Institut für Lebensmittel- und Ressourcenökonomik

Modelling the dynamics of land use & intensification in the Kilombero Valley, Tanzania

D i s s e r t a t i o n
zur
Erlangung des Grades
Doktor der Agrarwissenschaften
(Dr.agr.)

der
Landwirtschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Bisrat Haile Gebrekidan
aus

Alemaya, Äthiopien

Bonn 2021

Vorsitzende	Prof. Dr. Mariele Evers
Erstgutachter	Prof. Dr. Thomas Heckelei
Zweitgutachter	Prof. Dr. Mathias Becker
Fachnahes Mitglied	Prof. Dr. Jan Börner
Tag der mündlichen Prüfung	11 Dezember 2020

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

Acknowledgements

I am indebted to the many people that have helped me to make this research possible. First, I am incredibly grateful to my supervisor Prof. Dr. Thomas Heckelei. His guidance, unparalleled support & encouragement were critical to the completion of the thesis & my ambitious pursuits. Thank you for making your office always open no matter how busy you are.

Second, I would like to thank Prof. Dr. Mathias Becker, who agreed to act as a second examiner for this thesis & for the time invested in reviewing. I also take this opportunity to thank Prof. Dr. Jan Börner & Prof. Dr. Mariele Evers, who consented to be part of my doctoral committee.

I am deeply indebted to Dr. Sebastian Rasch for sharing his programming experience, many useful scientific discussions, and friendly conversations. The completion of my dissertation would not have been possible without his support & nurturing—Thank you, particularly for taking the time to read all my drafts that have culminated in this thesis.

I want to thank the staff of the Institute of Food and Resource Economics(ILR); the colleagueship provided an enjoyable & friendly atmosphere. Specially, I would like to thank Kristin Nitsche, Maria-Th. Hähnel & Birgit Klein for all the administrative support they offered to me. Thank you, Wolfgang Brül, for supporting me with my never-ending hardware & software requests.

This research would not have been possible without the assistance of many individuals in Tanzania that helped me to navigate the many technical obstacles to my fieldwork & providing me additional data. Special thanks to Prof.Salome Misana, Dr.Flister Mombo, Giuseppe Daconto, Irene Madiavale. I would also like to express my sincere thanks to the local extension service of Kilombero and Ulanga districts, the farmers who participated in this study, and my field Assistants.

I gratefully acknowledge the financial support of: (1) The GlobE: Wet-lands in East Africa project funded by the German Federal Ministry of Education and Research (FKZ: 031A250 A-H), with additional funding provided by the German Federal Ministry of Economic Cooperation & Development; (2) The CRC/Transregio 228:

Future Rural Africa: Future-making and social-ecological transformation (Project number: 328966760) funded by Deutsche Forschungsgemein-schaft (DFG).

I wish to acknowledge the support & great love of my family: my mother, father, brother, and my two sisters. Special thanks go to Etsegenet Befekadu & Embiet Kasahun as they kept me going on & this work would not have been possible without their support.

I have been very fortunate to come across good friends. I am thankful to all who provide me valuable advice, discussion, corporation & friendship throughout my stay in Bonn.

Kurzfassung

Die vorliegende Dissertation analysiert Determinanten und Konsequenzen der Entscheidungen von Landwirten bezüglich verschiedener Intensivierungsformen, Landnutzungen sowie deren temporalen Dynamiken mittels der Erstellung eines zeitlich und räumlich expliziten Agentenmodells des Kilombero Valley Floodplain (KVF) in Tansania. Die Dissertation ist in eine Einleitung und vier Hauptkapitel unterteilt. Letztere beantworten jeweils eine spezifische Forschungsfrage.

Die Einleitung motiviert zunächst die Forschungsfragen und gibt einen kurzen Überblick über die Studienregion (KVF) und erläutert die Datenquellen und Erhebungsmethoden. Es folgt eine Zusammenfassung der vier Hauptkapitel anhand der spezifischen Ziele, methodischen Ansätze und Hauptergebnisse sowie der Beiträge zur Literatur. Dies wird durch die Beschreibung der Begrenzung der Analysen sowie einen Forschungsausblick abgeschlossen.

Das zweite Kapitel stellt einen systematischen Überblick über landnutzungsbasierte Agentenmodelle dar. Sich am Rahmenkonzept „MRPOTATOEHEAD“ orientierend, identifiziert der Literaturüberblick alle gemeinsamen Modellkomponenten und deren unterschiedliche Ausprägungen in den spezifischen Modellen. Darüber hinaus werden die Besonderheiten der Modellen oder Fallstudien diskutiert. Die Ergebnisse zeigen, dass die Modelle sich hinsichtlich Skalierung, Detailstufe der sozialen- und biophysikalischen Dimensionen sowie in den angewandten Entscheidungsroutinen unterscheiden. Die Ausrichtung der Modelle auf die spezifischen Studiengegenden sowie Agrarnutzungssysteme schränken die allgemeinen Schlußfolgerungen ein.

Das dritte Kapitel konzentriert sich auf die Charakterisierung der Heterogenität von Landwirten in KVF und beleuchtet die Diversität ihrer Landnutzungs- und Unterhaltsstrategien mittels einer attributbasierten Typologie. Eine Kombination von „principal component analysis, hierarchical clustering“ und „K-means clustering“ stellt den methodischen Ansatz zur Umsetzung der Zielsetzung dar. Drei Farmtypen wurden identifiziert: „Monocrop rice producer“, „Diversifier“ und „Agropastoralist“. Die Beiträgeg des Kapitels umfassen: 1) die erste umfassende Klassifizierung von landwirtschaftlichen Haushalten im KVF, 2) die Hervorhebung von aktuellen landwirtschaftlichen Praktiken sowie die Generierung von Informationen welche für farmtypenspezifische Eingriffe benötigt werden und 3) eine quantitativ robustere sowie konsequentere Methodik zur Konstruktion sowie Validierung von Typologien.

Die resultierenden Farmtypen können in zukünftiger Forschung als Prototypen und zur Parametrisierung von Agentenmodellen genutzt werden.

Das vierte Kapitel untersucht die Entscheidungen von Landwirten im Kontext unabhängiger Determinanten durch die Betrachtung von vier im Tal genutzten Optionen (Nutzung von verbessertem Saatgut, Düngung, kleinteilige Bewässerung und schnelleren Wiederanbau). Das Kapitel präsentiert einen neuen Modellierungsansatz um nach Handlungsalternativen differenzierte Erklärungsvariablen sowie deren Wechselwirkung zu identifizieren. Hierzu wird eine Kombination aus einem „Baysian Belief Network“ (BBN), „design of experiments“ sowie multivariaten Regresssionsbäumen vorgeschlagen. Diese Methode ermittelt strategiespezifische Faktoren. Auch wenn die Wahl jeder Option durch die Kovariate unterschiedlich beeinflusst wird, spielen der Zugang zu nicht-landwirtschaftlichem Einkommen, Marktzugang und die Topographie der Flächen grundlegende Rollen über alle Intensivierungsoptionen hinweg.

Das fünfte und letzte Kapitel präsentier einen breiten Analyseansatz mit der Entwicklung eines räumlich und zeitlich expliziten, empirischen Agentenmodells welches die potentiellen Effekte von Migration und Infrastrukturbildung im KVF simuliert. Dieses Kapitel basiert auf den drei vorhergehenden. Die Simulationsergebnisse zeigen, dass Intensivierung in der Langzeitbetrachtung limitiert ist. Bei unkontrollierter Migration reagieren Landwirte eher mit Landexpansion als mit Intensivierung. Darüber hinaus ist ein zu vernachlässigbarer Effekt von verbesserter Transportinfrastruktur und der damit zusammenhängenden Reduzierung von Transportkosten auf Intensivierung und Nutzpflanzenproduktion im Tal zu verzeichnen.

Schlüsselwörter: *Intensivierung, Landnutzung, Subsahara-Afrika, Tansania, Kilombero-Tal, Agenten-basierte Modellierung, Bayesian Belief Network, Typologie*

Abstract

This thesis intends to examine the determinants and results of farmer's decisions to uptake different paths of intensification, land use, and its dynamics over time by building a spatially and temporally explicit agent-based model in the Kilombero Valley Floodplain (KVF) in Tanzania. The thesis is structured into an introduction and four main chapters, each answering a specific research question.

First, the introduction motivates the study and states the research questions. A brief context of the study site (KVF) and the different data sources and collection methods is given. Then, the four main chapters are summarized regarding the specific objectives, the approach and the main results as well as the contributions to the literature. It concluded with general limitations and outlook.

The second chapter provides a systematic review of agent-based models of land use in agricultural systems. Guided by the MRPOTATOHEAD framework, the review identifies the modeling components that appear in all the models and, how they are represented. Moreover, their peculiarities in the specific model or case study are discussed. The results show that models are unique in terms of scale, level of detail in both human and biophysical dimensions and the employed decision-making routines. Also, models are tailored to a particular study area and farming system under consideration.The targeted design of the models for specific study-regions and agricultural land uses restrict the derivation of generalizing conclusions.

The third chapter focuses on the characterization of farmers' heterogeneity in KVF and elicits the diversity of their land use and livelihood strategies through an attribute-based typology. The approach applied to achieve this objective is a combination of principal component analysis, hierarchical clustering, and K-means clustering. Three farm types were identified: "Monocrop rice producer", "Diversifier", and "Agropastoralist". The chapter's contributions comprise the following: (1) it offers the first concise classification of farm households in KVF (2) highlights current agricultural practices and provides vital information needed for targeted interventions per farm type (3) uses a quantitatively more rigorous and robust way to construct and validate typologies. The resulting farm types can be used in further research as a basis for building prototype farms and to parametrize agent-based models.

The fourth chapter investigates farmers' choices of intensification strategies alongside interdependent determinants by focusing on four options (use of improved seed, fertilizer application, small-scale irrigation and increasing frequency of planting) practiced in the valley. The chapter proposes a new modeling approach to identify option-specific determinants and their interdependence by combining a Bayesian Belief Network (BBN), design of experiments, and multivariate regression trees. The method has provided us with strategy-specific factors. Although the choice of each option is affected differently by covariates under consideration, access to non-farm income, access to market, and topography of the plot play essential roles across intensification options.

The fifth and final chapter takes a broad-based approach by developing a spatially and temporally explicit empirical agent-based model and simulates the potential effects of in-migration and infrastructure development in the KVF. This chapter builds upon the first three chapters. Simulation results show that intensification is limited in the long run and farmers engage in land expansion rather than intensification with uncontrolled immigration into the valley. More so, access to better road infrastructure and corresponding reduction in transport cost shows a negligible effects on trends of intensification and crop production in the valley.

Keywords: *Intensification, landuse, Sub-Sharan Africa, Tanzania, Kilombero Valley, Agent Based Models, Bayesian Belife Network, Typology*

Contents

List of Figures

List of Tables

List of Abbreviations

ASDS-II	**A**griculture **S**ector **D**evelopment **S**trategy -II
API	**A**pplication **P**rogramme **I**nterface
ASS	**A**griculture **S**ample **S**urvey
AWS	**A**mazon **W**eb **S**ervices
BBN	**B**ayesian **B**elief **N**etwork
BTC	**B**elgium **T**anzania **C**orporation
BTC	**C**onditional **P**robability **T**ables
CV	**C**oefficient **of** **V**ariation
DEM	**D**igital **E**levation **M**odel
DFID	**D**epartment **f**or **I**nternational **D**evelopment
DPSIR	**D**rivers **P**ressures **S**tate changes **I**mpacts and **R**esponses
DOE	**D**esign **of** **E**xperiment
EM	**E**xpectation **M**aximization
FA	**F**armer **A**gent
FAO	**F**ood and **A**griculture **O**rganization of the United Nations
GCA	**G**ame **C**ontrolled **A**rea
GCP	**G**oogle **C**loud **P**latform
GDP	**G**ross **D**omestic **P**roduct
GIS	**G**eographic **I**nformation **S**ystem
GLC	**G**lobal **L**and **C**over
GOT	**G**overnment **of** **T**anzania
GUI	**G**raphical **U**ser **I**nterface
Ha	**H**ectars
IA	**I**ndividual **A**gent
KILORWEMP	**KI**lombero and **LO**wer **R**ufiji **W**etlands **E**cosystem **M**anagement **P**roject
KPL	**K**ilombero **P**lantation **L**imited
KVF	**K**ilombero **V**alley **F**loodplain
LC	**L**and **C**ell
LSAI	**L**arge **S**cale **A**gricultural **I**nvestment
LU	**L**and **U**se
MASL	**M**eter **A**bove **S**ee **L**evel
MCRP	**M**onocrop **R**ice **P**roducer
MRPOTATOHEAD	**M**odel **R**epresenting **P**otential **O**bjects **T**hat **A**ppear in the **O**ntology of **H**uman **A**ction and **D**ecision
NGOs	**N**one **G**overnmental **O**rganizations
NOLH	**N**early **O**rthogonal **L**atin **H**ypercube
ODD+D	**O**verview, **D**esign **C**oncepts and **D**etails + **D**ecision

PA	**P**arcel **A**gent
PCA	**P**rincipal **C**omponent **A**nalysis
RA	**R**egion **A**gent
SAGCOT	**S**outhern **A**gricultural **G**rowth **C**orridor **o**f **T**anzania
SD	**S**tandard **D**eviation
SES	**S**ocio-**E**cological **S**ystem
SSA	**S**ub-**S**aharan **A**frica
TAN	**T**ree **A**ugmented **N**aive
TLU	**T**ropical **L**ivestock **U**nit
TNBS	**T**anzania **N**ational **B**ureau of **S**tatistics
TSh	**T**anzanian **Sh**iling
TWI	**T**opographic **W**etness **I**ndex
UML	**U**nified **M**odelling **L**anguage
UNDP	**U**nited **N**ations **D**evelopment **P**rogramme
URT	**U**nited **R**epublic **o**f **T**anzania
USAID	**U**.S. **A**gency **f**or **I**nternational **D**evelopment
WA	**W**ard **Agent**
WetABM	**W**etland **A**gent **B**ased **M**odel

Chapter 1

Overview of the thesis

1.1 Motivation

Like most sub-Saharan African countries, agriculture is the mainstay of Tanzania economy. Its role is manifested in several dimensions, including being a source of food, employment, and export earnings (ERM, 2012).With most of the population residing in rural areas where poverty and deprivation are widespread , the sector has sufficient scale and growth linkages to influence economic development and lift the majority from poverty (Diao, Hazell, & Thurlow, 2010).

In the last two decades, the government of Tanzania has formulated various agricultural policies and plans with different scope and focus towards attaining its long-term development objectives (Bassi, Casier, Pallaske, Perera, & Uzsoki, 2018). Specifically, these plans including the "Big result Now"(aimed at socio-economic development and agricultural expansion), Kilimo Kwanza and Southern Agricultural Growth Corridor of Tanzania (SACGOT) focusing on agricultural development and food security, altogether reecho the need for doubling the volume of farm exports and foreign exchange earnings.

While most of these policies and plans are usually implemented at the national level, they are targeted towards hotspot areas with high resource availability and potential for increasing productivity. One such area with a particular interest in agricultural production and intervention since the colonial era is the Kilombero Valley Floodplain wetland(KVF) (Sulle, 2020).

KVF is the most extensive low-altitude freshwater wetland in East Africa, covering approximately 6,300km^2 (Dinesen, 2016). Due to its unique biodiversity, ecology, and international importance, it was designated as a Ramsar site in 2002 (Wilson, McInnes, Mbaga, & Ouedaogo, 2017). KVF contains almost 75% of the threatened species of puku antelope, African Elephant, and three endemic birds species, to mention few (Lyon et al., 2015; Wilson et al., 2017). At the same time, the availability of fertile soil, supported by the delivery of nutrients from the annual seasonal flooding, has made the valley conducive to agricultural production and hence a target for smallholder farmers, commercial estates, and the government (Wilson et al., 2017). Thus, the valley is a typical example of competing needs that requires reconciliation between agricultural development and environmental protection (ERM, 2012; Milder, Buck, & Hart, 2013). Due to population increase, migration, and traditional farming practices, the valley has witnessed the conversion of the wetland area to agricultural land on an unprecedented scale (Leemhuis et al., 2017; Msofe et al., 2019).

Smallholder farmers account for the majority of agricultural production and cultivated area in the valley and have thus a unique role to play in the utilization and management of the wetland ecosystem (Milder, Buck, Hart, Scherr, & Shames, 2013; Mombo, Speelman, Huylenbroeck, Hella, & Moe, 2011). With limited land for agricultural expansion, the government and non-governmental organizations have been pushing towards agricultural intensification (Government of Tanzania, 2016). The efforts are usually targeted by mimicking the "green revolution" experience of East Asian countries. Interventions including expansion and improvement of an irrigation system, development of crop varieties, extension service, improved agricultural land use plans, and enhancing accessibility to the market are being forwarded to the Valley (ERM, 2012). Recently through the SACGOT initiative, the GOT has significantly emphasized the need for agricultural intensification by expanding the role of multinational companies and creating synergetic relations with smallholders to increase productivity and commercial production in the Valley (SAGCOT, 2012).

However, these past efforts haven't provided the required level of agricultural intensification (both in absolute and relative to other countries) to raise the living standards of small-scale farmers and increase the countries food self-sufficiency (Government of Tanzania, 2016). For instance, the analysis based on Agriculture Sample Survey for the year 2007 (TNBS, 2009) for Kilombero and Ulanga districts indicates that 23 percent of the survey households used improved seed, 35 percent cropped multiple times a year, 28 percent applied chemical fertilizer, and 3 percent used irrigation. This low level of intensification in the valley raises a critical question of how smallholder farmers in floodplain farming systems make intensification decisions and what socio-economic, biophysical, and institutional factors are important in choosing one intensification strategy over the other.

The analytical framework necessary for such an analysis needs to capture the complexity of the floodplain farming system. This complexity lies in the heterogeneity of the farmers, the multiple available options and the market landscape , and by the stochastic nature of the environment with which farmers interact (Nolan, Parker, Van Kooten, & Berger, 2009; Rindfuss et al., 2008). One of the approaches with proven capability of capturing the complexity of this coupled human-environment system is Agent-Based Modeling (ABM) (Entwisle, Malanson, Rindfuss, & Walsh, 2008; D. Parker, 2003). ABM is used as a computational methodology for formalizing and analyzing complex social systems on many scales, ranging from small groups of individuals to organizations and larger systems (Cioffi-revilla, 2011). They model actors' decisions and interactions in a sequential manner with explicit time-steps. They are preferable in economics when "the complex relationships between agent heterogeneity, interactions, and cross-scale feedbacks render traditional equilibrium-based models analytically intractable" (Nolan et al., 2009, .p 419).

1.2 Objective and research questions

The motivation for this thesis originates from the "GlobE – Wetlands in East Africa" (FKZ: 031A250A-H) project. The project was funded by the Federal Ministry of Education and Research of Germany (BMBF). It is one of many projects with an interdisciplinary approach to tackle the central goal of the National Research Strategy BioEconomy 2030 (BMBF, 2013), which is to secure the global food supply. The GlobE – Wetlands project and the participating African and German partners focused on wetland systems in East Africa (Kenya, Rwanda, Tanzania, Uganda). They researched the possibilities to reconcile future food production with environmental protection. As one of the main study sites within the project, the thesis takes the Kilombero valley as a case study.

The primary objective of this thesis is to examine the determinants and results of farmer's decisions to take different paths of intensification and its trend over time by building a spatially and temporally explicit agent-based model. In achieving our objective, we put forward four different research questions that contribute to both conceptual and empirical understanding of intensification and land use in the KVF:

1. Which elements and model components of ABMs are generally required to appropriately analyze land use and intensification decisions of smallholder farmers at the landscape scale?
2. What are the main characteristics and current farming practices of smallholder farmers in Kilombero Valley?
3. How do smallholder farmers make intensification decisions, and what are the main determinants they consider in their decision?
4. What is the potential effect of immigration and access to road infrastructure on the state of intensification, land use, and agricultural production in KVF?

1.3 Study site and data

1.3.1 Study site

Location: The valley is positioned at the foot of the Great Escarpment of East Africa in the southern half of Tanzania, about 300 km from the coast (Kato, 2007; Nindi, Maliti, Bakari, Kija, & Machoke, 2014) and lies between longitudes 34.563° and 37.797°E and latitudes 7.654° and 10.023°S (See Figure 1.1) (Wilson et al., 2017). It conversations about 11,600 km ^{2}, with a total length of 250 km and a width of up to 65 km. The floodplain is surrounded by the Udzungwa mountains in the northwest and the Mbarika Mountains and Mahenge Highlands in the southwestern parts (Lyon et al., 2015). The peak elevation drops from more than 1,800 masl to about 300 masl in a few kilometers. Generally, the floodplain is humid, with high temperatures ranging from 26°C to 32°C. While the relative humidity in the mountains is between 70 – 87%, the lowlands experience 58 – 85% humidity with average potential evaporation of 1800 mm (Wilson et al., 2017). KVF is a typical fertile alluvial floodplain with loamy, clay, clay loamy and sandy soils and is an essential source of nutrients and sediment (Milder, Buck, & Hart, 2013; Nindi et al., 2014).

Hydrology: The Kilombero valley forms part of the four principal sub-basins of the Rufiji River Basin and comprises several rivers and seasonally flooded marshes and

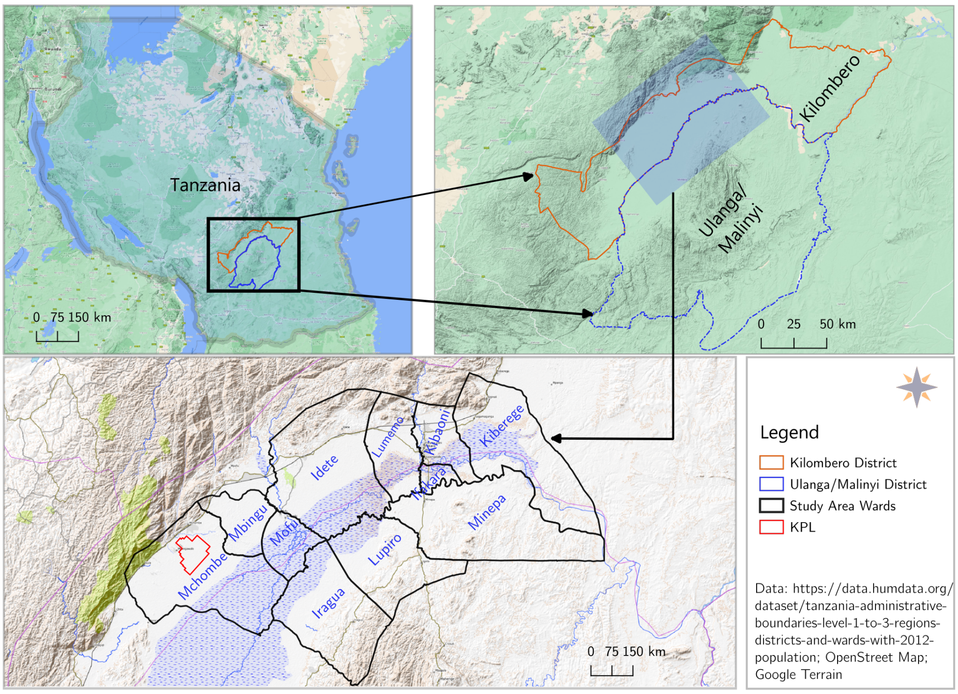


FIGURE 1.1: Location of study site

swamps (Dinesen, 2016).The valley receives annual precipitation between 1200 and 1400mm. The rainy season spans between December and April, while the dry season is between June and September (Wilson et al., 2017). The seasonal change in water dynamic is vast, and the plains sometimes become inundated during the wet season, while it dries up during the dry season except for the rivers and river margins as well as areas with permanent swamps and water bodies (Kato, 2007; Ntongani, Munishi, More, & Kashaigili, 2014).

Conservation: The KVF is of global, regional and national importance in terms of ecology and biodiversity. It comprises several conservation areas, including game-controlled area, Selous Game Reserve, and the Kilombero valley Ramsar site (Dinesen, 2016; Nindi et al., 2014). The valley contains a diverse flora of around 350 plant species, including endemic and threatened species (ERM, 2012). Since 1956, the Kilombero floodplain and adjacent woodland areas have been designated as a Game Controlled Area (GCA) and since 2002 as a Ramsar site (RAMSAR, 2002). Due to low enforcement of protection zoning (Munishi, Chuwa, Kilungu, Moe, & Temu, 2012), the Kilombero GCA has been managed by the Belgium Tanzania Corporation

(BTC) and the European Union, in partnership with the SAGCOT (KILORWEMP, 2017). Efforts are underway to redefine the borders of the GCA and create Wildlife Management Areas (Blache, 2019). These conservation areas in the Kilombero valley are reserved for tourist hunting and, therefore, not directly used by villagers (Blache, 2019).

Population and livelihood: According to the 2012 National census, the floodplain is home to more than 673,000 people (TNBS, 2013). The majority of the population lives in rural areas with low population density. Mang' ula and Ifakara are the two most populated Divisions in Kilombero, with a population density of 22 persons/km^2. The high population density in these two towns attributed to the fact that Ifakara is a district capital and Mang' ula has a large-scale sugar cane plantation (ERM, 2012).

Immigration into the valley has increased dramatically due to the perceived availability of high quality and cheap farmland. Conflicts between pastoralists and farmers over land use are a chronic and widespread problem, which has resulted in injury and litigation disputes (MALF, 2015; Nindi et al., 2014).

The KVF has a diverse ethnic profile. Ndamba, Mbunga, and Pogoro are considered native to the valley and arrived in the early 19th century from Malawi. Other groups who migrated to the valley include the Sagara (central Tanzania), Hehe (Iringa), Ndedeule (Zambia), Sukuma (Mwanza), Ngoni (Southern Tanzania), Ngindo (Rufiji), and Chaga (Kilimanjaro) (ERM, 2012).

Within the floodplain, socio-economic drivers generate many productive activities, primarily for farming (Kato, 2007; Wilson et al., 2017). Important activities include agriculture and forestry, urbanization and transport, flood protection, hydropower production, navigation, and recreation, that all, but in different ways, add pressure to the floodplain ecosystem (Wilson et al., 2017). And in recent years, a rapid increase in agricultural land use has been observed (Jones et al., 2012). According to the 2007 Agriculture sample survey, most of the district's land in Ulanga and Kilombero was used for the temporary annual crop planted in monoculture with paddy and maize being the dominant ones. The valley contributed close to 70 percent of the regional planted area under paddy rice. Notably, livestock production has increased in the valley since 2006. The natives generally do not keep livestock, and most of the livestock are owned by either pastoralists or agropastoralists who migrated into the Valley (ERM, 2012).

1.3.2 Data sources and collection

Both primary and secondary data sources were used to answer research questions 2,3, and 4. The core data source is a detailed household survey in 21 villages in two districts of the Kilombero Valley, Ulanga and Kilombero. In total, 304 farm households were interviewed to provide information on the farming systems in terms of resource use and management as well as their relevance for the livelihoods of the households. The household selection was based on a multi-stage sampling strategy. First, 11 wards were purposively selected based on the occurrence of floodplain farming. In the second stage, 21 villages were randomly selected using probabilities proportional to size within the wards. In the final stage, households were randomly selected from the list provided by village leaders. A GIS approach incorporating the land use map from the Global land cover (GLC30) (Jun, Ban, & Li, 2014), the administrative boundaries, and the 2012 census data from the Tanzania statistics office (TNBS, 2013) was used to estimate the total population size in the study area. To capture the heterogeneity of the biophysical characteristics of the study area, geospatial data was collected from different sources and processed. The primary geospatial data are land use map for 2014 (Leemhuis et al., 2017), Digital Elevation Model (DEM) at 90m resolution (Jarvis, Hannes, & Andy, 2008) and proximity raster maps. The DEM was the basis for generating other raster layers, including the Topographic Wetness Index (TWI) and elevation. We obtained the administrative ward boundaries from the Kilombero district land and settlement office. Proximity raster (to road, market, and river) are based on the open-source database OpenStreetMap (OpenStreetMap Contributors, 2017) and pre-processed using QGIS (QGIS Development Team, 2020).

In addition to the two core data sources, the 2007/08 Agricultural Sample Census was used in part of our analysis. The data is a survey of smallholders and large scale farmers representative at the national level and conducted by the government of Tanzania through the National Bureau of Statistics (TNBS, 2009). For our analysis, a subset of 810 smallholder farmers was extracted for Kilombero and Ulanga district.

1.4 Structure and contributions

Each research question is answered in four self-contained chapters constituting the main chapters 2 to 5. This section summarizes the main chapters and contributions of the thesis. For each chapter, I provide the objective, approach, and main findings. Also, it will highlight how each chapter is related to the subsequent chapters.

1.4.1 Review of Agent-Based Models of land use and intensification

The second chapter of the thesis provides a systematic review of agent-based models of land use in agricultural systems. ABMs have become the standard to explore the dynamics within the farming system at multiple scales. Given their ability to capture heterogeneity, out-of-equilibrium dynamics, and a dynamic representation of the environment, the number of studies utilizing ABM has risen in the last two decades. ABMs have become an accepted method in many disciplines, and their disciplinary (or at least system-specific) applications have matured considerably. This chapter aims to provide a systematic review and comparison of ABMs that will allow us to identify the process and elements that are peculiar to different models and those that are essential to understanding the dynamics of agricultural land use. Moreover, the review explores how either explicitly or implicitly intensification decisions are included in the respective models.

Eight purposively selected models based on the criterion that the primary agent is a farmer or farm household, and the models are developed and applied in the time frame of the last two decades are included in the review. Besides, we filtered out models where either their source code is not available for inspection, or a comprehensive documentation is missing. MRPOTATOHEAD (Model Representing Potential Objects That appear in the Ontology of Human Action and Decision) (D. C. Parker, Brown, Polhill, Deadman, & Manson, 2008) guided our review. The framework categorizes model elements into six conceptually related dimensions: information and data, interface to other models, demographics, land-use decision, land exchange, and model operation.

Results show that models primarily vary in terms of Information and data requirement, the interaction between agents and their environment, how the biophysical process is modeled, land use decisions as well as the software applied. In general, the reviewed ABMs are developed for purposively selected case study areas and empirically grounded with data collected through both quantitative and qualitative approaches. The models reviewed are applied on different scales ranging from village level to regional or catchment scale, with the largest being MPMAS, which is used for an area of 3,779km^2 and 34,691 farm households in Ghana (Schreinemachers & Berger, 2011). Out of the eight models we reviewed, CATCHSCAPE (Becu, Perez, Walker, Barreteau, & Le Page, 2003) is a small-scale ABM applied for a catchment of 43.6km^{2}with 2,600 agents. ABMs are capable of representing a large number of agents and spatial extent (up to the national level) with the right computational resources and complexity (Parry & Bithell, 2012). Our ABM presented in chapter 5 of the thesis, WetABM, goes beyond the scales of the models reviewed. By leveraging on parallelization (when appropriate) and cloud computing, WetABM models an area of 5,200 km^2 with 38,000-51,000 farmer agents.

Interaction between agents is an integral part of ABMs. The reviewed models vary with respect to how they captured the interaction between agents and between agents and the environment. Interaction between agents is modeled through social networks, spatial neighborhoods, competition for common-pool resources, and input and output markets. On the other hand, interaction with the environment is modeled by integrating with biophysical models that capture the environmental change caused directly or indirectly by the agents' actions.

Depending on the objective, data availability, and level of complexity, modelers flexibly design the decision-making routines of the agents. The reviewed studies use either optimization approaches or rule-based heuristics. Optimization-based decision-making assumes rational decision-makers and is implemented using mathematical programming models or genetic algorithms (Schreinemachers & Berger, 2006). Rule-based heuristics simplifies the decision making by using empirically derived if-then rules. In recent years, machine learning algorithms are proposed to represent agent decision making and drive rules for agent behavior (DeAngelis & Diaz, 2019). Following Sun and Müller (2013) work, we use empirically trained and validated Bayesian belief networks within WetABM to drive probabilistic agent behavior. This would have been analytically and computationally difficult under other decision-making models, for instance mathematical programming (Britz & Wieck, 2014).

1.4.2 Characterizing farming system and farmers in Kilombero Valley Floodplain

A key characteristic of agriculture in sub-Saharan Africa is an enormous heterogeneity at all levels: countries, subnational regions, villages or communities, individuals (Dercon & Gollin, 2014).

Similarly, there are many different types of agricultural producers and farm households in KVF, which relates not only to differences in agro-ecology, market conditions, legal frameworks or institutional arrangements but also to the way they manage and allocate household resources (labor, land, fertilizers, machinery, technology, etc.) to agricultural production (Kato, 2007; Mombo, Speelman, Hella, & Van Huylenbroeck, 2013; Saravia Matus, Cimpoeis, & Ronzon, 2013). Understanding heterogeneity across farm types through typologies is considered as both a 'requirement' and a 'tool' in the analysis of farm-households capacity to increase output and yields in an environmentally sustainable manner while taking into account economically viable pathways (Bidogeza, Berentsen, De Graaff, & Oude Lansink, 2009; Gebauer, 1987; Saravia Matus et al., 2013).

Also, smallholder farmers in KVF are not a uniform group but vary in demographics, land use, market participation, resource endowment, and psychographic factors. By combining Principal Component Analysis, Kmeans, and agglomerative hierarchical clustering, this thesis in chapter 3 identifies three different farm types. The majority of the farmers in the valley are mono crop rice producers characterized by their higher levels of land allocation to rice, market participation, and labor use. The second farm type identified are Diversifiers. Households in this group are similar to the mono-crop producers but with a significantly larger share of land allocated to maize and vegetables in addition to rice. Moreover, the share of hired labor is relatively small because less land is allocated to high labor-intensive crops such as rice. The third group of farmers identified is Agropastoralists. Household in this group lives both from crop production and livestock keeping. Furthermore, they also own significantly more land and have a higher per capita income. The large farm size is attributed to their aggressive land clearing strategy in the bottom valley after migrating to the area (Bamford, Ferrol-Schulte, & Smith, 2010; Mwamfupe, 2015).

The typology was also validated using secondary data obtained from the Agriculture sample survey of Tanzania (TNBS, 2009). For reproducibility and detailed characteristics of farmers in KVF, this chapter is supplemented by an online appendix found at `https://bsrthyle.github.io/FarmTypolgyV5/`. The results from this chapter contribute to the scant literature on the characterization of farmers in the floodplain and, in principle, shed light on current agricultural practices and provide vital information needed by identifying farm development trajectories and target appropriate interventions per farm type. More so, the typology is used as a basis for building prototype farms and to parametrize agent-based model in the chapter 4 of the thesis.

1.4.3 Modeling intensification decision in Kilombero Valley Floodplain

Chapter 4 explores the intensification decision of smallholder farmers in the Kilombero valley. By taking four land saving intensification options practiced in the valley: (1) use of chemical fertilizers, (2) use of improved seed, (3) use of small-scale irrigation systems, and (4) increasing frequency of planting, the chapter sheds light on how farm households make their intensification decisions when multiple pathways are available and highlight the different factors driving these choices.

In addition to providing the first study on agricultural intensification decisions in KVF, this chapter offers a novel methodological contribution: The use of a Bayesian Belief Network (BBN) in combination with design of experiments, and multivariate regression trees the approach takes uncertainty into account and provides a white-box approach that can be updated by stakeholders when new data is available. This could not be achieved by the traditional one-dimensional statistical models.

