers are scheduled or "binned" into one of the 40 simulation ticks.

The land use, capital and labour trajectories for a single household are shown in Figure 5.11. In this scenario, $\lambda = 10$, so an average of 10 offers are being made per month. The value of each of these offers is also 10 (though unrelated to the value of λ). While the linear programming agent can make the revenue-optimizing choice, the heuristic agent makes an estimate. As discussed in Section 4.4.4 on page 64, the decision tree agent estimates its income per year per unit of agriculture contributing labour and accepts employment if the offered wage is higher than the opportunity cost of lost labour. This heuristic assumes that income is linearly proportional to labour, which may be true under ideal circumstances if the household is constrained by labour (see Figure 5.10, right panel).

Figure 5.11 introduces a plot of labour availability, which is a calculation of the amount of available contributing labour for agriculture. It is calculated by including the amount of labour provided by members of the household working locally on the farm. It is increased by immigration and decreased by emigration, thereby illustrating the impact of off-site employment on labour availability.

Due, in part, to the economic opportunities presented to the household, there are marked differences between the linear programming agent and the decision tree agent. Again, the linear programming agent exhibits its tendency to act in extremes, in both its land use and migration decisions (Figures 5.12 and 5.13, respectively). The migration trajectory of this type of agent occurs in three phases, with emigration

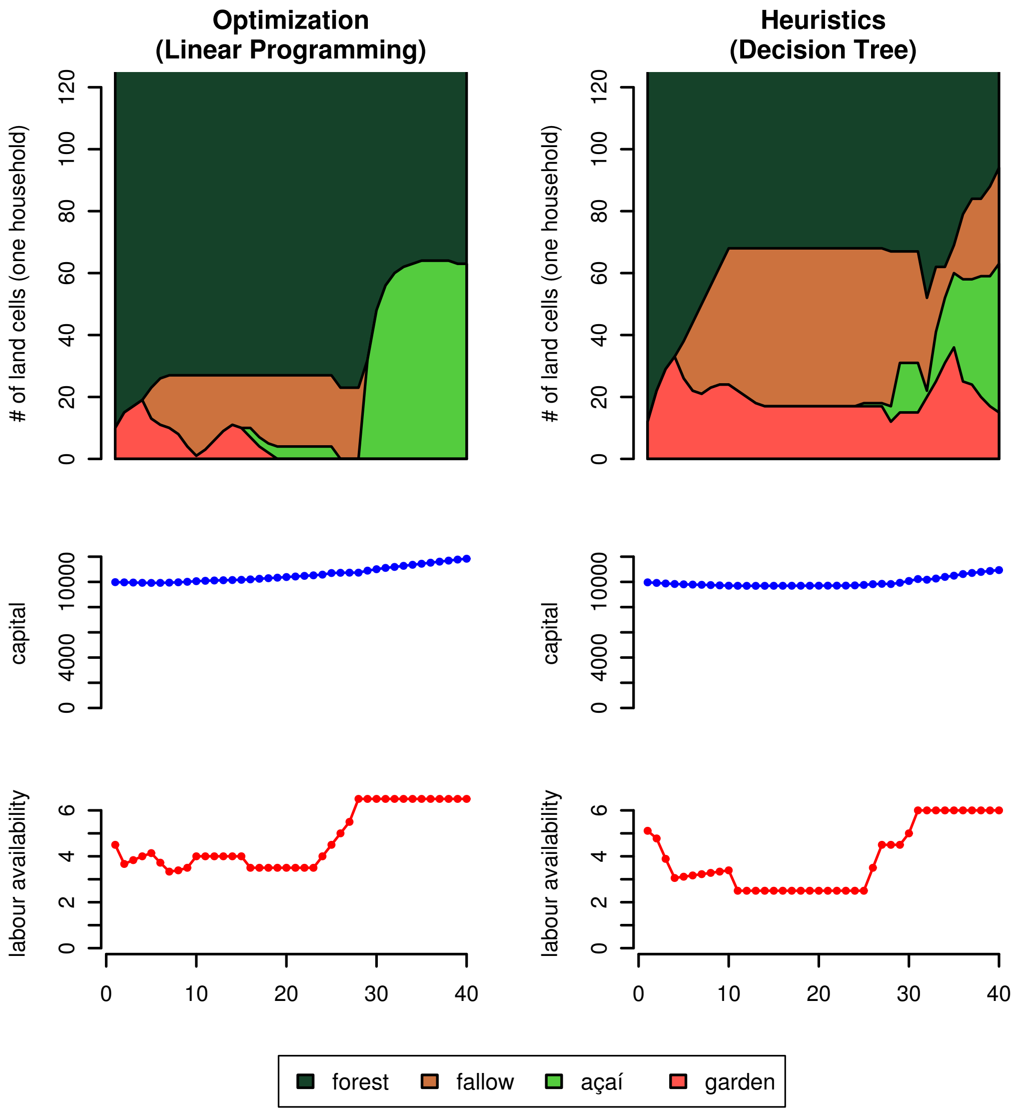


Figure 5.11: Single household trajectories, multi-sited

followed by a period of non-agriculture, then immigration back to the farm. In this particular scenario, off-site employment is most profitable from the start of the simulation, but there is a lag due to limited employment availability. During this lag, the household uses its members to invest in gardens, before most household members emigrate. Eventually, rising prices allow agriculture to become more profitable than off-site employment, so LP agents recall their members back to the farm.

In contrast, decision tree agents have a much more complex pattern, with some households practicing circular migration. A household agent estimates the opportunity cost of labour by dividing last year's income by last year's available labour.

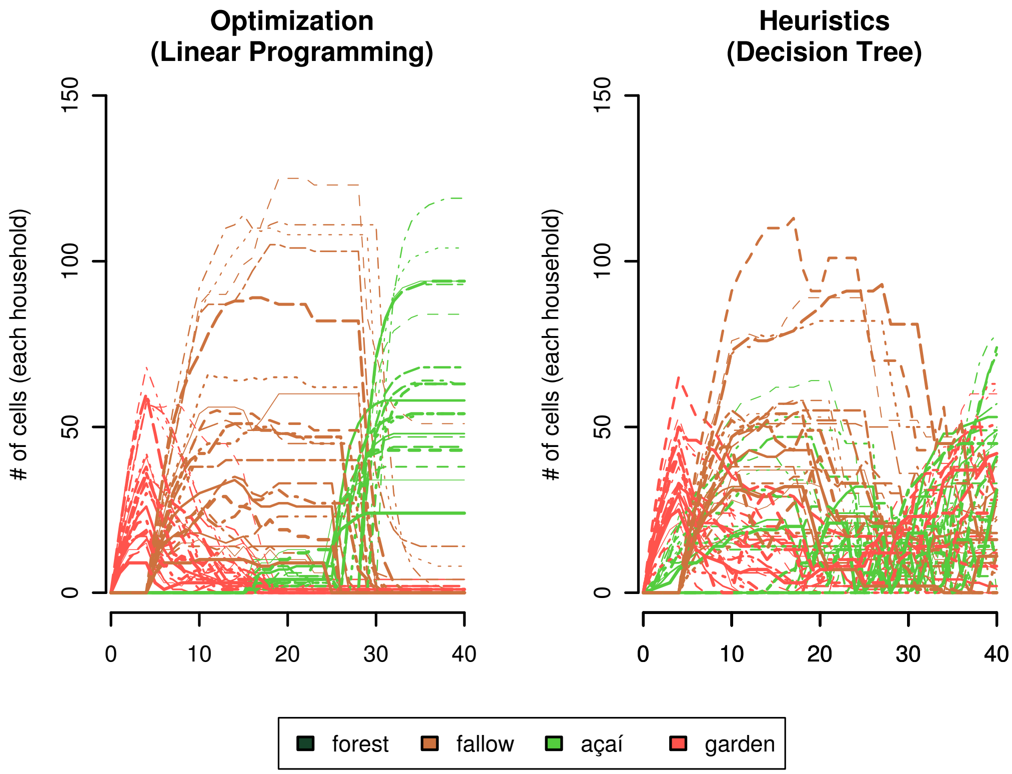


Figure 5.12: Multiple household land use trajectories, multi-sited

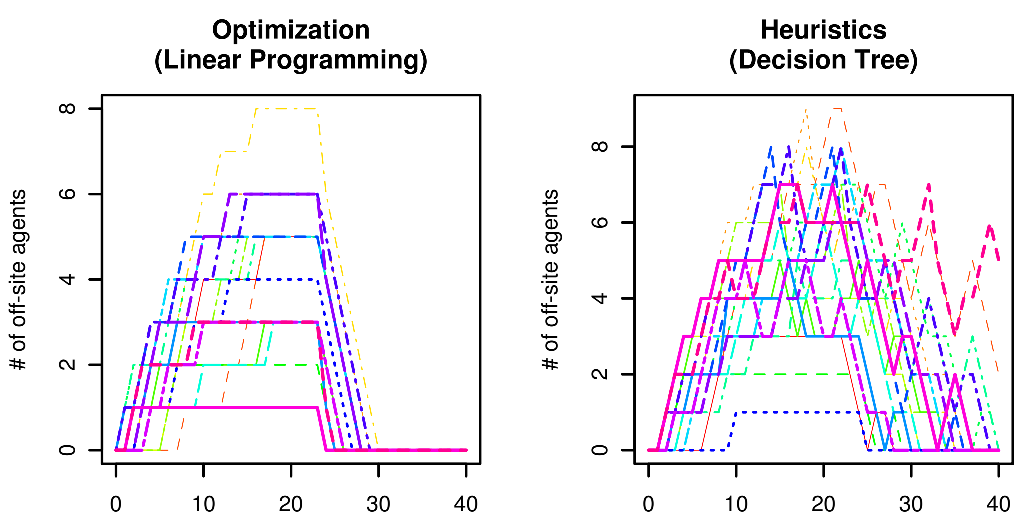


Figure 5.13: Multiple household migration trajectories

Circular migration occurs as the estimated opportunity cost of labour becomes roughly equal to the agents' wage. Since the estimated payoff of agricultural labour varies closely with the offered wage, individuals move back and forth, attempting to maximize their revenue. Since they calculate their opportunity cost of labour based on the number of agents at the farm and the net income in the past year only, their estimate varies as agents emigrate and immigrate. Circular migration does not occur with linear programming agents, who estimate their cost of labour differently. Their estimate does not vary around the offered wage.

In this scenario, decision tree households practice diverse resource strategies, cultivating gardens, managing açaí and pursuing off-site employment. Açaí becomes more common relative to other goods towards the end of the simulation, approximately corresponding to the years 1995–2010. In other scenarios, heuristic agents were much more loss averse, preferring to maintain existing cells rather than performing other actions. However, these agents consider employment before planning land use actions. The loss in labour results in an increased propensity to abandon cells as the smaller labour pool becomes insufficient to perform necessary maintenance. As agents return, the household is free to select new land uses based on current prices, leading to a much more dynamic land use trajectory.