Sensitivity analysis with BBN provides the main determinates of intensification decision, yet, it is limited when it comes to option-specific determinants. To overcome this limitation, we contribute to the literature on sensitivity analysis by applying a

Design of Experiment (DOE) and regression trees, which makes it possible to identify the relative importance of the determinants for each option under consideration.

Although the choice of each option is affected differently by the covariates under consideration crop choice, access to non-farm income, access to market, and topography of the plot play essential roles across all options. Choosing cropping multiple times is explained by variation in total labor available during the year, commercialization index, topographic wetness index, income, and distance to the central market. The variations in the probabilities of fertilizer application are also affected by the topographic wetness index if the farmer is diversifier, age, commercialization, and distance to the market. The use of improved seeds is influenced by the share of non-farm income, age, household size, distance to the market, and farm size. The probability of using irrigation and fertilizer application is affected by proximity to the market, farm size, the share of non-farm income, and the topographic wetness index. The variation in probabilities of use of irrigation is affected by variation in topographic wetness index, non-farm income, farm size if the farmer is of type subsistence, and availability of labor.

1.4.4 Immigration, access to infrastructure and intensification in Kilombero Valley Floodplain

The fifth and final chapter provides the development of an empirical agent-based model tailored to the socio-economic and biophysical characteristics of the KVF. Besides, the chapter provides the potential effect of migration and access to road infrastructure on the dynamics of intensification, land use, and income through simulation.

By combining the analysis made in the first three chapters of the thesis, this chapter leverages the information processed in those chapters. The agent-based model called WetABM is built using modular concepts that capture the different complex and dynamic systems at play in the valley. In relation to the various agent-based models reviewed in the first chapter, WetABM has some commonalities and differences. The main characteristics of WetABM is that it is empirically grounded and as well spatially and temporarily explicit.

One of the peculiar characteristics of the model is its large-scale ABM with almost 38 000-51,000 agents. Here, large-scale ABM is associated with high computational complexity rather than functional complexity. By leveraging on recent advances in computing (parallel and cloud computing), WetABM captures the full range of heterogeneity of farmers in Kilombero valley. Moreover, WetABM models the intensification decision of farmers using empirically trained and validated Bayesian belief networks similar to the work of Sun and Müller (2013).

Upscaling of sampled data to parametrize agent attributes within ABM is an important and challenging task (Smajgl & Barreteau, 2014). Our model provides multiple approaches (1) using Bayesian sampling, (2) sampling from empirical distributions fitted for each farm type identified in chapter two of this thesis.

There are multiple Interactions between the farmers through the endogenous output market, dynamically updating income aspiration and competition for land expansion. From a technical perspective, WetABM provides a modular design that will have an advantage for further extension and modification. For example, the land allocation submodule can be substituted by more complex routines without modifying the rest of the codebase. Or the scheduling of events can be extended to reactive, and the time steps can be modified from monthly to yearly with a slight modification. The details of WetABM are documented on supplementary ODD+D documentation at `https://bsrthyle.github.io/ODD-DforWetABM/`. The documentation provides a thorough model description and implementation.

In-migration to the valley and access to the market is considered one of the main drivers of land use and intensification. Yet no study captures the potential effect of these exogenous changes on the state of land use, intensification trends, and agricultural production. Within WetABM, the two exogenous changes are captured in a simplified manner. While immigration is considered an annual increase in the rate of immigration, road improvement is proxied by the reduction of transport costs to the market. The main finding from the baseline scenario shows that intensification (proxied by the number of farm households using one or more options) declines over the long run. However, rice and maize production will increase mainly due to the rise in land allocation to the two crops and land expansion. With continuous immigration into the valley resulting in increased population density and keeping the current level of protected area management effectiveness, farmers engage more in inland development than intensification. The crop area increases by 37 % compared to the baseline, and conversion is mainly concentrated in the Ulanga district, where the population density is lower than the Kilombero district. However, we also observe many farm households using improved seed variety, small-scale irrigation, and multiple crops per season compared to the baseline scenario. Our simulation result for reducing transportation cost as a surrogate for enhanced road infrastructure shows a negligible effect on intensification and agricultural production trends.

1.5 Limitation and outlook

The Kilombero valley floodplain is a complex environment both in terms of socio-economic activity and environmental dynamics. Although we tried to present a relevant socio-economic study for the area, the following limitations have to be considered, and further research might provide a more comprehensive insight into drivers and dynamics of agricultural production and smallholder livelihood in the valley. One of the main limitations of this study is that it did not capture all the biophysical elements and processes that are also important for understanding some of the dynamics in the valley. Especially, flooding patterns and associated risks will ultimately affect farmers' land use and intensification decisions. Besides, the intensification decision taken by the farmer will also have feedback on the quantity and quality of water resources in the valley. Moreover, our modeling approach doesn't take into account weather variation and future climate change scenarios. Future research is needed to consider more comprehensive feedback between the farmers' decisions and the biophysical environment. One possible approach is the ongoing work to integrate the WetABM with SWAT (Soil and water assessment tool) similar to the work of Daloğlu et al. (2014).

The second limitation is concerning the market dynamics for input and output. We have tried to capture the output market for rice and maize through endogenous price formation. However, the input market is entirely missing. One of the main difficulties in taking the input market is the informal market structure in the valley. Mobile input suppliers backed by global multinational companies and government involvement in the supply of inputs have made it complex enough to model it within WetABM and require further study.

1.6 References

Bamford, A., Ferrol-Schulte, D., & Smith, H. (2010). The status of the Ruipa corridor between the Selous Game Reserve and the Udzungwa Mountains. *Report to Frontier Tanzania, Dar es Salaam, Tanzania.*

Bassi, A. M., Casier, L., Pallaske, G., Perera, O., & Uzsoki, D. (2018). *The Sustainable Asset Valuation of the Southern Agricultural Growth Corridor of Tanzania (SAGCOT) Initiative : A focus on irrigation infrastructure Summary of Results* (Tech. Rep. No. December 2018). Retrieved from https://www.iisd.org/project/SAVi-sustainable-asset-valuation-toolhttps://www.iisd.org/sites/default/files/publications/savi-tanzania-sagcot-initiative.pdf

Becu, N., Perez, P., Walker, A., Barreteau, O., & Le Page, C. (2003). Agent based simulation of a small catchment water management in northern Thailand Description of the CATCHSCAPE model. *Ecological Modelling*, *170*(2-3), 319–331.

Bidogeza, J. C., Berentsen, P. B. M., De Graaff, J., & Oude Lansink, A. G. J. M. (2009). A typology of farm households for the Umutara Province in Rwanda. *Food Security*, *1*(3), 321–335. Retrieved from http://link.springer.com/10.1007/s12571-009-0029-8 doi: 10.1007/s12571-009-0029-8

Blache, A. (2019). Grabbing land, catching votes!. land and the 2015 election campaign in kilombero district, tanzania. *Les Cahiers d'Afrique de l'Est/The East African Review*(53).

BMBF. (2013). *National research strategy bioeconomy 2030 our route towards a biobased economy* (Tech. Rep.). Federal Ministry of Education and Research (BMBF). Retrieved 2020-01-15, from https://www.pflanzenforschung.de/application/files/4415/7355/9025/German_bioeconomy_Strategy_2030.pdf

Britz, W., & Wieck, C. (2014). Analyzing structural change in dairy farming based on an agent based model. *Technical Paper Institute for Food and Resource Economics University of Bonn.*

Cioffi-revilla, C. (2011). Geographic Information Systems (GIS) and Spatial Agent-Based Model (ABM) Simulations for Sustainable Development. , 1–22.

Daloğlu, I., Nassauer, J. I., Riolo, R. L., Scavia, D., Dalog ˘lu, I., Iverson Nassauer, J., ... Scavia, D. (2014). An integrated social and ecological modeling framework—impacts of agricultural conservation practices on water quality. *Ecology and Society*, *19*(3). doi: 10.5751/ES-06597-190312

DeAngelis, D. L., & Diaz, S. G. (2019). *Decision-making in agent-based modeling: A current review and future prospectus* (Vol. 6). Frontiers Media S.A. doi: 10.3389/fevo.2018.00237

Dercon, S., & Gollin, D. (2014). Agriculture in african development: theories and strategies. *Annu. Rev. Resour. Econ.*, *6*(1), 471–492.

Diao, X., Hazell, P., & Thurlow, J. (2010). The role of agriculture in African development. *World development*, *38*(10), 1375–1383.

Dinesen, L. (2016). Kilombero Valley Floodplain (Tanzania). In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The wetland book: Ii: Distribution, description and conservation* (pp. 1–8). Dordrecht: Springer Netherlands.

Entwisle, B., Malanson, G., Rindfuss, R. R., & Walsh, S. J. (2008). An agent-based model of household dynamics and land use change. *Journal of Land Use Science*, *3*(1), 73–93. doi: 10.1080/17474230802048193

ERM. (2012). Southern Agricultural Growth Corridor of Tanzania (SAGCOT): Strategic Regional Environmental and Social Assessment. RFP Number: a/082/hq/S/Sagt/pmu/7, ERM Project Number: 0159588. July 2012. (July), 144.

Gebauer, R. H. (1987, jan). Socio-economic classification of farm households —conceptual, methodical and empirical considerations. *European Review of Agricultural Economics*, *14*(3), 261–283. Retrieved from https://academic.oup.com/erae/article-lookup/doi/10.1093/erae/14.3.261 doi: 10.1093/erae/14.3.261

Government of Tanzania. (2016). *National Agriculture Policy* (Tech. Rep.). Ministry of Agriculture Food Security and Cooperatives. Retrieved from http://www.tzdpg.or.tz/fileadmin/documents/external/national_development_frameworks/ASDP2_Final_Document_20_May._2016__after_edit__1_.pdf

Jarvis, A., Hannes, R., & Andy, N. E. G. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Jones, T., Bamford, A. J., Ferrol-Schulte, D., Hieronimo, P., McWilliam, N., &

Rovero, F. (2012). Vanishing Wildlife Corridors and Options for Restoration: A Case Study from Tanzania. *Tropical Conservation Science*, *5*(4), 463–474.

Jun, C., Ban, Y., & Li, S. (2014). China: Open access to Earth land-cover map. *Nature*, *514*(7523), 434.

Kato, F. (2007). Development of a major rice cultivation area in the kilombero valley, tanzania. *African Study Monographs*(36), 3–18.

KILORWEMP. (2017). *Technical advisory services for biodiversity conservation and wetland management in Kilombero valley RAMSAR site, Tanzania:Pastoralism Diagnostic Study* (Tech. Rep.). Retrieved from https://kilomberovalley.files.wordpress.com/2019/02/kilorwemp-2017-scoping{_}imp{_}gmp.pdf

Leemhuis, C., Thonfeld, F., Näschen, K., Steinbach, S., Muro, J., Strauch, A., … Diekkrüger, B. (2017, aug). Sustainability in the Food-Water-Ecosystem Nexus: The Role of Land Use and Land Cover Change for Water Resources and Ecosystems in the Kilombero Wetland, Tanzania. *Sustainability*, *9*(9), 1513. Retrieved from http://www.mdpi.com/2071-1050/9/9/1513 doi: 10.3390/su9091513

Lyon, S. W., Koutsouris, A., Scheibler, F., Jarsjö, J., Mbanguka, R., Tumbo, M., … van der Velde, Y. (2015). Interpreting characteristic drainage timescale variability across Kilombero Valley, Tanzania. *Hydrological Processes*, *29*(8), 1912–1924. doi: 10.1002/hyp.10304

MALF. (2015). *Tanzania Livelihood Baseline Profile: Kilombero-Ulanga-Lusewa Paddy, Maize, and Cassava Livelihood Zone* (Zone Discription No. 43). Ministry of Agriculture Livestock and Fisheries: Agriculture Department.

Milder, J. C., Buck, L. E., & Hart, A. (2013). *Applying an Agriculture Green Growth Approach in the SAGCOT Clusters: Challenges and Opportunities in Kilombero, Ihemi and Mbarali* (Tech. Rep.). Dar es Salaam: SAGCOT Centre.

Milder, J. C., Buck, L. E., Hart, A. K., Scherr, S. J., & Shames, S. A. (2013). A framework for Agriculture Green Growth: Greenprint for the Southern Agricultural Growth Corridor of Tanzania (SAGCOT). *Dar es Salaam: SAGCOT Centre*.

Mombo, F., Speelman, S., Hella, J., & Van Huylenbroeck, G. (2013, nov). How characteristics of wetlands resource users and associated institutions influence the sustainable management of wetlands in Tanzania. *Land Use Policy*, *35*, 8–15. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264837713000732 doi: 10.1016/J.LANDUSEPOL.2013.04.010

Mombo, F., Speelman, S., Huylenbroeck, G. V., Hella, J., & Moe, S. (2011). Ratification of the Ramsar convention and sustainable wetlands management : Situation analysis of the Kilombero Valley wetlands in Tanzania. *Journal of Agricultural Extension and Rural Development*, *3*(9), 153–164.

Msofe, N. K., Sheng, L., Lyimo, J., Msofe, N. K., Sheng, L., & Lyimo, J. (2019, jan). Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability*, *11*(2), 505. Retrieved from http://www.mdpi.com/2071-1050/11/2/505 doi: 10.3390/su11020505

Munishi, P. K. T., Chuwa, J. J., Kilungu, H., Moe, S. R., & Temu, R. P. C. (2012). Management Effectiveness and Conservation Initiatives in the Kilombero Valley Flood Plains Ramsar Site, Tanzania. *Tanzania Journal of Forestry and Nature Conservation*, *81*(2), 1–10.

Mwamfupe, D. (2015). Persistence of farmer-herder conflicts in tanzania. *International Journal of Scientific and Research Publications*, *5*(2), 1–8.

Nindi, S. J., Maliti, H., Bakari, S., Kija, H., & Machoke, M. (2014). Conflicts Over Land and Water Resources in the Kilombero Valley Floodplain, Tanzania. *African Study Monographs*, *50*(October 2014), 173–190.

Nolan, J., Parker, D., Van Kooten, G. C., & Berger, T. (2009). An overview of computational modeling in agricultural and resource economics. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, *57*(4), 417–429.

Ntongani, W. A., Munishi, P. K. T., More, S. R., & Kashaigili, J. J. (2014). Local Knowledge on the Influence of Land Use/Cover Changes and Conservation Threats on Avian Community in the Kilombero Wetlands, Tanzania. *Open Journal of Ecology*, *4*(September), 723.

OpenStreetMap Contributors. (2017). *Planet dump retrieved from https://planet.osm.org* . https://www.openstreetmap.org.

Parker, D. (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.

Parker, D. C., Brown, D. G., Polhill, J. G., Deadman, P. J., & Manson, S. M. (2008). Illustrating a new 'conceptual design pattern'for agent-based models and land use via five case studies: the MR POTATOHEAD framework. *Agent-based Modeling in Natural Resource Management, eds. AL Paredes and CH Iglesias, Valladolid, Spain.*

Parry, H. R., & Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In *Agent-based models of geographical systems* (pp. 271–308). Springer.

QGIS Development Team. (2020). *QGIS Geographic Information System. Open Source Geospatial Foundation Project.* Retrieved from http://qgis.osgeo.org

RAMSAR. (2002). *Information Sheet on Ramsar Wetland: The Kilombero Valley Floodplain.* Retrieved from https://rsis.ramsar.org/RISapp/files/RISrep/TZ1173RIS.pdf

Rindfuss, R., Entwisle, B., Walsh, S., Li, A., Badenoch, N., Brown, D., ... Verburg, P. (2008). Land use change: Complexity and comparisons. *Journal of Land Use Science*, *3*(1), 1–10. doi: 10.1080/17474230802047955

SAGCOT. (2012). *The SAGCOT Greenprint: A Green Growth Investment Framework for the Southern Agricultural Growth Corridor* (Tech. Rep.).

Saravia Matus, S., Cimpoeis, D., & Ronzon, T. (2013). Literature Review and Proposal for an International Typology of Agricultural Holdings.

Schreinemachers, P., & Berger, T. (2006). Land use decisions in developing countries and their representation in multi-agent systems. *Journal of Land Use Science*, *1*(1), 29–44. doi: 10.1080/17474230600605202

Schreinemachers, P., & Berger, T. (2011). An agent-based simulation model of human-environment interactions in agricultural systems. *Environmental Modelling and Software*, *26*(7), 845–859.

Smajgl, A., & Barreteau, O. (2014). *Empirical Agent-Based Modelling - Challenges and Solutions - Volume 1, The Characterisation and Parameterisation of Empirical Agent-Based Models | Alexander Smajgl | Springer.* Springer-Verlag New York. doi: 10.1007/978-1-4614-6134-0

Sulle, E. (2020). Bureaucrats, investors and smallholders: contesting land rights and agro-commercialisation in the southern agricultural growth corridor of tanzania. *Journal of Eastern African Studies*, *14*(2), 332–353.

Sun, Z., & Müller, D. (2013). A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. *Environmental Modelling and Software*, *45*, 15–28. doi: 10.1016/j.envsoft.2012.06.007

TNBS. (2009). National Sample Census Of Agriculture 2007/08. , *V*, 1–345. Retrieved from https://www.nbs.go.tz/tnada/index.php/catalog/7

TNBS. (2013). *2012 population and Housing Census. Population Distribution by Administrative Areas* (Tech. Rep.). National Bureau of Statistics.

Wilson, E., McInnes, R., Mbaga, D. P., & Ouedaogo, P. (2017). Ramsar Advisory Mission report 83: United Republic of Tanzania. (1173), 85. Retrieved from https://www.ramsar.org/document/ramsar-advisory-mission-report-83-united-republic-of-tanzania-2016

Chapter 2

Review of selected Agent-Based Models of land use applied in an agricultural system

Abstract: The chapter presents a systemic review of 8 agent-based models to provide a comprehensive overview and formality of modeling practices and elements essential to understanding the dynamics of agricultural land use. Guided by a MRPOTATOHEAD framework, the review makes several inferences concerning the modeling components that are apparent in all the models, how they are represented, and their peculiarity to either a specific model or case study. Although almost all models reviewed tried to capture the underlying biophysical aspect of land use and farmers decision-making mechanism in an integrated manner, we observe the difference in terms of data,type of specific biophysical process modeled, the decision-making algorithm and programming toolkit used.

Keywords: *Agent-based model, land use, decision making, agriculture, farming, MRPOTATOHEAD*

2.1 Introduction

Different uses of land and anthropogenic transformation to meet the ever-increasing human needs for food and fiber have significant effects on the functioning of the ecosystem through changes in the carrying capacity and usability of the landscape at a local as well as at the global level. The globalization, and the further increase of population, lifestyle changes, progress in technology and infrastructure, and changes in industrial production patterns have further accelerated land transformation (Hubacek & Vazquez, 2002). The last century saw land-use change occurring at unprecedented speed and on a global scale. As argued by Duke and Wu (2014), the current surge in the rate of land conversion and changes are perhaps among the most prevalent socio-economic forces affecting ecological and economic systems and human well-being. Land-use change is a complex process that results from the interaction of decision making at different scales ranging from the individual farmer who decides on his plot to a global market that influences the farmer's decision making (Lambin, Geist, & Lepers, 2003; Verburg & Lesschen, 2006).

Given the challenges and complexity of balancing the competing needs for land, the study of land-use and land cover change is of extreme importance for both current and future discussions on climate change, biodiversity and food security (Lambin & Geist, 2008; D. Parker, 2003; Verburg, Schot, Dijst, & Veldkamp, 2004). Over the years, the land-use (LU) research community has made a substantial effort to provide improved measurement and modeling frameworks to enhance a solid understanding of both the theoretical and empirical foundation of land-use change (see Agarwal, Green, Grove, Evans, & Schweik, 2002; Brown, Walker, Manson, & Seto, 2004). However, the the complexity of causes, processes, and impacts of land change has demanded researchers to look for models capable of simulating the major socio-economic and biophysical driving forces in conjunction with interactions on several spatial and temporal scales (Lambin, Geist, & Rindfuss, 2006).

In recent years, agent-based models (ABM) have emerged as a useful tool for exploring the dynamics of land-use at multiple scales (Brown et al., 2004). ABM's can represent the behavior of human actors more realistically (accounting for bounded rationality, heterogeneity, interactions, evolutionary learning and out of equilibrium dynamics) and can combine this with a dynamic representation of the spatial environment

(Filatova, Verburg, Parker, & Stannard, 2013).

Originating from the fields of Complexity, Chaos, Cybernetics, Cellular Automata, and Computer science (Heath, Hill, & Ciarallo, 2009; Huigen & Fischer, 2003), ABM is a computational methodology for formalizing and analyzing complex social systems on many scales, ranging from small groups of individuals to organizations and larger systems (Cioffi-revilla, 2011).

When applied to the study of land-use change, ABM usually combines two general components: First, the cellular component represents the physical landscape in which agents are situated and interact. Second, an agent-based component representing the autonomous agents and their land-use decision-making routine (D. Parker, 2003; Schreinemachers & Berger, 2011). Agents interact with each other and their environment, resulting in emergent (bottom-up) macroscopic properties. Interactions can be direct, such as communication and physical interaction, or indirect via multiple-pathway feedbacks and from aggregate outcomes (Cioffi-revilla, 2011; Heckbert, Baynes, & Reeson, 2010).

By avoiding the top-down approach imposed by conventional mathematical or econometric models, ABMs have provided modelers with the flexibility to account for nonlinear dynamics and a non-global equilibrium or disequilibria of the system under consideration. However, this flexibility and advantage of ABMs also points to its limitations.With respect to the strand of empirical ABM, the flexibility will leads to ad hoc modeling practices with modelers trying to replicate reality as much as possible (Zimmermann, Heckelei, & Domínguez, 2009). Consequently, the diversity in modeling approaches makes the comparison of generated results over different ABMs of land-use (ABM-LU) change problematic as the underlying methodological basis has not been compared in the first place.

Although relatively young compared to other modeling paradigms of land-use change, recent efforts have contributed to the evolution of ABMs to ease the design and behavioral analysis as well as to make it handy for policy evaluation as the models increase in sophistication and ability (Matthews, Gilbert, Roach, Polhill, & Gotts, 2007). To explore the main modeling paradigms and best practices, researchers have made a series of reviews and comparisons of ABMs. For a general review of ABM applied for a land-use change please see (Matthews et al., 2007; D. Parker, 2003; Villamor et al., 2011), and for a comparison of different ABM of land use (see D. C. Parker, Entwisle, et al., 2008a; Polhill, Parker, Brown, & Grimm, 2008), for review of their application in urban residential choice (see Huang, Parker, Filatova, & Sun, 2014), an overview of computational models in agriculture and resource economics (see Nolan, Parker, Van Kooten, & Berger, 2009).

These prevalent comparisons of ABMs-LU covering diverse systems might also be owned to the fact that recently agent-based modeling itself has become an accepted method in many disciplines, and their disciplinary (or at least system-specific) applications have matured lately. Thus, we are now able to compile information on empirical ABM-LU , explicitly targeting the agricultural system.

Through reviewing and comparing eight agent-based models of land-use change in agriculture, we aim to contribute to this growing area of research through a more in-depth understanding of ABMs of land use by thematically dissecting those processes commonly implemented in all models, those that are customized to particular case studies, and that are essential for explaining the land-use change. This aims at building a knowledge base of agricultural land use ABMs and enabling a better comparison of results with those generated by traditional agricultural simulation models. Moreover, highlighting similarities and differences in modeling approaches can inform future modelers in terms of best practices and can thus contribute to an increased comparability of results generated by competing ABM-LU.

The work of Nolan et al. (2009) provides a broader overview by comparing different computational modeling approaches that are applied in agriculture systems (stochastic and dynamic programming, optimization, and ABMs). However, our work makes an explicit contribution by solely focusing on the review and comparison of empirical ABMs of LU in the agriculture system, which is only now becoming a feasible task due to the recent developments mentioned above.

The scope of our review is deliberately restricted to a particular subset of ABM: where the models are built for agrarian systems with farmers as the central decision-making units. Two main reasons are behind this choice of scope for our review. First, agricultural land-use change is one of the leading forces behind current land transformation and evolution we observe at a local or global scale. Given the need for meeting the growing demand for food, understanding land-use change in agricultural systems provides significant relevance. Second, as already mentioned above, most of the land use that is observed on the global scale is the result of interactions–of individual models of farmers' decisions at a local or micro level. Furthermore, farmers are the primary agents who decide (or not), given their knowledge, preferences, or interactions, to which particular use the land should be allocated. Besides, models are filtered using criteria that they have to be recent (likely to be used in some form in the 2000s), be empirical models (uses empirical data), and that the documentation is easily accessible. The review concentrates on the general semantics and architecture of the model. However, since most of these models are often applied to different case studies with or without significant extension, we will highlight the main additions and applications of the original model.

The remaining part of the chapters proceeds as follows: The second section begins by laying out existing frameworks for reviewing and comparing ABMs. The third section is devoted to the review of the models and breaking down their basic structure and semantics using the MRPOTATOHEAD framework. The fourth section presents our discussion focusing on the four key themes considered important in designing empirical ABMs. Finally, the last section concludes this chapter.

2.2 Framework for reviewing ABMs of land-use

In reviewing and comparing ABM of LU, it is prudent to have a framework to guide us in eliciting the specific elements of commonality and peculiarities among the models. One of the criticisms stated against ABMs is the lack of rigor and standards in critical modeling aspects such as model descriptions, calibration of parameters, and verification (Grimm et al., 2010; D. C. Parker, Brown, Polhill, Deadman, & Manson, 2008; Polhill et al., 2008). To this end, some scholars have suggested protocols, ontological structures, and standardization methods for developing models and comparing agent-based models.

The first framework used to compare ABM models is the ODD protocol (Overview, Design concepts, and Details) from ecological literature (Grimm et al., 2010). Although, as the name indicated, the ODD protocol is designed explicitly as a model documentation protocol to communicate the basic structure of the model, its scales, processes, schedule, and how it was designed, some researchers have used it to review and compare ABM models. For example, Polhill et al. (2008) used the ODD protocol to compare three ABM-LU (FEARLUS and ELMM, SOME, and SLUDGE). As authors acknowledged, the ODD protocol, which is designed for general model documentation, falls short in two dimensions as a framework to review ABM of land-use and land cover change. First, it's more concerned towards model code (how entities, processes, and schedules are implemented) rather with conceptual frameworks. Second, although the ODD protocol is sufficient to communicate the model structure, it misses a very significant element of ABM domain: the human decision environment (Grueau, 2013). However, to overcome the latter limitation, Müller et al. (2013) have extended the ODD protocol to include the human dimension aspect.

Another framework is MRPOTATOHEAD (Model Representing Potential Objects That appear in the Ontology of Human Action and Decision) (D. C. Parker, Brown, et al., 2008). The framework involves the creation of a standard design pattern at the conceptual level to enable the comparison of agent-based models of LU (Grueau, 2013; D. C. Parker, Brown, et al., 2008).

According to D. C. Parker, Brown, et al. (2008) MRPOTATOHEAD is a conceptual design framework which segments many of the principal elements generally used in ABM models into related themes which each model can optionally use and provides a universal medium for comparison and potential collaboration (D. C. Parker, Entwisle, et al., 2008a). In the first use of the framework, the authors compared five diverse ABM-LU models. Given the limitation of ODD for model comparison and our objective of unveiling commonalities and unique elements across models, we opted for MRPOTATOHEAD as a guiding framework for our review.

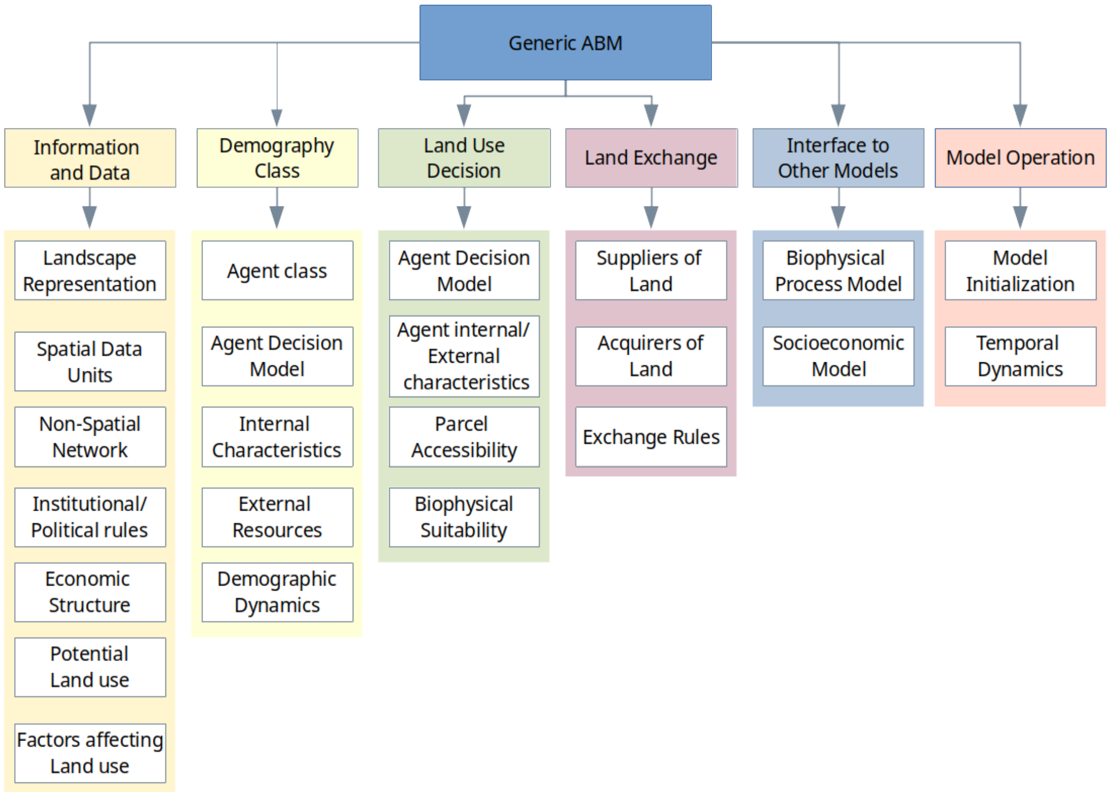


FIGURE 2.1: An illustration of MRPOTATOHEAD framework

(Model Representing Potential Objects That appear in the Ontology of Human Action and Decision) framework. Adapted from (D. C. Parker, Entwisle, et al., 2008a)

As shown in Figure 2.1, elements appearing in the model are categorized into six conceptually related dimensions: Information and Data, Demographics, Land-use decision, Land exchange, Interface to other models, and Model operation.

The ***Information and data class*** contain spatial information and data structures that are included in the model. How is the landscape represented? Is it an empirically calibrated model or a theoretically abstract model? What type of data structure is used for landscape representation (raster or vector data)? How is the parcel structured and its relation to the agent? What data layers, both spatial data layers and non-spatial network layers, are included? What kind of neighborhood effects are considered? Did the model include any institutional and political rules and economic rules and constructs?

The ***Demographic class*** covers the main characteristics and decision-making mechanisms of the agents in the model. Who is the central decision-making unit in the model? What kind of decision-making routine and strategy is included in the model? Which internal characteristics are relevant and considered in the agent profile? Which aspect of demographic dynamics (in and out-migration, aging, social structural change, reproduction, and life-cycle dynamics) are considered?

The ***Land-Use decision class*** is closely related to household decision-making class. In addition to the internal structure of the agent, biophysical constraints and suitability of specific land for the intended use are taken in to account.

The ***Land Exchange class*** describes how an agent's access to land is modeled. Is there any land market (either endogenous or exogenous market) or externally induced settlement or institutional structure that determines access to the land?

The ***Interface to other model class*** describes the coupling of the agent-based model with other biophysical or social dynamics model. Since ABMs are appealing for their ability to integrate the biophysical and human system processes, most ABM of land-use is coupled with one or more sub-models. How are the models linked (loose coupling or tight coupling)? What are the intra-feedback and inter feedback loop between them?

The ***Model operation class*** defines the model's initialization (before running the simulation) Is it based on empirical data or hypothetical data, and what are the respective data sources? At which specific time step is the model initialized. Additionally, this class also contains event scheduling and temporal dynamics of the model.

2.3 Systemic review of the selected models

the section summarizes the eight models by characterizing them through the six thematic classes discussed in the previous section. In no way are these models exhaustive of the available models; instead, they are representative of the different modeling practices that are followed by agent-based modelers. The eight models considered for review are LUCITA, MP-MAS, LUDAS, Velbuna, et al., PAMPAS, SAMBA-GIS, CHANOS, and CATCHSCAPE. As suggested in D. C. Parker, Entwisle, et al. (2008b), when the discussion of the class is closely related, then classes are combined. For example, land-use decision-making class and land exchange class can be combined under the same section.

2.3.1 LUCITA

LUCITA is an agent-based simulation of land-use change initially developed by Lim, Deadman, Moran, Brondizio, and McCracken (2002) to explore the effect of different socio-economic drivers on land-use patterns in a region of the Amazon rain-forest near Altamira, Brazil. The model has been modified and applied to different case studies in the same geographic area by (Cabrera, Deadman, & Brondizio, 2010; Deadman, Robinson, Moran, & Brondizio, 2004).

Information and data class: LUCITA operates on a spatially referenced abstract raster landscape. Three grids are representing the landscape: land cover, soil quality, and property parcels (are combinations of cells with an average size of 100 ha). The cells within the grids represent 1 ha spatial resolution and are geo-referenced with a common origin. The property parcels are fixed during the simulation run, and each farmer owns one property.

Interface to other models: The biophysical processes internally computed in the model. A process-based model governs the impacts of deforestation on soil properties, the relationship between soil fertility and successful crop yields, and the effect of soil properties on the rates of natural reforestation. Soil changes through clearing and burning practices, and soil-depletion and crop-yield prediction are determined by regression equations derived from another model called KPROG2 (Lim et al., 2002).

Demographic class: Each agent in the model represents a colonist family who migrates to the region over time. The agents are characterized by the composition of the family (age, sex, fertility, and mortality rate), available family and male labor pools, available liquid capital, and the land-use strategies that they are capable of implementing. The household demographic parameters are randomly assigned to the agents using normal distributions specific to their particular agent type.

Land use decision and land exchange: In a given year, households make decisions regarding clearing of land, the burning of deforested area, production of crops, and the harvest of those crops. These decisions are constrained by the amount of labor and capital endowment of the household. Besides, the ability of the farmer to clear and burn new land is determined by a set of clearing preferences that are exogenously set by the modeler. Once the household clears the patch of land, he will decide regarding which crop should be planted based on previous experiences. This decision-making process is governed by a classifier system. Each agent has eight distinct land-use strategies, or rules, which are represented by binary strings. Utilizing a genetic algorithm, these solution chromosomes are competing with one another for selection. Thus, households decide based on past performance by selecting the most productive land-use strategies. In the later versions of the model, the decision making was changed to a simple heuristic or rule-based decision tree (Cabrera et al., 2010; Deadman et al., 2004) to reduce the complexity attached to the genetic algorithm. The heuristics follow simple if-then steps, selecting crop choices depending on their subsistence requirement, soil quality, capital endowment, and available labor.

Model operation class: The original version of the LUCITA (Lim et al., 2002) was developed using the Swarm simulation system (Minar, Burkhart, Langton, Askenazi, et al., 1996). However, the later version (Cabrera et al., 2010; Deadman et al., 2004) is modeled utilizing the RePast Simulation framework (North et al., 2013). Each run is set up for 30 iterations, such that the first iteration represents 1971 and the beginning of colonization in the area. Households are allocated to 100-hectare plots on a grid representing an area of 15 km by 20 km. At each iteration, 50 households are assigned to the plots.

2.3.2 MPMAS

MPMAS (Mathematical Programming Based Multi-Agent System), developed by Berger (2001) is an agent-based model for simulating a land-use change in agriculture and forestry. According to the authors, the main feature of this model is its methodological embeddedness in the discipline of agricultural economics as it represents farmer decision making by whole-farm mathematical programming. MPMAS is a generic model with a modular architecture that can be easily configured to apply to various spatial scales, ranging from a village community to a large region within a country, depending on research interest and data availability. To this end, it has been applied in several empirical studies across different regions and scales. For instance, to model the impact of climate change on land-use and farm incomes in the Swabian

Jura, southwest Germany (Troost & Berger, 2014), adaptation to climate variability in Ghana (Wossen & Berger, 2015) and Ethiopia (Berger et al., 2015), the impact of hybrid maize varieties and better access to farm credit on poverty and environmental sustainability around Lake Victoria, Uganda (Schreinemachers & Berger, 2011).

Information and data class: MPMAS operates on a raster landscape represented by grid cells of each 0.5 ha. Other spatial information was organized as a layer storing the location of plots and farmstead and soil property. Farm households are allowed to own multiple cells (up to five).

Interface to other models: The MPMAS has been integrated with different biophysical models according to the research interest under investigation usually through either modules for water flows and soil nutrient changes or at runtime to an existing external process-based biophysical models using unique interfaces , such as MONICA (Latynskiy, Berger, & Troost, 2014), ExpertN (Troost & Berger, 2014), CROPWAT (Wossen & Berger, 2015), Tropical Soil Fertility calculator(TSPS) (Schreinemachers & Berger, 2011).

Demographic class: MPMAS captures the demographic dynamics of farm agents by using probabilities to update the fertility, mortality, marriage, and household composition.

Land-use decision class: Farm household is the central decision-making unit in the model. Heterogeneity among farm households was captured using the location of the agents' farmsteads, the location of their fields, the individual household composition (age, sex, and labor supply), and available resources such as cash, livestock, tree orchards, farm equipment, and specific agent characteristics (membership to a particular group).

The interaction between agents and their environment is captured by updating yield coefficients in a decision matrix. Agent's decisions affect the environment through crop choices, investments in infrastructure, and the use of chemical fertilizers. The feedback between the environment and the agents is captured via crop yield. Agents also interact with each other through either land or water markets or technology diffusion.

Agents make decisions on production, investment, and consumption level by choosing optimal land-use, and resource allocation given their resource endowment. The decision-making of individual farm households is modeled by repeatedly solving a constrained optimization problem wherein they maximize their expected net farm and non-farm income. For investment decisions, agents maximize their expected long-term average levels of net farm and non-farm income. Contrary, consumption decisions are based on optimal short-term levels.

Model operation class: The source code for MPMAS written in the object-oriented programming language C++. Modular extensions that can be used to link to other applications. Microsoft Excel is used to store input and output data in the earliest version. Later versions utilize the relational database format.

2.3.3 LUDAS

LUDAS is an agent-based land-use model developed at the Center for Development Research (Le, 2005). The model is a coupled human landscape system. The initial model was applied to an upland watershed in the Aluoi district of the central coast of Vietnam. Other versions of the model were applied by different authors to analyze different case studies. GH-LUDAS is one alteration exploring the impact of population growth, climate change and policy intervention on income and land-use in the Atkankwidi Catchment of Upper East Ghana (Schindler, 2009), SRL-LUDAS analyzed land management and vulnerability to natural hazards in Balapitiya and Maduganga, Sri Lanka (Kaplan, 2011) and LB-LUDAS simulated the temporal and spatial scale effects of payments for ecosystem services and their tradeoffs in the Jambi province of Indonesia (Villamor, Le, Djanibekov, van Noordwijk, & Vlek, 2014).

Information and data class: The landscape in LUDAS is represented by a GIS raster layer. Each cell has a 30m by 30m resolution, containing attributes and ecological response function or sub-models. The cells store biophysical spatial data (e.g., terrain condition, land cover, accessibility to rivers/streams) including economic, spatial variables (proximate distance to roads), and institutional (e.g., owner, village territory, protection zoning class) variables as well as the history of the cell properties.

Interface to other models: LUDAS does not directly link to any external biophysical model. However, an ecological response mechanism of the landscape is represented by internal sub-models of agricultural and forest productivity dynamics, and a cellular automate sub-model of land-cover transition.

Demographic class: A household profile is represented by an array of household socio-economic variables (e.g., educational status, household size, labor, land endowment, income, etc.) and variables measuring accessibility's of the household to specific policies. Household profiles are dynamic over time with different degrees. Annual income and land endowment change as a result of yearly household land-use action. Policy-related variables change in response to an exogenous change of policy variables. Some demographic variables, such as household age, advances regularly over time, but the household's ethnicity and size, are stable with small stochastic variance. The agents are grouped in different classes of a livelihood-based household typology.

Land-use decision class: The decision-making mechanism is implemented by a decision-making module. Using the household characteristics, their perceived landscape organization, and the characteristics of other agents. The decision program implements reflex and bounded-rational decision-making mechanisms when choosing either farm location or forest product collection, respectively. It assumes that household agents behave reactively according to production rules when deciding where to collect forest products. Moreover, they are assumed to likely select options returning optimal utility when looking for a location for cultivation. Although the decision mechanism is the same for all agents, decision outcomes are diverse as the agent's profile and event structure of utility functions are individual-specific.

LUDAS captures the interaction between the farmers and their environment through two different linkages: First, via a tenure relation that regulates access and usage of their respective land resources. Second, through a perception response loop that models both physical and information flow between the environment and the farmers. Farmers perceive the biophysical conditions and potential benefits around them. Through practicing land-use activities, the household agents modify the structure of the spatial organization in their environments.

The model also includes externally induced policy influences. The policy interventions affect the system through three different paths: 1) through policy-related variables of household agents, 2) through institutional variables of landscape agents, and 3) through directly modifying interaction rules. For example, protection zoning modifies spatial ownership and the perception of the farmers about the spatial organization.

Model operation class: LUDAS is developed in the Netlogo model framework (Tisue & Wilensky, 2004). The setup of the initial state of the system involves importing the sampled household data, up-scaling of households, and generating of managed land parcels. The model runs on an annual basis for 20-30 years.

2.3.4 PAMPAS

PAMPAS has been developed by Bert et al. (2011) to explore land-use and tenure as well as a structural change of agriculture in Argentina pampas which is one of the most fertile agricultural regions in the world.

Information and data class: The model operates on an abstract grid. Each grid cell represents a farm of variable size characterized by size, tenure regime, soil type, operator, land allocation, and aspiration level (the gross margin the farmer aspires to achieve from his farm within a specific year). The environment also contains topographical relations among farmers through Moore neighborhoods, which makes it spatially explicit.

Interface to other models: PAMPAS is loosely linked with an external crop growth model called Decision Support System for Agrotechnology Transfer (DSSAT) to simulate Physical yields for different crop choices. CERES and CROPGRO are used to simulate maize, wheat and soybean yield as a function of soil type, crop genetic character, and daily weather.

Demographic class: Farm households or family businesses are the main decision-making entities in the model. Entities are characterized by an attribute of total operated farms and area, operational status, working capital, and position in the social network. In the current version of the model, the dynamics of the life cycle of a specific individual is not included. Instead, agents exit the farming in case of illiquidity.

Land-use decision class: Farmers decide on their land allocation intending to maintain or increase their working capital or to expand the cropped area. They adjust their land allocation by comparing their economic achievement with dynamically changing aspiration thresholds. They adjust their aspiration level (AL) every production cycle depending on their previous satisfaction, on the expected weather condition, as well as output and input prices. Once they update their AL, they will decide to either increase their farmland area, to maintain the same land area, to reduce or even to quit farming. Farmers acquire new land only through renting it from less satisfied farmers. The land rental market is based on exogenous prices rather than on endogenously evolving land prices.

Model operation class: PAMPAS is implemented in REPAST (The Recursive Porous Agent Simulation Toolkit) software framework (North et al., 2013). Before each run of the simulation, the model is initialized for farmers and their farm from a relational database and randomly allocates them to the grid. Simulation loops represent one cropping season (one year) and run for 100 iterations with 1900 being the starting year.

2.3.5 SAMBA-GIS

SAMBA GIS is a participatory model developed by Castella, Trung, and Boissau (2005) that combines role-playing games, agent-based modeling, and geographic information systems. The model was applied to different villages in the mountainous "Bac Kan" province in Vietnam. The land-use pattern at the district level emerges from land users interacting at the village level. A typology of villages was created to capture their diversity and joint trajectories of land-use drivers. Villages vary based on their accessibility of livelihood options and availability of other sources of income and specific development projects.

Information and data class: The landscape in SAMBA-GIS is represented as a grid that can take either an abstract grid or the real landscape through GIS layers of the villages. A land cover map of 1990 and soil data are the two layers that are included to initialize the model. Also, each cell of the raster map stores parameters for yields of cultivated crops, remoteness from a residential area, and suitability of the cell for paddy, maize, and upland crops. A single cell represents 1km^{2}equivalent of the land on the ground.

Demographic and land-use decision class: Small scale subsistence farm households are the central agents in the model. They are characterized by each household members' labor force availability, land-use choices, available capital, and ownership of livestock. Farmer's decision to allocate their production factors to different land-use options is modeled as heuristic behavior. The rules were constructed based on a participatory role-playing game and a literature survey from the same case study areas. For instance, the primary assumption of the decision-making process is that farmers make decisions about their production activity based on their capacity to secure a minimum rice requirement for their household. It is only after self-sufficiency is achieved that farmers pursue other livelihood options. The other decision hypothesis is that farmers' decision making is entirely based on their access to rice land, not significantly determined by cultural values and ethnic background. The decision-making routine is universal across households, and heterogeneity emerges as a result of their profile and access to rice land. Besides, agricultural production and productivity vary with the soil and topography conditions along with socio-economic characteristics. Concerning the environment, the cells in the model are transformed as a result of farmers cropping choices, natural growth dynamics of vegetation, and livestock grazing.

Model operation class: SAMBA-GIS is implemented in CORMAS (Common pool Resources and Multi-Agent Systems) modeling platform (Bousquet, Bakam, Proton, & Le Page, 1998). The model was initialized with secondary data and survey on population, ethnicity, the number of buffaloes, presence of projects (e.g., reforestation or development projects), and GIS layers for land use, soils, and accessibility. The model runs on yearly bases for ten steps (from 1990-2001).

2.3.6 CHANOS

CHANOS is an agroecosystem agent-based model developed by Mialhe, Becu, and Gunnell (2012) to explore the decision-making process of investor agents and farmer's cropping choices and the resulting land-use pattern. The model was applied to the Pampanga delta in the Philippines, an area dominated by rice and aquaculture.

Information and data class: CHANOS operates on an abstract raster landscape with two GIS layers: one for land-use in 1970 and the other for elevation. Each cell takes the three categories of water bodies, agricultural land, and natural habitat. The primary function of the water body cells is to transport dissolved salt to other cells, whereas natural habitat cells store the land cover type and suitability to aquaculture conversion. The agricultural land represents the farms, each supporting a unique cropping system (rice, rotation of rice and aquaculture, or aquaculture).

Interface to other models: CHANOS is not linked to an external sub-model. However, it contains two internal sub models: environmental sub-models which simulates the annual subsidence of plots by allocating farm salinity as a function of the farm's distance from the nearest saltwater and investment sub-model taking into account investor's decision process to convert land to farming under the constraint of no fragmented farm and investor's capacity.

Demographic and land-use decision class: Farmers and investors are the two central decision-making units. Investors characterized by their willingness and capacity to invest have an objective to acquire new land when return to their investment is positive. A farmer is the other agent who owns a single farm with the primary objective to attain satisfaction and certainty. The diversity of farmers is captured based on their attribute and their class of behavior. The attributes include a cropping system (rice, seasonal rotation of rice and aquaculture, aquaculture), income, spending satisfaction, uncertainty, and the number of people they know (a proxy for social capital). Based on objectives and cognitive strategies, three different agent behaviors were imposed in the model: rational, collective mind, and bounded rational. Rational agents optimize either their immediate profit (short term strategy) or gain stability in profit (medium-term strategy). In addition to optimizing immediate profit, Collective mind agents also have an objective to imitate the same cropping system as their network peers and neighbors and obey the government guidelines. The bounded rational agents have three more objectives to the already mentions ones (maximize profit, gain stability of profit, imitation, obey government guidelines), they choose the simplest of rice or aquaculture cropping systems, supply staple foods and secure a stable income. Production levels for each cropping system were estimated using a linear regression (production level of rice and aquaculture as dependent variable and salinity levels as independent variables).

Model operation class: CHANOS is built on Netlogo ABM framework. The environment was initiated with land use, elevation, and topographic map from 1970. The farms were initiated based on a survey through an iterative clustering method. For the current application, the model runs for 40 years starting from 1970, each run representing one season (the equivalent of six months).

2.3.7 CATCHSCAPE

CATCHSCAPE has been developed by Becu, Perez, Walker, Barreteau, and Le Page (2003) to explore the impact of upstream water management on the downstream farming viability under different irrigation management options in Mae Uam catchment in northern Thailand.

Information and Data class: CATCHSAPE operates on an abstract grid to represent the whole catchment of 43.6km^2. Different layers at a different level of organization are included to systematically represent the catchment as realistic as possible. A combination of information on soil texture, soil depth, and the slope was represented as land units, while Land Use classified as paddy, Upland, and the forest is also incorporated as a layer.

Interface to other models: The biophysical dynamics are internally simulated using a water balance model that provides runoff and water storage as a function of rainfall and irrigation, a hydraulic model to control water flow and distribution into canals and plots using river network node.

Demographic and land-use class: Farmer crop choice decision is based on constrained optimization, where farmers maximize their profit given their constraint in cash, labor, and water availability. However, the decision of cultivating paddy is based on heuristics, as the authors argued; rice cultivation is mostly motivated by social-cultural preference than profit maximization. During the rainy season, farmers decide on the amount of paddy to produce up to attain their minimum household consumption, given the availability of cash and labor. During the dry season, they also decide to allocate part of their labor to an off-farm activity or not, as rice is mostly cropped during the rainy season. In terms of land-use dynamics, farmers also have to decide on either to buy a new plot, install irrigation in the rain-fed land or to convert forest plots into upland plot depending on investment cost and local policy control.

Model operation class: CATCHSCAPE is implemented in the CORMAS platform (Bousquet et al., 1998). Initially, the model runs over ten years, and the scenarios have been repeated 20 times to estimate the variability of the results. At each time step, successive phases of parameter update, cropping decision, farming activity, biophysical dynamics, crop harvesting, irrigation planning, and land dynamics are implemented.

2.3.8 Velbuna et al.

Valbuena, Verburg, Veldkamp, Bregt, and Ligtenberg (2010) developed a conceptual agent-based model to explore how the diversity of farmer's decision making affects the structure of the landscape at a regional level. The conceptual model is then applied to a case study in Netherland and Australia.

Information and data class: The model operated on a grid landscape where parcels (the combinations of one or more cells) are the decision-making units in the model. Each cell in the grid representing 1 ha has an attribute of field number, field size, owner, land-use type, production per hectare, distance to the residence, and suitability for agriculture. While the land-use type and ownership are derived from a real land-use map and cadastral map respectively, the suitability of the land to agriculture is estimated based on a logistic regression between the current use of land for agriculture (dependent variable) and location conditions (soil characteristics and land consolidation process).

Interface to other models: Although the model captures the biophysical process through the estimation of land suitability internally, there is no other external model that is explicitly linked to it.

Demographic and land-use class: The conceptual framework of the model outlines that farmer's decision making is affected by both; internal and external factors. Internal factors are those aspects that are related to a farmer's willingness and ability to act. Ability refers to conditioning factors and options farmers have at some specific point in time. Willingness relates to intentions, values, or preferences for choosing individual options. Farmer's decisions are also affected by factors external to the farm. The factors can be either compulsory or/and voluntary (government policies, credit, demand from outside the system).

In order to simplify farmer's diversity, an agent typology was created based on their willingness and ability. Five different agent groups were identified in the region: hobby, conventional, diversifier, expansionist-conventional, and expansionist diversifier.

Land use decision is based on a probabilistic decision-making process (probabilitly matrix). The decision processes which can be either discrete or continuous are similar across farmers but different between different agent types through distinctive probability assignments. For example, the farm expansion decision is divided into three different options: buy, keep, or sell land. Given these options under the three decision-making processes, a probability is assigned to each option to differentiate between the agent types.

Model operation class: The model is implemented using NetLogo (Tisue & Wilensky, 2004). Parameterization of agents and their attributes was based on survey and census data as well as cadastral and land cover maps. The temporal extent considered was 15 years, where each run represented a single year.

2.4 Discussion

ABMs, like any other model, should be an abstraction of reality that is complex enough to emulate the essential features of the real-world phenomena, yet simple enough to be tractable (North et al., 2013; D. Parker, 2003). Based on the reviewed models several inferences can be made concerning the classes that are apparent in all the models, how they are represented, and their peculiarity to either a specific model or case study. Although almost all models reviewed tried to capture the underlying biophysical process and decision-making mechanism in an integrated manner, we observe the difference in terms of, information and data requirement, type of specific biophysical process modeled, the decision-making algorithm and programming toolkit used.

2.4.1 Information and data requirement

Due to their complex structure and micro-level processes, empirical grounding of agent-based model can be data intensive (Smajgl & Barreteau, 2014).

Data for parameterization of the agents and the environment comes from different sources, including surveys, census, role playing games, focus group discussion and geographic information systems. Looking at the landscape representation, either an abstract or empirical raster grid emerges as the most prominent to landscape representation. Modelers also include different layers of the grid to represent the environment. Land use and soil grid are the commonly used data layers that provide not only the information for the agent decision making but also serve as a platform for the biophysical process modeling. The dominance of raster representation compared to vector data is attributed to computational efficiency, availability of processed and ready to use raster layers and their easy integration to widely used software frameworks (e.g REPASTJ).

Models reviewed here are applied on different scales ranging from village level to regional or catchment scale. with the most large scale being MPMAS has the largest (spatial extent and number of agents) application, which is applied for an area of 3,779km^2 and 34,691 farm households in Ghana. As the scale of the model increase, modelers have different options to upscale data from sample to population including monte-carlo simulation, typology or census. In addition to availability of data and computational power,the scale of application is, of course, dependent on the objective of the study and heterogeneity of the agents and the environment.

2.4.2 Modeling the biophysical process

The capability for capturing the biophysical process depends on how insightful and relevant the key landscape processes are modeled in a spatiotemporally explicit manner (Villamor et al., 2011). All models under review have represented the landscape and model some biophysical processes. However, integration can be either including the biophysical model internally (tight coupling) (LUDAS, Velbunea, et al.; CHANOS) or the human subsystem can be linked to an external biophysical model through information sharing(loose coupling)(MPMAS, PAMPAS). The biophysical process models (hydrological process, the land cover change process, subsidence, or soil nutrient dynamics) are specific to the objective of each reviewed models. Process based crop growth models that simulate potential yields given different determining factors are the standard external models that are included in the models reviewed above.

2.4.3 Land use decision making

Since all the models reviewed are purposively are selected based on the criteria of developed for the agricultural system, farmers are the central decision-making units. The farmers can be either individual farmers or farm households that make the day to day decisions of their farm management. A Range of characteristics and endowments are used to capture the heterogeneity of farmers. By creating a typology, it is possible to reduce the complexity of a farmer's decision making, mainly when the model is applied at a regional scale with limited farm-level data (LUDAS and Valbuena). Typologies reduce the diversity of farmers and farming strategies through artificially grouping farmers using some specific criteria.

An important finding of our review is that models are relatively limited in modeling land exchange between farm households except for MPMAS. As also noted by D. C. Parker, Brown, et al. (2008); D. C. Parker, Entwisle, et al. (2008b), existing ABMs do not include an endogenous land market. However, this does not mean that models must always have a land market. For instance, in most developing countries, the land is exchanged under imperfect market condition, and access to land can be based on different institutional settings that are not congruent with dynamics on open markets.

The decision-making mechanism of the agent can be categorized into two broader groups; optimization and heuristics. Modelers implemented optimization through, mathematical programming (MPMAS and CATCHSCAPE) or genetic algorithm (LUCITA). Heuristic decision making is implemented by simple if-then rules to mimic the decision-making procedure of the agents. Both approaches have their advantage and disadvantage. As Schreinemachers and Berger (2011) reasoned, although intuitive, straightforward and transparent, heuristic modelers face the problem of lacking information regarding alternatives and the difficulty of coping with large numbers of rules. On the other hand, optimization has the advantage of representing heterogeneity, and a large number of decisions can be incorporated. However, optimizations are criticized for assuming a fully rational decision maker implicitly and for the black box character of the algorithms (Cabrera et al., 2010; Deadman et al., 2004). Interestingly apparent is limited application of statistical methods to drive empirically based agent behavior.

2.4.4 Modeling framework and software

Different simulation toolkits and software are used to develop agent-based models. Modelers are flexible in choosing between these toolkits and frameworks, depending on their capability and functionality they provide. Most of the available tools are based on object-oriented programming paradigm. The object-oriented paradigm provides a very suitable medium for the development of agent-based models. In particular, it provides modularity useful for developing a virtual simulation laboratory. The models under review in this chapter are implemented on different programming toolkits (Netlogo, Repast, CORMAS). Framework choices are influenced by the complexity of the model and programming language experience of the modeler. For a complete overview and comparison of software, toolkit refer to (Kravari & Bassiliades, 2015).

2.5 Conclusion

ABM is an exciting addition to existing models aiming to understand the dynamics of agricultural land use through a bottom-up approach. Our review highlights the flexibility ABM provides modelers to incorporate different processes and elements into their model depending on their objective, data availability and essential attributes of their study area. However, caution is advised not to let the flexibility lead to ad-hoc modeling practices not supported neither by theory or literature. Our review has also provided a comprehensive overview of modeling techniques and essential elements. A model aimed at understanding the dynamics of land use and change is expected to have specific peculiar properties:- spatial explicit, heterogeneous agents, sensing and prediction, decision-making mechanism, and complex interaction. All the eight models we reviewed take a spatial distribution of the landscape feature either through georeferenced raster layer or abstract grid.

However, all most all models implement random land allocation (plots to farmers) mechanism. With the increasing availability of spatial data and cadastral maps, future models can provide a more realistic tenure relation between agents' resource endowment and the landscape. Interactions of relatively moderate complexity already occur in some of the models reviewed above (MPMAS, LUDAS, and PAMPUS), in which social learning trigger actions, but an extension to modes of social interactions derived by empirical insights (e.g., social media analysis) or motivated by theory (e.g., specific network typologies) are still possible using ABM. Given the potential offered by ABM, increased computing power, and availability of scalable software frameworks and a new wealth of empirical data regarding eco- and social systems, future studies may contribute by expanding model elements including empirically or theoretically motivated interaction typologies, trade-offs between farming and non-farming activities, disaggregated markets for output and input, endogenous price formation and even identifying causal relationships. All before mentioned elements are scant in the current state of ABMs of land-use change as found by this review.

2.6 References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. *Apollo The International Magazine Of Art And Antiques*, 62.

Becu, N., Perez, P., Walker, A., Barreteau, O., & Le Page, C. (2003). Agent based simulation of a small catchment water management in northern Thailand Description of the CATCHSCAPE model. *Ecological Modelling*, *170*(2-3), 319–331.

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., Wossen, T., Troost, C., Latynskiy, E., Tesfaye, K., & Gbegbelegbe, S. (2015). Adaptation of farm-households to increasing climate variability in {Ethiopia}: {Bioeconomic} modeling of innovation diffusion and policy interventions. In *Agriculture in an intercinnected world. {Milano}*.

Bert, F. E., Podestá, G. P., Rovere, S. L., Menéndez, Á. N., North, M., Tatara, E., ... Toranzo, F. R. (2011). An agent based model to simulate structural and land use changes in agricultural systems of the argentine pampas. *Ecological Modelling*, *222*(19), 3486–3499.

Bousquet, F., Bakam, I., Proton, H., & Le Page, C. (1998). Cormas: common-pool resources and multi-agent systems. In *International conference on industrial, engineering and other applications of applied intelligent systems* (pp. 826–837).

Brown, D. G., Walker, R., Manson, S., & Seto, K. (2004). Modeling Land-Use and Land-Cover Change. *Land Change Science, Political Ecology, and Sustainability: Synergies and Divergences*, 1–17.

Cabrera, a. R., Deadman, P. J., & Brondizio, E. S. (2010). Exploring the Choice of Decision Making Method in an Agent Based Model of Land Use Change. *International Environmental Modelling and Software Society*. Retrieved from http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Castella, J. C., Trung, T. N., & Boissau, S. (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).

Cioffi-revilla, C. (2011). Geographic Information Systems (GIS) and Spatial Agent-Based Model (ABM) Simulations for Sustainable Development. , 1–22.

Deadman, P., Robinson, D., Moran, E., & Brondizio, E. (2004). Colonist household decisionmaking and land-use change in the Amazon Rainforest: An agent-based simulation. *Environment and Planning B: Planning and Design*, *31*(5), 693–709.

Duke, J. M., & Wu, J. (2014). IntroductionLand as an Integrating Theme in Economics. In *Handbook of Land Economics.* Oxford University Press. Retrieved from //www.oxfordhandbooks.com/10.1093/oxfordhb/9780199763740.001.0001/oxfordhb-9780199763740-e-022

Filatova, T., Verburg, P. H., Parker, D. C., & Stannard, C. A. (2013). Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental Modelling and Software*, *45*, 1–7.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, *221*(23), 2760–2768.

Grueau, C. (2013). Towards a domain specific modeling language for agent-based modeling of land use/cover change. In *International Conference on Conceptual Modeling* (pp. 267–276).

Heath, B., Hill, R., & Ciarallo, F. (2009). A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, *12*(4), 9.

Heckbert, S., Baynes, T., & Reeson, A. (2010). Agent-based modeling in ecological economics. *Ecological Economics Reviews*, *1185*, 39–53.

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, *41*(4), 661–689.

Hubacek, K., & Vazquez, J. (2002, March). *The economics of land use change* (IIASA Interim Report). IIASA, Laxenburg, Austria. Retrieved from http://pure.iiasa.ac.at/id/eprint/6770/

Huigen, M., & Fischer, G. (2003). Agent based modelling in land use and land cover change studies. *Interim Report, Laxenburg, IIASA*(September 2003).

Kaplan, M. (2011). Agent-based modeling of land-use changes and vulnerability assessment in a coupled socio-ecological system in the coastal zone of Sri Lanka. *Ecology and Development series*, *77*, 228.

Kravari, K., & Bassiliades, N. (2015). A survey of agent platforms. *Journal of Artificial Societies and Social Simulation*, *18*(1), 11.

Lambin, E. F., Geist, H., & Rindfuss, R. R. (2006). Introduction: local processes with global impacts. In *Land-use and land-cover change* (pp. 1–8). Springer.

Lambin, E. F., & Geist, H. J. (2008). *Land-use and land-cover change: local processes and global impacts.* Springer Science & Business Media.

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. *Annual review of environment and resources*, *28*(1), 205–241.

Latynskiy, E., Berger, T., & Troost, C. (2014). Assessment of P Olicies for Low-Carbon Agriculture by Means of Multi-Agent Simulation. In *7th International Congress on Environmental Modelling and Software. San Diego. CA (USA): Ames, DP, Quinn, NWT, Rizzoli, AE (Eds.).*

Le, Q. B. (2005). Multi-Agent System for Simulation of Land-use and Land Cover Change. *A Theoretical Framework and Its First Implementation for an Upland Watershed in the Central Coast of Vietnam. Ecology and Development Series*, *29*(291.).

Lim, K., Deadman, P., Moran, E., Brondizio, E., & McCracken, S. (2002). *Agent-Based Simulations of Household Decision Making and Land Use Change near Altamira, Brazil.*

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology*, *22*(10), 1447–1459. doi: 10.1007/s10980-007-9135-1

Mialhe, F., Becu, N., & Gunnell, Y. (2012). An agent-based model for analyzing land use dynamics in response to farmer behaviour and environmental change in the Pampanga delta (Philippines). *Agriculture, Ecosystems and Environment*, *161*, 55–69.

Minar, N., Burkhart, R., Langton, C., Askenazi, M., et al. (1996). The swarm simulation system: A toolkit for building multi-agent simulations.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., ... Schwarz, N. (2013). Describing human decisions in agent-based models–ODD+D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48.

Nolan, J., Parker, D., Van Kooten, G. C., & Berger, T. (2009). An overview of computational modeling in agricultural and resource economics. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, *57*(4), 417–429.

North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M., & Sydelko, P. (2013). Complex adaptive systems modeling with repast simphony. *Complex adaptive systems modeling*, *1*(1), 3.

Parker, D. (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.

Parker, D. C., Brown, D. G., Polhill, J. G., Deadman, P. J., & Manson, S. M. (2008). Illustrating a new 'conceptual design pattern'for agent-based models and land use via five case studies: the MR POTATOHEAD framework. *Agent-based Modeling in Natural Resource Management, eds. AL Paredes and CH Iglesias, Valladolid, Spain*.

Parker, D. C., Entwisle, B., Rindfuss, R. R., Vanwey, L. K., Manson, S. M., Moran, E., … others (2008a). Case studies, cross-site comparisons, and the challenge of generalization: comparing agent-based models of land-use change in frontier regions. *Journal of Land Use Science*, *3*(1), 41–72.

Parker, D. C., Entwisle, B., Rindfuss, R. R., Vanwey, L. K., Manson, S. M., Moran, E., … Malanson, G. (2008b, jan). Case studies, cross-site comparisons, and the challenge of generalization: comparing agent-based models of land-use change in frontier regions. *Journal of land use science*, *3*(1), 41–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19960107http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2786205 doi: 10.1080/17474230802048151

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

Schindler, J. (2009). A multi-agent system for simulating land use and land cover change in the Atankwidi catchment of Upper East Ghana.

Schreinemachers, P., & Berger, T. (2011). An agent-based simulation model of human-environment interactions in agricultural systems. *Environmental Modelling and Software*, *26*(7), 845–859.

Smajgl, A., & Barreteau, O. (2014). *Empirical Agent-Based Modelling - Challenges and Solutions - Volume 1, The Characterisation and Parameterisation of Empirical Agent-Based Models | Alexander Smajgl | Springer*. Springer-Verlag New York. doi: 10.1007/978-1-4614-6134-0

Tisue, S., & Wilensky, U. (2004). Netlogo: A simple environment for modeling complexity. In *International conference on complex systems* (Vol. 21, pp. 16–21).

Troost, C., & Berger, T. (2014). Dealing with uncertainty in agent-based simulation: Farm-level modeling of adaptation to climate change in southwest Germany. *American Journal of Agricultural Economics*, *97*(3), 833–854.

Valbuena, D., Verburg, P. H., Veldkamp, A., Bregt, A. K., & Ligtenberg, A. (2010). Effects of farmers' decisions on the landscape structure of a Dutch rural region: An agent-based approach. *Landscape and Urban Planning*, *97*(2), 98–110.

Verburg, P. H., & Lesschen, J. P. (2006). Agriculture, climate and future land use patterns: potential for a simulation-based exploration. In *Agriculture and climate beyond 2015* (pp. 5–32). Springer.

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

Villamor, G. B., Le, Q. B., Djanibekov, U., van Noordwijk, M., & Vlek, P. L. G. (2014). Biodiversity in rubber agroforests, carbon emissions, and rural livelihoods: An agent-based model of land-use dynamics in lowland Sumatra. *Environmental Modelling and Software*, *61*, 151–165.

Villamor, G. B., van Noordwijk, M., Le, Q. B., Lusiana, B., Matthews, R., & Vlek, P. L. G. (2011). Diversity deficits in modelled landscape mosaics. *Ecological Informatics*, *6*(1), 73–82. doi: 10.1016/j.ecoinf.2010.08.003

Wossen, T., & Berger, T. (2015). Climate variability, food security and poverty: Agent-based assessment of policy options for farm households in Northern Ghana. *Environmental Science and Policy*, *47*, 95–107.

Zimmermann, A., Heckelei, T., & Domínguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants.

Environmental science & policy, *12*(5), 601–618.

Chapter 3

Characterizing farmers and farming system in Kilombero Valley Floodplain, Tanzania[†]

Abstract: Recognizing the diversity of farmers is crucial for the success of agricultural, rural, or environmental programs and policies aimed at sustainable use of natural resources. In this study, based on survey data collected in the Kilombero Valley Floodplain (KVF) in Tanzania, we design a typology of farmers to describe the range of farm types and farming systems systematically and to understand their livelihood and land use behavior. The KVF is the largest, low-altitude, seasonally-flooded, freshwater wetland in East Africa. Despite its values, KVF is a very fragile ecosystem threatened by current and future human interventions. We apply multivariate statistical analysis (a combination of Principal Component Analysis and Cluster analysis) to identify farm groups that are homogenous within and heterogeneous between groups. Three farm types were identified: "Monocrop rice producer", "Diversifier", and "Agropastoralist". Monocrop rice producers are the dominant farm types accounting for 65 percent of the farm households in the valley, characterized by more than 80 percent of the land allocated to rice high and showing strong market participation and high utilization of labor. Diversifiers, on the other hand, allocate more land to maize, and vegetables. Agropastoralists account for 7 percent of the surveyed farmers and differ from the other two groups by on average larger land ownership, a combination of livestock and crop production, and larger household sizes. This typology represents the diversity of farmers in KVF concerning their land use and livelihood strategy and will allow to target policy interventions. Besides, it may also inform further research about the diverse landscape of floodplain farming through the classification and interpretation of different socio-economic positions of farm households.

Keywords: *Kilombero valley, Tanzania, farmer typology, principal component analysis,hierarchical clustering,farmer diversity*

[†]This chapter is published in journal of Sustainability as Gebrekidan BH, Heckelei T, Rasch S. Characterizing Farmers and Farming System in Kilombero Valley Floodplain, Tanzania. Sustainability. 2020; 12(17):7114. https://doi.org/10.3390/su12177114

3.1 Introduction

The Kilombero Valley Floodplain (KVF) in Tanzania is the largest, low-altitude, seasonally-flooded, freshwater wetland in East Africa. The valley was designated as a Ramsar site in 2002 due to its international, national and regional importance for a wide array of ecosystem services: waterflow regulation, fisheries, dry-season grazing, tourism, and hunting. Besides, it is part of the "Southern Agricultural Growth Corridor," an area earmarked for future investments in agricultural development (ERM, 2012; Milder, Buck, & Hart, 2013).