Comparing the multi-sited case with the single-sited case, Figures 5.12 and 5.7 respectively, employment results in decreased land use for much of the simulation run. For much of the period when gardens were otherwise greatly cultivated in the case of single-sited households (Figure 5.7), employed agents were abandoning much of their land. Eventually, these two scenarios begin to converge as agents are recalled to the farm due to rising prices of goods, with the exception of heuristic households which maintain diversity.

5.4.1 Sensitivity Analysis

Since there are two variables which jointly determine the impact of employment on economic welfare, 4 plots will be prepared: One variable is kept fixed at two distinct values, while the sensitivity of the model to the other variable is plotted. This process is repeated, swapping the fixed variable for the other variable.

First, the sensitivity of the model to the wages will be analyzed. Two pairs of data will be produced, performing two scenarios for each decision-making method. Keeping the offer arrival rate fixed at 10 and 20 offers per year, and plotting capital against wage, Figures 5.14 and 5.15 are produced. In all cases, there appears to be a linear relationship between capital and wage, as expected. However, for linear programming agents, the relationship is piecewise. This trend is similar to the one illustrated in Figure 5.3. Again, this is due to the revenue-maximizing nature of the objective function: Since low offers offer little value in comparison with eventual revenues from agriculture and *costs are not taken into account*, low offers are not accepted.

The heuristic agent uses an alternative approach: It compares an eligible agent's wage with the estimated loss of net income. When net income is negative, the "loss" of net income is negative. There is perceived to be a gain from the removal of the individual from the labour pool. In this case, the household agent will even accept zero-valued employment offers, thus mitigating their loss. In these cases, decision tree agents outperform revenue-maximizing linear programming agents.

One household in Figure 5.14 performs consistently, regardless of the offer value. This particular household does not have sufficient labour to accept employment. Households across each pair of parameter sweep plots have the same initial demographic composition, so this household appears in every case in Figure 5.14.

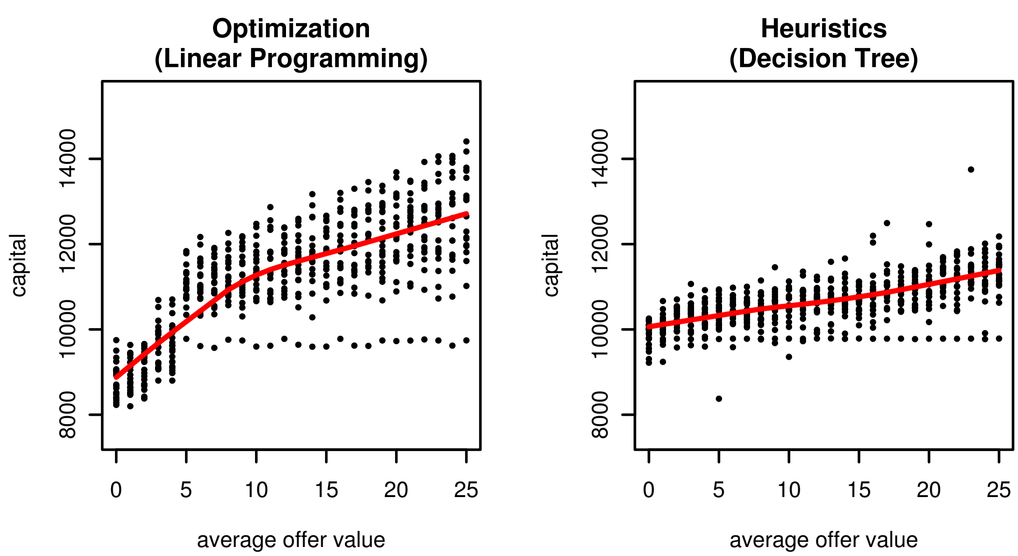


Figure 5.14: Sensitivity to the value of offers, averaging 10 offers per year

Plotting the economic welfare of agents by the ease of access to employment, it is apparent that capital correlates with employment availability. Figure 5.16 shows cases where the offer value is less than the LP agent's acceptable value, as shown by the left portions of Figures 5.14 and 5.15. The plot shows that heuristic agents outperform revenue-optimizing agents when there is available employment.

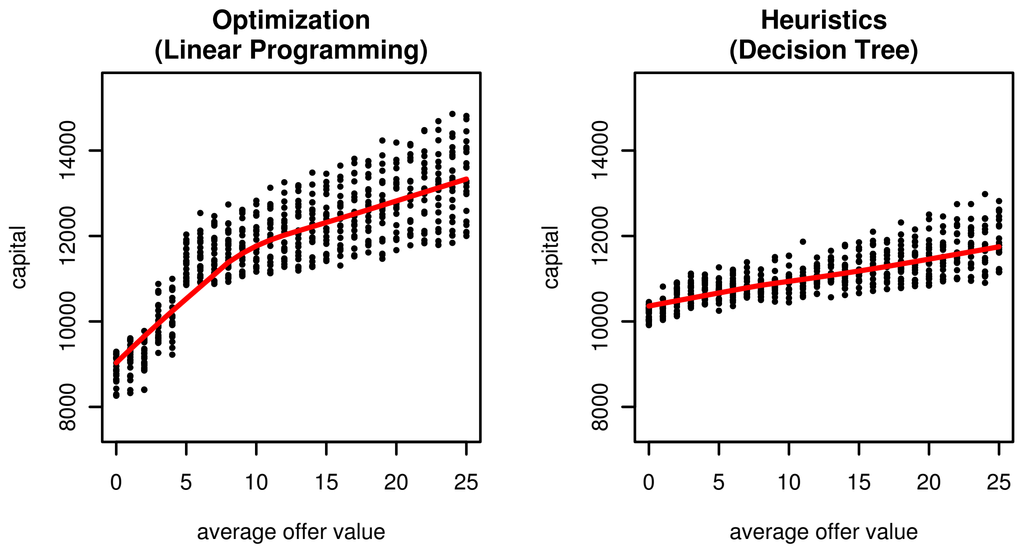


Figure 5.15: Sensitivity to the value of offers, averaging 20 offers per year

In Figure 5.17, the offer value is acceptable to LP agents, at least for the first part of the simulation before rising prices of goods make them more profitable. In these cases, revenue-optimizing agents outperform heuristic agents.

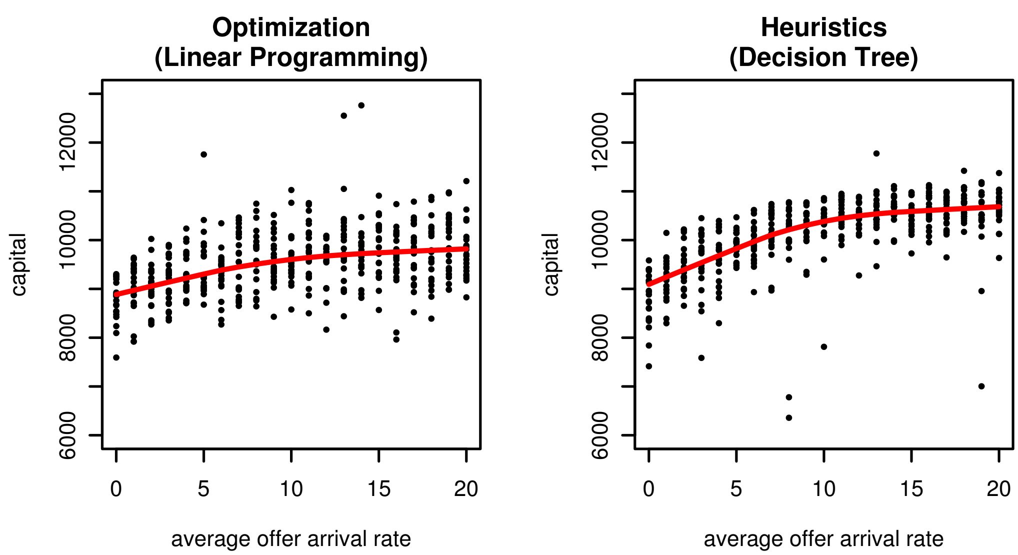


Figure 5.16: Sensitivity to the offer arrival rate, value of 4

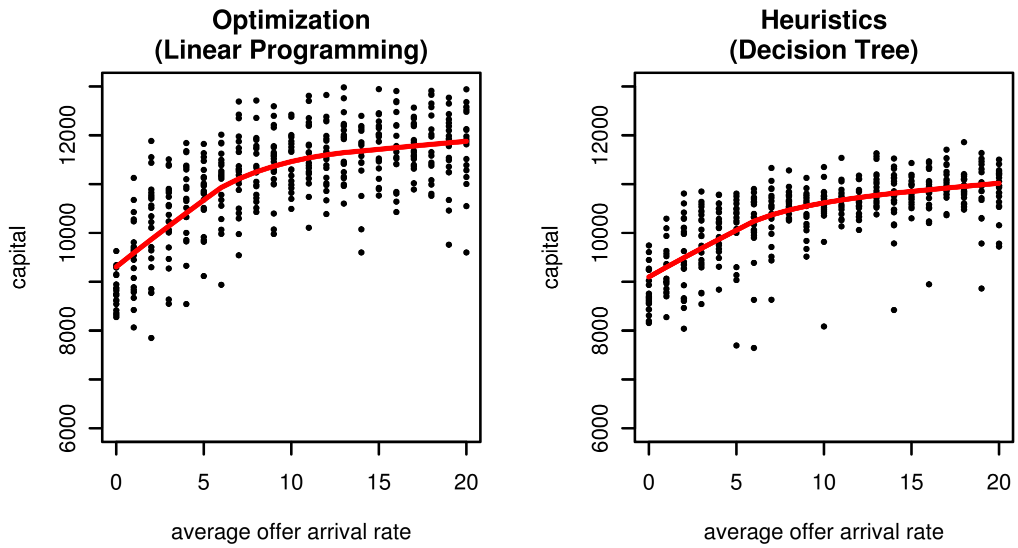


Figure 5.17: Sensitivity to the offer arrival rate, value of 10

Economic disparity

Economic opportunities available to some, but not to others, result in economic disparity. Box and whisker plots, in Figures 5.18 and 5.19, have been prepared for the data shown in Figures 5.14 and 5.17. Based on the differences between the richest and poorest households, these plots illustrate the economic disparity resulting from wage and employment availability.