Despite its values, KVF is a very fragile ecosystem threatened by human interventions. Conversion to cropland and excessive exploitation by improperly planned development activities in the valley is and will continue to have severe, adverse, and irreversible impacts on its capacity to provide services in the future (ERM, 2012). In both neighboring districts (Ulanga and Kilombero), population density has been increasing steadily. As a result, productive agricultural land is scarce, and clearing wetland vegetation for crop farming is impossible. The problem is further aggravated by intense competition between smallholder farmers, migrating pastorals, large scale commercial ventures, governmental and non-governmental conservation groups (Bamford, Ferrol-Schulte, & Smith, 2010; Dinesen, 2016; Kato, 2007; Milder, Buck, & Hart, 2013; Nindi, Maliti, Bakari, Kija, & Machoke, 2014). Many studies have provided evidence for the perilous situation the smallholders are in, from the degradation of ecosystems to the fragility of their livelihoods (Kangalawe & Liwenga, 2005; Milder, Buck, & Hart, 2013; Mombo, Speelman, Kessy, Hella, & Van Huylenbroeck, 2012; Msofe et al., 2019; Ronald, Dulle, & Honesta, 2014) characterized by persisting food insecurity and high inequality. The government of Tanzania has recognized the need for increasing smallholder welfare and achievement of economic growth and poverty reduction through sustainable intensification pathways (ERM, 2012; Jenkins, 2012; Schnitzer & Azzarri, 2014). Backed by international donors (DFID, USAID, UNDP, FAO, Norwegian Embassy) and multinational companies (Bayer CropScience, Monsanto, Syngenta, Yara, Unilever, Nestle, SAB Miller, and others), the government has shown renewed interest to invest in both large scale and smallholder farmers in KVF (Martin-Prével, Frédéric, & The Oakland Institute, 2016). Efforts have been aiming at removing critical obstacles through increasing supply and efficiency of input use, training and capacity building, finance, infrastructure, value-chains, and markets (Jenkins, 2012; Milder, Buck, & Hart, 2013; Milder, Buck, Hart, Scherr, & Shames, 2013; New Markets Lab & SAGCOT, 2017).

However, there are many different types of farm households in KVF, which differ in terms of the available natural resource base, the dominant pattern of farm activities, household livelihoods and the way they allocate household resources (labor, land, fertilizers, machinery, technology, etc.) to agricultural production (Kato, 2007; Mombo, Speelman, Hella, & Van Huylenbroeck, 2013; Saravia Matus, Cimpoeis, & Ronzon, 2013). Diversity among farmer households in terms of resource endowment, land size, and household characteristics will have an implication on how they will respond and benefit from policies and investments.

Such diversity among farmers, has received increased interest from the public and private sector in recent years. The latter especially became aware of Sub-Saharan Africa (SSA), where the majority of the population is rural, and agriculture is considered the engine of growth. Generally, SSA's farming systems are highly heterogeneous and are driven by a complex set of socio-economic and biophysical factors (AGRA, 2013, 2017; Dixon, Gulliver, Gibbon, & Hall, 2001). Such heterogeneity has important policy implications as Garrity, Dixon, and Boffa (2012, .p 51) argues that, "the diversity of farming systems in Africa is greater than in any other part of the world. . . and generic policy assessments related to resource management or production are usually inappropriate and are often downright misleading." Yet, initial efforts to understand the diversity of farmers in the SSA are based on distinct points of polarization, including crop production vs. livestock breeding, food crops vs. cash crops, subsistence farming vs. market-oriented (Saravia Matus et al., 2013) rather than on more contextualized typologies. As a result, international development programs and national policymakers have struggled to "reconcile their recognition of heterogeneity and complex systems, with the reductionist inclinations that come with a focus on large scale, or even on global priorities" (Whitfield, Dixon, Mulenga, & Ngoma, 2015, .p 6). This struggle can possibly be resolved by adding more contextualized types from case study research to the empirical wealth on farmer diversity upon which more profound and largescale generalizations can be built in the future.

The case is not different in KVF, where blanket policies and interventions are implemented. For example, Osabuohien, Efobi, Herrmann, and Gitau (2019) reported that a Large-Scale Agricultural Investment (LSAI) scheme, as promoted by the SACGOT initiative, exhibits a negative association with the welfare of female-headed households, and they recommend specific targeting of potential beneficiaries. Similarly, Herrmann (2017) denotes considerable heterogeneity among households in terms of benefits from the effect of out-grower schemes under SAGCOT. Land rich outgrowers benefit more than land-poor ones, and farmers under sugarcane outgrower schemes are benefiting more than those under rice outgrower schemes. Moreover, land poor and landless households are more benefiting from wage employment rather than from outgrower projects. A case study from a program initiated by Kilombero Plantation Limited (KPL) and "Feed the Future Tanzania NAFAKA" on Sustainable Rice Intensification (SRI) also shows that farm households with higher labor supply were able to increase their income due to the implementation of SRI (Nakano, Tanaka, & Otsuka, 2018).

To this end, understanding farmer diversity through typologies is now considered as a 'requirement' and a 'tool' in the analysis of farm households capacity to increase output and yields in an environmentally sustainable manner while taking into account economically viable pathways (Bidogeza, Berentsen, De Graaff, & Oude Lansink, 2009; Gebauer, 1987; Saravia Matus et al., 2013). Generating a typology means "reducing the assumed or known variety of different types of farm households concerning their sources of livelihood and their 'socio-economic status' into a reasonably small number of groups which — in some respect — can be treated as a unit" (Gebauer, 1987, .p 262-263).

There is a vast number of studies conducted to characterize farmers through typologies. The aims of these studies vary which also determine the type of the methodological approach, the variable selection, and the characterization of the identified groups. Typologies are constructed to generally understand the farming systems (Guiomar et al., 2018; Köbrich, Rehman, & Khan, 2003), explore land use and intensification(Bidogeza et al., 2009; Goswami, Chatterjee, & Prasad, 2014; Kuivanen, Michalscheck, et al., 2016; Takeshima et al., 2013; Valbuena, Verburg, & Bregt, 2008), technology adoption(Berre et al., 2017), livelihood strategy (Kuivanen, Alvarez, et al., 2016; Pacini et al., 2014; Pienaar & Traub, 2015; Tittonell et al., 2010), vulnerability to climate change and environmental assessment (Andersen, Elbersen, Godeschalk, & Verhoog, 2007; Daloglu, Nassauer, Riolo, & Scavia, 2014; Hazeu et al., 2011; Nin-Pratt, ElDidi, & Breisinger, 2018; Shukla, Agarwal, Sachdeva, Kurths, & Joshi, 2019). Although there are attempts to provide an international typology of farmers (see Saravia Matus et al., 2013), it is often constructed for a specific case study site (country or region). Wezel et al. (2014) and Therond, Duru, Roger-Estrade, and Richard (2017) provide a comprehensive review of the development of farming system typologies, illustrate those that include environmental aspects, and consider their broader setting.

In this paper, we develop a typology of farmers in KVF that captures their heterogeneity and elicit the diversity of farm-households that might be expected to exhibit different land-use behavior and livelihood strategies. By combining Principal Component Analysis (PCA) and Clustering (Alvarez, Paas, Descheemaeker, Tittonell, & Groot, 2014; Hansen & Jaumard, 1997; Husson, Le, & Pages, 2017), we classify farm households into homogenous groups facing similar constraints, incentives, and other exogenous factors. The reasons why the characterization of farm households through a robust typology in KVF stands appealing are threefold: (1) Despite the aforementioned renewed interest for agricultural intensification in KVF by the government, there is no concise classification scheme (except the smallholder farmer vs. large-scale commercial ventures narrative) that would form the basis to understanding how different farm households are likely to respond to changes in policy and environment. (2) The different types of farm households identified also shed light on current agricultural practices and provide vital information needed for targeted interventions per farm type (Alvarez et al., 2018; Saravia Matus et al., 2013).(3) The resulting farm types can be subsequently used in further research as a basis for building prototype farms (Alvarez et al., 2018) as case study objects and to parametrize agent-based models, similar to those of Daloglu et al. (2014); Q. B. Le, Park, and Vlek (2010); Valbuena et al. (2008); Villamor et al. (2011). Besides, our paper provides two methodological contributions. First, we use a combination of Hierarchical clustering and K-means clustering to elicit better and robust clusters (Husson et al., 2017). This will avoid the problem of local minima associated with K-means clustering. Second, based on independent data, we validate the stability of the groups we identified. Thus, we contribute methodologically by outlining a quantitively more rigorous way to construct typologies.

The remaining part of the chapter is structured as follows. The next section introduces the study site, data, and variable selection, and the methodological approach used in the construction of farm typologies. Section 3 presents the results, discussion, a validation exercise and policy implications. The final section concludes the chapter.

3.2 Material and method

3.2.1 Data and variable selection

The data used in the current study was collected using a household survey in 21 villages in two districts of the Kilombero Valley, Ulanga and Kilombero. In total, 304 farm households were interviewed using a structured questionnaire with an extensive set of questions that were selected to discover the farming system in terms of resources, land use, and sources of livelihoods. The selection of households to be interviewed was based on a multi-stage sampling strategy. In the first stage, 12 wards were purposively selected based on the occurrence of floodplain farming. In the second stage, 21 villages were randomly selected within the wards. In the final stage, households were randomly selected from the list provided by each village's leader. The number of interviewees per village ranges from 5 in smaller villages to 15 in the biggest. A GIS coverage incorporating the land use map form GLC30 (Jun, Ban, & Li, 2014), the administrative boundary and the 2012 census data (National Bureau of Statistics, 2013) from the Tanzania statistics office was used to estimate the boundaries and total population size in the study area. From the sample survey, we selected those variables considered most relevant to explain the livelihood strategy and land use of farmers in KVF. Using the Sustainable Livelihood framework (Ian Scoones, 1998), we selected 12 variables (that can be mapped into human, physical, natural, and financial capital) considered to shape people's livelihood strategies. Besides, we added three variables for farmer's land use decision and crop choices (percentage of the total cultivated land allocated for rice, maize, and vegetables). The descriptive statistics of the key variables used for the typology are presented in Table 3.1.

Variable	Description	Unit	Mean (SD)	CV
Age	The age of the household head	years	46.53 (12.92)	0.28
Household size	Number of individuals in the household	number	5.12 (2.15)	0.42
Share of rice	Percentage of the total cultivated land allocated to rice	%	78.77 (23.93)	0.3
Share of maize	Percentage of the total cultivated land allocated to maize	%	13.92 (21.19)	1.52
Farm size	The size of farm land owned	ha	2.61 (2.78)	1.06
TLU	Total Tropical livestock unit	TLU	1.46 (6.56)	4.49
Percent hired	Share of labor hired	%	37.25 (33.21)	0.89
Commercialization index	An index of commercialization	index	47.05 (24.8)	0.53
Expenditure on Agro-inputs	Overall input intensity (Fertilizer, seed and agro-chemicals) (ha^{-1})	TSh	64984.41 (349645.18)	5.38
Distance river	Distance from plot to the nearest river	km	2.61 (3.68)	1.41
Off Farm income	Percentage of Income from non-farm sources	%	9.78 (21.74)	2.22
Share of Vegetable	Percentage of the total cultivated land allocated to vegetables	%	3.8 (11.92)	3.14
Per capita income	Per capita income per year	TSh (000)	516.37 (1124.84)	2.18
Total labor person days	Total labor use in the farm (ha^{-1})	Man-days	317.36 (337.07)	1.06
Years of schooling	Total number of years in school	years	6.37 (2.56)	0.4

Note: SD=Standard Deviation; CV = Coefficient of Variation;TSh=Tanzanian shilling; ha^{-1} = per hectare; n=300

TABLE 3.1: Descriptive statistics for the variables included in typology constructions

3.2.2 Methods of typology construction

There are two broader strands of methodologies that can be used to construct a typology. The first category comprises qualitative constructions of typologies, also known as subjective methods of classification (Köbrich et al., 2003). They rely on literature and on the knowledge and judgment of the researcher in interpreting patterns to define the specific partition of different groups (Iraizoz, Gorton, & Davidova, 2007; Pienaar & Traub, 2015; Saravia Matus et al., 2013). Although they are more descriptive than explanatory (Köbrich et al., 2003), qualitative methods provide a fast determination of relevant farm types based on a small number of characteristics. Examples of studies in this category include (Andersen et al., 2007; Daloglu et al., 2014; Schmitzberger et al., 2005; Valbuena et al., 2008). The most notable statistical approaches applied include Principle Component Analysis (PCA), Multi-dimensional Scaling (MDS), Multiple-Correspondence Analysis (MCA), and Factor Analysis for dimension reduction and Hierarchical or Non-Hierarchical Clustering. Some of the studies that apply a quantitative approach include (Köbrich et al., 2003; Takeshima, 2016; Takeshima & Edeh, 2013). The most notable statistical approaches applied include Principle Component Analysis (PCA), Multi-dimensional Scaling (MDS), Multiple-Correspondence Analysis (MCA), and Factor Analysis for dimension reduction and Hierarchical or Non-Hierarchical Clustering. Some of the studies that apply a quantitative approach include (Bidogeza et al., 2009; Goswami et al., 2014; Iraizoz et al., 2007; Kuivanen, Michalscheck, et al., 2016; Nin-Pratt et al., 2018; Pacini et al., 2014; Pienaar & Traub, 2015; Singh, Dorward, & Osbahr, 2016; Takeshima, 2016; Tittonell et al., 2010). (Alvarez et al., 2014; Kuivanen, Michalscheck, et al., 2016; Lacoste, Lawes, Ducourtieux, & Flower, 2018) on the other hand, provide a comparison and discuss the complementarity of quantitative and qualitative approaches. A multivariate approach that combines Principal Component Analysis (PCA) and both hierarchical and partitioning clustering is used in this study. PCA is a multivariate statistical technique that linearly transforms a large number of independent variables into smaller, conceptually more coherent set of variables called principal components (Dunteman, 1989). Components account for decreasing proportions of the total variance of the original variables. The first component being the best linear combination of variables that accounts for the highest share of the variance in the data than any other linear combination. And the second component is then the second-best linear combination of variables from the residual variance subject to the constraint that its orthogonal to the first component. The process continues to extract components until all of the variances are accounted for (Hair, Black, Babin, Anderson, & Tatham, 2014).

Performing PCA involves several steps. (1) we check the validity of our sample data for PCA using Bartlett's test of sphericity to test the statistical significance that the correlation matrix has significant correlations among at least some of the variables (Hair et al., 2014). (2) variables are then standardized (converted to z scores) to avoid an inappropriately strong influence of variables with large variance (Husson et al., 2017). (3) The next specifies similarities between two different observations using Euclidean distance(Husson et al., 2017).(4) Using the commonly employed latent root criterion(Kaiser's-Guttman Rule), we extract components having eigenvalues greater than 1 (Dunteman, 1989; Hair et al., 2014). We use the PCA to separate signal and noise in the original dataset. Maintaining the extracted components representing the essential information and applying the clustering on the PCA without the noise leads to a stable and more precise cluster (Husson, Josse, & Pages, 2010).

In order to support the aim of combining strong heterogeneity between the types while showing homogeneity within a group, we perform the cluster analysis on the retained components from the PCA. Cluster analysis, also called Q analysis, typology construction, unsupervised pattern recognition, or numerical taxonomy, is a group of multivariate techniques whose primary purpose is to segment objects based on the characteristics they possess (Everitt, Landau, Leese, & Stahl, 2011; Hair et al., 2014). The two most commonly used clustering methods are Hierarchical Clustering and Partitioning. Hierarchical Clustering consists of a series of partitions which proceed either by a series of successive subdivisions (Divisive hierarchical method) or mergers of observations into groups (Agglomerative hierarchical approach). The agglomerative hierarchical approach starts with as many clusters as observations. In each subsequent step, the two most similar clusters are combined to build a new aggregate cluster (Hair et al., 2014). A divisive hierarchical method, on the other hand, starts with an initial single group of observations and successively dividing into sub-groups such that objects in one group are dissimilar to objects in the other group (Hair et al., 2014; Härdle & Simar, 2013). In contrast to hierarchical methods, partitioning clustering does not involve the treelike construction process. Instead, they work by portioning the data into a user-specified number of clusters and then iteratively reassigning observations to clusters until some numerical criterion is met (Everitt et al., 2011; Hair et al., 2014; Husson et al., 2017).

In this study, we combined Agglomerative Hierarchical Clustering and K-means Clustering. The rationale for combining the two methods is discussed in detail in (Hair et al., 2014; Husson et al., 2010, 2017). The Agglomerative Hierarchical Clustering is used to select the number of clusters and profile cluster centers using Ward's minimum-variance method. This method allows us to decompose the total inertia (total variance) in between and within-group variance. The total inertia can be decomposed (Husson et al., 2010, p.4):

$$\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}(x_{iqk}-\overline{x}_k)^2=\sum_{k=1}^{K}\sum_{q=1}^{Q}I_q(x_{qk}-\overline{x}_k)^2+\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}(x_{iqk}-\overline{x}_{qk})^2 \quad (3.1)$$

with x_{iqk} the value of the variable k for the individual i of the cluster q, $\overline{x}_{qk}$ the mean of the variable k for cluster q, $\overline{x}_k$ the overall mean of variable k and Iq the number of individuals in cluster q.

A division into N clusters is made when the increase of between-inertia between N - 1 and N clusters is much higher than the one between N and N + 1 clusters. In the next step K-means clustering is performed, using the seed points and number of clusters from the hierarchical tree to provide more accurate and improved cluster memberships. Both the PCA and clustering methods are implemented using FactoMineR: A Package for Multivariate Analysis (S. Le, Josse, & Husson, 2008) and Factoextra: Extract and Visualize the Results of Multivariate Data Analyses (Kassambara & Mundt, 2016) in R statistical software (R Core Team, 2018).

3.3 Result and discussion

Based on the methodology outlined in section 2, cluster analysis on the principal components was performed to understand the diversity of farm households in KVF based on their livelihood strategy and land use. In the following section, a descriptive analysis of the variables in the cluster analysis is presented.

3.3.1 Descriptive statistics

The average household size in our sample was 5 (SD=2.15, n=300) with a minimum of 2 members and a maximum of 11 members. Forty-four percent of respondents have a family size of fewer than four members, which can be considered as a small family. Furthermore, 41% are medium-sized with 5-8 numbers of members. 12% of households in the sample are extended families, with more than eight members. Most of the households in the surveyed villages obtain their livelihood from agriculture. Crop production, mainly rice and maize, are the essential crops both for home consumption and income generation. Some households also integrate crop production with livestock rearing. Although income from farming is the dominant livelihood strategy for the majority of the farmers, 26% of the households have received some form of non-farm income, accounting for close to 10% of their total annual income. The most common sources for non-farm income in the area include remittances, rental of land, brick selling, and small business shops. The amount of land to which a household has access and the terms on which it utilizes that land are factors that influence its decisions on how to use the land resources to earn a livelihood. The average farm size in the valley is 2.6 hectares (sd= 2.8). Farmers typically own multiple parcels, with 62% of them holding two or more parcels. Usually, one large parcel is located in the seasonally flooded area which is used for rice and maize production and the smaller plots are often in proximity of the homesteads. Households plant some vegetables for home consumption on the latter.

Paddy rice is the dominant crop cultivated in the area, usually prioritized both for its local consumption and income-generating potential. On average, farmers allocate 80% of their land for rice production, 13% to maize. And some farmers also produce vegetables, cassava, and other permanent crops and fruits. Farmers market different proportions of their crops for cash. The survey result shows that, on average, 60% of the rice and maize cultivated is sold for cash and that the remaining 40% is retained for home consumption. Farmer commercialization index, which is a composite index of farmer's total crop sales to total crop cultivation, is 46% in the valley. The marketing channel is characterized by a large number of small traders operating between the farmer and the rice mills or maize market located in Ifakara (the district market center). The local traders buy small quantities directly from farmers and transport them to mills where it is milled and the rice sold to inter-regional traders, local retailers or directly to consumers.

Having sufficient labor is a key factor for the livelihood of households in the valley. Labor is provided either by household members or hired from the local labor pool. The result shows that hiring and exchanging labor occur frequently in the area. 94% of surveyed households have hired laborers to help with different stages of cultivation, the majority being hired during land preparation and cultivation stages. On average, 63% of the total man-day is provided by family labor, and the remaining 37% is from hired labor.

3.3.2 Principal component analysis

Once the variables are standardized, and outliers are identified and removed, we checked the validity of our sample data for PCA using Bartlett's test of sphericity. The significant value of the test [Chi-Square= 1060.663, p=0.0] shows that the correlation matrix has significant correlations among at least some of the variables (Hair et al., 2014) and we can proceed to PCA.

In total, 15 variables were included in the PCA, and based on the latent root criterion (eigenvalue greater than 1), we extracted six components as input for the cluster analysis (Table 3.2). The six components together account for 66.56% of the total variance in the original data set. Table 3.2 also shows the correlation between the variables and each component. The bold values identify the top three strongly correlated variables with the respective PC. The first component (PC1) accounts for 16% of the variance, and it is positively correlated with farm household size, farm size in ha, and tropical livestock unit owned by the household. Hence, the PC1 represents the resource endowment of the household. The second component, which accounts for 14.4% of the total variance, is positively correlated with the share of land allocated to rice and the size of the farm owned by the household. Also, it is negatively correlated with the share of land allocated to maize and vegetables. Generally, the second component represents the land use decision of the farm household. PC3 explains 11.23% of the variance, and it is strongly correlated with per capita income, percentage of income from non-farm activity, and percent of labor hired. Hence, PC3 represents the financial capital of the farm household. PC4, on the other hand, explains 9.17% of the variance in the original data, and it is correlated with total expenditure on agrochemical inputs, access to the river, and the percentage of land allocated to vegetables. PC5 and PC6 account for 8.5% and 7.2% of the total variance, respectively. While PC5 is highly correlated with age of the household head, years of schooling, and share of land allocated for vegetables, PC6 is associated with per capita income, market participation, and distance from the river. These six components were used in subsequent cluster analysis.

Variables	Correlation between a variable & a principal component					
	PC1	PC2	PC3	PC4	PC5	PC6
Age of household head	0.4087	0.0818	-0.2588	0.0385	**-0.6458**	-0.1076
Household size	**0.5634**	0.3491	0.0777	0.1578	0.2715	-0.1077
Share of land allocated to rice	-0.4839	**0.7854**	-0.3324	0.016	0.0316	0.0328
Share of land allocated to maize	0.5246	**-0.6634**	0.2328	-0.2004	0.2237	0.1673
Farm size owned in Ha	**0.5328**	**0.432**	0.4265	0.0602	0.1217	-0.1854
Tropical livestock unit	**0.5591**	0.3199	0.0082	0.29	0.3321	0.0207
Share of hired labor	-0.4772	0.1536	**0.4861**	0.054	-0.1049	-0.2018
Commercialization index	-0.4606	-0.1168	0.0823	0.3672	0.2311	**-0.3844**
Total expenditure in agro-inputs (000 Tsh)	-0.2082	0.0056	0.2158	**0.628**	0.0495	0.2937
Distance from the nearest river in Km	0.0882	-0.0691	0.1854	**0.5091**	-0.0691	**0.4159**
Share of Off farm income	-0.1143	0.1656	**0.4908**	-0.2506	-0.2908	0.2696
Share of land allocated to vegetables	-0.0153	-0.3791	0.3349	**0.4172**	**-0.394**	-0.3816
Income per capita	-0.078	0.2869	**0.5297**	-0.1394	-0.1602	**0.4079**
Total labor person days per year	-0.3336	-0.2853	-0.4229	0.1884	0.1535	0.3579
Years of schooling	-0.4332	-0.1294	0.4455	-0.2871	**0.3907**	-0.1092
Eigenvalues	2.23	2.01	1.57	1.28	1.19	1.02
Cumulative explained variance	16	30.33	41.56	50.72	59.27	66.56

TABLE 3.2: Six principal components with loading, eigenvalues and cumulative explained variance

3.3.3 Cluster analysis

Using the hierarchical and k-means clustering, a three-cluster solution was obtained. Figure 3.1 provides the tree-based representation of the observation, also known as a dendrogram. Moreover, partitioning in three clusters is represented on the scatter plot produced by the first two principal components, and the dots (representing farmers) are colored according to their cluster group (Figure 3.2). The cluster dendrogram shows explicitly three different farm groups identified by the cluster analysis. Table 3.3 presents the variables that discriminate each cluster group.

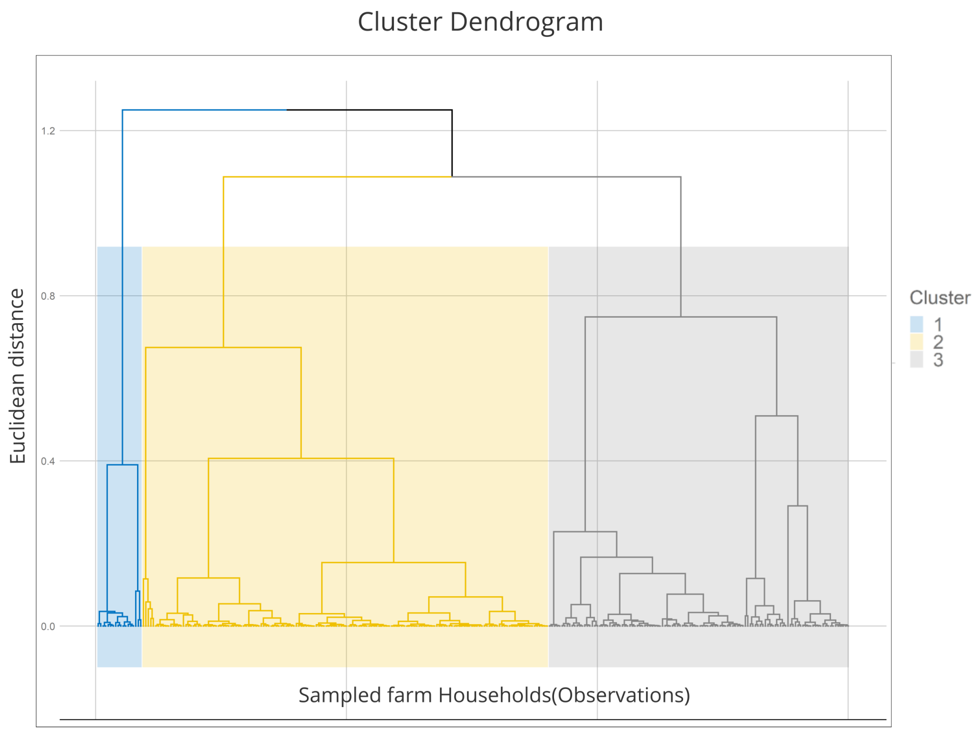


FIGURE 3.1: Dendrogram of three farm types in KVF

Note: The cluster dendrogram is the standard ways of representing the hierarchical relations and allocation of samples in to groups. The cluster is based on agglomerative hierarchical clustering with Euclidean distance as the similarity measure and Ward's linkage strategy(n= 300)

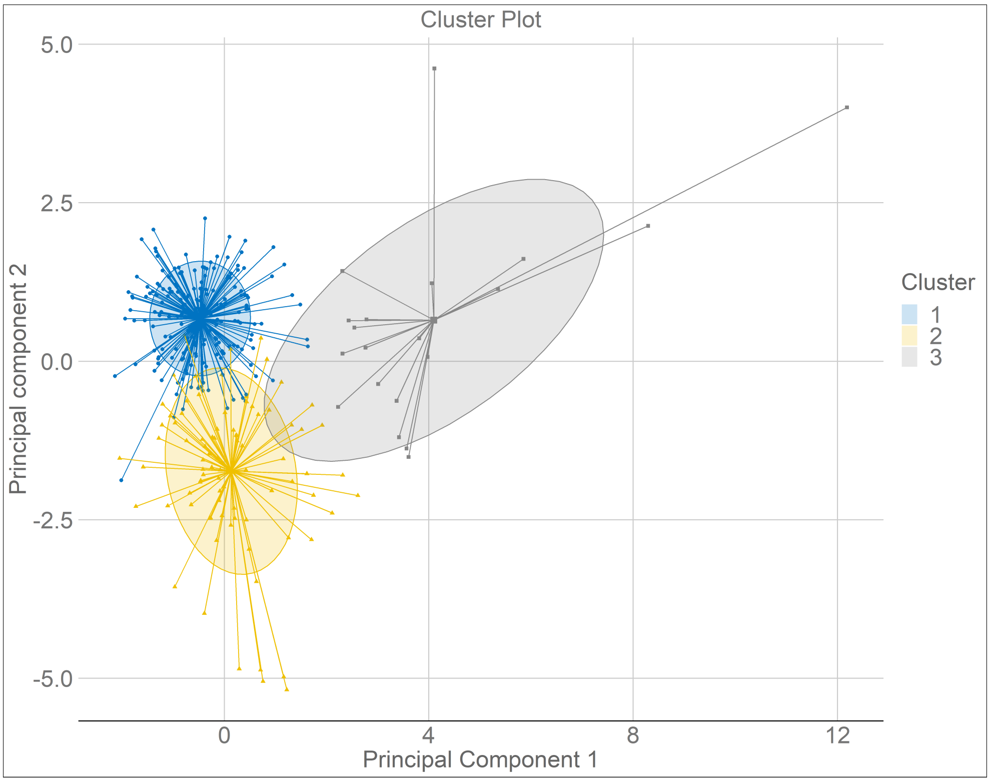


FIGURE 3.2: Distribution of the farmer household in three groups projected on the one and two dimensional plane.

Note:Principal component 1 and 2 are the first two components from PCA that captures 30.33% of the variation.

Cluster I account for 68.4% of the farm households in KVF. Share of rice, share of hired labor and the household commercialization index are significantly and positively associated with the first cluster. Given the importance of these variables, we labeled the first cluster as "Monocrop rice producers" (MCRPs), with almost 92 percent of their land allocated to rice (compared to 79% for all farmers). Considering the main crop, they tend to have a larger share of hired labor and higher input intensity. Almost 50 percent of their rice harvests are sold to the market to cover the costs of inputs and basic household needs. In terms of livelihood, they are dependent on farm income without livestock integration. They own less land with an average of 1.97 hectares compared to an average of 2.5 hectares in the study site. Although there is limited off-farm income opportunity, monocrop rice producers also receive income from non-farming activities.

Cluster II accounts for 25.2 % of the sampled farm households. Share of land allocated to maize, rice, and vegetables as well as share or hired labor are most significantly associated with cluster two. Hence, we labeled the second cluster of farmers as "Diversifiers." Diversifiers are different from the other two groups, mainly in terms of their land-use decision. Although the highest share of land is allocated to rice (47%), they also produce maize (40%) and vegetables (10 %). Households in this group mainly rely on family labor, with only 24 % of the labor provided by wage labor.

Cluster III comprises 6.4 % of the farm households. The third cluster is strongly associated with farm size, TLU, Household size, and per capita income. Given the mix of farming and livestock keeping, we labeled it as "Agropastoralists." The Agropastoralists own relatively more land and TLU, have larger household sizes, and earn larger per capita income relative to their peers in the valley. Moreover, they are characterized by lower market participation (crop) and lower labor person-days per year per hectare. Agropastoralists are recently migrated farmers from other parts of the country who have cleared new land for cultivation of crops and livestock keeping. One possible explanation for the lower market participation (commercialization index of 31 compared to the overall average 47) is the large household size, which might require them to keep a significant portion of their output for home consumption.

Figure 3.3 provides the box plots for the characterization of the three farm groups. To test if there is a significant difference between the groups, a pairwise mean comparison is conducted. As shown in the plots, there is a significant difference between the Agropastoralist and Diversifier types in terms of Farm size (3.3A), land allocated to crops (3.3 B,C&D), household size (3.3 F), TLU (3.3 H) and per capita income (3.3 J). Similarly, the result show significant difference between Agropastoralists and MCRP in terms of Farm size (3.3A), land allocated to crops (3.3 B,C&D), commercialization index (3.3 E), household size (3.3 F), share of hired labor (3.3 G), TLU (3.3 H) and per capita income (3.3 J). Looking at the difference between MCRP and diversifiers, there is a significant difference between the two farm groups in Farm size (3.3A), land allocated to crops (3.3 B,C&D), commercialization index (3.3 E),the share of hired labor (3.3 G) and age of the household head (3.3 I).

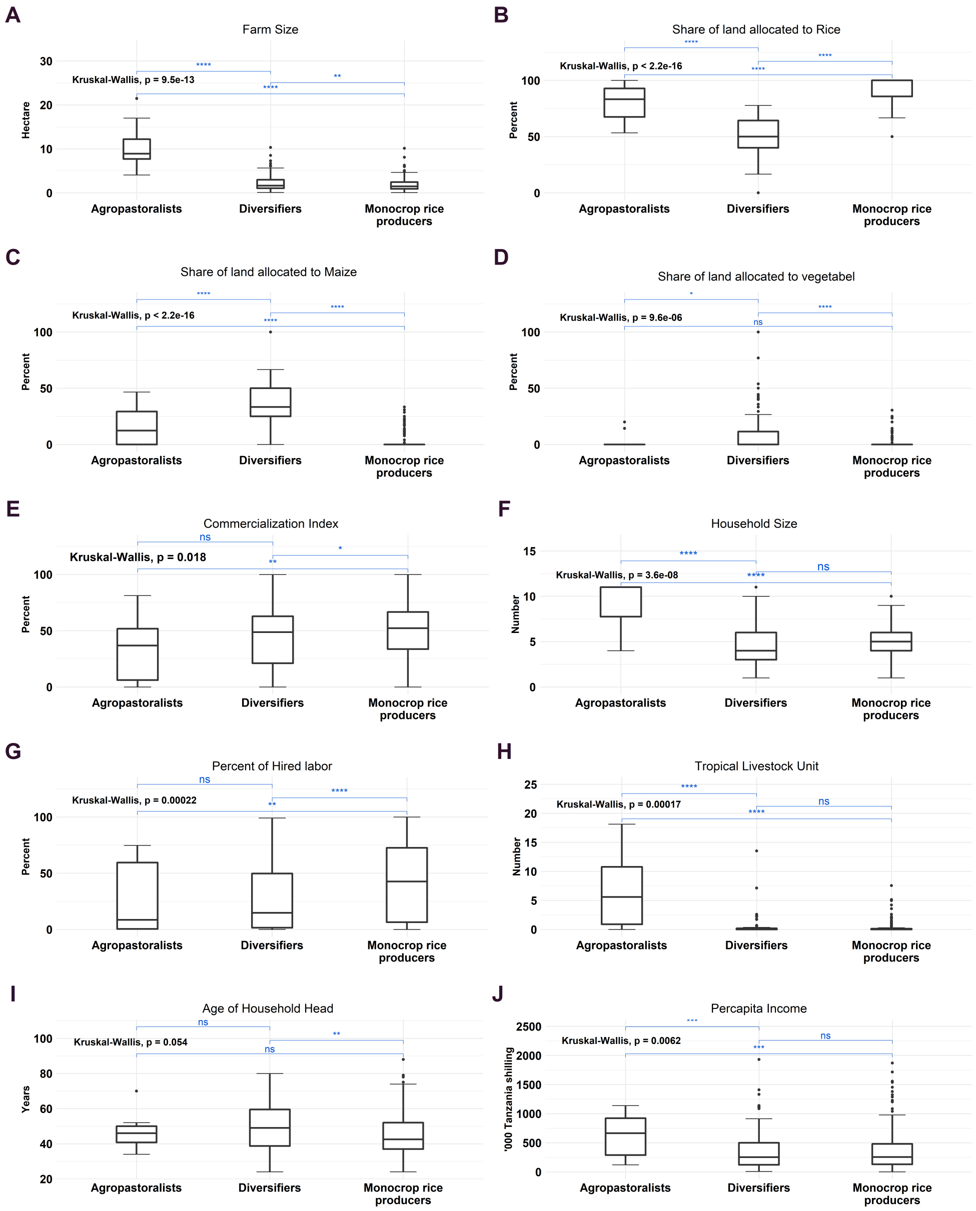


FIGURE 3.3: Box plots of farmer characterization in KVF by main variables.

Note: The horizontal lines between the box plots shows if there is significant differences in mean of a particular variable between groups (pairwise mean comparison) at different significant levels $[ns : p > 0.05][* : p <= 0.05][** : p <= 0.01][*** : p <= 0.001][**** : p <= 0.0001]$. *Kruskal-Wallis test is a non-parametric test to compare samples from two or more groups of independent observations,* $p < 0.05$ *is considered as significant*

3.3.4 Validation of Typology

In order to check the validity and stability of the clusters identified above, we conduct a validation exercise using the 2007 Agriculture sample survey (ASS) of Tanzania (TNBS, 2009). The data contains 810 observations across 54 villages in Kilombero and Ulanga districts. The selection of the variables and algorithms are the same as in the above analysis [1]. The typology from the new data set also reveals the same pattern as the one we found from our survey. The same number of clusters are identified, and the main variables that discriminate the clusters are the same Figure 3.5. Besides, the typology from the 2007 agricultural sample survey also shows other interesting differences between farm types. For example, the distance of the main farmer field from the river is significantly higher for Diversifiers relative to their peers of Monocrop rice producers and Agropastoralists. This might explain why diversifiers can allocate a relatively larger share of land to maize.

[1]However, the ASS data misses two important variables, Per capita income and amount of labour used in crop production

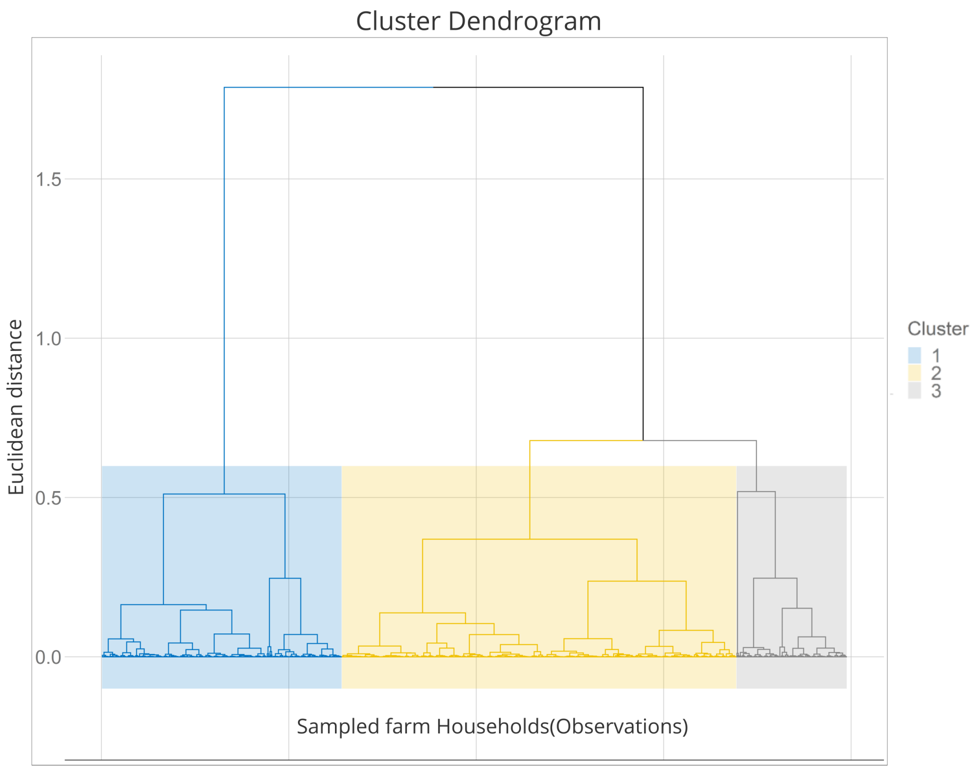


FIGURE 3.4: Dendrogram of three farm types in KVF (Validation data)

Note:The cluster is based on agglomerative hierarchical clustering with Euclidean distance as the similarity measure and Ward's linkage strategy. The validation is based on a data from Agriculture Sample Survey (2007) (TNBS, 2009) (n= 800)

3.3.5 Policy implications

Effective development strategies and policies seeking to harmonize future food production and environmental sustainability in KVF should be systematically targeted and thus need to take into consideration the challenges and opportunities associated with different farm types. The current agricultural policy of Tanzania is addressed in several government strategies and policy documents, including the Agriculture Sector Development Programme-II (ASDS-II), KILIMO KWANZA Resolve, the Tanzania Food Security Investment Plan and the Southern Agriculture Growth Corridor of Tanzania (SAGCOT) (ERM, 2012).

ASDS-II and SAGCOT are two agricultural programs with direct implications for KVF. Although the two policy interventions represent different priorities (smallholder farmers and large scale commercial ventures, respectively), both policies envision to increase agricultural production and reduce rural poverty through training and information on agricultural technology by extension services, building infrastructure including small-scale irrigation, road and warehouses, and integration of smallholder farmers into value chains (URT, 2013).

To date, these policies have tended to ignore the diversity of smallholder farmers, their needs, and constraints (KILORWEMP, 2017; NRGF, 2017; Wineman, Jayne, Isinika Modamba, & Kray, 2020). Effective development strategies and plans seeking to harmonize future food production and environmental sustainability in KVF should be systematically targeted and thus need to take into consideration the challenges and opportunities associated with different farm types. The variety of farm households identified through our typology can form a basis for prioritizing existing policies and for targeting future intervention to a specific farming system. For instance, the ASDS-II has vowed to increase access to agricultural mechanization services, including tractors, power tillers, weeder, and harvesters, etc., in collaboration with the private sector (URT, 2013, p.71). The monocrop rice producer could benefit from such interventions that prioritize access to labor-saving technologies and innovations, as they use significantly more family and wage labor for land preparation, weeding, and harvesting of rice. Although the adoption of more diverse cropping systems depends fundamentally on the hydrological regime of a particular farm, Monocrop rice producers and Agropastoralists could benefit from policies and interventions targeting transition towards agroecology through temporal and spatial diversification of cropping practices (rotation, multiple cropping, and intercropping) accompanied by water management practices. This will help them to spread production and income risk over a broader range of crops and to reduce vulnerability to exogenous shocks. Both Monocrop rice producers and Diversifiers earn their income mainly from a single source (crop production). Thus, they could also benefit from efforts towards income diversification into non and off-farm activities and from increased credit access for investing in diversified production systems. Since all the farmers still use traditional farming practices, they could benefit from access to low cost, environmentally friendly, and improved farming technologies as envisioned in both

ASDS-II and SAGCOT (ERM, 2012; URT, 2016). That will allow them to increase their productivity, which might in turn reduce the speed and scale of the current transformation of natural ecosystems into agricultural production. Finally, the Agropastoralist have not been actively engaged in the current policy landscape (URT, 2011) and they require additional attention. Poor infrastructure and insecurity increase the costs and risks of commercialization for Agropastoralists located in remote areas. They are less able to respond to terms of trade and sell less of their surplus production. Interventions through road infrastructure (especially between the isolated settlements and the main road) as envisioned in SAGCOT (AgDevCo & Prorustica, 2011, p. 19) might benefit Agropastoralists. As conflicts between the Agropastoralist and the crop farmers are increasing in recent years (Bergius, Benjaminsen, Maganga, & Buhaug, 2020), sustainable rangeland management that ensures mobility and connectivity to key natural resources and takes in to account the carrying capacity of the floodplain (as foreseen in ASDS-II (URT, 2016, p.21)) might benefit both the farmers and the environment.

3.4 Conclusion

In this study, we attempted the first classification and characterization of farm households in KVF using cross-sectional data collected in 2015. By combining principal component analysis, hierarchical clustering, and K-means clustering, we segment farmers by a purely data-driven approach into groups exhibiting similarity within and differences between them based on their livelihood and land use. Moreover, we provide an inductive generalization (Gebauer, 1987) through a concise characterization of the groups and assign appropriate meanings to them.

Our result shows an easily comprehensible typology with three representative farm types that capture the main aspects of the heterogeneity. The majority of the farmers in the valley are Monocrop rice producers who are characterized by their higher land allocation to rice, market participation, and labor use. The second farm type identified is called Diversifier. Households in this group are similar to the Monocrop rice producers in some respect but show a significant difference in terms of using relevant acreage for maize and vegetables in addition to rice. More so, the share of hired labor is relatively small, due to less emphasis on labor-intensive rice production.

The third group of farmers is identified as Agropastoralists. Households in this group pursue their livelihood by combining crop production with livestock keeping. Furthermore, they also own significantly more land and have a higher per capita income. Our validation based on a completely independent dataset shows a similar classification and characterization of farmers, which indicates that a combination of PCA, hierarchical, and K-means clustering provides stable clusters. Understanding the diversity of farmers in KVF is essential for any effort geared towards increasing production and reduction of poverty in the region. Recognition of this diversity may avoid a lack of success and unintended consequences of policy measures caused by ignoring the specific constraints and circumstances of each farm type. Besides, the farm typology will help us to define particular agent types and to appropriately parameterize behavioral models for future research of land use and intensification in KVF.

3.5 References

AgDevCo, & Prorustica. (2011). *Southern agricultural growth corridor of tanzania investment blueprint.* (Tech. Rep.). AgDevCo. Retrieved 2020-08-13, from www.sagcot.com/uploads/media/Invest-Blueprint-SAGCOT_High_res.pdf

AGRA. (2013). *African agriculture status report : Focus on staple crops* (Tech. Rep.). Alliance for a Green Revolution in Africa(AGRA). Retrieved from https://agra.org/wp-content/uploads/2017/09/agra_africa_agriculture_status_report_2013.pdf

AGRA. (2017). *African agriculture status report : The business of smallholder agriculture in sub-saharan africa(issue 5)* (Tech. Rep.). Alliance for a Green Revolution in Africa(AGRA). Retrieved from https://agra.org/wp-content/uploads/2017/09/Final-AASR-2017-Aug-28.pdf

Alvarez, S., Paas, W., Descheemaeker, K., Tittonell, P., & Groot, J. (2014). Typology construction , a way of dealing with farm diversity General guidelines for Humidtropics. *Report for the CGIAR Research Program on Integrated Systems for the Humid Tropics.*, *Plant Scie*(December), 1–37.

Alvarez, S., Timler, C. J., Michalscheck, M., Paas, W., Descheemaeker, K., Tittonell, P., ... Groot, J. C. J. (2018, may). Capturing farm diversity with hypothesis-based typologies: An innovative methodological framework for farming system typology development. *PLOS ONE*, *13*(5), e0194757. Retrieved from http://dx.plos.org/10.1371/journal.pone.0194757 doi: 10.1371/journal.pone.0194757

Andersen, E., Elbersen, B., Godeschalk, F., & Verhoog, D. (2007, feb). Farm management indicators and farm typologies as a basis for assessments in a changing policy environment. *Journal of environmental management*, *82*(3), 353–62. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0301479706002489http://www.ncbi.nlm.nih.gov/pubmed/17126987 doi: 10.1016/j.jenvman.2006.04.021

Bamford, A., Ferrol-Schulte, D., & Smith, H. (2010). The status of the Ruipa corridor between the Selous Game Reserve and the Udzungwa Mountains. *Report to Frontier Tanzania, Dar es Salaam, Tanzania.* Retrieved from https://www.researchgate.net/publication/268350008_THE_STATUS_OF_THE_RUIPA_CORRIDOR_BETWEEN_THE_SELOUS

_GAME_RESERVE_AND_THE_UDZUNGWA_MOUNTAINS

Bergius, M., Benjaminsen, T. A., Maganga, F., & Buhaug, H. (2020). Green economy, degradation narratives, and land-use conflicts in tanzania. *World Development*, *129*, 104850.

Berre, D., Corbeels, M., Rusinamhodzi, L., Mutenje, M., Thierfelder, C., & Lopez-Ridaura, S. (2017). Thinking beyond agronomic yield gap: Smallholder farm efficiency under contrasted livelihood strategies in malawi. *Field crops research*, *214*, 113–122.

Bidogeza, J. C., Berentsen, P. B. M., De Graaff, J., & Oude Lansink, A. G. J. M. (2009). A typology of farm households for the Umutara Province in Rwanda. *Food Security*, *1*(3), 321–335. Retrieved from http://link.springer.com/10.1007/s12571-009-0029-8 doi: 10.1007/s12571-009-0029-8

Daloglu, I., Nassauer, J. I., Riolo, R. L., & Scavia, D. (2014). Development of a farmer typology of agricultural conservation behavior in the american corn belt. *Agricultural Systems*, *129*, 93–102. doi: 10.1016/j.agsy.2014.05.007

Dinesen, L. (2016). Kilombero Valley Floodplain (Tanzania). In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The wetland book: Ii: Distribution, description and conservation* (pp. 1–8). Dordrecht: Springer Netherlands.

Dixon, J., Gulliver, A., Gibbon, D., & Hall, M. (2001). *Farming systems and poverty: improving farmers' livelihoods in a changing world* (Tech. Rep.). The World Bank. Retrieved from http://documents.worldbank.org/curated/en/126251468331211716/Farming-systems-and-poverty-improving-farmers-livelihoods-in-a-changing-world

Dunteman, G. H. (1989). *Principal components analysis* (No. 69). Sage.

ERM. (2012). *Southern Agricultural Growth Corridor of Tanzania (SAGCOT): Strategic Regional Environmental and Social Assessment. RFP Number: a/082/hq/S/Sagt/pmu/7, ERM Project Number: 0159588* (Tech. Rep. No. July). Environmental Resources Management. Retrieved from http://documents1.worldbank.org/curated/en/968601468312048448/pdf/E30750V30REV0A00Box382146B00PUBLIC0.pdf

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis 5th addition2011* (5th ed. ed.; David J. Balding et al., Eds.). West Sussex: John Wiley & Sons, Ltd Registered. doi: 10.1007/BF00154794

Garrity, D., Dixon, J., & Boffa, J.-M. (2012). Understanding african farming systems. *Food Security in Africa: bridging research and Practise*, 1–50.

Gebauer, R. H. (1987, jan). Socio-economic classification of farm households —conceptual, methodical and empirical considerations. *European Review of Agricultural Economics*, *14*(3), 261–283. Retrieved from https://academic.oup.com/erae/article-lookup/doi/10.1093/erae/14.3.261 doi: 10.1093/erae/14.3.261

Goswami, R., Chatterjee, S., & Prasad, B. (2014). Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: study from coastal West Bengal, India. *Agricultural and Food Economics*, *2*(1). doi: 10.1186/s40100-014-0005-2

Guiomar, N., Godinho, S., Pinto-Correia, T., Almeida, M., Bartolini, F., Bezák, P., ... Wästfelt, A. (2018). Typology and distribution of small farms in Europe: Towards a better picture. *Land Use Policy*, *75*(May 2017), 784–798. doi: 10.1016/j.landusepol.2018.04.012

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Multivariate Data Analysis* (7th ed.). Essex CM20 2JE: Pearson Education Limited. Retrieved from www.pearsoned.co.uk

Hansen, P., & Jaumard, B. (1997). *Cluster analysis and mathematical programming* (Vol. 79) (No. 1-3). Springer-Verlag. doi: 10.1007/BF02614317

Härdle, W. K., & Simar, L. (2013). *Applied multivariate statistical analysis*. doi: 10.1007/978-3-642-17229-8

Hazeu, G. W., Metzger, M. J., Mücher, C. A., Perez-Soba, M., Renetzeder, C., & Andersen, E. (2011). European environmental stratifications and typologies: An overview. *Agriculture, Ecosystems & Environment*, *142*(1-2), 29–39. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167880910000253 doi: 10.1016/J.AGEE.2010.01.009

Herrmann, R. T. (2017). Large-scale agricultural investments and smallholder welfare: A comparison of wage labor and outgrower channels in tanzania. *World Development*, *90*, 294 - 310. Retrieved from http://www.sciencedirect.com/science/article/pii/S0305750X16300390 doi: https://doi.org/10.1016/j.worlddev.2016.10.007

Husson, F., Josse, J., & Pages, J. (2010). *Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing* *data?* (Vol. 18; Tech. Rep. No. 1). Applied Mathematics Department. doi: 10.2478/v10037-010-0003-0

Husson, F., Le, S., & Pages, J. (2017). *Exploratory multivariate analysis by example using R*. New York: A Chapman & Hall Book.

Ian Scoones. (1998). *Sustainable Rural Livelihood:- A Framework For Analysis* (Working Paper No. 72). Institute of Development Studies. Retrieved from http://www.rlarrdc.org.in/images/Sustainable%20Rural%20Livelihhods%20Working%20Paper.pdf

Iraizoz, B., Gorton, M., & Davidova, S. (2007, mar). Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain. *Agricultural Systems*, *93*(1-3), 143–169. Retrieved from https://www.sciencedirect.com/science/article/pii/S0308521X06000813 doi: 10.1016/J.AGSY.2006.05.002

Jenkins, B. (2012). *Mobilizing the Southern Agricultural Growth Corridor of Tanzania: A Case Study* (Tech. Rep.). The CSR Initiative at the Harvard Kennedy School. Retrieved from http://growafrica.com/wp/wp-content/uploads/2012/04/16.-Harvard-SAGCOT-Case-study.pdf

Jun, C., Ban, Y., & Li, S. (2014). China: Open access to Earth land-cover map. *Nature*, *514*(7523), 434.

Kangalawe, R. Y. M., & Liwenga, E. T. (2005). Livelihoods in the wetlands of Kilombero Valley in Tanzania: Opportunities and challenges to integrated water resource management. *Physics and Chemistry of the Earth*, *30*(11-16 SPEC. ISS.), 968–975. doi: 10.1016/j.pce.2005.08.044

Kassambara, A., & Mundt, F. (2016). Factoextra: extract and visualize the results of multivariate data analyses. *R package version*, *1*(3).

Kato, F. (2007). Development of a major rice cultivation area in the Kilombero Valley, Tanzania. *African Study Monographs*, *36*(March), 3–18. doi: 10.14989/68498

KILORWEMP. (2017). *Technical advisory services for biodiversity conservation and wetland management in Kilombero valley RAMSAR site, Tanzania:Pastoralism Diagnostic Study* (Tech. Rep.). Retrieved from https://kilomberovalley.files.wordpress.com/2019/02/kilorwemp-2017-scoping{_}imp{_}gmp.pdf

Köbrich, C., Rehman, T., & Khan, M. (2003, apr). Typification of farming systems for constructing representative farm models: Two illustrations of the application of multi-variate analyses in Chile and Pakistan. *Agricultural Systems*, *76*(1), 141–157. Retrieved from https://www.sciencedirect.com/science/article/pii/S0308521X02000136 doi: 10.1016/S0308-521X(02)00013-6

Kuivanen, K. S., Alvarez, S., Michalscheck, M., Adjei-Nsiah, S., Descheemaeker, K., Mellon-Bedi, S., & Groot, J. C. (2016). Characterising the diversity of smallholder farming systems and their constraints and opportunities for innovation: A case study from the Northern Region, Ghana. *NJAS - Wageningen Journal of Life Sciences*, *78*, 153–166. Retrieved from http://dx.doi.org/10.1016/j.njas.2016.04.003 doi: 10.1016/j.njas.2016.04.003

Kuivanen, K. S., Michalscheck, M., Descheemaeker, K., Adjei-Nsiah, S., Mellon-Bedi, S., Groot, J. C. J., & Alvarez, S. (2016). A comparison of statistical and participatory clustering of smallholder farming systems - A case study in Northern Ghana. *Journal of Rural Studies*, *45*(August), 184–198. doi: 10.1016/j.jrurstud.2016.03.015

Lacoste, M., Lawes, R., Ducourtieux, O., & Flower, K. (2018, mar). Assessing regional farming system diversity using a mixed methods typology: the value of comparative agriculture tested in broadacre Australia. *Geoforum*, *90*, 183–205. Retrieved from https://www.sciencedirect.com/science/article/pii/S001671851830023X doi: 10.1016/J.GEOFORUM.2018.01.017

Le, Q. B., Park, S. J., & Vlek, P. L. G. (2010). Land Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. 2. Scenario-based application for impact assessment of land-use policies. *Ecological Informatics*, *5*(3), 203–221.

Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software, Articles*, *25*(1), 1–18. Retrieved from https://www.jstatsoft.org/v025/i01 doi: 10.18637/jss.v025.i01

Martin-Prével, A., Frédéric, M., & The Oakland Institute. (2016). *The Unholy Alliance, Five Western Donors Shape a Pro-Corporate Agenda for African Agriculture.* Retrieved from https://www.oaklandinstitute.org/sites/oaklandinstitute.org/files/unholy_alliance_web.pdf

Milder, J. C., Buck, L. E., & Hart, A. (2013). *Applying an Agriculture Green Growth Approach in the SAGCOT Clusters: Challenges and Opportunities in Kilombero, Ihemi and Mbarali* (Tech. Rep.). Dar es Salaam: SAGCOT Centre. Retrieved from https://ecoagriculture.org/wp-content/uploads/2015/08/

Applying-an-AGG-Approach-in-the-SAGCOT-Clusters-June-2013.pdf

Milder, J. C., Buck, L. E., Hart, A. K., Scherr, S. J., & Shames, S. A. (2013). A framework for Agriculture Green Growth: Greenprint for the Southern Agricultural Growth Corridor of Tanzania (SAGCOT). *Dar es Salaam: SAGCOT Centre*. Retrieved from https://ecoagriculture.org/wp-content/uploads/2015/08/A-Framework-for-AGG-Greenprint-for-SAGCOT-June-2013.pdf

Mombo, F., Speelman, S., Hella, J., & Van Huylenbroeck, G. (2013, nov). How characteristics of wetlands resource users and associated institutions influence the sustainable management of wetlands in Tanzania. *Land Use Policy*, *35*, 8–15. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264837713000732 doi: 10.1016/J.LANDUSEPOL.2013.04.010

Mombo, F., Speelman, S., Kessy, J., Hella, J., & Van Huylenbroeck, G. (2012). Determinants of access patterns to goods and services from wetlands in Tanzania and the impact on sustainable wetland management. *African Journal of Agricultural Research*, *7*(41), 5585–5593. Retrieved from http://www.academicjournals.org/AJAR doi: 10.5897/AJAR12.1243

Msofe, N. K., Sheng, L., Lyimo, J., Msofe, N. K., Sheng, L., & Lyimo, J. (2019). Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability*, *11*(2), 505. Retrieved from http://www.mdpi.com/2071-1050/11/2/505 doi: 10.3390/su11020505

Nakano, Y., Tanaka, Y., & Otsuka, K. (2018). Impact of training on the intensification of rice farming: evidence from rainfed areas in tanzania. *Agricultural Economics*, *49*(2), 193–202.

National Bureau of Statistics. (2013). 2012 Population and housing census; Population Distribution by Adminstrative Areas. *National Bureau of Statistics*.

New Markets Lab, & SAGCOT. (2017). A Legal Guide to Strengthen Tanzania's Seed and Input Markets. , 148. Retrieved from http://sagcot.co.tz/index.php/mdocs-posts/a-legal-guide-to-strengthen-tanzanias-seed-and-input-markets/

Nindi, S. J., Maliti, H., Bakari, S., Kija, H., & Machoke, M. (2014). Conflicts Over Land and Water Resources in the Kilombero Valley Floodplain, Tanzania. *African Study Monographs*, *50*(October 2014), 173–190.

Nin-Pratt, A., ElDidi, H., & Breisinger, C. (2018). Farm households in Egypt: A typology for assessing vulnerability to climate change. (September), 53 pages. Retrieved from http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/132808

NRGF. (2017). *Natural resource governance in kilombero cluster and the sagcot initiative: An assessment of key issues and recommendations for action* (Tech. Rep.). CEESP and IUCN.

Osabuohien, E. S., Efobi, U. R., Herrmann, R. T., & Gitau, C. M. (2019). Female labor outcomes and large-scale agricultural land investments: Macro-micro evidencefrom tanzania. *Land Use Policy*, *82*, 716 - 728. Retrieved from http://www.sciencedirect.com/science/article/pii/S0264837718313279 doi: https://doi.org/10.1016/j.landusepol.2019.01.005

Pacini, G. C., Colucci, D., Baudron, F., Righi, E., Corbeels, M., Tittonell, P., & Stefanini, F. M. (2014). Combining multi-dimensional scaling and cluster analysis to describe the diversity of rural households. *Experimental Agriculture*, *50*(3), 376–397. doi: 10.1017/S0014479713000495

Pienaar, L., & Traub, L. N. (2015). Understanding the smallholder farmer in South Africa: Towards a sustainable livelihoods classification. In *International conference of agricultural economists.* Milan, Italy. Retrieved from https://ageconsearch.umn.edu/bitstream/212633/2/Pienaar-UnderstandingthesmallholderfarmerinSouthAfrica-1233.pdf

R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.r-project.org/

Ronald, B., Dulle, F., & Honesta, N. (2014). Assessment of the Information Needs of Rice Farmers in Tanzania: a Case Study of Kilombero District, Morogoro. *Library Philosophy & Practice*, 1–33.

Saravia Matus, S., Cimpoeis, D., & Ronzon, T. (2013). *Literature Review and Proposal for an International Typology of Agricultural Holdings* (Tech. Rep.). Retrieved from http://www.fao.org/publications/card/en/c/790a6b2e-7bbe-4e0a-af20-18d310463ae0/

Schmitzberger, I., Wrbka, T., Steurer, B., Aschenbrenner, G., Peterseil, J., & Zechmeister, H. (2005, jun). How farming styles influence biodiversity maintenance in Austrian agricultural landscapes. *Agriculture, Ecosystems & Environment*, *108*(3), 274–290. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167880905000563 doi: 10.1016/J.AGEE.2005.02.009

Schnitzer, P. ., & Azzarri, C. (2014). *Tanzania Agricultural Snapshot 2007/8.* Retrieved from http://harvestchoice.org/sites/default/files/downloads/publications/WorkingPaper-Tanzania{_}5.pdf

Shukla, R., Agarwal, A., Sachdeva, K., Kurths, J., & Joshi, P. (2019). Climate change perception: an analysis of climate change and risk perceptions among farmer types of indian western himalayas. *Climatic Change*, *152*(1), 103–119.

Singh, C., Dorward, P., & Osbahr, H. (2016). Developing a holistic approach to the analysis of farmer decision-making: Implications for adaptation policy and practice in developing countries. *Land Use Policy*, *59*, 329–343. Retrieved from http://dx.doi.org/10.1016/j.landusepol.2016.06.041 doi: 10.1016/j.landusepol.2016.06.041

Takeshima, H. (2016, dec). Understanding Irrigation System Diversity in Nigeria: A Modified Cluster Analysis Approach. *Irrigation and Drainage*, *65*(5), 601–612. Retrieved from http://doi.wiley.com/10.1002/ird.1973 doi: 10.1002/ird.1973

Takeshima, H., & Edeh, H. (2013). Typology of Farm Households and Irrigation Systems: Some Evidence from Nigeria. *Ssrn*(April). doi: 10.2139/ssrn.2286862

Takeshima, H., Nin-Pratt, A., Diao, X., Nin-Pratt, A., Diao, X., Nin-Pratt, A., & Diao, X. (2013, oct). Mechanization and agricultural technology evolution, agricultural intensification in Sub-Saharan Africa: Typology of agricultural mechanization in Nigeria. *American Journal of Agricultural Economics*, *95*(5), 1230–1236. Retrieved from https://academic.oup.com/ajae/article-lookup/doi/10.1093/ajae/aat045 doi: 10.1093/ajae/aat045

Therond, O., Duru, M., Roger-Estrade, J., & Richard, G. (2017). A new analytical framework of farming system and agriculture model diversities. a review. *Agronomy for Sustainable Development*, *37*(3), 21.

Tittonell, P., Muriuki, A., Shepherd, K., Mugendi, D., Kaizzi, K., & Okeyo, J. (2010). The diversity of rural livelihoods and their influence on soil fertility in agricultural systems of East Africa–A typology of smallholder farms. *Agricultural systems*, *103*(2).

TNBS. (2009). National Sample Census Of Agriculture 2007/08. , *V*, 1–345. Retrieved from https://www.nbs.go.tz/tnada/index.php/catalog/7

URT. (2011). *(sagcot) investment project public notice. re-disclosure of environmental and social management framework (esmf).* (Tech. Rep.). Prime Minister's Office.

Retrieved 2020-08-13, from http://documents1.worldbank.org/curated/en/684491468165274986/pdf/SFG1663-REVISED-EA-P125728-Box396313B-PUBLIC-Disclosed-10-4-2016.pdf

URT. (2013). *National Agriculture Policy* (Tech. Rep.). Ministry of Agriculture Food Security and Cooperatives. Retrieved from http://www.fao-ilo.org/fileadmin/user_upload/fao_ilo/pdf/ICA_MLW_and_TZ/NATIONAL_AGRICULTURAL_POLICY-2013.pdf

URT. (2016). *Agricultural sector development strategy-ii 2015/2016–2024/2025* (Tech. Rep.). United Republic Of Tanzania. Retrieved 2020-08-13, from http://www.tzdpg.or.tz/fileadmin/documents/external/national_development_frameworks/ASDP2_Final_Document_20_May._2016__after_edit__1_.pdf

Valbuena, D., Verburg, P. H., & Bregt, A. K. (2008, oct). A method to define a typology for agent-based analysis in regional land-use research. *Agriculture, Ecosystems and Environment*, *128*(1-2), 27–36. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167880908001370 doi: 10.1016/j.agee.2008.04.015

Villamor, G. B., van Noordwijk, M., Le, Q. B., Lusiana, B., Matthews, R., & Vlek, P. L. G. (2011). Diversity deficits in modelled landscape mosaics. *Ecological Informatics*, *6*(1), 73–82. doi: 10.1016/j.ecoinf.2010.08.003

Wezel, A., Casagrande, M., Celette, F., Vian, J.-F., Ferrer, A., & Peigné, J. (2014). Agroecological practices for sustainable agriculture. a review. *Agronomy for sustainable development*, *34*(1), 1–20.

Whitfield, S., Dixon, J. L., Mulenga, B. P., & Ngoma, H. (2015). Conceptualising farming systems for agricultural development research: cases from eastern and southern africa. *Agricultural Systems*, *133*, 54–62.

Wineman, A., Jayne, T. S., Isinika Modamba, E., & Kray, H. (2020). The changing face of agriculture in tanzania: Indicators of transformation. *Development Policy Review*, *n/a*(n/a). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/dpr.12491 doi: 10.1111/dpr.12491

3.A Appendix

3.A.1 Table of Discramintory variables

	Cluster Mean	Overall Mean	Cluster SD	Overall SD	P.value
Cluster I [MCRPs] [68.4%]					
Share of land allocated to rice	91.47	79.18	11.02	23.91	0.00
Percent of labor hired	43.41	36.80	33.66	33.13	0.00
Commercialization index	49.87	47.05	23.45	24.75	0.00
Total labor person-days ($ha^{-1}year^{-1}$)	322.11	297.39	259.15	241.17	0.01
Total expenditure on Agro-inputs (000 Tsh) (ha^{-1})	36.358	29.998	92.565	79.086	0.05
Household size	4.81	5.05	1.64	1.98	0.00
Share of land allocated to vegetables	1.42	3.53	5.22	11.51	0.00
Farm size in Ha	1.97	2.37	1.51	2.17	0.00
TLU	0.23	0.68	0.67	2.24	0.00
Share of land allocated to maize	3.41	13.85	7.41	21.37	0.00
Cluster II [Diversifier] [25.2%]					
Share of land allocated to maize	39.66	13.85	24.43	21.37	0.00
Share of land allocated to vegetables	9.98	3.53	19.84	11.51	0.00
Percent of labor Hired	24.41	36.80	27.43	33.13	0.00
Share of land allocated to rice	47.00	79.18	19.87	23.91	0.00
Cluster III [Agropastoralist] [6.4%]					
TLU	7.07	0.68	5.36	2.24	0.00
Farm size in Ha	7.67	2.37	3.06	2.17	0.00
Household size	8.44	5.05	2.22	1.98	0.00
Share of land allocated to maize	23.99	13.85	17.67	21.37	0.04
Total labor person-days ($ha^{-1}year^{-1}$)	151.53	297.39	121.15	241.17	0.01
Years of schooling	4.89	6.44	3.45	2.47	0.01
Commercialization index	31.09	47.05	23.95	24.75	0.00
Percent of labor hired	14.81	36.80	22.86	33.13	0.00

Table 3.3: Cluster description by main discriminatory variables

3.A.2 Box Plot of Validation cluster

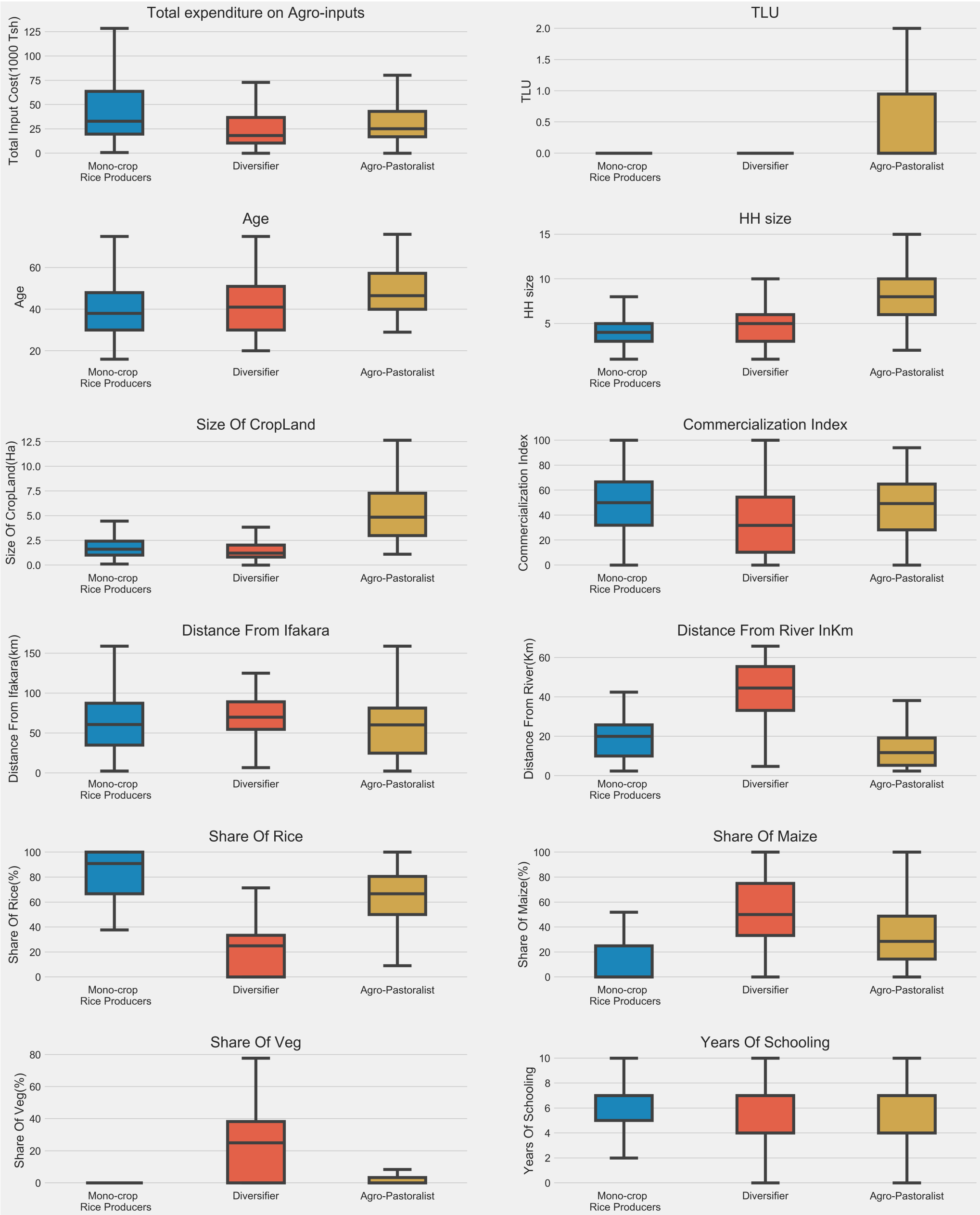


FIGURE 3.5: Box plots of main variables by farm type from the validation study.