Figure 5.18 plots the sensitivity to wage, with a fixed average offer rate. As the wage increases, households which are able to supply off-site labour are rewarded well. However, the employment availability's effect on economic disparity (Figure 5.19) is not as strong as expected. The spread of the interquartile range is only slightly larger between offer arrival rates of approximately 1–5/year. Although outliers do illustrate a very small number of households being left behind, this is not necessarily a result of a lack of employment: Similar numbers of households were also significantly poorer than others in the variable price scenario without off-site employment (Figure 5.8). These households face a significant shortage in either land or labour. Conversely, Figure 5.18 illustrates a small number of outliers which perform significantly better than the majority of agents. These households have a high supply of labour and are able to reap the rewards of off-site employment, gaining a larger advantage when wages are high.

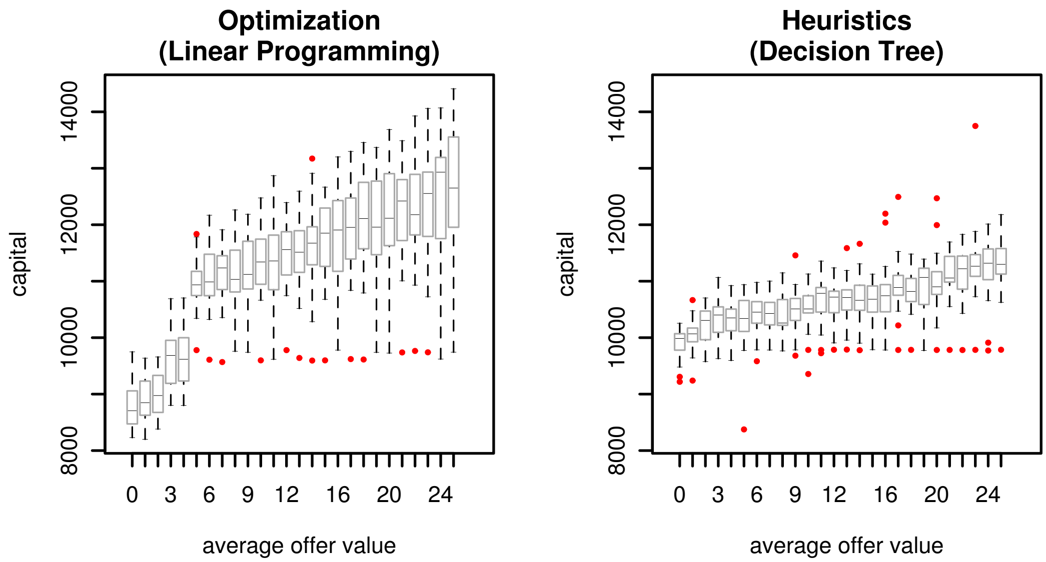


Figure 5.18: Box-whisker plot of sensitivity to wage, 10 offers/year

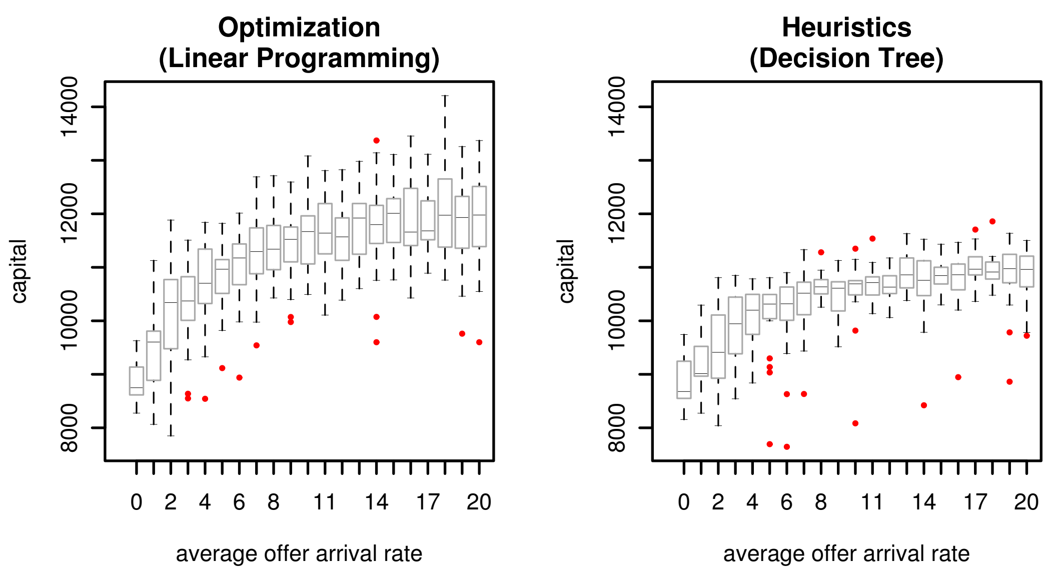


Figure 5.19: Box-whisker plot of sensitivity to employment availability, wage of 10/year

5.5 Spatial land allocation

The decision-making of all agents in MARIA are handled in two stages: non-spatial and spatial. Lacking spatial data, forest is assumed to be initially homogeneous and agents first make their decisions based on largely non-spatial factors. This issue may be addressed in a more integrated spatial decision-making process once better spatial data becomes available.

In the meantime, spatial land allocation algorithms, developed in Section 4.4.6, were derived from very general qualitative descriptions of land allocation practices. These algorithms are run after the first stage of decision making, once the agent has taken an inventory of land uses and has specified a desired inventory. Spatial land allocation algorithms assign cell transitions—between the existing and desired inventories—to the most appropriate locations.

Figure 5.20 illustrates snapshots of a run which attempts to cluster similar land uses together by maximizing the number of similar nearest neighbours. The displayed cells constitute a riverine property, surrounded by water to the northeast and unowned cells along other borders. These snapshots are taken from the decision tree case in Figure 5.1, an example of a transition from gardens to açaí. Among multiple cells with the same number of nearest neighbours, the tie is broken by minimizing the distance to the house.

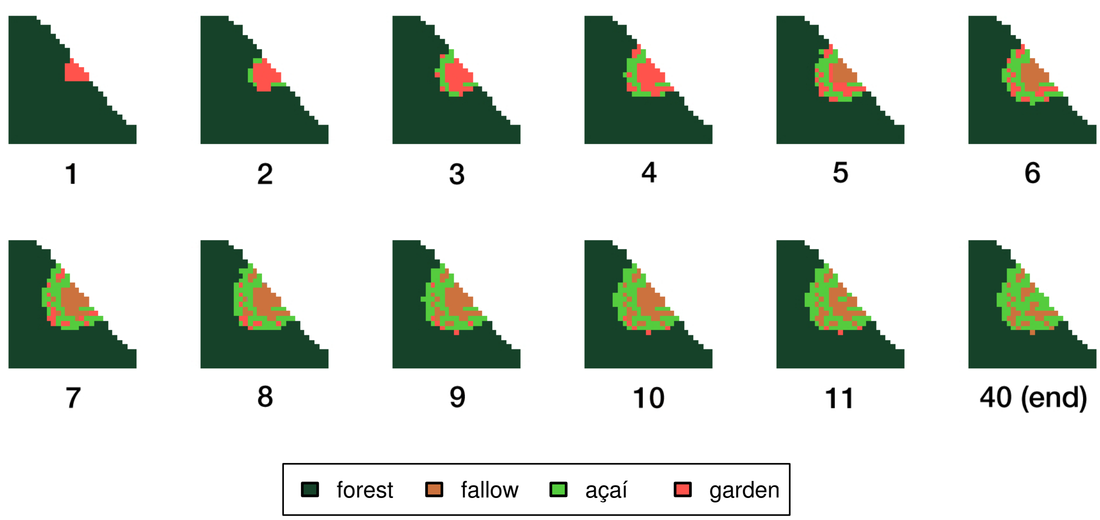


Figure 5.20: Series of spatial land use changes, numbered by step

This arrangement results in shifting cultivation in concentric circles, followed by patches of açaí in a steady state. Fallowed land can only be converted to intensively-managed açaí after a number of years since last cultivation, since the agent must wait until the density of açaízals becomes sufficiently high. The circular pattern of land use was not modelled explicitly, but is a result of minimizing the Euclidean distance to the house.

Such preliminary heuristics were adopted for illustrative purposes only, based on the non-spatial land use trajectories discussed previously in this chapter. These spatial algorithms ignore areas unsuitable for cultivation, such as circular areas known as *bolotas*, as well as other factors which may influence the choice of location. More detailed spatial input data is required to pursue a better spatial model of land use change.

5.6 Chapter summary

All results were presented in pairs, comparing linear programming, representing optimization, with decision trees, representing fast and frugal heuristics. The results were presented in three stages, introducing complexity at each stage. For each stage, a sensitivity analysis was run to identify driving variables of land use change and economic success.

First, a scenario was run such that prices and labour availability were constant, since agents did not emigrate. This stage served to illustrate the differences between optimizing agents and heuristic agents. The heuristic agent, using revenue-weighted probabilistic choice, exhibited a natural drift toward açaí production from gardens, as açaí management is more permanent than the shifting cultivation used for gardens. The sensitivity analysis of this scenario, performed by modifying the selling price of açaí goods relative to other goods and costs, showed that non-forward-thinking revenue-optimizing agents were too short-sighted to realize that gardens would be more profitable than açaí management in cases where the revenue per cell of açaí is slightly higher than that of manioc. With this error, optimizing agents only performed as poorly as heuristic agents, which had performed consistently across the parameter sweep. Profits should be linearly proportional to selling prices (though the trend may be a piecewise continuous linear function): This was true for the heuristic agents throughout the sweep.

Introducing variable prices based on açaí and agroforestry products results in similar land use trajectories as the constant price scenario, but capital resources of the agents are slightly more variable. As profits rise and costs stay constant, agents transition from year-over-year losses to gains. While these gains and losses are linearly proportional to selling prices, sensitivity analyses indicate that larger revenue-maximizing households are more affected by the changes. The sensitivity analyses were performed by multiplying the selling prices of *all* goods by some scaling factor.

Revenue-maximizing households did not consider the costs of their actions, so large households would invest their plentiful labour resources into large managed areas, which led to loss as maintenance costs were paid.

Finally, an employer agent was introduced into the model to entice agents to emigrate. A single employer agent was used to model all employment made available to the community of Paricatuba. This agent generated offers at random based on an exponential distribution with a constant *average* arrival rate. Each of these offers came with a monetary value, an annual wage, provided that the accepting agent would cease to perform agricultural labour and move off-site. When coupled with the variable price scenario, agents are enticed to pursue employment while prices are low, resulting in a smaller footprint as compared with the original non-employment scenario. As prices rise, agents are recalled to the farm.

Spatial land use allocation algorithms were briefly discussed, mainly as illustrative tools. Applying a heuristic to minimize the distance to the agents' house in a weighted linear combination results in a land use pattern in concentric circles. The heuristic which maximizes similar land uses produces some degree of patching. However, these heuristics were based largely on qualitative descriptors and a more detailed spatial model including undesirable areas is required.

Chapter 6

Discussion

6.1 Introduction

Reviewing the objectives introduced in Chapter 1, the modular architecture and implementation of an initial version of MARIA are defined in Chapter 4. Taking advantage of polymorphism, the modularity of this approach has been proven with the implementation of alternative decision-making approaches. This approach can be extended to implement alternative farming arrangements, such as a sharecropping arrangement with separate landowner and sharecropper agents. Chapter 5 covered the remaining objectives. The suitability of alternative decision-making approaches is discussed throughout the chapter. Spatial algorithms are briefly discussed at the end of Chapter 5. The effects of one driver of land use change—market prices—are illustrated in the results chapter as well. Labour requirements are presented as a constraint to land use change, most notably when labour is reduced as a result of off-site employment.

This chapter presents a discussion of the model's design, results and future work. The chapter begins by addressing the debate between rationality and bounded rationality. The limitations of the model are then discussed, in terms of its design, the input data provided and the outputs produced. The chapter concludes with potential land use research applications utilizing MARIA or similar approaches.

6.2 Bounded rationality

The previous chapter presented optimizing and boundedly rational agents in parallel, highlighting the similarities and differences between them. The "rational" agents possessed unlimited cognitive capacity, as implemented through linear programming, but were revenue-optimizing rather than profit-optimizing, a shortcoming which led them to perform worse than heuristic agents in some cases. Furthermore, rational and boundedly rational, heuristic agents were not forward thinking: Rational agents optimized their revenue based on their current state and immediate potential actions. Similarly, heuristic agents were also limited to their current state, their memory of last year's production and immediate actions. However, heuristic agents did not attempt to find an optimal solution, but approximated economic payoffs and weighted alternative feasible options by potential revenue.

In terms of financial success, the optimizing agent should have procured more capital than the heuristic agent. However, in terms of profit, this did not occur in many cases, especially when year-over-year losses were incurred. Since the optimizing agents were *revenue*-maximizing, they would allocate capital and labour resources without considering costs. Similarly, the heuristic agents were also revenue-maximizing, but performed better than the optimizing agents. This is due to the stochastic choice used by the heuristic agents, which would choose a mixture of optimal and nearly-optimal solutions with high probability, avoiding extreme behaviour. Overall, while the heuristic agents' profits were not as high as optimizing agents' profits in the best scenarios, the heuristic agents' losses were also tempered in poorer scenarios.

In Maria, and in general, which type of agent results in a more realistic model? Without more detailed data, it is difficult to gauge which model is more realistic. Overall, the optimizing agents were too homogeneous in their actions, managing an entire landscape of one sole (managed) land use classification. In contrast, the real-world community of Paricatuba managed and cultivated diverse resources instead of producing a monocultured environment. Perhaps, even if by chance, optimizing methods are better suited for modelling the community of Marajó-Açu, but sharecroppers characterizing this area were not modelled explicitly: All households in Maria are single-unit decision makers, unlike in a sharecropping arrangement where there are multiple levels of decision-making, the landowner and the sharecropper.

The heuristic agents in this initial version of MARIA created a more diverse landscape, overall, but slowly drifted toward açaí production despite constant exogenous factors. The permanence of açaí management led it to become a "terminal" land use, where açaí cells were not reallocated for reasons other than labour shortages. In general, this is true in the real-world system as well. However, neither decision-making model characterized the diverse land use patterns in Paricatuba, with the exception of heuristic households in the scenario allowing off-site employment (Figures 5.11 and 5.12). Within this scenario, the dynamic nature of labour availability caused by circular migration encouraged agents to abandon their açaízals and reallocate these cells at a later date. All other scenarios converged on a homogeneous landscape of a single, most profitable land use, usually açaí. This is reflective of an overall trend toward açaí and serves to illustrate one driver of change (or rather, constancy)—the permanence of açaízals—but fails to capture the land use diversity of the area.

In summary, the differences between each *tested* decision-making method—linear programming and decision trees with probabilistic choice—are presented in Table 6.1. Many of these points can be generalized to the class of algorithms they represented, whether rationally optimizing or heuristic. Table 6.1 organizes the differences into specific categories. The decision-making bias, for optimizing agents, is typically predictable as the chosen outcome is the one which maximizes the objective function. Heuristic biases may be unexpected: In Chapter 5, decision tree agents tend to drift toward açaí, a bias which was not readily apparent during design.

	Optimizing (Linear programming)	Heuristic (Decision tree)
Cognition	unlimited	simple or bounded
Bias	predictable	may be unexpected
Implementation	easy	ease varies by complexity
Exogenous issues	can identify	can identify
Endogenous issues	cannot identify	can identify
Computation speed	slow	fast
Info. requirements	all available	limited
Info. limitations	linear constraints, and linear objectives	numeric
Note	compatible with economic models	

Table 6.1: Decision-making method comparison

To describe the other categories in Table 6.1, the ease of implementation refers to the code complexity as well as the complexity of design decisions. Linear programming agents require a well-formed optimization problem, which should be readily apparent, given the agents' constraints and objectives. Simple heuristic agents may be easy to implement, but complex agents emulating cognitive processes may present difficulty. Exogenous and endogenous issues refer to inefficiencies external and internal to agent decision-making, respectively. Perfect decision-makers, approximated by optimizing agents, should not present internal cognitive inefficiencies, but heuristic agents may identify limitations to cognitive processes. In terms of informational requirements, optimizing agents require all constraint and objective variables, while frugal heuristic agents ignore many of these variables to approximate a solution with less resources. Obviously, the informational limitation of both models, and any computational model, is that numeric variables are required. Linear programming decision-makers have a further requirement that constraints and objectives can be expressed as linear inequalities and expressions, respectively.

Broadly speaking, which type of agent provides a more useful model? Schreinemachers and Berger (2006) argue that optimizing methods are useful for modelling inefficiencies with factors exogenous to the agent, while heuristic methods are useful for modelling inefficiencies of the decision-maker. With this in mind, both models are useful. Analyzed in parallel, as was done in Chapter 5, each method provided a useful anchor with which to compare the other method. The ideal optimizing model—a forward-thinking profit-maximizing agent—provides an upper bound on an agent's potential success. Even the revenue-optimizing model provided a point of comparison with the heuristic model, a similar upper bound in many cases. With this upper-bound, the level of inefficiency of the heuristic method can be measured. Since traditional economic models assume purely rational decision-makers, the comparison of rational to boundedly rational agents can provide one gauge of how closely a system may align with an economic model.

6.3 Limitations of the model

There are a number of limitations of the model, ranging from the scope of its design to the amount and quality of input data available at the time of writing. These limitations were related to resource constraints as well as a desire to develop a simple model without overreaching complexity, where outcomes would be difficult to attribute to causative factors. The model was designed in order to capture certain aspects of the local population, such as their land use activities and economic choices. However, the scope of the model was limited to characterize trends from certain key events, such as rising açaí prices and migration. This intended limitation was also implied in the structure of the Chapter 5, as successively-introduced scenarios enabled new features of the model and increased complexity. Through this method of "controlled complexity", simpler models were used to explain fundamental aspects of the model, such as differences between the decision-methods employed. The knowledge gained from the simpler models, including the biases introduced by certain methods, could then be applied to more complex models. Specific limitations in the overall model are described in this section.

6.3.1 Model design

Scope

Within the real-world environment of the Amazonian *várzea*, the current version of Maria encompasses only a specific subset of farmers. Agents in Maria are rather homogeneous in comparison with local farmers, as only the characteristics of the majority of households in Paricatuba—smallholder farmers—have been used to develop and parameterize the model. Other types of households include sharecroppers and those living in co-operative communities. Unequal access to markets and economic institutions such as brokers and middlemen have not been modelled. The current implementation assumes that all agents have equal access to the same market and sell their goods at the same price. Agents, however, do not have equal access to the employment market, instead receiving opportunities by chance.

Instead of including all details into a complex initial version of the model, a general architecture has been developed in which sharecropping and cooperative agents can be implemented as alternative `Household` agents, utilizing the social network layer and messaging system to facilitate agent communication. The existing environmental model is sufficient for these agents, except in the cases involving upland terrain. Upland models would require new agricultural activities and land cover transition rules to be integrated into the model, as none have been implemented in the current version. The human model allows for the replacement of smallholder household agents with implementations of alternative arrangements, such as sharecroppers.

At the household level, the scope of the model limited practices to açaí management and garden cultivation. Other land use activities, such as fishing and shrimping, have not been included. Subsistence activities and costs, as a whole, have also been excluded. This is similar to the approach taken in Lucita (Lim, 2000), where subsistence activities were unknown. However, Lucita did include subsistence costs as some value—a constant defined per adult and per child—deducted from each household's capital each year. This reflected the fact that households near Altamira could purchase subsistence goods from nearby markets. The effect of subsistence activities can be implicitly modelled by reducing the amount of contributing labour for (commercial) agriculture, but this method does not model subsistence uses of the landscape.

Environment

The biophysical model encapsulated by the environmental context was designed to be as simple as possible, while retaining sufficient complexity to reflect real-world constraints. In particular, it presents household agents with biophysical constraints expressed by finite cultivation periods and, when applicable, soil recovery times. Furthermore, the environmental model punishes lack of maintenance by reducing the yield, forcing agents to perform maintenance in order to assume a full harvest. Ideally, such an environmental model would be based on a mathematical model such as KPROG2 used by Lucita (Fearnside, 1986; Lim, 2000). However, other models have demonstrated sufficient complexity for their needs based on far simpler environmental models. For example, Fearlus used bitstrings which changed at random over either time or space (Polhill et al., 2001). While Fearlus was not a case-specific model, it did sufficiently describe the differences between the decision models evaluated. Likewise, Maria's current environmental model is sufficient for highlighting the differences between its decision models. However, a case can be made that a more detailed environmental model could illustrate effects of decision-making on yield and other environmental outcomes.

In the absence of more detailed biophysical characteristics, numerical assumptions were made. Effectively, these assumptions were based on timings. These simplifications prevented agents from performing more complex cultivation practices, since they were unable to attribute poor yield to factors outside timing. This model also assumes that household agents perform the same cultivation practices between similar land uses, a simplifying assumption that was assisted by institutional homogeneity within the community of Paricatuba. Labour requirements and capital requirements of preparing and maintaining each land use were assumed to be the same across all households, as all households were assumed to have access to similar technology.