Note: Based on the 2007 Agricultural Sample Census (TNBS, 2009). [n=800]

Chapter 4

Modeling farmers intensification decisions with a Bayesian Belief Network: The case of the Kilombero Floodplain in Tanzania†

Abstract: Kilombero Valley floodplain in Tanzania is one of the leading agriculture hotspot areas in Africa with significant interest from government and donors. Small-holder farmers in the valley are under continuous pressure to intensify their agriculture production and intensification is often considered as a rule rather than an option. Nevertheless, it remains below what is possible. Our objective in this chapter is to investigate farmers' choices of intensification and their interdependent determinants by focusing on four options practiced in the valley. We propose a new modeling approach to identify option-specific determinants and their interdependence by combining a Bayesian Belief Network (BBN), design of experiments, and multivariate regression trees. Our approach complements, existing lower dimension statistical models by taking uncertainty into account and by providing an easily updatable "white box" model structure. The BBN is constructed and calibrated using survey data from 304 farm households. Results show that choices of intensification options are driven by crop choices, access to market as well as capital and plot characteristics. Our novel approach of combining a data-driven Bayesian Belief Network (BBN) with regression trees identifies factors determining farmers' prioritization of one strategy over the other.

Keywords*: Intensification, BBN, Agriculture, Tanzania, Kilombero Valley, Regression Tree, design of experiment*

†This chapter is submitted to Agricultural Economics journal as Gebrekidan, BH., T. Heckelei and S. Rasch : Modeling Farmers Intensification Decisions with a Bayesian Belief Network: The case of the Kilombero Floodplain in Tanzania

4.1 Introduction

Achieving food security while promoting sustainable development is at the top of priorities for governments in Africa. Agriculture offers a significant potential for poverty reduction and inclusive growth particularly with almost 60 percent of the population depending on it and by being the primary source of income for 90% of the rural population (AFDB, 2015; Diao, Hazell, & Thurlow, 2010; Kanu, Salami, & Numasawa, 2014; World Bank, 2007). Governments, donors, and private enterprises recognize the importance of increasing the productivity of the sector. Yet, the sector has remained stagnant and in need of modernization and intensification with the yield gap for cereals being wide compared to other regions (Jayne, Mather, & Mghenyi, 2010; Johansson & Abdi, 2019; Van Ittersum et al., 2016).

Tanzania is quite a typical case. The agricultural sector is the backbone of the country's economy and a key driver for rural development. The sector continues to employ around 80 percent of the total workforce and provides livelihoods to more than 70 percent of the population; it contributes to approximately 95 percent of the national food requirements (Mwimo et al., 2016). Besides, the sector contributes 28 percent of the gross domestic product and accounts for about 27 percent of export earnings (Milder, Buck, Hart, Scherr, & Shames, 2013; Mwimo et al., 2016; USAID, 2019; WFP, 2019). However, the sector remains subsistence mostly, with population growth surpassing production growth, food self-sufficiency declining, and malnutrition remaining high (WFP, 2019).

The government pursues a policy of increasing domestic agricultural production-driven either by a shift to large scale commercial farms or by improved productivity of smallholders through providing opportunity and access to resources (Coulson, 2015; URT, 2013). The government has sought to align these goals through different policy statements and national visions, including Kilimo Kwanza (Agriculture First), the Southern Agricultural Growth Corridor of Tanzania (SAGCOT), and Big Results Now (Coulson, 2015).

One focal area for the government in its bid to transform the country into a sustainable food basket is the wetland of the Kilombero Valley Floodplain. The low altitude plain with alluvial deposits has a productive natural resource base with fertile land, reliable water availability, and extensive pastures (Bamford, Ferrol-Schulte, & Smith, 2010; Nindi, Maliti, Bakari, Kija, & Machoke, 2014). In the past, smallholders in the floodplain have enjoyed abundant land to increase their agricultural production by bringing new wetland areas and marginal lands under cultivation (see Figure 4.1) (Leemhuis et al., 2017; Msofe et al., 2019).

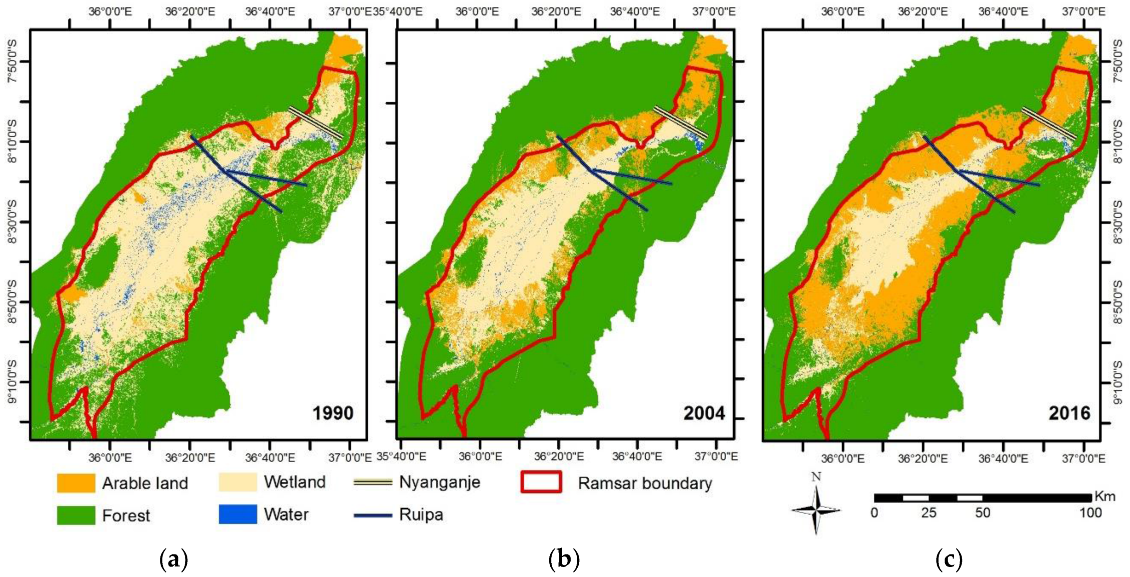


FIGURE 4.1: LULC maps of the Kilombero floodplain for 1990 (a), 2004 (b) and 2016 (c)

(Leemhuis et al., 2017, p.8)

However, this type of land-use change is associated with various negative environmental consequences, such as loss of habitat as well as of above and underground biodiversity (Jones et al., 2012). More so, increased immigration and population growth brought the expansion of agricultural land to its limit, and agricultural intensification has become the rule rather than an option (Binswanger-Mkhize & Savastano, 2017; Kajisa, 2016; Otsuka & Place, 2013). Hence, backed by the broader community of non-government organizations and private multinational companies, the government has been promoting the use of optimized/high-quality inputs, adoption of new technologies or mechanization and value-chain development as a means to increase productivity and closing yield gaps for generating sustainable and inclusive growth (Agra, 2016; Binswanger-Mkhize & Savastano, 2017; Otsuka & Larson, 2016b). The potential for yield increase in the Kilombero Valley Floodplain is evident from the current average rice and maize yield of farmers in the valley being 1.2 tons per hectare, with more than 50 percent of farmers generating less than 1 ton per ha (Figure 4.2). Which is also substantially lower than 10 to 11 tons per hectare of potential attainable yield under improved management practices and input-intensive rice and production systems in Tanzania (Nakano, Tanaka, & Otsuka, 2018; Senthilkumar, Tesha, Mghase, & Rodenburg, 2018).

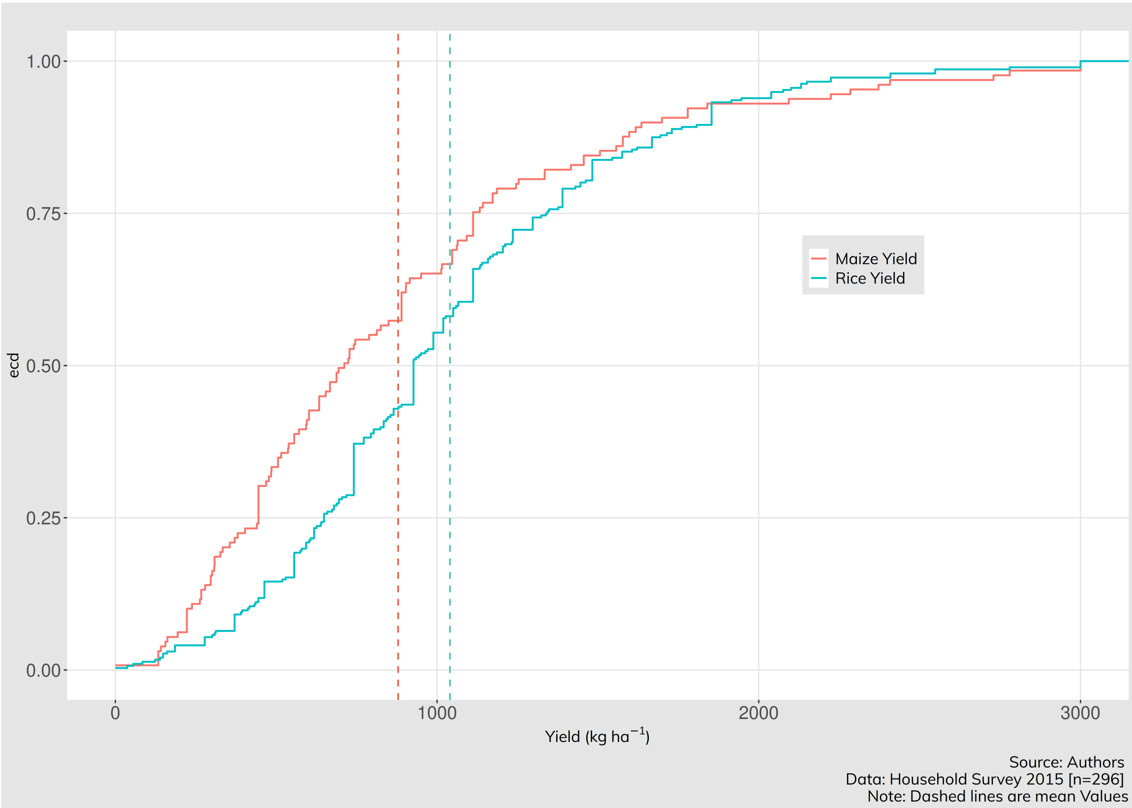


FIGURE 4.2: Empirical cumulative distribution of rice & maize yields in the KVF

While efforts are underway to accelerate and intensify production through increased uptake of improved technologies and greater use of inputs, adoption of these options by smallholder farmers have been disappointing so far (although increasing gradually) (Sheahan & Barrett, 2014). Feder, Just, and Zilberman (1985) and Foster and Rosenzweig (2010) reviewed the vast amount of literature on the adoption of different technologies in developing countries. A sub-strand of this literature body deals with farmers' intensification decisions and factors driving relevant technology choices (Abay et al., 2016; Erenstein, 2006; Headey, Dereje, & Taffesse, 2014; Holloway & Lapar, 2007; Howley, O. Donoghue, & Heanue, 2012; Kassie, Teklewold, Jaleta, Marenya, & Erenstein, 2015; Okike et al., 2001; Schelhas, 1996; Shriar, 2000, 2001; Wainaina, Tongruksawattana, & Qaim, 2016).

It is apparent from these sources that the choices farmers make are not simple reflexive responses to external drivers, but are instead the result of complex decision making. Farmers decide by conditioning on socio-economic characteristics of the farm households, infrastructure, and existing institutions as well as of the agro-ecological context (Wainaina et al., 2016). However, neither the current state of agricultural intensification nor context or locally specific factors determining farmers' choices of particular intensification options are known for KVF (Milder, Buck, & Hart, 2013; Milder, Buck, Hart, Scherr, & Shames, 2013; Nakano, Kajisa, & Otsuka, 2016).

It is in this light that further analysis of the determinants of farmers' decisions to uptake different paths of intensification and land management practices is required. As noted by Vanlauwe, Coyne, et al. (2014); Vanlauwe, Van Asten, and Blomme (2014), pathways towards intensification in Africa will require a broad-based approach considering a diversity of pathways adapted to local agro-ecological conditions, crop choice and cropping patterns, the farmer's ability and willingness to invest, and specific to institutional settings. Recent studies have also shown farmers' perceived constraints and that benefits play significant roles in their decision to adopt a particular option (Alomia-Hinojosa et al., 2018; Ntshangase, Muroyiwa, & Sibanda, 2018; Yamano, Rajendran, & Malabayabas, 2015). Thus, it is integral to uncover factors that drive farmers' prioritization of options and their perceived constraints. The analysis of pathway-specific and locally relevant determinants of intensification in KVF might allow institutions to draft future policies that will increase productivity through ecologically sustainable and pro-poor pathways.

The chapter offers two novel contributions to the existing literature. First, we investigate how farm households make their intensification decisions when multiple pathways are available and highlight the different factors driving these choices in the context of a sensitive ecological landscape with high yield potential. Here, we focus on four land-saving intensification options practiced in the valley: (1) use of chemical fertilizers, (2) use of improved seed, (3) use of small-scale irrigation systems, and (4) increasing frequency of planting. Second, we propose using a Bayesian belief network (BBN) as an analytical tool alternative to existing, typically lower-dimensional models (e.g., logit and probit models, decision tree). BBNs have the advantage of explicitly accounting for uncertainty and allow integrating a wide range of input data, including expert knowledge. They are easily adaptable in terms of structure and dependencies between different influencing factors (Korb & Nicholson, 2010; Sun & Müller, 2013). More so, an integral part of our approach is to exemplify how a combination of BBN, design of experiments, and regression trees are used to identify strategy-specific determinants.

The organization of the chapter is as follows. After providing a brief overview of the motivation for the methodological approach in section 2, section 3 introduces the study site along with data and variable selection. Section 4 briefly explains the empirical model. Section 5 will present the results and discussion of intensification strategy choices, and section 6 concludes the chapter.

4.2 The motivation of the methodological approach

Different modeling tools have been proposed to understand individual decision making concerning intensification options. The first type models the intensification decision as a binary choice problem of adopting a single intensification option or not. Prevalent approaches include probit models (Abay et al., 2016), logistic regression models (Erenstein, 2006; Okike et al., 2001; Perz, 2003) and decision trees (Gladwin, 1989) to mention a few (see (Besley & Case, 1993) for review of modeling farmer adoption decision). The second branch of models recognize the choice of intensification decisions as a process involving the joint application of multiple practices, or they consider the option as one potential strategy out of many that can be modeled by multivariate models including multinomial probit (MNP) (Dorfman, 1996; Kassie et al., 2015; Wainaina et al., 2016) or multinomial logit selection (MNLS) models (Kassie et al., 2018; Khonje, Manda, Mkandawire, Tufa, & Alene, 2018; Teklewold, Kassie, Shiferaw, & Köhlin, 2013). We propose a Bayesian Belief Network (BBN) as an alternative tool to model the multivariate setting. BBN, also known as Bayesian Net, Causal Probabilistic Network, Bayesian Network, or simply Belief Network, is a probabilistic graphical modeling tool that allows for knowledge representation and support for reasoning under uncertainty (Kjaerulff & Madsen, 2012; Korb & Nicholson, 2010; Pearl, 2009). Like other graphical models, the nodes represent stochastic variables and the arcs direct dependencies based on process understanding, statistical, or other types of associations between the linked variables (Chen & Pollino, 2010). More formally, the Bayesian network can be described as Directed Acyclic Graph (DAG) which defines a factorization of a joint probability distribution over the variables, where the directed links of the DAG give the factorization. Specifically, for a DAG, $G = (V, E)$, where V denotes a set of nodes and E a set of directed links (or edges) between pairs of the nodes, a joint probability distribution, $P(X_V)$, over the set of (typically discrete) variables X_v indexed by V can be factorized as

$$P(X_V) = \prod_{v \epsilon V} P(X_v | X_{pa(v)}) \tag{4.1}$$

where $X_{pa(v)}$ is a set of parent nodes for variable X_v, for each node v an element of V (Kjaerulff & Madsen, 2012).

BBN uses Bayes theorem and probability calculus to represent a causal linkage between two connected stochastic variables. For instance $X \rightarrow Y$, where X directly influences Y, we need to drive the posterior probability distribution $P(X|Y = y)$ using the prior distribution $P(X)$ and the conditional probability distribution $P(Y|X)$.

In addition to jointly considering multiple pathways of intensification choices, three further advantages of BBN's motivate this methodological choice:

(a) given our limited understanding regarding farmer's decision-making process and the relevance of random events, we opted to use probability theory to deal with uncertainty explicitly. The Bayesian approach to uncertainty ensures that the system as a whole remains consistent and offers a direct way to apply the model to data (Koski & Noble, 2011). As BBNs are joint probability distributions, uncertainty is propagated through the model and presented in the final results. Contrary to deterministic models, the probabilistic representation of knowledge in BBN prevents overconfidence in the strength of responses obtained by simulating changes in one or more variables of interest (Uusitalo, 2013).

(b) Unlike other 'black box' models, BBN provides generality and formalism of displaying relationships clearly and intuitively (Daly, Shen, & Aitken, 2011; Margaritis, 2003) as well as making them amenable to analysis and modification by experts and stakeholders (Sun & Müller, 2013; Uusitalo, 2013). If data exists, they are also able to incorporate qualitative beliefs and attitudes of stakeholders along with quantitative data.

(c) BBNs are easily updated as more data become available.

BBNs emerged from artificial intelligence and are widely used in diverse domains, including medicine, environmental modeling, natural resources management, and forecasting (Daly et al., 2011; Korb & Nicholson, 2010; Uusitalo, 2013). There are only a few BBN applications to farm management (see Drury, Valverde-Rebaza, Moura, & de Andrade Lopes, 2017) for a review of BBN applications in agriculture). Cain (2001) uses a BBN to explore the determinants of crop yield, and similarly, Prishchepov, Ponkina, Sun, and Müller (2019) used a BBN to examine the determinants of wheat yields in Siberia. (Sun & Müller, 2013) choose a BBN for modeling the binary choice of participating in a scheme with payments for ecosystem services and combine it with opinion dynamics in an agent-based modeling framework to simulate land conversion patterns. In a similar work, Frayer, Sun, Müller, Munroe, and Xu (2014) developed a BBN to analyze drivers of the decision to plant trees on former cropland. Aalders (2008) and Celio and Grêt-Regamey (2016) built a BBN to incorporate farmers' choices of different land-use options. Rasmussen et al. (2016) developed a large scale BBN tool for risk management in EU agriculture using Farmers Agricultural Data Network data (FADN). Pope and Gimblett (2017) used BBN in combination with agent-based modeling to explore the different ranching strategies farmers choose under varying environmental conditions. To the best of our knowledge, there is no other study applying a BBN to model farmers' adoption of intensification options.

4.3 Context and data

4.3.1 Study site

This study was conducted in the Kilombero Valley Floodplain (KVF), Tanzania. The low altitude plain with alluvial deposits has a productive natural resource base with fertile land, reliable water availability, and extensive pastures (Bamford et al., 2010; Nindi et al., 2014). Located in the Ulanga and Kilombero districts in southern Tanzania, it forms one of the four principal sub-basins of the Rufiji River Basin. It comprises several rivers and seasonally flooded marshes and swamps (Dinesen, 2016). The seasonal change in water levels is substantial, and the plains are fully inundated during the wet season, while the water fully retracts to rivers and river margins as well as areas with permanent swamps and water bodies (Kato, 2007; Ntongani, Munishi, More, & Kashaigili, 2014).

The Valley lies at the foot of the Great Escarpment of East Africa in the southern half of Tanzania, about 300 km from the coast (Kato, 2007; Nindi et al., 2014). It covers an area of about $11{,}600^2$, with a total length of 250 km and a width of up to 65 km. The altitude within the valley is about 300m above sea level. Generally, the floodplain is humid with high temperatures ranging from 26°C to 32°C. While the relative humidity in the mountains is between 70 – 87%, the lowlands experience 58 – 85% humidity with average potential evaporation of 1800 mm (Msofe et al., 2019; Wilson, Mcinnes, Mbaga, & Ouedraogo, 2017). The KVF is a typical, fertile alluvial floodplain with loamy, clay, clay loamy, and sandy soils and is an essential source of nutrients and sediment (Milder, Buck, & Hart, 2013; Nindi et al., 2014). The KVF also offers considerable ecological value as it comprises the Kilombero Game Controlled Area with approximately 7000^2 and the Kilombero valley Ramsar site, which covers $7{,}976^2$ Dinesen (2016); Nindi et al. (2014).

As one of Africa's most extensive wetlands, the Kilombero Valley has a long history of productive activities, primarily farming (International Water Management Institute (IWMI), 2014; Kato, 2007). The floodplain contributes substantially to the livelihood of more than 500 thousand people living there (National Bureau of Statistics, 2013) by providing crops, fish, drinking water, forest products, and fuelwood (Mombo, Speelman, Huylenbroeck, Hella, & Moe, 2011). In recent years a rapid increase in agricultural land use has been observed Jones et al. (2012). Immigration into the valley has increased dramatically due to the perceived availability of high quality and cheap farmland. Conflicts between pastoralists and farmers over land use are a chronic and widespread problem, which has resulted in injury and litigation disputes (Dinesen, 2016; MALF, 2015; Nindi et al., 2014).

4.3.2 Data

The core data source for our analysis is a household survey in 21 villages in two districts of the Kilombero Valley, Ulanga and Kilombero. In total, 304 farm households were interviewed to provide information on the farming systems in terms of resource use and management as well as their relevance for the livelihoods of the households. The household selection was based on a multi-stage sampling strategy. First, 11 wards were purposively selected based on the occurrence of floodplain farming. In the second stage, 21 villages were randomly selected using probability to population size within the wards. In the final stage, households were randomly selected from the list provided by village leaders. A GIS approach incorporating the land use map form GLC30 (Jun, Ban, & Li, 2014), the administrative boundary, and the 2012 census data from the Tanzania statistics office was used to estimate the boundaries and total population size in the study area. To drive the biophysical characteristics of farmer plots (slope, elevation, roughness), we used a Digital Elevation Model(DEM) at 90m resolution from SRTM (Jarvis, Nelson, & Guevara, 2008).

4.3.3 Variable selection

Generally, factors influencing the choice of intensification strategy can be grouped into three broad categories. Those are 1) socio-economic characteristics of the farm households, 2) infrastructure and institutional factors, and 3) the agro-ecological context of the farm (Wainaina et al., 2016). Studies often find that household resource endowments in terms of labor, land and capital, risk behavior and social capital affect the choice of a particular strategy (Erenstein, 2006; Feder et al., 1985; Ghadim, Pannell, & Burton, 2005; Okike et al., 2001). For example, increasing the frequency of cropping is often constrained by the availability of labor. Human capital gained through education, training, and experience influences the choices farmers to make (Kijima, 2016; Wainaina et al., 2016). Infrastructure and institutions like distance to nearest markets or access to credit and agricultural extension are shown to matter in many empirical studies (Feder et al., 1985; Kassie et al., 2018, 2015; Teklewold et al., 2013). Finally, biophysical characteristics of the farm plots such as slope, hydrological regime, soil characteristics affect the choice of one strategy over another (Khonje et al., 2018; Nkonya, Schroeder, & Norman, 1997; Sirén, 2007).

To avoid unnecessary model complexity, we use a data mining technique to select variables that are most important in explaining variation in choice of intensification. In the first step, we created a binary variable encoding a farmer as either an intensifier (if the farmer adopted at least one option) or a non-intensifier. Then we run a random forest algorithm between the binary choice variable and a range of explanatory variables using Scikit-learn (machine learning library in python) (Pedregosa et al., 2011). Random forest algorithms have recently become popular in variable selection. Please see (Genuer, Poggi, & Tuleau-Malot, 2015; J. Rogers & Gunn, 2006; J. D. Rogers & Gunn, 2005; Sandri & Zuccolotto, 2006) for details of their use and application in variable selection.

Our variable of interest (target node), intensification, is treated as a discrete node that contains four intensification options (use of improved seed variety, small scale irrigation, fertilizer application, and multiseason farming) and all possible combinations of these strategy bundles. Also, a state which captures the absence of an intensification strategy is included. Since only five strategy sets were observed in our data set, the unobserved strategy combinations are represented by one state called others. Here, we can take advantage of BBNs ability to update the conditional probabilities once new data become available.

We used per capita income as a surrogate for a resource endowment and availability of capital. It comprises income from Agriculture (farming and fishing), income from off-farm activities, income from land rental, and from brick making. We also include farmer type to represent the farming system followed by a particular household. The farm type variable is a typology constructed through a combination of principal component analysis and hierarchical clustering in order to stratify farmers into clusters that are homogeneous according to their livelihood and land use (Husson, Lê, & Pagès, 2017). In order to capture the quality and hydrological characteristics of the farm, we generated a topographic wetness index using a digital elevation model of our study area (slope and upslope contributing area). The index provides an indication of the relative wetness within the catchment and is highly correlated with soil moisture and ground level water (Sörensen, Zinko, & Seibert, 2006). Expected prices will affect crop choices and expected income. We included prices received by the household for rice and maize (which is dependent on the distance from the market) as expected prices for the two crops. Due to a lack of past price data, here we assume farmers expect what they received during the planting period. Our distance variable measures the distance from the farm to the nearest big market in km. Measuring the distance from the farm rather than from homestead takes in to account the access and cost of transport from snd to the farm. While homesteads are clustered around the single main road, farms which are located bottom valley are quite far from homesteads.

Although BBNs are generally capable of handling continuous nodes, we followed a common practice of discretizing all continuous variables into smaller classes to reduce the complexity and high computational requirement of training the BBN (Frayer et al., 2014; Sun & Müller, 2013). There are several ways to discretize continuous variables. We partition our continuous variables using a heuristic method called 'equal frequencies' where the variable is transformed to K equal lengths or width (Clarke & Barton, 2000; Nojavan A., Qian, & Stow, 2017).

Table 4.1 presents the descriptive statistics of our target node and 15 evidence nodes that are included in the final network. Regarding the target node (intensification choice), around 62% of our sample households did not intensify their production, whereas 38% of the households have adopted one or more of the intensification options. 12% of farmers have used improved seed varieties, 8% are planting in both short and long rainy season, 7% use chemical fertilizers, 6.73% use irrigation, and chemical fertilizer combined, while only 3.7% use irrigation. Rice is the dominant crop planted by farmers in our survey. Around 42% of the farmers' plant rice as a mono-crop, and 27% produce rice in combination with maize, and 7.7% produce vegetables in addition to that. The majority of households are small scale mono-crop producers (60%), 30% are small scale farmers that diversify in terms of their land use, and 10% represent large scale agopastoralists who practice crop production and livestock keeping. 70% of the farmers own less than 3 hectares of cropland and 18% between 3 and 6 hectares. Around 12 % own more than 6 hectares. We can also see from the table that farmers generally participate in the output market, 41% of the farmers marketed between 30 and 60% of their output to the market and 34% sold more than 60 percent of their production. Only 33 percent of our sampled households have access to credit, and on average non-farm income accounts 8 percent of the total income. Households are, on average, 22 km from the nearest big market located in the town of Ifakara. The average household size for the entire sample was

5 (SD=2.18), with a minimum of 2 members and a maximum of 11 members. 44% of respondents have a family size of fewer than four members, which can be considered as a small family. And 41% are medium-sized with 5-8 members. 12% of households in the sample are extended families, with more than eight members.

	Variable	Unit	Mean (SD)
1	Age of the Household Head	Year	46.41 (12.82)
2	Household Size	Number	5.13 (2.18)
3	Farm Size	Hectares	2.64 (2.83)
4	Share of Hired Labor	Percent	36.41 (32.98)
5	Commercialization Index	Percent	46.87 (24.99)
6	Total Labor	Man-Days(Ha^{-1})	321.94 (343.58)
7	Share of Non-farm Income	Percent	8.1 (18.37)
8	Topographic Wetness Index	Index	20.73 (4.49)
9	Distance to the nearest big Market	Km	22.1 (16.44)
10	Maize Price	(Tsh)(kg^{-1})	382.01 (425.44)
11	Rice Price	(Tsh)(kg^{-1})	1232.38 (260.49)
12	Income	000' (Tsh)	501.58 (1048.67)
	Categorical Variables		Frequency (%)
13	Intensification options 1. Apply Fertilizer 2. Apply Improved Seed 3. Crop Multiple Times 4. None 5. Use Irrigation 6. Use Irrigation + Fertilizer Application		21 (7.1%) 35 (11.8%) 24 (8.1%) 186 (62.6%) 11 (3.7%) 20 (6.7%)
14	Farm Type 1. Agro-Pastoralist 2. Diversifier 3. Mono-Crop Rice Producers		21 (7.1%) 81 (27.3%) 195 (65.7%)
15	Credit Access 1. No 2. Yes		199 (67.0%) 98 (33.0%)
16	Crop Choice 1. Maize 2. Rice 3. Rice + Maize 4. Rice + Maize + Vegetables 5. Vegetables + Rice		5 (1.7%) 148 (49.8%) 100 (33.7%) 20 (6.7%) 22 (7.4%)

TABLE 4.1: Descriptive statistics for the variables included in the final network

4.4 Methodology to empirically specify, validate and interpret the BBN

Once variables are selected, the empirical BBN can now be specified, validated, and interpreted. Figure 4.3 presents the workflow. The specification or learning of a BBN involves two steps, structure (1) and parameter (2) learning, which correspond to model selection and parameter estimation in classic statistical models, respectively (Koller & Friedman, 2009; Nagarajan, Scutari, & Lèbre, 2013). The learning is followed by k-fold stratified cross-validation (3) and a sensitivity analysis allowing to interpret the final specification of the BBN (4). In the following section, we will highlight each step-in detail.

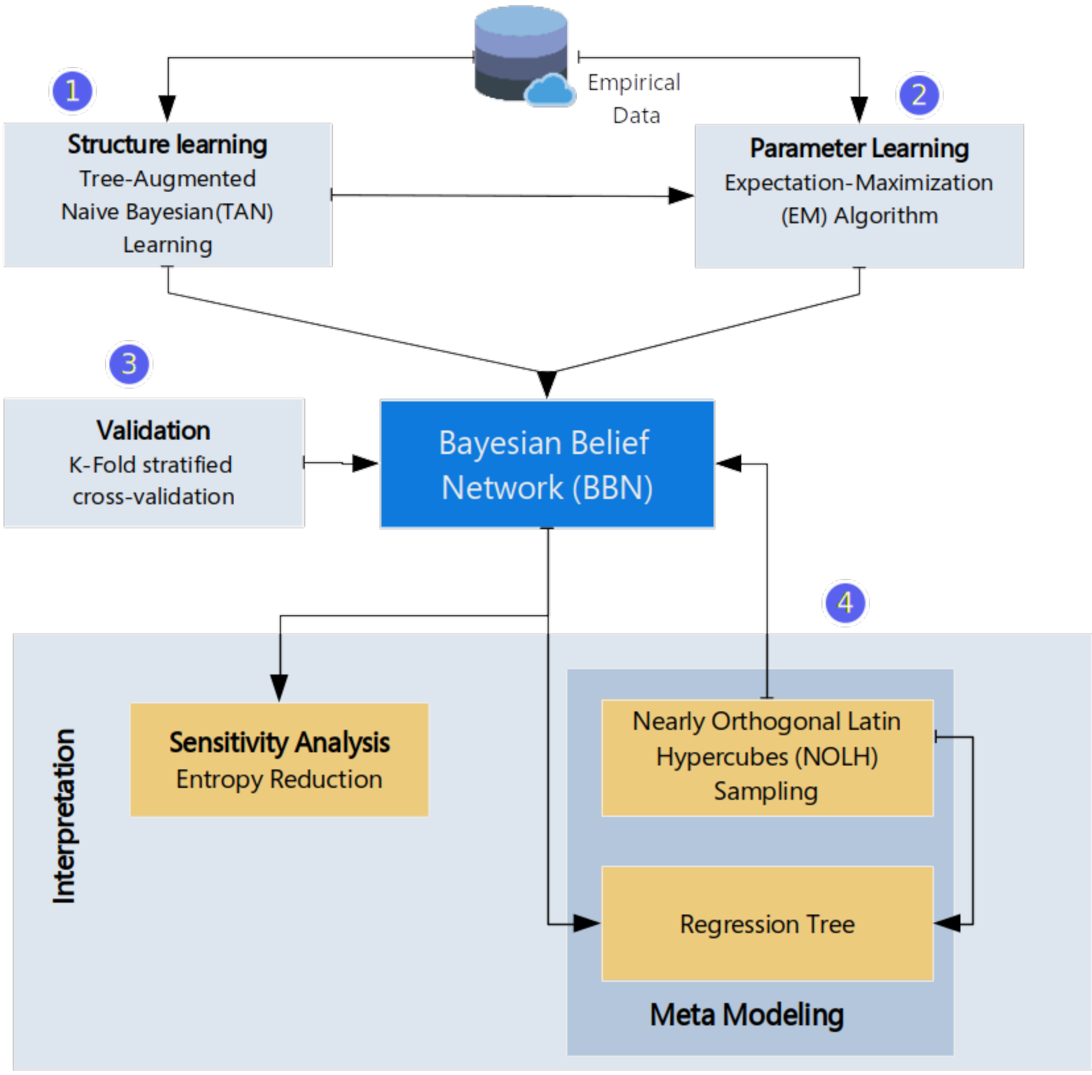


FIGURE 4.3: Workflow of BBN modelling, DOE & Regression Tree

4.4.1 Structure learning

There are generally two different approaches to build the structure of a BBN. The structure can be learned through knowledge engineering from experts, literature, and theory or learned from empirical data. In this study, we use the data-based approach. However, following Sun and Müller (2013), we augmented the empirical approach with theory and plausibility, as explained further below.

The empirical approach of learning a BBN involves finding dependencies between variables that result in a distribution as close as possible to the observed data in the probability space. For a detailed explanation of algorithms learning the structure of a Bayesian network from data (see Koller & Friedman, 2009; Nielsen, Keil, & Zeller, 2013).

There exist two general classes of algorithms for learning the structure of a Bayesian network from a data:

Constraint-based structure learning: The approaches view a Bayesian structure learning network as a representation of interdependencies. Based on statistical tests (such as chi-squared or mutual information), the approach chooses conditional dependence and independence and uses these relationships as constraints to construct a BBN (Koller & Friedman, 2009; Neapolitan, 2010). Algorithms under this category include 'inductive causation,' 'grow-shrink,' and 'incremental association.'