The current model cannot differentiate between the broad category of "maintenance" and the actions which comprise it, such as thinning, weeding and pruning. A more advanced model could allow farming agents to perform these individual actions, but would require appropriate variables to be implemented within the environmental model, such as detailed land cover, soil variables, and ultimately, yield.

Decision-making

Outside of the rational vs. boundedly rational debate, the decision-making structure of each household has been assumed to be in an "all for one" arrangement. That is, the household makes its decisions as a single unit, rather than from a collaboration or competition among its members. A recent work by Siqueira (2009) analyzes the internal decision-making structure of a household, specifically female influence on household decision-making. Siqueira argues that the decision-making model is male-oriented, but that females in the household have a degree of input, varying by land tenure or informal education. A more detailed decision-making model can be integrated into this simulation, but this would shift its focus toward micro-simulation, whereas a simpler heuristic model may be sufficient for exploring macro-drivers of land use change.

Agents' knowledge and technology were not modelled explicitly. The effect of agent knowledge on crop yields was expressed in other models, such as MameLuke, where outputs were based on agents' years of experience, as a proxy for knowledge (Huigen, 2004). Such an approach can be applied to this model, if field studies indicate that experience is a driver of crop output. With regards to technology, all agents are assumed to possess similar technology and technological advancements do not occur over time. A future model may address this limitation through trade of capital for reduced labour requirements. For example, the purchase of a chainsaw would reduce clearing time, but at the cost of purchase, fuel and maintenance, and the purchase of a motorboat may make direct transportation to market feasible or would reduce transportation costs.

6.3.2 Input data

The model's results were limited by a lack of detailed spatial data. Assumptions were made to fill the gaps. Spatially, the underlying landscape was assumed to be initially homogeneous. The source DEM, derived from an SRTM raster at a resolution of

90 m, was deemed far too coarse to capture local topographic variation. Likewise, detailed classified land cover maps were unavailable to parameterize the model's initial state, so the entirety of the model landscape was initially classified as unmanaged floodplain forest. While this was a reasonable assumption, as most of the chosen area was comprised of floodplain forest or intensively-managed agroforestry, there was no differentiation made between açaí palm and other species of vegetation. Both the lack of fine-grained elevation data and vegetation data prevent the identification of areas unsuitable for intensive management.

A very significant limitation of the model comes from its use of experimentally-determined parameters. Labour and cost functions were not known, so these were derived experimentally by driving the model through coarse- and fine-grained parameter sweeps. "Acceptable" model runs were deemed to be those which resulted in steady-state outcomes, particularly related to a household's capital and spatial footprint, based on steady-state inputs. Invalid model runs resulted in tremendously wealthy or completely decapitalized households. Other invalid runs resulted in unrealistic scenarios of complete deforestation. Although the model produces outputs which appear to make sense and are valid within the experimental frame, it is important to remember that many of the labour and cost parameters were experimentally-derived from the model itself. Ideally, these would be parameterized from data acquired through detailed field studies, estimating the input parameters using methods such as maximum likelihood estimation.

Simple parameter estimation based on the existing model may still result in an unrealistic model. Production functions, which determine the number of cells managed or harvested based on inputs and technology, were assumed to be linear for simplicity and in order to be compatible with a linear programming approach. A production function is often specified as $Q = f(X_1, X_2, \ldots, X_n)$, where Q is the output, X_i is the quantity of some input i, and f is the production function, related to technology. Inputs to the production function, in this model, are capital, labour, and available land (soil fertility). Linear production functions are assumed to be unrealistic, as they assume constant returns to scale. Alternative production functions include quadratic and Cobb-Douglas, which can model diminishing returns (Cobb and Douglas, 1928).

Surveys from 2007 at various communities in the Amazonian *várzea*, including Paricatuba, Marajó-Açu, and Praia Grande, have been conducted by the research team at Indiana University-Bloomington and are currently being codified for use in models. The surveys contain information on migration cycles, deforestation and production.

These surveys can be utilized as a snapshot of each household state at the time the surveys were collected. Due to the historic (past) timeframe of the model, rather than predictive (future), these surveys should not be used for initial parameterization, but rather, to verify and validate the model. At some time in the future, if similar surveys are conducted again, the 2007 surveys could be used to parameterize the model, with the future surveys used for verification and validation.

6.4 Future use of Maria

The decision-making model employed in future versions of Maria and other models may take a parallel approach, comparing alternative decision-making methods in the same manner as this thesis. Similarly, such a method would expose the biases of each decision-making model, providing the modeller with insight on endogenous and exogenous inefficiencies. However, this method may also reduce model confidence from a reader's point of view, since it may be unclear which model best represents the true system. Alternatively, a hybrid method, such as the one taken by Jager et al. (2000), may allow the agent to select a decision-making method based on environmental factors such as relative welfare and aspiration levels. *Homo psychologicus* agents in this model used aspiration levels (relative to peers) and uncertainty to determine which strategy to take: deliberation (optimization), imitation, repetition (autoimitation) and social comparison, where the agent chooses the better perceived outcome of imitation or repetition. Another, more simple method a future model may employ is the selection of just one of the decision-making models discussed. However, it is important to note the biases generated by the decision-making model, as they may significantly affect the model outcome.

With the development of the model framework, expert input and local knowledge can be integrated into the model to develop scenarios more closely tied to the study area. The model can be used as a collaborative tool to test assumptions in a virtual environment and share results among experts. As is, the Maria model is still in its infancy, requiring additional empirical data to improve and assess its quantitative performance (Axtell and Epstein, 1994). From a software usability perspective, Maria remains lacking, requiring a trained user to modify simple scalar parameters or a programmer to perform more complex changes. A limited set of complex parameter changes have been abstracted to scalar parameters, such as the optimizing vs. heuristic decision-method selection. This selection may be made at the parameter level, by setting the appropriate parameters (`percentHeuristicHouseholds`, for example).

However, the addition of new decision-methods or slight alterations of existing methods require a programming skill-set. The Repast Simphony environment offers a limited flowchart solution for the definition of agent behaviours, but there is no explicit support for this in Maria. As is, the Repast Simphony development environment can create black box models, where numeric parameters can be modified, but code cannot be modified. In the meantime, before modelling becomes more accessible without a programming skill set, collaboration between a modeller and expert is recommended throughout the design and implementation processes in order to develop a model eventually suitable for black box experimentation.

There has been specific interest from research teams in Columbia University and Indiana University-Bloomington regarding the role of multi-sited households in the area: Such arrangements are a result of economic linkages with households inside and outside the study area. Within the scope of the model, these multi-sited households have only been integrated for emigrant agents, but these represent only one type of multi-sited arrangement. Other arrangements include relationships with upland farmers for fishing rights and access to riverine transportation (Brondízio, 2008). Such arrangements may necessitate the modelling of the needs and desires of upland farmers, currently outside the scope of this initial model.

The role of middlemen as a source of employment and transportation arrangement is significant. Middlemen act on behalf of up to 90 % of açaí producers in some areas (Brondízio, 2008), providing services in the face of rapid spoilage, isolation and lack of transportation. Middlemen, as well as transportation costs, are outside the scope of the model and may present significant additional complexity. How does the presence or absence of middlemen affect açaí price? How does the viability of *marretagem* as an enterprise affect the labour pool? These are some of the questions that should be addressed as middlemen are integrated into the simulation environment.

Through the database, existing model outputs detail state variables for each agent and cell at each time step. While further applications of these output variables have not yet been discussed, the model's outputs can imply far more than its raw variables. For sites of secondary succession, a regression equation derived by Brondízio (2008, App. 3.1) can be used to calculate the average stand height based on land use history. In this equation (simplified from Brondizio's original equation), the average stand height is predicted based on site age (x_a) and previous land use types. x_p and x_m are Boolean variables indicating that the site was either abandoned pasture or an abandoned mechanized field, land uses which are not integrated into the current model:

$$Y = -0.0556x_a^2 + 1.56x_a + 1.879x_p - 1.116x_m + 0.184$$

These variables, namely the last land use and the site age, are available as model outputs, allowing the modeller to determine average stand height over the model using the provided equation. This has not been done, however, for the particular reason that this equation is specified only for long fallow swidden cycles, which are not necessarily followed in the model. Care should be taken to avoid misrepresenting output data, so caution has been taken in deriving conclusions from model outputs. However, a more refined model could potentially provide such detailed conclusions, within a degree of uncertainty.

Is additional complexity justified or even useful? There is a temptation to include several possible economic activities, household arrangements and other external influences on decision-making into the model. This increases the number of variables to be analyzed, and further complicates the verification and validation processes. While it makes for a "more realistic" model when all variables are included, the model may be as difficult to analyze as the real system, negating the value of model preparation. A modeller should be cautious when adding complexity to a designed system to ensure that it remains analyzable without presenting unnecessary complications.

APPENDICES

Appendix A

Distributed computing

A.1 Parameter sweep distribution

The runtime performances of early iterations of the model were found to be bound by processor speed. While a single simulation run can be performed in the matter of seconds, several thousands of runs require a significant amount of time to complete. By utilizing multiple processor cores or multiple computers, the computation time of large parameter sweeps can be significantly reduced.

In the current version of Repast Simphony, as shown in Figure A.1, parameter sweeps cannot be distributed outside of a process. The entirety of a batch run defined by an XML file or Groovy script must be run on one single process and is thus limited to one CPU core. Repast Simphony currently includes support for distributing work units *within* a run, which would require Terracotta-specific code to encapsulate work units within the model. For instance, independent agent actions may be executed in parallel, but synchronization must occur between time steps. Repast mailing list discussions indicate that distributed batch runs are to be implemented in the future.

To take advantage of multi-core systems and multi-node clusters with the current version of Repast Simphony, parameter sweep definitions must be broken up into smaller units. Since Repast Simphony is an open source application, it was feasible to reuse code. In particular, the `XMLSweeperProducer` class is used to construct a tree-based parameter sweeper from an XML file as listed above. The parameter sweeper is fed into an instance of the `BatchRunner` class, which manipulates the input variables and runs the model according to the parameter sweeper specification.

Instead of directly running the `BatchRunner` class with a produced parameter sweeper, a batch pre-processor was created. This pre-processor takes one or more

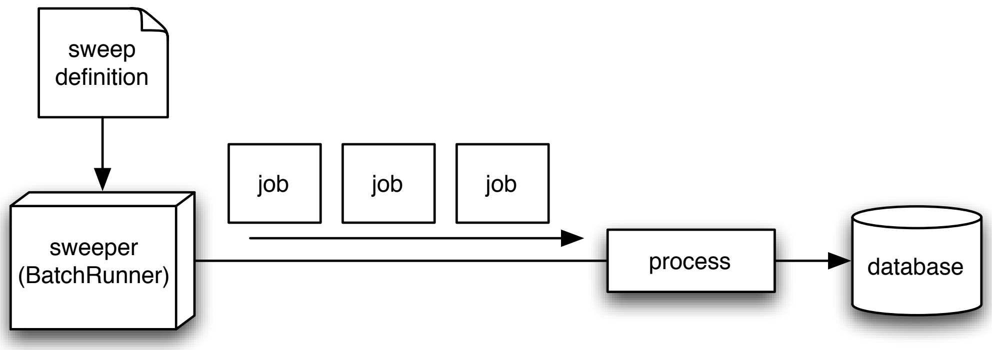


Figure A.1: Serial parameter sweep in Repast Simphony

batches of runs, which would otherwise be processed by one computer node, and repackages them into several batches which can be distributed among multiple nodes.

The pre-processor is implemented using Repast Simphony's `XMLSweeperProducer` and other Simphony classes. It accepts any number of parameter sweep definition files to produce a list of parameter sets. The parameter list is then used to create several batches consisting of one run each. Whereas the original parameter sweep in the XML example in Listing 4.3 produced one batch of 9 runs, 9 batches of one run each would be produced with this method. Each batch can then be distributed as a work unit to a set of worker processes.

A.2 The Master-Worker pattern

The Master-Worker pattern is a common method for distributing embarrassingly parallel processes like Monte Carlo simulations in a metacomputing environment, such as an ad-hoc cluster of networked workstations (Basney et al., 1999; Goux et al., 2000). A process is embarrassingly parallel if it can be trivially divided into segments with no shared state between them. A parameter sweep falls under this category, since each run is independent of other runs. Using Terracotta (Terracotta, Inc., 2008), a Java Virtual Machine clustering framework, the Master-Worker pattern was implemented to run parameter sweeps across multiple cores in a single workstation or across many nodes in a cluster. As described by Heymann et al. (2000), since task sizes are similar (as parameter values should not impact the execution time of a run), good efficiency and speedup can be attained even as the number of worker nodes approaches the number of tasks.

The Master-Worker pattern may be implemented using a single shared queue. This queue is a message pipe. A "master" process may place a work unit, such as a set of parameters, on the queue. At the other end, worker processes poll the queue until work is found. A worker process consumes the work and returns a result. The master process may poll its work units to check if they have been processed already. If the work units are guaranteed to be processed in linear order, only the first submitted work unit needs to be polled. However, if the units are not processed in linear order or if some units are lost, more work units must be polled. This may be the case if some nodes complete work faster than others. Alternatively, the worker process may place the completed work on another shared queue so that the master does not need to poll its own work units. Instead, master processes may poll this *return* queue to check for completed work, through which work units may be returned in any order. Given a timeout or another failure detection scheme, work units may be marked as "failed". The master may then re-submit these work units or report failure.

Under the Terracotta framework, objects can be shared between nodes. Changes to the objects must be replicated to each node. However, the cost required to synchronize shared objects between nodes can be high, in terms of network latency or bandwidth, especially if the shared objects are large, modified frequently or shared between many nodes. Furthermore, nodes reading a shared object must acquire shared locks while a node writing to a shared object must acquire an exclusive lock. Locks ensure that data is not lost or misread due to concurrent access, by ensuring that multiple writers do not write data to the same object, as all but the last-written change to the object would be overwritten. In another case, if a reader is given access to an object as it is being written, the reader may retrieve partially-written, corrupt data.

If a node wishes to acquire a (shared) read lock and an exclusive write lock is in place, it must wait until the write lock has been released, though other shared locks do not prohibit the acquisition of a new lock. If a node wishes to acquire a write lock, it must wait until all shared and exclusive locks have been released.

To minimize the amount of synchronization which must take place, objects are shared with as few parties as possible. Each master and worker are given two exclusive queues, an incoming queue and an outgoing queue. In contrast, given a single shared queue and many worker processes, many of these processes would contend for exclusive read/write access to the shared queue, resulting in increased latency as processes wait for access. Figure A.2 illustrates the distributed batch run in this arrangement. Masters and workers enter and exit the system freely, whether by job completion, in the case of master processes, or unexpected failure.

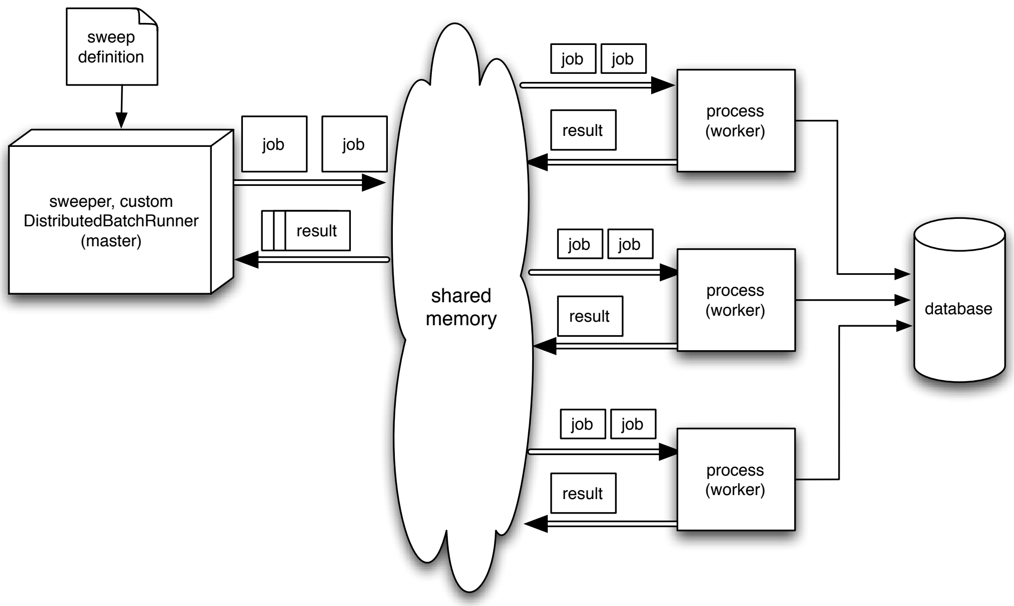


Figure A.2: Distributed run based on the Master-Worker pattern

On startup, a master places its work units on its outgoing queue and waits for completed work on its incoming queue. Each worker listens for incoming work at startup and upon completion of each work unit. Each worker waits for work on its incoming "scheduling" queue and places the results on its outgoing queue. Work is scheduled by placing work units onto the scheduling queue. More specifically, when a master joins, assuming it has defined its work units *a priori*, it places all of its work units on its outgoing queue. It then schedules its work from its outgoing queue to any available worker nodes. Work units may remain on its outgoing queue if workers are unavailable or if all scheduling queues are full. When a new worker joins the cluster, it checks all masters' outgoing queues to see if there is work available and will schedule any available work onto its own queue.

This implementation of the Master-Worker pattern is fault-tolerant. If a worker node fails, its incoming and outgoing queues continue to exist, but the results of its currently processing job will be lost. In this case, another idle worker (when available) will distribute the work from the lost worker's remaining schedule, including the job which was currently running. Adopting *at least once* message semantics, in the rare occasion that a worker completes a job without reporting it (having failed between the two steps), the job will be re-processed. If all other work queues become full as the lost worker's scheduled jobs are redistributed, the remaining schedule becomes "orphaned", to be adopted by a new worker or a worker which becomes idle, whichever happens first. Thus, all jobs are guaranteed to be processed despite worker failures.

If a master aborts, all jobs in its outgoing queue are cancelled, though immediately scheduled and running jobs are allowed to complete.

Appendix B

Parameters

Parameter	Description
runlength	The number of simulation ticks (years) per model run. (default = 40)
numHouseholds	The number of households. (default = 21)
numPersons	The number of individuals in the population. (default = 144)
numOffers	The number of urban employment offers made in each simulation tick, if constant.
lambdaOffers	The rate parameter of random employment offer generation, if exponentially distributed.
acaiPrice	The selling price of açaí, if constant over time.
maniocPrice	The selling price of manioc, if constant over time.
timberPrice	The selling price of timber, if constant over time.
priceStreamMultiplier	A scaling factor applied to variable selling prices read from input filestreams.
acaiLabour	Labour requirement of new açaí cells.
maniocLabour	Labour requirement of new garden cells.
maintainAcaiLabour	Labour requirement of açaí cell maintenance.
maintainManiocLabour	Labour requirement of garden cell maintenance.
harvestAcaiLabour	Labour requirement of açaí cell harvest.
harvestManiocLabour	Labour requirement of garden cell harvest.
harvestTimberLabour	Labour requirement of timber extraction per cell.
acaiCost	Capital requirement of new açaí cells.
maniocCost	Capital requirement of new garden cells.
forestFallowCost	Capital requirement of forest fallow cell maintenance.
maintainAcaiCost	Capital requirement of açaí cell maintenance.
maintainManiocCost	Capital requirement of garden cell maintenance.
percentOptimizingHouseholds	Proportion of linear programming household agents in the population.
percentHeuristicHouseholds	Proportion of decision-tree household agents in the population.

Table B.1: Parameter descriptions

Parameter	Value
runlength	40
numHouseholds	21
numPersons	144
numOffers	0
lambdaOffers	N/A
acaiPrice	0.0003
maniocPrice	0.002
timberPrice	N/A (infeasible)
priceStreamMultiplier	N/A (prices are constant)
acaiLabour	0.2
maniocLabour	0.4
maintainAcaiLabour	0.1
maintainManiocLabour	0.2
harvestAcaiLabour	0.5
harvestManiocLabour	0
harvestTimberLabour	N/A (infeasible)
acaiCost	1
maniocCost	3
forestFallowCost	0
maintainAcaiCost	2
maintainManiocCost	0.04
percentOptimizingHouseholds	100%, 0%
percentHeuristicHouseholds	0%, 100%

Table B.2: Constant price scenario parameters

Parameter	Value
acaiPrice	N/A (not constant—read from file)
maniocPrice	N/A (not constant—read from file)
priceStreamMultiplier	0.000003

Table B.3: Variable price scenario parameters, as differences from Table B.2

References

Albisser, G. and B. Lehmann (2007). Agent-based modeling of structural change in agriculture: Experiences from the implementation of cooperation and collaboration. In *I Mediterranean Conference of Agro-Food Social Scientists*, Barcelona, Spain.

Aspinall, R. J. (2008). Basic and applied land use science. In R. J. Aspinall and M. J. Hill (Eds.), *Land Use Change: Science, Policy and Management*. CRC Press.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Axelrod, R. (2003). Advancing the art of simulation in the social sciences. *Japanese Journal for Management Information System 12*(3), 1–19.