Score-based structure learning: This is an optimization-based search that treats a Bayesian network analogous to specifying a statistical model. It produces potential candidates of Bayesian networks, calculates a score for each candidate, and returns a candidate with the highest score (Kjaerulff & Madsen, 2012; Nielsen et al., 2013).

For our study, we adopted a Tree-Augmented Naive (TAN) Bayesian network, which is a variant of score-based structure learning (Friedman et al., 1997). It relaxes the naive Bayes attribute independence assumption by imposing constraints on the network structure and chooses the tree that maximizes the likelihood of the training data (Koller & Friedman, 2009; NorsysSoftwareCorp, 2016; Zheng & Webb, 2010). According to Friedman et al. (1997), learning the structure of a network using TAN embodies a good trade-off between quality of estimation and computational complexity. Once the structure of the network is constructed with TAN, we iteratively modify the structure based on theoretical knowledge, as suggested by (Frayer et al., 2014; Neapolitan, 2010; Prishchepov et al., 2019; Sun & Müller, 2013). During this process, the direction of some of the links is reversed (for example, the link from prices to distance to market was changed from distance to market to prices), and structure learning is repeated with the newly constrained link there after. The structure was developed with Netica (5.4) (NorsysSoftwareCorp, 2016) [1].

[1]Netica provides a number of simplifying tasks for the modeler including a high visual capability to display the network and advanced algorithms to learn the structure and parameters of the

4.4.2 Parameter learning

As for the structure learning of BBN, there are also several alternative ways of parameter learning, i.e., generating estimates for the Conditional Probability Tables (CPTs). In this study, the probabilities were derived from the survey data using a maximum likelihood procedure. We opted for learning the CPTs from data to reduce the number of conditional probability entries required from experts or literature and also to learn objective rather than subjective probabilities. There are three commonly used algorithms to solve the underlying maximum likelihood problem: Count learning, expectation-maximization (EM), and gradient descent (Frank, 2015; NorsysSoftwareCorp, 2016). Although we explored all the three algorithms, our final network is based on EM learning as it converges more robustly and provides better calibration to our data.

4.4.3 Validation of the BBN

Validating the BBN is vital to ensure the quality of the model. There are several ways to check the validity of the constructed BBN, both quantitatively and qualitatively. Qualitative ones check the validity using an expert opinion (Celio & Grêt-Regamey, 2016; Frank, 2015). Quantitative validation uses a test data set not employed in parameter learning to check the quality of predictions of the target variable. We validate our BBN using a quantitative validation technique. We perform five-fold cross-validation by partitioning our data into five disjoint subsets followed by an iterative validation for the five sets using confusion matrix metrics. To take the unbalanced nature of our target node into account and to ensure that all the states are equally represented in the split, we use the stratified cross-validation tool from Caret (Classification And Regression Training) (Kuhn, 2008).

4.4.4 Sensitivity analysis

We use sensitivity analysis to measure changes in the probabilities of the target node with changes in states of one of the input nodes (Pollino, Woodberry, Nicholson,

network. Using the Java API for the construction of BBNs provides an advantage in terms of transparency, reproducibility and easy integration with other modeling tools of interest as it resulted in Java source code. The full documentation and Java code, Python code for Regression tree and the data are stored in Github (`https://bsrthyle.github.io/RepoForBBNpaper/`)

Korb, & Hart, 2007). The sensitivity analysis serves to identify the significant and informative variables that affect the choice of intensification options (Sun & Müller, 2013). Since the input nodes required for the sensitivity analysis contain discrete values, the Entropy Reduction (Mutual Information) method is used here to determine the sensitivity of the BBN model's output to variations in a particular input parameter. The entropy reduction method computes the expected reduction in entropy of the target node (i.e., the increase of information) due to findings at another child node. It is calculated as (Marcot, Steventon, Sutherland, & McCann, 2006; NorsysSoftwareCorp, 2016; Pearl, 1988):

$$I = H(Q) - H(Q|F) = \sum_{q} \sum_{f} \frac{P(q,f)\log_2[P(q,f]}{P(q)P(f)} \tag{4.2}$$

where $H(Q)$ and $H(Q|F)$ are the entropy of Q before and after any new findings respectively.

4.4.5 Meta-modelling

Although the sensitivity analysis provides insight regarding the importance of influencing factors for the choice of an intensification strategy, it does not tell us how the probability of each strategy is influenced by the factors included in our model. To deliver such information, we introduce a new approach of global sensitivity analysis combining a BBN simulation based on the Design of Experiment (DEO) with a meta-modeling approach. To generate sample configurations of the evidence nodes, we utilize Nearly Orthogonal Latin Hypercube sampling (NOLH) to draw from the probability distributions at the evidence nodes (Sanchez, 2005). We provide the drawn values to the network as evidence, and we record the probabilities of each strategy for each sample point. To summarize the effect of the different nodes on the variation of the probabilities of each strategy, we followed a meta-modeling approach by applying a regression tree for each of the options (Coutts & Yokomizo, 2014). The advantages of the regression tree are the incorporation of non-linear interactions, the minimal assumptions about the structure of the data needed, their robustness to outliers, and the implicit handling of variable selection (Coutts & Yokomizo, 2014; Kuhn & Johnson, 2013). Besides, it provides an analogy for natural rule induction from the results. The regression trees were implemented using Scikit-learn (machine learning tool in python) (Pedregosa et al., 2011).

4.5 Result and discussion

Figure 4.4 presents the final learned Bayesian belief network determining intensification decisions in KVF. The structure of the network highlights some essential relations between variables conditioned on the choice of intensification strategy. By construction (TAN structure learning algorithms), all variables are associated with the choice of intensification strategy. Also, some of the input nodes are correlated. For example, the age of the household head is correlated with farmers' commercialization index, access to credit, and crop choices. The share of hired labor is also associated with the size of cropland and household size. The topographic wetness index of the plot is mainly affecting the crop choices of the farmer.

Moreover, the distance from the nearest big market directly influences the expected prices of rice and maize and the amount of output sold to the market. The size of cultivated land is directly associated with a share of hired labor and total labor use in man-days. Another association exists between per capita income, market participation, the percentage of hired labor, and access to a non-farm source of income.

The BBN also shows the posterior probabilities learned from the data using the EM algorithm. The values of the probabilities are in line with the data, which shows that the parameter learning algorithm can learn the underlying joint distribution and conditional probability tables given the network structure. However, to fully assess the quality of both structure and parameter learning, we also conducted a fivefold stratified cross-validation (see model validation section above). The resulting confusion matrix from the five cross-validations resulted in an average error rate of 45% with incorrectly predicting the choice of intensification strategy.

Figure 4.5 presents the results obtained from the sensitivity analysis. The figure compares the contribution of each variable to the total expected entropy reduction of our target node *ChoiceOfIntensificationStrategy*. A closer inspection of Figure 4.5 indicates that crop choices have the most considerable influence on the choice of

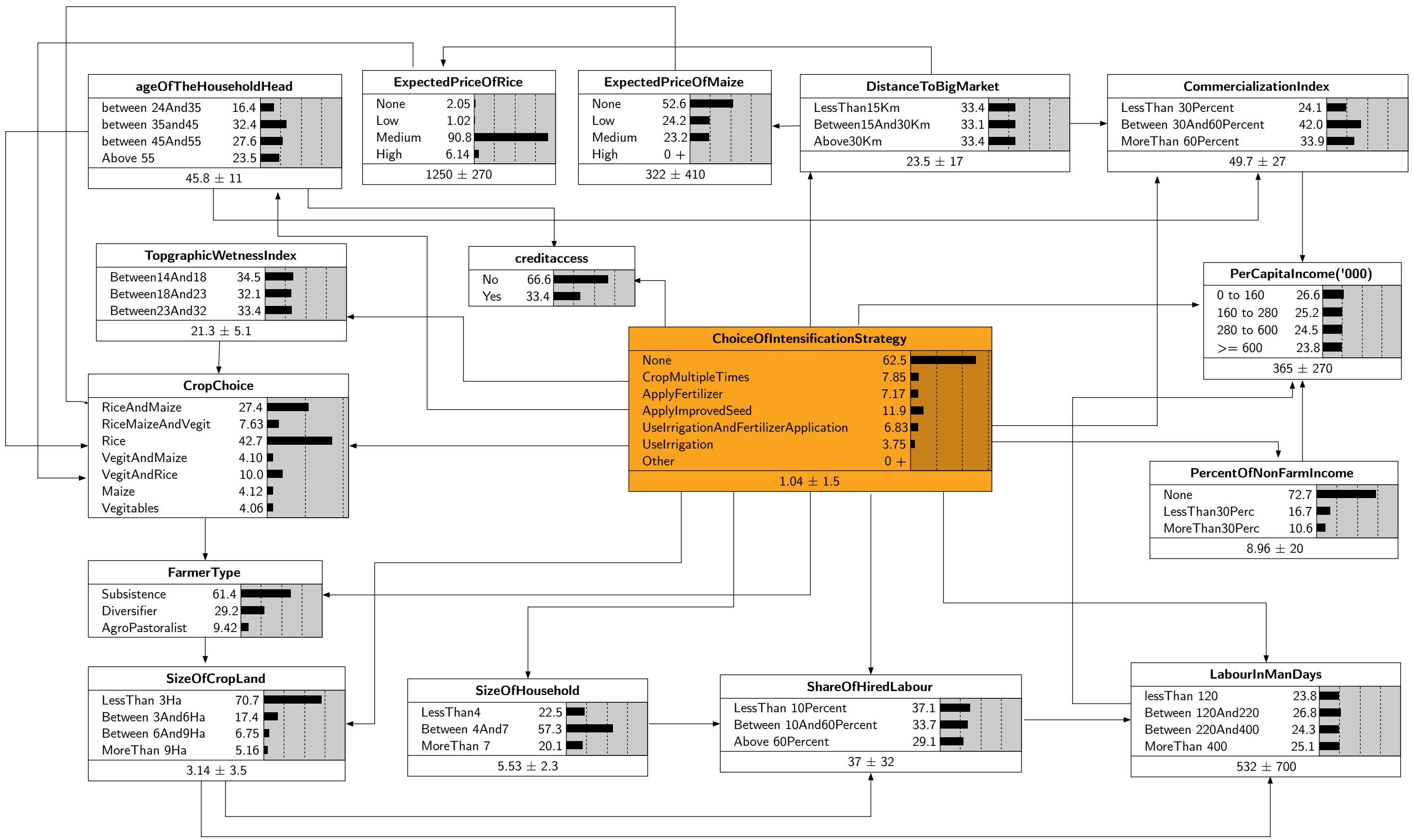


FIGURE 4.4: A BBN of intensification decision in KVF

Note: In order to reduce the complexity of the CPT and easy learning of the parameters, the links leave the target node rather than pointing to it. This is ok as BBN uses Bayesian inference to make prediction or diagnostics (NorsysSoftwareCorp, 2016) intensification strategy with a 5.87% reduction in entropy, followed by the distance from the nearest market with a 3.8% variance reduction. Also, per capita income, the share of income from the non-farm activity, age, farmer type, the percentage of hired labor and topographic wetness index, size of cropland, and household size influence the variation in the choice of intensification strategy. All other variables show less than a 1% reduction in entropy. Given that rice is the main crop produced in the area, the variations in crop choice are more or less dependent on the mixed cropping of rice with maize and vegetable. Per capita income is a surrogate for farmer's endowment and their ability to invest additional resources required for adopting new strategies.

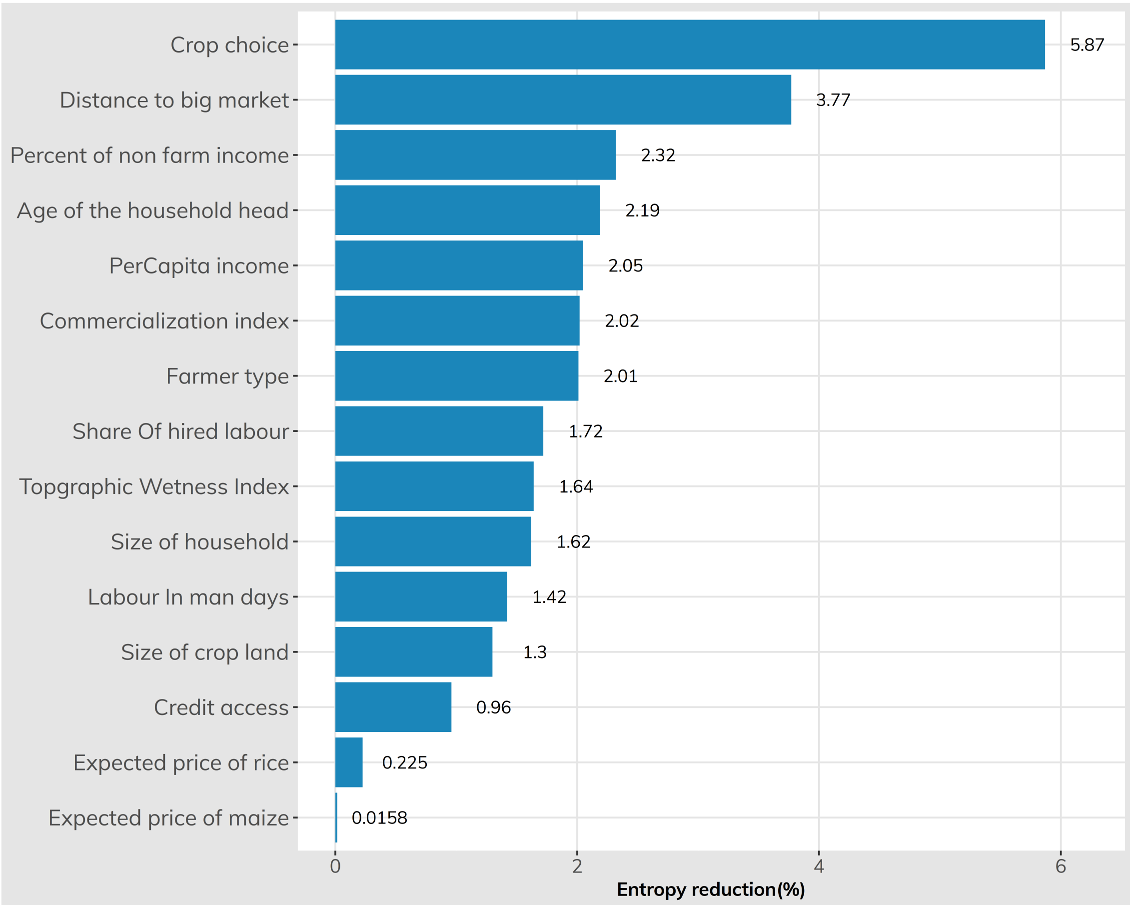


FIGURE 4.5: Sensitivity analysis of intensification decision

Distance from the farm to the nearest market and commercialization also have a strong influence on the choice of the intensification strategy. This is in line with

Erenstein (2006), who found that access to the market has an effect on both access to key input and output markets and thereby significantly affects intensification. Although the availability of family and hired labor is considered a crucial determinant of intensification choice (Lee, 2005; Wainaina et al., 2016), the sensitivity analysis shows only a moderate influence in our case.

A subsequent global sensitivity and meta-modeling analysis uncover factors leading to the choice of a particular intensification strategy. A general finding from the meta-modeling analysis is that each option is influenced differently by the determinants. Although the variations in the probabilities of options are affected by a common set of variables, the magnitude and order of the effects are different across the given option. Figure 4.6 shows a variable importance plot for the regression trees for each intensification strategy. The variable importance is calculated based on the (normalized) total reduction of the residual sum of squares (RSS) brought by a specific variable. A large value indicates a vital predictor (James, Witten, Hastie, & Tibshirani, 2013).

More so, the feature importance from the regression tree reveals that the variation in probabilities of choosing cropping multiple times is captured by variation in total labor available during the year, commercialization index, topographic wetness index, income, and distance to the central market. The variations in the probabilities of fertilizer application are only affected by the topographic wetness index if the farmer is diversifier, age, commercialization, and distance to the market. The use of improved seeds is influenced by the share of non-farm income, age, household size, distance to the market, and farm size. The probability of using irrigation and fertilizer application is affected by proximity to the market, farm size, the share of non-farm income, and topographic wetness index. The variation in probabilities of use of irrigation is affected by variation in topographic wetness index, non-farm income, farm size if the farmer is of type subsistence, and availability of labor.

The regression trees allow us to see how specific attributes of farmers contribute to the probability of choosing a particular strategy. Table 4.2 presents the ranges of variable values, which lead to the highest and lowest probability of selecting a specific option. We opted to present those extreme cases in order to draw a picture of variable importance and critical thresholds. A complete representation of the regression tree for each option is provided in appendex 4.9,4.8,4.11,4.12,4.10.

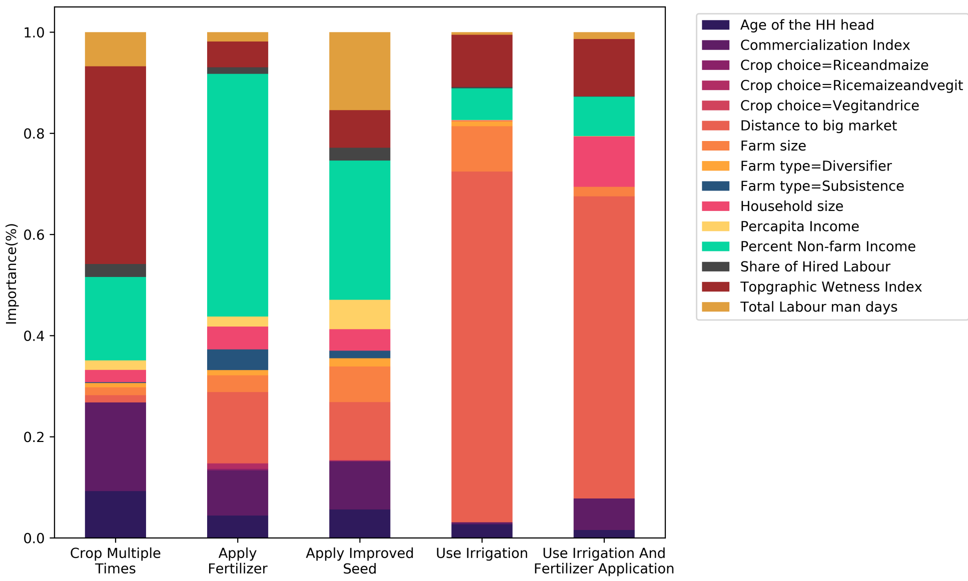


FIGURE 4.6: Variable importance for intensification decisions

Note: The data underlining the regression trees are based on posterior probabilities from BBN using 257 sample points generated through NOLH design that captured the structure of the conditional distribution between determinants.

A farm household with the highest probability of cropping multiple seasons is characterized by fields located in relatively wetter areas (higher than 18.9 TWI), relatively far from the market (more than 44km), and is selling between 36 to 66 percent of her output. Contrary, a farm with the lowest probability of cropping multiseason is located in a relatively drier area (below 18.9 wetness index), sells between 0% and 66% of crop output, is located in a distance less than 44 km from the market, has below 615 labor man-days and higher income (above 228,824 Tsh). Households with the highest probability of adopting improved seed, are characterized by less than five household members , which is located more than 23 km away from the biggest market and cultivates more than 10 ha. A farm household with the lowest probability of adopting improved seed variety has more than five household members, a household head younger than 51 years, sells less than 72 % of his crop output, and owns more than 9 ha of land.

As expected, farm households with the highest probability of using small scale irrigation are characterized by available labor of more than 110 man-days. They are not mono crop rice producers and have a topographic wetness index of less than 28. On the other hand, farm households with the lowest probability of using small scale irrigation are characterized by the availability of more than 110 labor man-days, are non-mono-crop rice producer, have a TWI of less than 28.7, receiving less than 31 percent of income from non-farm sources and a household size greater than 5. Farm households with farm size greater than 3.56 hectares and a TWI of less than 17.425 will have a 27 percent probability of using fertilizer(highest probability). Furthermore, those who cultivate less than 3.56 ha and are located less than 35 km from the market will have the lowest probability (7 percent) of using fertilizer. Farm households with the highest probability (22 percent) to combine small-scale irrigation with fertilizer application receive more than 38% of non-farm income, have a farm size greater than 7 hectares, and are located less than 22 km from the market.

Cropping multiple times	
High Probability = **0.338**	Low Probability = **0.107**
Topographic wetness index $>$18.9	Topographic wetness index $>$18.9
Commercialization Index $<$66.6%	Commercialization Index $<$66.6%
Distance from market $>$44 km	Distance from market $\leqq$ 44 km
Commercialization Index $<$36.14%	Total labor (man days) $\leqq$ 615
	Income ('000 Tsh) $>$228
Improved seed	
High Probability = **0.282**	Low Probability = **0.051**
Household size $<$5	Household size $>$4
Distance to the market $>$=23 km	Age of household head $\leqq$ 51 years
Farm size $>$10 ha	Commercialization Index $<$72 %
	Farm size $>$9 ha
Small-scale irrigation	
High Probability = **0.058**	Low Probability = **0.008**
Labor in Man-days $>$110	Labor in Man-days $>$110
Farm Type $\neq$ Mono-crop rice producer	Farm Type $\neq$ Mono-crop rice producer
Topographic wetness index $>$28.78	Topographic wetness index $<$28.78
	Share of non-farm income $<$30 %
	Household size $>$5
Fertilizer Application	
High Probability = **0.276**	Low Probability = **0.051**
Farm size $>$5 ha	Farm size $\leqq$ 5 ha
Topographic wetness index $<$28	Distance to the market $\leqq$ 35 km
Irrigation and fertilizer	
High Probability= **0.223**	Low Probability = **0.0228**
Share of non -farm income $>$38.13%	Share of non -farm income $\leqq$ 38.13%
Farm size $>$7.4 ha	Topographic wetness index $<$28
Distance to the market $>$22 km	Farm Type $\neq$ diversifier
	Share of non-farm income $>$30 %

TABLE 4.2: Factors discriminating highest & lowest probability of choosing a specific intensification option

Finally, we also qualitatively examined the perceived constraints of intensification by farmers in the valley. Figure 4.7 presents the main constraints reported by the farmers for each option. Regarding improved seed adoption, almost half of the farmers reported high market demand for traditional variety as the main reason for not using improved seed. Farmers also reported high prices (20.7 %), lack of financial capital (15.4%), high transaction cost (9.47%), and uncertainty of quality (5.33%) as the main reason for not using improved seed varieties. Reported reasons for not cropping multiple seasons include erratic rainfall (51.1%) and labor shortage (36.2%). The majority of the farmers reported a high price of fertilizer as the primary constraint for adoption. The main reasons for not using fertilizer are the high cost of fertilizers (54.8%) and the belief that their soil is fertile (38.7%). Lack of irrigation tools (49.1%), limited access to water (25.7%), enough water on the plot (21.6 %), and labor shortage (3.59 %) are reported as the main reason for not using irrigation.

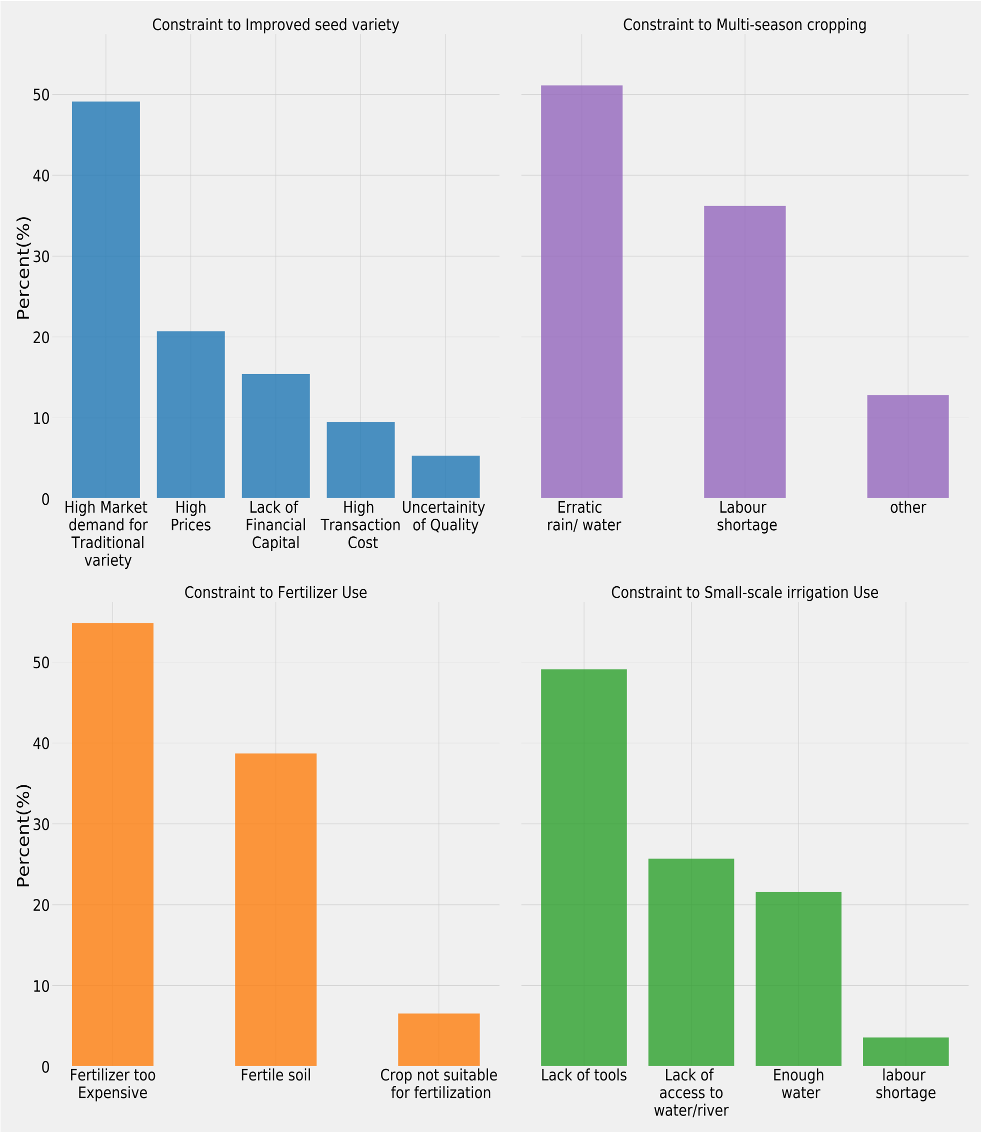


FIGURE 4.7: Perceived constrain for specific intensification option

4.6 Conclusion

Increasing population pressure, efforts to protect the fragile wetland biome and the need to raise food production will stimulate smallholder farmers in the KVF to shift from expansion of land to intensification through the adoption of land-saving options. Although the adoption of these options has been increasing gradually (Otsuka & Larson, 2016a), they are still relatively small in number, and there is significant room for intensification by smallholders in the valley. We presented a systematic analysis designed to explore the different factors that affect the choice of intensification strategy in the KVF by considering a diversity of options that are used in floodplain production systems. Our data-driven BBN enables us to represent complex interactions and dependencies between different factors and intensification strategy choices while taking uncertainties into account. Through our augmented sensitivity analysis, we elucidated the individual determinants of choices for improved seed variety, fertilizer application, small scale irrigation, and of increasing the frequency of cropping in KVF. We included different covariates capturing plot and household characteristics, access to market, and agro-ecological conditions. Our novel approach of combining a data-driven Bayesian Belief Network (BBN) with regression trees has provided us with strategy specific factors. Although the choice of each option is affected differently by covariates under consideration, access to non-farm income, access to market, and topography of the plot play essential roles across options.

Smallholder farmers who did not adopt any of the intensification options, despite their willingness to increase their yield, have expressed different reasons. The main reason for not cropping multiseason is insufficient rain or lacking access to water and labor shortage. Higher fertilizer prices and higher consumer demand for traditional seedlings are reported as the main reasons for not using fertilizer and improved seed, respectively — access to river network and tools for irrigation limits farmers to irrigate their plot. There are some limitations to our study, and one should interpret our findings in light of these limitations. They arise from a limited dataset for training and validation of our BBN. Also, combinations of intensification options are rarely observed in our sample. Farmers might have other options that they uptake, which are not considered in our study. The BBN presented in this study is static and does not take dynamics into account. To this end, a further study is envisioned that combines the BBN with a spatially explicit agent-based model (ABM) that considers the heterogeneity of the farmers, their interaction among themselves and with the floodplain. Combining BBN with an ABM reduces the computational challenges of ABMs by providing probabilistic agent rules. Representing BBN nodes as state variables in the ABM will provide temporal dynamics to the BBN approach (Kocabas & Dragicevic, 2013; Sun & Müller, 2013). Since the project under which this study is conducted is still ongoing, we will update the BBN when more data is available and also verify it with stockholders in the valley.

4.7 References

Aalders, I. (2008). Modeling Land-Use Decision Behavior with Bayesian Belief Networks. *Ecology and Society*, *13*(1). doi: 10.5751/ES-02362-130116

Abay, K. A., Berhane, G., Taffesse, A. S., Koru, B., Abay, K., & Others. (2016). Understanding farmers' technology adoption decisions: Input complementarity and heterogeneity. *Ethiopia Strategy Support Program Working Paper*, *82*.

AFDB. (2015). *African Development Report 2015: Growth, Poverty and Inequality Nexus: Overcoming Barriers to Sustainable Development* (Tech. Rep.).

Agra. (2016). *Africa Agriculture status report:Progress towards Agricultural Transformation in Africa* (Tech. Rep.). Green Revolution in Africa (AGRA).

Alomia-Hinojosa, V., Speelman, E. N., Thapa, A., Wei, H.-E., McDonald, A. J., Tittonell, P., & Groot, J. C. J. (2018, jan). Exploring farmer perceptions of agricultural innovations for maize-legume intensification in the mid-hills region of Nepal. *International Journal of Agricultural Sustainability*, *16*(1), 74–93. Retrieved from https://www.tandfonline.com/doi/full/10.1080/14735903.2018.1423723 doi: 10.1080/14735903.2018.1423723

Bamford, A., Ferrol-Schulte, D., & Smith, H. (2010). The status of the Ruipa corridor between the Selous Game Reserve and the Udzungwa Mountains. *Report to Frontier Tanzania, Dar es Salaam, Tanzania*.

Besley, T., & Case, A. (1993). Modeling technology adoption in developing countries. *The American Economic Review*, *83*(2), 396–402.

Binswanger-Mkhize, H. P., & Savastano, S. (2017). Agricultural intensification: The status in six African countries. *Food Policy*, *67*, 26–40. Retrieved from http://dx.doi.org/10.1016/j.foodpol.2016.09.021http://www.ncbi.nlm.nih.gov/pubmed/28413244http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5384439 doi: 10.1016/j.foodpol.2016.09.021

Cain, J. (2001). Improvements in Natural Resources Water Sector and Beyond. *Ecology*, *44*(0), 124. Retrieved from http://nora.nerc.ac.uk/9461/

Celio, E., & Grêt-Regamey, A. (2016). Understanding farmers' influence on land-use change using a participatory Bayesian network approach in a pre-Alpine region in Switzerland. *Journal of Environmental Planning and Management*, *59*(11), 2079–2101. doi: 10.1080/09640568.2015.1120713

Chen, S. H., & Pollino, C. A. (2010). Guidelines for good practice in Bayesian network modelling. *Environmental Modelling and Software Modelling*, 16. Retrieved from http://www.iemss.org/iemss2010/papers/S00/S.00.07.GuidelinesforgoodpracticeinBayesiannetworkmodelling-SERENACHEN.pdfhttp://www.iemss.org/iemss2010/papers/S00/S.00.07.GuidelinesforgoodpracticeinBayesiannetworkmodelling-SERENACHEN.pdf

Clarke, E. J., & Barton, B. A. (2000). Entropy and MDL discretization of continuous variables for Bayesian belief networks. *International Journal of Intelligent Systems*, *15*(1), 61–92.

Coulson, A. (2015). Small-scale and large-scale agriculture: Tanzanian experiences. In M. Stahl (Ed.), *In looking back, looking ahead – land, agriculture and society in east africa* (pp. 44–73). Nordic Africa Institute.

Coutts, S. R., & Yokomizo, H. (2014). Meta-models as a straightforward approach to the sensitivity analysis of complex models. *Population Ecology*, *56*(1), 7–19. doi: 10.1007/s10144-013-0422-1

Daly, R., Shen, Q., & Aitken, S. (2011). Learning Bayesian networks: approaches and issues. *The knowledge engineering review*, *26*(2), 99–157.

Diao, X., Hazell, P., & Thurlow, J. (2010). The role of agriculture in African development. *World development*, *38*(10), 1375–1383.

Dinesen, L. (2016). Kilombero Valley Floodplain (Tanzania). In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The wetland book: Ii: Distribution, description and conservation* (pp. 1–8). Dordrecht: Springer Netherlands.

Dorfman, J. H. (1996). Modeling multiple adoption decisions in a joint framework. *American Journal of Agricultural Economics*, *78*(3), 547–557.

Drury, B., Valverde-Rebaza, J., Moura, M.-F., & de Andrade Lopes, A. (2017, oct). A survey of the applications of Bayesian networks in agriculture. *Engineering Applications of Artificial Intelligence*, *65*, 29–42. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0952197617301513https://www.sciencedirect.com/science/article/abs/pii/S0952197617301513

Erenstein, O. (2006). Intensification or extensification? Factors affecting technology use in peri-urban lowlands along an agro-ecological gradient in West Africa. *Agricultural Systems*, *90*(1-3), 132–158. doi: 10.1016/j.agsy.2005.12.005

Feder, G., Just, R. E., & Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic development and cultural change*, *33*(2), 255–298.

Foster, A. D., & Rosenzweig, M. R. (2010, sep). Microeconomics of Technology Adoption. *Annual Review of Economics*, *2*(1), 395–424. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev.economics.102308.124433 doi: 10.1146/annurev.economics.102308.124433

Frank, S. (2015). *Expert-based bayesian network modeling for environmental management*. Retrieved from https://books.google.de/books?id=GDHisgEACAAJ

Frayer, J., Sun, Z., Müller, D., Munroe, D. K., & Xu, J. (2014). Analyzing the drivers of tree planting in Yunnan, China, with Bayesian networks. *Land Use Policy*, *36*, 248–258.