Axtell, R. and J. Epstein (1994). Agent-based modeling: Understanding our creations. *The Bulletin of the Santa Fe Institute 9*, 28–32.

Banks, J. (1998). *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. Wiley-Interscience.

Basney, J., R. Raman, and M. Livny (1999). High throughput Monte Carlo. In *Proceedings of the Ninth SIAM Conference on Parallel Processing for Scientific Computing*, San Antonio, Texas.

Batty, M. (1997). Cellular automata and urban form: A primer. *Journal of the American Planning Association 63*(2), 266–274.

Berger, T. (2001). Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics 25*(2–3), 245–260.

Berger, T., M. Goodchild, M. A. Janssen, S. M. Manson, R. Najlis, and D. C. Parker (2001). Methodological considerations for agent-based modeling of land-use and land-cover change. In *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*, Volume 6 of *LUCC Report Series*, Bloomington, Indiana.

Berger, T. and D. C. Parker (2001). Examples of specific research: Introduction. In *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*, Volume 6 of *LUCC Report Series*, Bloomington, Indiana.

Boschetti, L., D. Roy, and C. Justice (2008). Using NASA's World Wind virtual globe for interactive internet visualization of the global MODIS burned area product. *International Journal of Remote Sensing 29*(11), 3067–3072.

Bousquet, F. and C. Le Page (2004). Multi-agent simulations and ecosystem management: A review. *Ecological Modelling 176*(3-4), 313–332.

Box, G. and N. Draper (1987). *Empirical Model-building and Response Surface*. Wiley.

Brondízio, E. (2004). From staple to fashion food: Shifting cycles and shifting opportunities in the development of the açai palm fruit economy in the Amazon Estuary. In D. J. Zarin, J. R. R. Alavalapati, F. E. Putz, and M. Schmink (Eds.), *Working Forest in the Neotropics*, Chapter 19, pp. 339–365. Columbia University Press.

Brondízio, E. (2009). Agriculture intensification, economic identity, and shared invisibility in Amazonian peasantry: Caboclos and colonists in comparative perspective. In *Amazon Peasant Societies in a Changing Environment*, Chapter 8, pp. 181–214. Springer Science and Business Media.

Brondízio, E., C. A. M. Safar, and A. Siqueira (2002). The urban market of Açai fruit (Euterpe oleracea Mart.) and rural land use change: Ethnographic insights into the role of price and land tenure constraining agricultural choices in the Amazon estuary. *Urban Ecosystems 6*, 67–97.

Brondízio, E. and A. Siqueira (1997). From extractivists to forest farmers: Changing concepts of caboclo agroforestry in the amazon estuary. *Research in Economic Anthropology 18*, 233–280.

Brondízio, E. S. (2008). *The Amazonian Caboclo and the Açaí palm: Forest Farmers in the Global Market*, Volume 16 of *Advances in Economic Botany Monograph Series*. New York: New York Botanical Garden Press.

Brondízio, E. S., E. F. Moran, P. Mausel, and Y. Wu (1994). Land use change in the Amazon estuary: Patterns of caboclo settlement and landscape management. *Human Ecology 22*(3), 249–278.

Buttrey, S. E. (2005). Calling the lp_solve linear program software from R, S-PLUS and Excel. *Journal of Statistical Software 14*(4), 1–13.

Byron, M. (1998). Satisficing and optimality. *Ethics 109*, 67–93.

Castella, J.-C., T. N. Trung, and S. Boissau (2005). Participatory simulation of land-use changes in the northern mountains of Vietnam: The combined use of an agent-based model, a role-playing game, and a geographic information system. *Ecology and Society 10*(1), 27.

Chuvieco, E. (1993). Integration of linear programming and GIS for land-use modelling. *International Journal of Geographical Information Science 7*(1), 71–83.

Cleary, D. (1993). After the frontier: Problems with political economy in the modern Brazilian Amazon. *Journal of Latin American Studies 25*(2), 331–349.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*, 829—836.

Cleveland, W. S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician 35*(1), 54.

Cobb, C. W. and P. H. Douglas (1928). A theory of production. *The American Economic Review 18*(1), 139–165.

Couclelis, H. (2001). Why I no longer work with agents: A chellenge for ABMs of human-environment interactions. In *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*, Volume 6 of *LUCC Report Series*, Bloomington, Indiana.

Dantzig, G. B. (2002). Linear programming. *Operations Research 50*(1), 42–47.

Deadman, P., K. Lim, D. Robinson, E. Moran, E. Brondízio, and S. McCracken (2001). Lucita – multi-agent simulations of land-use change near Altamira, Brazil. In *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*, pp. 68–72.

Deadman, P., D. Robinson, E. Moran, and E. Brondízio (2004). Colonist household decisionmaking and land-use change in the Amazon Rainforest: An agent-based simulation. *Environment and Planning B 31*, 693–710.

Demazeau, Y. and J. P. Müller (1991). From reactive to intentional agents. In Y. Demazeau and J. P. Müller (Eds.), *Decentralized Artificial Intelligence*, Volume 2. Elsevier North-Holland.

Dufour, D. L. (1990). Use of tropical rainforests by native Amazonians. *BioScience 40*(4), 652–659.

Eastman, J. R., H. Jiang, and J. Toledano (1998). Multi-criteria and multi-objective decision making for land allocation using GIS. In E. Beinat and P. Nijkamp (Eds.), *Multicriteria Analysis for Land-Use Management*. Kluwer Academic Publishers.

Environmental Systems Research Institute (1992). ESRI ArcWorld Database.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation 11*(4), 12.

Epstein, J. M. and R. Axtell (1996). *Growing Artificial Societies*. Cambridge: MIT Press.

Fearnside, P. (1986). *Human Carrying Capacity of the Brazilian Rainforest*. Columbia University Press.

Finin, T., Y. Labrou, and J. Mayfield (1997). KQML as an agent communication language. In J. M. Bradshaw (Ed.), *Software Agents*. MIT Press.

Foundation for Intelligent Physical Agents (2000). FIPA ACL message representation in string specification. Technical report, Foundation for Intelligent Physical Agents.

Friedl, M. and C. Brodley (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment 61*(3), 399–409.

Gamma, E., R. Helm, R. Johnson, and J. Vlissides (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. Reading, Mass.: Addison-Wesley.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game 'Life'. *Scientific American 223*(4), 120.

Gigerenzer, G. (2006). Bounded and rational. In R. J. Stainton (Ed.), *Contemporary Debates in Cognitive Science*, pp. 115–133. Oxford, UK: Blackwell.

Gigerenzer, G. and P. M. Todd (1999). *Simple Heuristics that Make Us Smart*. Evolution and Cognition Series. Oxford University Press.

Gilbert, N., J. C. Hawksworth, and P. Sweeney (2008). An agent-based model of the UK housing market. University of Surrey.

Gotts, N. M., J. G. Polhill, and A. N. R. Law (2003). Aspiration levels in a land use simulation. *Cybernetics and Systems 34*, 663–683.

Goux, J.-P., J. Linderoth, and M. Yoder (2000). Metacomputing and the master-worker paradigm. Preprint ANL/MCS-P792-0200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne.

Hare, M. and P. Deadman (2008). Further towards a taxonomy of agent-based simulation models in environmental management. In A. L. Paredes and C. H. Iglesias (Eds.), *Agent Based Modelling in Natural Resource Management*, Chapter 1, pp. 3–52. Universidad de Valladolid.

Hedden-Dunkhorst, B., M. Denich, K. Vielhauer, A. Mendoza-Escalante, J. Börner, T. Hurtienne, F. R. de Sousa Filho, T. D. de Abreu Sá, and F. de Assis Costa (2003). Forest-based fallow systems: A safety net for smallholders in the Eastern Amazon. In *The International Conference on Rural Livelihoods, Forests and Biodiversity*, Bonn, Germany.

Heymann, E., M. A. Senar, E. Luque, and M. Livny (2000). Evaluation of an adaptive scheduling strategy for master-worker applications on clusters of workstations. In *Proceedings of the 7th International Conference on High Performance Computing*, Volume 1970 of *Lecture Notes in Computer Science*, pp. 310–319.

Huigen, M. G., K. P. Overmars, and W. T. de Groot (2006). Multiactor modeling of settling decisions and behavior in the San Mariano Watershed, the Philippines: A first application with the MameLuke framework. *Ecology and Society 11*(2), 33.

Huigen, M. G. A. (2004). First principles of the MameLuke multi-actor modelling framework for land use change, illustrated with a Philippine case study. *Journal of Environmental Management 72*(1-2), 5–21.

IBGE (2007). Contagem da população 2007.

Jager, W. and M. Janssen (2003). The need for and development of behaviourally realistic agents. In *Multi-Agent-Based Simulation II*, Lecture Notes in Computer Science. Springer Berlin/Heidelberg.

Jager, W., M. A. Janssen, H. J. M. D. Vries, J. D. Greef, and C. A. J. Vlek (2000). Behaviour in commons dilemmas: Homo economicus and Homo psychologicus in an ecological-economic model. *Ecological Economics 35*(3), 357–379.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

Kirkpatrick, S., C. Gelatt, and M. Vecchi (1983). Optimization by simulated annealing. *Science 220*(4598), 671–680.

Lawler, E. and D. Wood (1966). Branch-and-bound methods: A survey. *Operations Research 14*(4), 699–719.

Lim, K. S.-W. (2000). Agent-based simulations of land use change in the Amazon. Master's thesis, University of Waterloo.

Liu, J., T. Dietz, S. R. Carpenter, M. Alberti, C. Folke, E. Moran, A. N. Pell, P. Deadman, T. Kratz, J. Lubchenco, E. Ostrom, Z. Ouyang, W. Provencher, C. L. Redman, S. H. Schneider, and W. W. Taylor (2007). Complexity of Coupled Human and Natural Systems. *Science 317*(5844), 1513–1516.

Manson, S. (2004). The SYPR integrative assessment model: Complexity in development. In B. L. Turner, D. Foster, and J. Geoghegan (Eds.), *Integrated Land-Change Science and Tropical Deforestation in the Southern Yucatán: Final Frontiers*, pp. 271–291. Oxford, United Kingdom: Clarendon Press of Oxford University Press.

Manson, S. (2006a). Land use in the southern Yucatán peninsular region of Mexico: Scenarios of population and institutional change. *Computers, Environment and Urban Systems 30*(3), 230–253.

Manson, S. M. (2001). Calibration, verification, and validation. In *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*, Volume 6 of *LUCC Report Series*, Bloomington, Indiana.

Manson, S. M. (2005). Agent-based modeling and genetic programming for modeling land change in the southern Yucatán peninsular region of Mexico. *Agriculture, Ecosystems & Environment 111*(1–4), 47–62.

Manson, S. M. (2006b). Bounded rationality in agent-based models: Experiments with evolutionary programs. *International Journal of Geographical Information Science 20*(9), 991.

Matthews, R. (2006). The people and landscape model (PALM): Towards full integration of human decision-making and biophysical simulation models. *Ecological Modelling 194*, 329–343.

Matthews, R. B., N. G. Gilbert, A. Roach, J. G. Polhill, and N. M. Gotts (2007). Agent-based land-use models: A review of applications. *Landscape Ecology 22*(10), 1447–1459.

MDA Federal (2004). Landsat GeoCover 1990/TM Edition Mosaics.

Menzies, N. K. (2007). *Our Forest, Your Ecosystem, Their Timber: Communities, Conservation, and the State in Community-based Forest Management*. Columbia University Press.

Millington, J., R. Romero-Calcerrada, J. Wainwright, and G. Perry (2008). An Agent-Based Model of Mediterranean Agricultural Land-Use/Cover Change for Examining Wildfire Risk. *Journal of Artificial Societies and Social Simulation 11*(4), 4.

Minar, N., R. Burkhart, C. Langton, and M. Askenazi (1996). The Swarm simulation system: A toolkit for building multi-agent simulations. Working Paper 96-06-042, Santa Fe Institute, Santa Fe.

Moran, E. (1981). *Developing the Amazon*. Bloomington: Indiana University Press.

Moulin, B. and B. Chaib-draa (1996). An overview of distributed artificial intelligence. In G. M. P. O'Hare and N. R. Jennings (Eds.), *Foundations of Distributed Artificial Intelligence*. Wiley-Interscience.

Muñiz-Miret, N., R. Vamos, M. Hiraoka, F. Montagnini, and R. O. Mendelsohn (1996). The economic value of managing the açaí palm (Euterpe oleracea Mart.) in the floodplains of the Amazon estuary, Pará, Brazil. *Forest Ecology and Management 87*, 163–173.

Murrieta, R. S., E. S. Brondízio, A. Siqueira, and E. F. Moran (1989). Estratégias de subsistência de uma população ribeirinha do Marajó-Açu, Ilha de Marajó, Brazil. *Boletim do Museu Paraen se Emilio Goeldi, Série Antropologia 8*(50), 147–163.

Murrieta, R. S. S., D. L. Dufour, and A. D. Siqueira (1999). Food consumption and subsistence in three Caboclo populations on Marajó Island, Amazonia, Brazil. *Human Ecology 27*(3), 455–475.

Nash, J. C. (2000). The (Dantzig) simplex method for linear programming. *Computing in Science and Engineering 2*(1), 29–31.

Nikolai, C. and G. Madey (2009). Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation 12*(2), 2.

North, M. J., N. T. Collier, and J. R. Vos (2006). Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modelling and Computer Simulation 16*(1), 1–25.

North, M. J. and C. Macal (2007). *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press.

North, M. J., E. Tatara, N. T. Collier, and J. Ozik (2007). Visual agent-based model development with Repast Simphony. In *Proceedings of the Agent 2007 Conference on Complex Interaction and Social Emergence*, Argonne. Argonne National Laboratory.

Nugent, S. (1993). *Amazonian Caboclo Society: An Essay on Invisibility and Peasant Economy*. Oxford: Berg Publishers.

Ormerod, P. and B. Rosewell (2009). Validation and Verification of Agent-Based Models in the Social Sciences. In *Epistemological Aspects of Computer Simulation in the Social Sciences*, Lecture Notes in Computer Science, pp. 130–140. Springer.

Pace, R. (1997). The Amazon Caboclo: What's in a name? *Luso-Brazilian Review 34*(2), 81–89.

Padoch, C., E. Brondízio, S. Costa, M. Pinedo-Vasquez, R. Sears, A. Siqueira, et al. (2008). Urban Forest and Rural Cities: Multi-sited Households, Consumption Patterns, and Forest Resources in Amazonia. *Ecology and Society 13*(2), 2526–2541.

Padoch, C., J. C. Inuma, W. D. Jong, and J. Unruh (1985). Amazonian agroforestry: a market-oriented system in Peru. *Agroforestry Systems 3*, 47–58.

Parker, D., D. Brown, J. G. Polhill, S. M. Manson, and P. Deadman (2008). Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: The MR POTATOHEAD framework. In A. L. Paredes and C. H. Iglesias (Eds.), *Agent-based Modelling in Natural Resource Management*, pp. 29–62. Valladolid, Spain: Universidad de Valladolid.

Parker, D. and V. Meretsky (2004). Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems and Environment 101*(2-3), 233–250.

Parker, D. C., T. Berger, and S. M. Manson (Eds.) (2001). *Agent-Based Models Of Land-Use and Land-Cover Change. Proceedings of an International Workshop October 4–7, 2001, Irvine, California, USA*.

Parker, D. C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers 93*(2), 314–337.

Peterman, R. M. and J. L. Anderson (1999). Decision analysis: A method for taking uncertainties into account in risk-based decision making. *Human and Ecological Risk Assessment 5*(2), 231–244.

Peters, C. M., M. J. Balick, F. Kahn, and A. B. Anderson (1989). Oligarchic forests of economic plants in amazonia: Utilization and conservation of an important tropical resource. *Conservation Biology 3*(4), 341–349.

Pettit, P. (1984). Satisficing consequentialism. In *Proceedings of the Aristotelian Society, Supplementary Volumes*, Volume 58, pp. 165–176. Blackwell Publishing on behalf of The Aristotelian Society.

Pignotti, E., P. Edwards, A. Preece, N. Gotts, and G. Polhill (2004). FEARLUS-G: A semantic grid service for land-use modelling. In *Proceedings of the ECAI-04 Workshop on Semantic Intelligent Middleware for the Web & the Grid*.

Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science 146*(3642), 347–353.

Polhill, J. G., N. M. Gotts, and A. N. R. Law (2001). Imitative versus nonimitative strategies in a land-use simulation. *Cybernetics and Systems 32*(1–2), 285–307.

Polhill, J. G., D. Parker, D. Brown, and V. Grimm (2008). Using the ODD protocol for describing three agent-based social simulation models of land-use change. *Journal of Artificial Societies and Social Simulation 11*(2), 3.

Pontius, R. G. and L. C. Schneider (2001). Land-cover change model validation by an ROC method for the Ipswich Watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment 85*(1–3), 239–248.

Poundstone, W. (1985). *The Recursive Universe*. William Morrow and Company.

Qudrat-Ullah, H. (2005). Structural validation of system dynamics and agent-based simulation models. In *Proceedings of the 19th European Conference on Modelling and Simulation*.

Quinlan, J. (1990). Decision trees and decision-making. *IEEE Transactions on Systems, Man and Cybernetics 20*(2), 339–346.

Railsback, S. F., S. L. Lytinen, and S. K. Jackson (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation 82*(9), 609–623.

Robinson, D., D. Brown, D. Parker, H. Wittmer, N. Gotts, P. Promburom, E. Irwin, and T. Berger (2007). Comparison of empirical methods for building agent-based models in land use science. *Journal of Land Use Science 2*(1), 31–55.

Robinson, D. T. (2003). Modelling farmer household decision-making and its effects on land use/cover change in the Altamira region, Párá, Brazil. Master's thesis, University of Waterloo.

Russell, S. and P. Norvig (2002). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*, 143–186.

Schreinemachers, P. and T. Berger (2006). Land use decisions in developing countries and their representation in multi-agent systems. *Journal of Land Use Science 1*(1), 29–44.

Schruben, L. W. (1982). Detecting initialization bias in simulation output. *Operations Research 30*(3), 569–590.

Silva, E. and K. Clarke (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems 26*(6), 525–552.

Silva-Fosberg, M. C. (1996). Occupation of the Amazon and the impact on Caboclo populations. In *95th American Anthropological Association Annual Meeting*, San Francisco.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review 63*(2), 129–138.

Simon, H. A. (1957). Rationality and decision making. In *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in a Social Setting*, pp. 196–206. New York: John Wiley & Sons.

Simon, H. A. (1976). *Administrative Behavior* (3rd ed.). New York: The Free Press.

Siqueira, A. (2009). *Women, Gender Relations and Decision-Making in Caboclo Households in the Amazon Estuary*, Chapter 10, pp. 241–257. Springer Netherlands.

Siqueira, A. D. (1997). *The Ecology of Food and Nutrition among Caboclo Populations in the Amazon Estuary*. PhD dissertation, Indiana University, Bloomington.

Siqueira, A. D., E. Brondízio, R. Murrieta, H. Silva, W. Neves, and R. B. Vietler (1993). Estratégias de subsistência da população ribeirinha do Igarapé do Paricatuba, Ilha do Marajó, Brasil. *Boletim do Museu Paraense Emilio Goeldi, Série Antropologia 9*, 153—170.

Siqueira, A. D., R. S. Murrieta, and E. S. Brondízio (2000). Land tenure, access to resources and food security in the Amazon Estuary. In *"Constituting the Commons: Crafting Sustainable Commons in the New Millenium", the Eighth Conference of the International Association for the Study of Common Property*, Bloomington, Indiana.

Takeyama, M. and H. Couclelis (1997). Map dynamics: Integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science 11*(1), 73–91.

Terracotta, Inc. (2008). *The Definitive Guide to Terracotta: Cluster the JVM for Spring, Hibernate and POJO Scalability*. Apress.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In *Handbook of Computational Economics*, Volume 2 of *Handbooks in Economics Series*, Chapter 16, pp. 831–880. Elsevier/North-Holland.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography 46*(2), 234–240.

Torrens, P. M. and D. O'Sullivan (2000). Cities, cells, and complexity: Developing a research agenda for urban geocomputation. In *Proceedings of the 5th International Conference on GeoComputation, University of Greenwich, Angleterre*.

Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology 5*, 207–232.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

Verburg, P., P. Schot, M. Dijst, and A. Veldkamp (2004). Land use change modelling: current practice and research priorities. *GeoJournal 61*(4), 309–324.

Vivid Solutions (2003). JTS Topology Suite technical specifications. Version 1.4.

Wallace, A. R. (1853). *Palm Trees of the Amazon: And Their Uses*. London: John Van Voorst.

Wilensky, U. (1999). *NetLogo*. Evanston, Illinois: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Xiang, X., R. Kennedy, G. Madey, and S. Cabaniss (2005). Verification and validation of agent-based scientific simulation models. In *Agent-Directed Simulation Conference*.

Ziegler, B. P. (1976). *Theory of Modelling and Simulation*. Wiley-Interscience.