Friedman, N., Geiger, D., Goldszmidt, M., Division, C. S., Park, M., Langley, P., ... Goldszmidt, M. (1997). Bayesian Network Classifiers *. *Machine learning*, *163*(2-3), 131–163.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, *7*(2), 19–33. Retrieved from https://hal.archives-ouvertes.fr/hal-01251924/

Ghadim, A. K. A., Pannell, D. J., & Burton, M. P. (2005, jul). Risk, uncertainty, and learning in adoption of a crop innovation. *Agricultural Economics*, *33*(1), 1–9. Retrieved from http://doi.wiley.com/10.1111/j.1574-0862.2005.00433.x doi: 10.1111/j.1574-0862.2005.00433.x

Gladwin, C. H. (1989). Modeling farmers' decisions to change: using cognitive science in the desgin of agricutural technology. *Social Science perspectives on managing agricultural technology*, 127–141. Retrieved from http://publications.iwmi.org/pdf/H005732.pdf

Headey, D., Dereje, M., & Taffesse, A. S. (2014). Land constraints and agricultural intensification in Ethiopia: A village-level analysis of high-potential areas. *Food Policy*, *48*, 129–141. Retrieved from http://dx.doi.org/10.1016/j.foodpol.2014.01.008 doi: 10.1016/j.foodpol.2014.01.008

Holloway, G., & Lapar, M. L. A. (2007). How big is your neighbourhood? Spatial implications of market participation among filipino smallholders. *Journal of Agricultural Economics*, *58*(1), 37–60. doi: 10.1111/j.1477-9552.2007.00077.x

Howley, P., O. Donoghue, C., & Heanue, K. (2012). Factors Affecting Farmers'

Adoption of Agricultural Innovations: A Panel Data Analysis of the Use of Artificial Insemination among Dairy Farmers in Ireland. *Journal of Agricultural Science*, *4*(6), 171–179. Retrieved from http://www.ccsenet.org/journal/index.php/jas/article/view/10687 doi: 10.5539/jas.v4n6p171

Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory multivariate analysis by example using r*. CRC press.

International Water Management Institute (IWMI). (2014). *Wetlands and people.* IWMI. Retrieved from http://www.iwmi.cgiar.org/Publications/Books/PDF/wetlands-and-people.pdf doi: 10.5337/2014.202

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-1-4614-7138-7 doi: 10.1007/978-1-4614-7138-7

Jarvis, A., Nelson, A., & Guevara, E. (2008). *Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).* Retrieved 2012-02-09, from http://srtm.csi.cgiar.org

Jayne, T. S., Mather, D., & Mghenyi, E. (2010, oct). Principal Challenges Confronting Smallholder Agriculture in Sub-Saharan Africa. *World Development*, *38*(10), 1384–1398. doi: 10.1016/j.worlddev.2010.06.002

Johansson, E. L., & Abdi, A. M. (2019). Mapping and quantifying perceptions of environmental change in Kilombero Valley, Tanzania. *Ambio*. doi: 10.1007/s13280-019-01226-6

Jones, T., Bamford, A. J., Ferrol-Schulte, D., Hieronimo, P., McWilliam, N., & Rovero, F. (2012). Vanishing Wildlife Corridors and Options for Restoration: A Case Study from Tanzania. *Tropical Conservation Science*, *5*(4), 463–474.

Jun, C., Ban, Y., & Li, S. (2014). China: Open access to Earth land-cover map. *Nature*, *514*(7523), 434.

Kajisa, K. (2016). On the Determinants of Low Productivity of Rice Farming in Mozambique: Pathways to Intensification. In *In pursuit of an african green revolution* (pp. 13–38). Tokyo: Springer Japan. Retrieved from http://link.springer.com/10.1007/978-4-431-55693-0{_}2 doi: 10.1007/978-4-431-55693-0_2

Kanu, B. S., Salami, A. O., & Numasawa, K. (2014). Inclusive growth: an imperative for african agriculture. *African Journal of Food, Agriculture, Nutrition and Development*, *14*(3), A33–A33.

Kassie, M., Marenya, P., Tessema, Y., Jaleta, M., Zeng, D., Erenstein, O., & Rahut, D. (2018, feb). Measuring Farm and Market Level Economic Impacts of Improved Maize Production Technologies in Ethiopia: Evidence from Panel Data. *Journal of Agricultural Economics*, *69*(1), 76–95. Retrieved from http://doi.wiley.com/10.1111/1477-9552.12221 doi: 10.1111/1477-9552.12221

Kassie, M., Teklewold, H., Jaleta, M., Marenya, P., & Erenstein, O. (2015). Understanding the adoption of a portfolio of sustainable intensification practices in eastern and southern Africa. *Land Use Policy*, *42*, 400–411. Retrieved from http://dx.doi.org/10.1016/j.landusepol.2014.08.016 doi: 10.1016/j.landusepol.2014.08.016

Kato, F. (2007). Development of a major rice cultivation area in the Kilombero Valley, Tanzania. *African Study Monographs*, *36*(March), 3–18. doi: 10.14989/68498

Khonje, M. G., Manda, J., Mkandawire, P., Tufa, A. H., & Alene, A. D. (2018, sep). Adoption and welfare impacts of multiple agricultural technologies: evidence from eastern Zambia. *Agricultural Economics*, *49*(5), 599–609. Retrieved from http://doi.wiley.com/10.1111/agec.12445 doi: 10.1111/agec.12445

Kijima, Y. (2016). On the Possibility of Rice Green Revolution in Rainfed Areas in Uganda: Impact Evaluation of a Management Training Program and Guidebook Distribution. In *In pursuit of an african green revolution* (pp. 65–89). Tokyo: Springer Japan. Retrieved from http://link.springer.com/10.1007/978-4-431-55693-0{_}4 doi: 10.1007/978-4-431-55693-0_4

Kjaerulff, U. B., & Madsen, A. L. (2012). *Bayesian networks and influence diagrams: a guide to construction and analysis* (Vol. 22). Springer Science & Business Media.

Kocabas, V., & Dragicevic, S. (2013). Bayesian networks and agent-based modeling approach for urban land-use and population density change: A BNAS model. *Journal of Geographical Systems*, *15*(4), 403–426.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Korb, K. B., & Nicholson, A. E. (2010). *Bayesian Artificial Intelligence, Second Edition* (2nd ed.). Boca Raton, FL, USA: CRC Press, Inc.

Koski, T., & Noble, J. (2011). *Bayesian networks: an introduction* (Vol. 924). John

Wiley & Sons.

Kuhn, M. (2008). Caret package. *Journal of Statistical Software*, *28*(5), 1–26.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 810). Springer.

Lee, D. R. (2005). Agricultural sustainability and technology adoption: Issues and policies for developing countries. *American journal of agricultural economics*, *87*(5), 1325–1334.

Leemhuis, C., Thonfeld, F., Näschen, K., Steinbach, S., Muro, J., Strauch, A., ... Diekkrüger, B. (2017, aug). Sustainability in the Food-Water-Ecosystem Nexus: The Role of Land Use and Land Cover Change for Water Resources and Ecosystems in the Kilombero Wetland, Tanzania. *Sustainability*, *9*(9), 1513. Retrieved from http://www.mdpi.com/2071-1050/9/9/1513 doi: 10.3390/su9091513

MALF. (2015). *Tanzania Livelihood Baseline Profile: Kilombero-Ulanga-Lusewa Paddy, Maize, and Cassava Livelihood Zone* (Zone Discription No. 43). Ministry of Agriculture Livestock and Fisheries: Agriculture Department.

Marcot, B. G., Steventon, J. D., Sutherland, G. D., & McCann, R. K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, *36*(12), 3063–3074. Retrieved from http://www.nrcresearchpress.com/doi/abs/10.1139/x06-135 doi: 10.1139/x06-135

Margaritis, D. (2003). *Learning bayesian network model structure from data* (Tech. Rep.). Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA461103.pdf

Milder, J. C., Buck, L. E., & Hart, A. (2013). *Applying an Agriculture Green Growth Approach in the SAGCOT Clusters: Challenges and Opportunities in Kilombero, Ihemi and Mbarali* (Tech. Rep.). Dar es Salaam: SAGCOT Centre.

Milder, J. C., Buck, L. E., Hart, A. K., Scherr, S. J., & Shames, S. A. (2013). A framework for Agriculture Green Growth: Greenprint for the Southern Agricultural Growth Corridor of Tanzania (SAGCOT). *Dar es Salaam: SAGCOT Centre*.

Mombo, F., Speelman, S., Huylenbroeck, G. V., Hella, J., & Moe, S. (2011). Ratification of the Ramsar convention and sustainable wetlands management : Situation analysis of the Kilombero Valley wetlands in Tanzania. *Journal of Agricultural Extension and Rural Development*, *3*(9), 153–164.

Msofe, N. K., Sheng, L., Lyimo, J., Msofe, N. K., Sheng, L., & Lyimo, J. (2019, jan). Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability*, *11*(2), 505. Retrieved from http://www.mdpi.com/2071-1050/11/2/505 doi: 10.3390/su11020505

Mwimo, L., Mbowe, W., Kombe, C., Kibesse, B., Mziya, M., Kazi, M., ... Ndunguru, E. (2016). *Contract farming schemes in tanzania: Benefits and challenges* (Tech. Rep. No. Working Paper Series: No 8). Bank of Tanzania. Retrieved from https://www.sustainableagtanzania.com/_webedit/uploaded-files/All%20Files/machinery/Contract%20Farming%20Schemes%20in%20Tanzania-%20Benefits%20and%20Challenges.pd

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian Networks in R.* Retrieved from http://link.springer.com/10.1007/978-1-4614-6446-4 doi: 10.1007/978-1-4614-6446-4

Nakano, Y., Kajisa, K., & Otsuka, K. (2016). On the Possibility of Rice Green Revolution in Irrigated and Rainfed Areas in Tanzania: An Assessment of Management Training and Credit Programs. In *In pursuit of an african green revolution* (pp. 39–64). Tokyo: Springer Japan. Retrieved from http://link.springer.com/10.1007/978-4-431-55693-0{_}3 doi: 10.1007/978-4-431-55693-0_3

Nakano, Y., Tanaka, Y., & Otsuka, K. (2018, mar). Impact of training on the intensification of rice farming: evidence from rainfed areas in Tanzania. *Agricultural Economics*, *49*(2), 193–202. Retrieved from http://doi.wiley.com/10.1111/agec.12408 doi: 10.1111/agec.12408

National Bureau of Statistics. (2013). *2012 Population and Housing Census. Population Distribution by Administrative Areas, Tanzania* (Tech. Rep.). National Bureau of Statistics.

Neapolitan, R. E. (2010). Learning Bayesian Networks. *International journal of data mining and bioinformatics*, *38*, 505–19. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21659034 doi: 10.1016/j.jbi.2010.03.005

Nielsen, T., Keil, A., & Zeller, M. (2013, may). Assessing farmers' risk preferences and their determinants in a marginal upland area of Vietnam: a comparison of multiple elicitation techniques. *Agricultural Economics*, *44*(3), 255–273. Retrieved from http://doi.wiley.com/10.1111/agec.12009 doi: 10.1111/agec.12009

Nindi, S. J., Maliti, H., Bakari, S., Kija, H., & Machoke, M. (2014). Conflicts Over Land and Water Resources in the Kilombero Valley Floodplain, Tanzania. *African Study Monographs*, *50*(October 2014), 173–190.

Nkonya, E., Schroeder, T., & Norman, D. (1997, jan). Factors Affectin adoption of improved maize seed and Fertilizer in Northern Tanzania. *Journal of Agricultural Economics*, *48*(1-3), 1–12. Retrieved from http://doi.wiley.com/10.1111/j.1477-9552.1997.tb01126.x doi: 10.1111/j.1477-9552.1997.tb01126.x

Nojavan A., F., Qian, S. S., & Stow, C. A. (2017). Comparative analysis of discretization methods in Bayesian networks. *Environmental Modelling and Software*, *87*, 64–71. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2016.10.007 doi: 10.1016/j.envsoft.2016.10.007

NorsysSoftwareCorp. (2016). *Netica-J Manual.* Retrieved from http://www.norsys.com/netica-j/docs/NeticaJ{_}Man.pdf

Ntongani, W. A., Munishi, P. K. T., More, S. R., & Kashaigili, J. J. (2014). Local Knowledge on the Influence of Land Use/Cover Changes and Conservation Threats on Avian Community in the Kilombero Wetlands, Tanzania. *Open Journal of Ecology*, *4*(September), 723.

Ntshangase, N. L., Muroyiwa, B., & Sibanda, M. (2018, feb). Farmers' perceptions and factors influencing the adoption of no-till conservation agriculture by small-scale farmers in Zashuke, KwaZulu-Natal province. *Sustainability (Switzerland)*, *10*(2), 555. Retrieved from http://www.mdpi.com/2071-1050/10/2/555 doi: 10.3390/su10020555

Okike, I., Jabbar, M. A., Manyong, V., Smith, J. W., Akinwumi, J. A., & Ehui, S. K. (2001). Agricultural intensification and efficiency in the West African savannahs: evidence from northern Nigeria. *Socio-economics and Policy Research Working Paper - International Livestock Research Institute*(33). Retrieved from http://search.ebscohost.com/login.aspx?direct=true{&}db=lah{&}AN=20013164805{&}site=ehost-live doi: http://dx.doi.org/10.1111/j.1471-6712.1987.tb00601.x

Otsuka, K., & Larson, D. F. (2016a). Conclusions: Strategies Towards a Green Revolution in Sub-Saharan Africa. In *In pursuit of an african green revolution* (pp. 183–194). Tokyo: Springer Japan. Retrieved from http://link.springer.com/10.1007/978-4-431-55693-0{_}9 doi: 10.1007/978-4-431-55693-0_9

Otsuka, K., & Larson, D. F. (Eds.). (2016b). *In Pursuit of an African Green Revolution.* Tokyo: Springer Japan. Retrieved from http://link.springer.com/10.1007/978-4-431-55693-0 doi: 10.1007/978-4-431-55693-0

Otsuka, K., & Place, F. (2013). Evolutionary changes in land tenure and agricultural intensification in sub-Saharan Africa. *United Nation University Working Paper*, *2014/051* (November), 47. doi: 10.1093/oxfordhb/9780199687107.013.015

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed., Vol. 79) (No. 2). New York, NY, USA: Cambridge University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Others (2011). Scikit-learn: Machine learning in Python. *The journal of Machine Learning*, *12*, 2825–2830. Retrieved from http://www.jmlr.org/papers/v12/pedregosa11a.html doi: 10.1007/s13398-014-0173-7.2

Perz, S. G. (2003). Social Determinants and Land Use Correlates of Agricultural Technology Adoption in a Forest Frontier : A Case Study in the Brazilian Amazon Author (s): Stephen G . Perz Published by : Springer Stable URL : http://www.jstor.org/stable/4603460 Accessed :. *Human Ecology*, *31* (1), 133–165. Retrieved from https://www.jstor.org/stable/pdf/4603460.pdf doi: http://dx.doi.org/10.1023/A:1022838325166

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. T. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software*, *22* (8), 1140–1152. doi: 10.1016/j.envsoft.2006.03.006

Pope, A., & Gimblett, R. (2017). Linking Participatory, Bayesian, and Agent-Based Modeling Techniques to Simulate Coupled Natural-Human System: A Case Study with Ranchers in Sonora, Mexico. In *Environmental modeling with stakeholders* (pp. 345–359). Springer.

Prishchepov, A. V., Ponkina, E., Sun, Z., & Müller, D. (2019, jan). Revealing the determinants of wheat yields in the Siberian breadbasket of Russia with Bayesian networks. *Land Use Policy*, *80*, 21–31. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264837718300401?via{%}3Dihub doi: 10.1016/J.LANDUSEPOL.2018.09.038

Rasmussen, L. V., Mertz, O., Christensen, A. E., Danielsen, F., Dawson, N., & Xaydongvanh, P. (2016). A combination of methods needed to assess the actual use of provisioning ecosystem services. *Ecosystem Services*, *17*, 75–86.

Rogers, J., & Gunn, S. (2006). Identifying Feature Relevance Using a Random Forest. In (pp. 173–184). Springer, Berlin, Heidelberg. Retrieved from http://link.springer.com/10.1007/11752790{_}12 doi: 10.1007/11752790_12

Rogers, J. D., & Gunn, S. R. (2005). Ensemble Algorithms for Feature Selection. In (pp. 180–198). Retrieved from http://link.springer.com/10.1007/11559887{_}11 doi: 10.1007/11559887_11

Sanchez, S. M. (2005). Nolhdesigns spreadsheet. *Software available online via http://diana. cs. nps. navy. mil/SeedLab*, 1430–1439.

Sandri, M., & Zuccolotto, P. (2006). Variable Selection Using Random Forests. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.), *Data analysis, classification and the forward search* (pp. 263–270). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/3-540-35978-8{_}30 doi: 10.1007/3-540-35978-8_30

Schelhas, J. (1996). Land Use Choice and Change : Intenszfiation and Diversification in the Lowland Tropics of Costa Rica. *Human Organization*, *55*(3), 298–306. doi: 10.17730/humo.55.3.k4v1050710073063

Senthilkumar, K., Tesha, B. J., Mghase, J., & Rodenburg, J. (2018, oct). Increasing paddy yields and improving farm management: results from participatory experiments with good agricultural practices (GAP) in Tanzania. *Paddy and Water Environment*, *16*(4), 749–766. doi: 10.1007/s10333-018-0666-7

Sheahan, M., & Barrett, C. B. (2014). Understanding the agricultural input landscape in sub-saharan africa: Recent plot, household, and community-level evidence. doi: 10.1596/1813-9450-7014

Shriar, A. J. (2000). Agricultural intensity and its measurement in frontier regions. *Agroforestry Systems*, *49*(3), 301–318. doi: 10.1023/A:1006316131781

Shriar, A. J. (2001). The dynamics of agricultural intensification and resource conservation in the buffer zone of the Maya Biosphere Reserve, Petén, Guatemala. *Human ecology*, *29*(1), 27–48.

Sirén, A. H. (2007). Population growth and land use intensification in a subsistence-based indigenous community in the Amazon. *Human Ecology*, *35*(6), 669–680. doi: 10.1007/s10745-006-9089-y

Sörensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions*, *10*(1), 101–112.

Sun, Z., & Müller, D. (2013). A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. *Environmental Modelling and Software*, *45*, 15–28. doi: 10.1016/j.envsoft.2012.06.007

Teklewold, H., Kassie, M., Shiferaw, B., & Köhlin, G. (2013). Cropping system diversification, conservation tillage and modern seed adoption in Ethiopia: Impacts on household income, agrochemical use and demand for labor. *Ecological Economics*, *93*. Retrieved from http://dx.doi.org/10.1016/j.ecolecon.2013.05.002 doi: 10.1016/j.ecolecon.2013.05.002

URT. (2013). *National Agriculture Policy* (Tech. Rep.). Dar es Salaam: Ministry of Agriculture food Security and Cooperative. Retrieved from http://www.fao-ilo.org/fileadmin/user{_}upload/fao{_}ilo/pdf/ICA{_}MLW{_}and{_}TZ/NATIONAL{_}AGRICULTURAL{_}POLICY-2013.pdf

USAID. (2019). *Agriculture factsheet: Tanzania* (Tech. Rep.). Retrieved from https://www.usaid.gov/sites/default/files/documents/1860/Agriculture{_}Fact{_}Sheet{_}April{_}2019{_}FINAL.pdf

Uusitalo, L. (2013). Advantages and challenges of Bayesian networks in environmental modeling networks in environmental modelling. (May 2007).

Van Ittersum, M. K., Van Bussel, L. G., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., ... Cassman, K. G. (2016, dec). Can sub-Saharan Africa feed itself? *Proceedings of the National Academy of Sciences of the United States of America*, *113*(52), 14964–14969. doi: 10.1073/pnas.1610359113

Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., ... Asten, P. V. (2014). Sustainable intensification and the African smallholder farmer. *Current Opinion in Environmental Sustainability*, *8*, 15–22. Retrieved from http://dx.doi.org/10.1016/j.cosust.2014.06.001 doi: 10.1016/j.cosust.2014.06.001

Vanlauwe, B., Van Asten, P., & Blomme, G. (2014). *Challenges and opportunities for agricultural intensification of the humid highland systems of sub-saharan africa.* Springer.

Wainaina, P., Tongruksawattana, S., & Qaim, M. (2016, may). Tradeoffs and complementarities in the adoption of improved seeds, fertilizer, and natural resource management technologies in Kenya. *Agricultural Economics*, *47*(3), 351–362. Retrieved from http://doi.wiley.com/10.1111/agec.12235 doi: 10.1111/agec.12235

WFP. (2019). *WFP Tanzania Country Brief* (Tech. Rep.). Retrieved from www.wfp.org/countries/Tanzania

Wilson, E., Mcinnes, R., Mbaga, P., & Ouedraogo, P. (2017). *Kilombero Valley, United Republic of Tanzania: Ramsar Advisory Mission Report* (Tech. Rep.). Retrieved from https://www.ramsar.org/sites/default/files/documents/library/ram83{_}kilombero{_}valley{_}tanzania{_}2016{_}e.pdf

World Bank. (2007). *World Development Report 2008.* The World Bank. doi: 10.1596/978-0-8213-6807-7

Yamano, T., Rajendran, S., & Malabayabas, M. L. (2015, oct). Farmers' self-perception toward agricultural technology adoption: evidence on adoption of submergence-tolerant rice in Eastern India. *Journal of Social and Economic Development*, *17*(2), 260–274. doi: 10.1007/s40847-015-0008-1

Zheng, F., & Webb, G. I. (2010). Tree Augmented Naive Bayes. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 990–991). Boston, MA: Springer US.

4.A Appendix

4.A.1 Regression trees for each intensification option

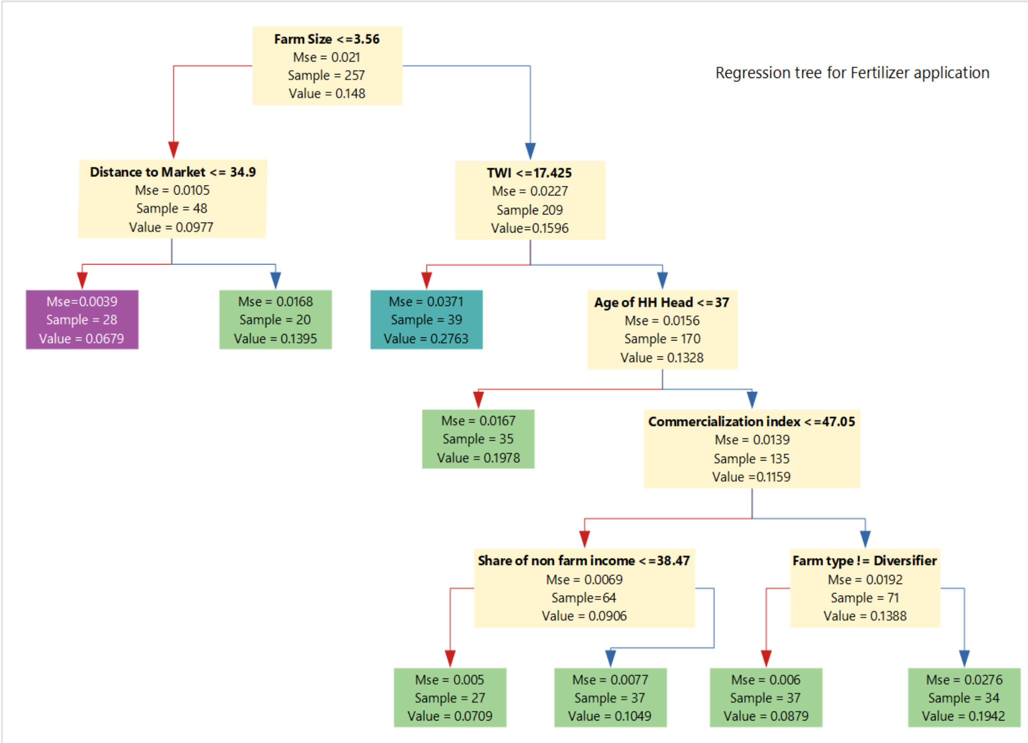


FIGURE 4.8: Regression tree for fertilizer application

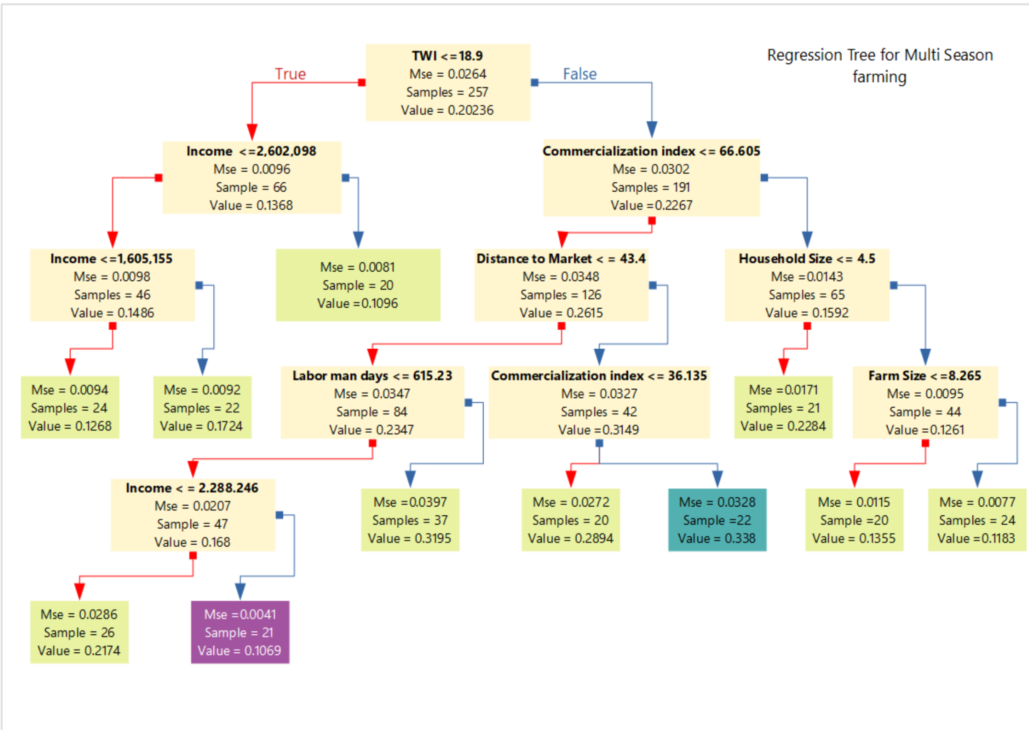


FIGURE 4.9: Regression tree for crop multiple times

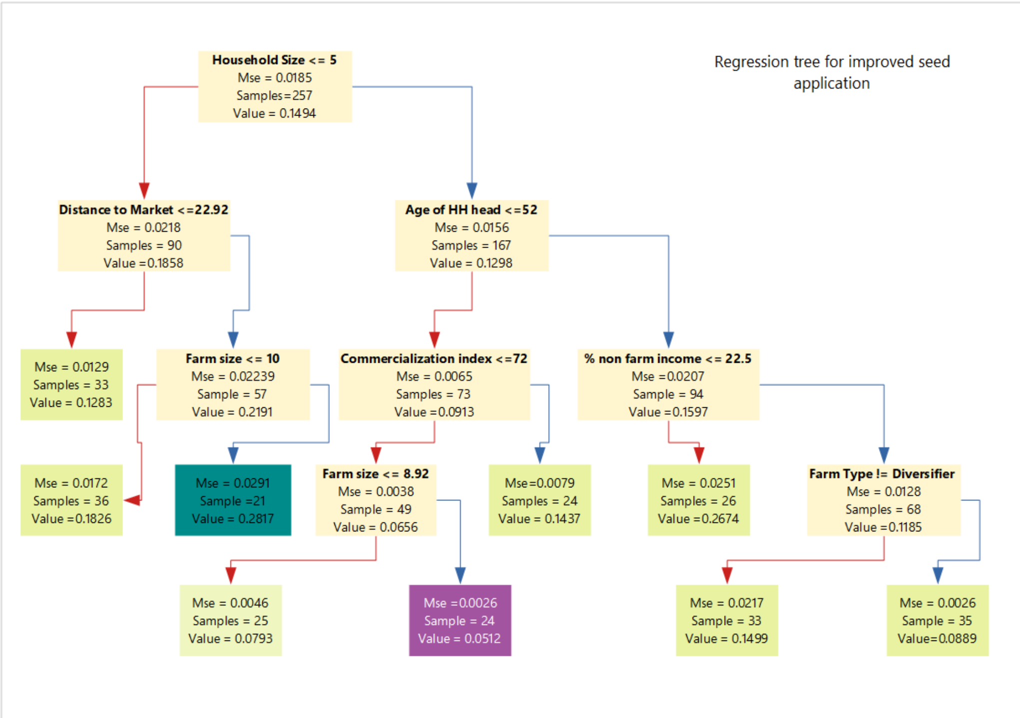


FIGURE 4.10: Regression tree for improved Seed

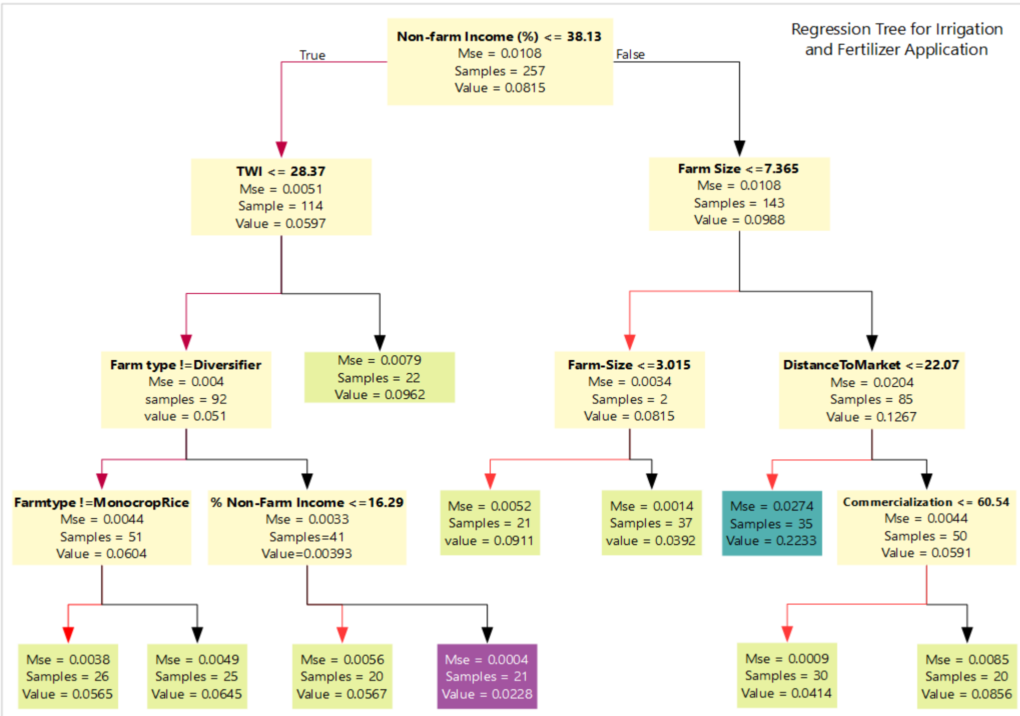


FIGURE 4.11: Regression tree for irrigation and fertilizer

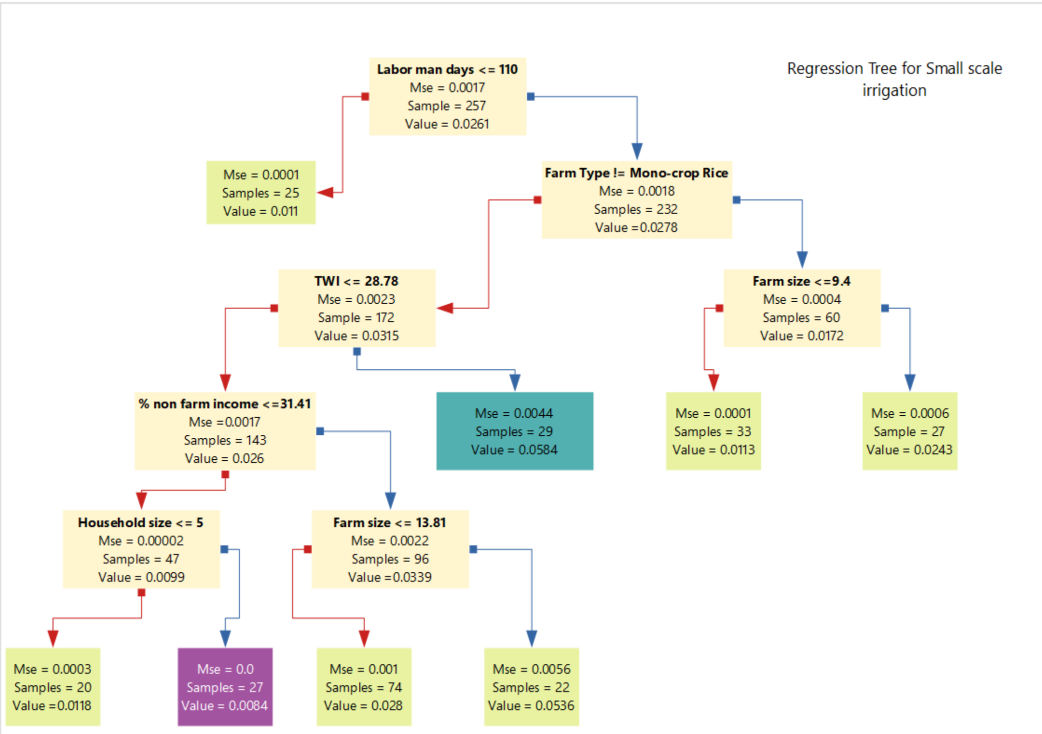


FIGURE 4.12: Regression tree for irrigation Use

Chapter 5

Migration, access to infrastructure and the pattern of land use and intensification in Kilombero Valley: An Agent-Based Modeling approach

Abstract: The Kilombero valley has undergone considerable change over time, and it presumed to continue even at a faster rate. Both global and proximate drivers are pushing the changes observed in the valley. Two of the main driving forces that are evident are the rise of the population through migration into the valley and the government's drive for intensification through establishing both physical and institutional infrastructure. This chapter aims to simulate the effect of a sustained increase in migration and access to better road infrastructure by building a spatially and temporally explicit agent-based model. The design of the ABM called *"WetABM"* is an approach to integrate the socio-economic and bio-physical elements of the KVF to examine the effects of the two exogenous changes on the dynamics of intensification, land use, and agricultural production dynamics. Three different scenarios "business as usual", "in-migration", and "improved road infrastructure" are discussed.

Keywords: *Intensification, ABM, Kilombero valley, land use, crop production, im-migration, road infrastructure*

5.1 Introduction

Floodplain wetlands have been used for agriculture for millennia and have "nurtured the development of many important cultures around the world" (Villar, 2014, .p 2). They are one of the most productive natural resources upon which rural economies depend, providing food and energy, medicine, building material, and dry season grazing (Rebelo, McCartney, & Finlayson, 2010). Due to their immense potential and diverse ecosystem services it provides, wetlands have been under pressure, and their drainage and reclamation for agriculture have become more widespread (Everard & Wood, 2017; Villar, 2014; Wood & van Halsema, 2008).

The pressure on wetlands arises both from local and macro drivers operating at different scales towards exploitation and degradation. Through the application of the drivers, pressures, state changes, impacts and responses (DPSIR) framework to 90 cases drawn from around the world Wood and van Halsema (2008) have identified in-migration, land shortage, government policy and plans to improve the national food security and local market are some of the many drivers of wetland degradation. As shown in Figure 5.1, these drivers will exert pressure on the functioning of the ecosystem through intensification and expansion of agriculture to the floodplains or wetlands.

Although both intensification and expansion of agriculture are considered as pressure to the Wetland, the development of agriculture through drainage of wetlands has a significant impact on both the ecosystem and the well-being of the community (Everard & Wood, 2017; Wood & van Halsema, 2008). On the other hand, intensification had the potential to reduce the expansion of agriculture to the wetlands by increasing output per unit area. As opined by the influential Millennium Ecosystem Assessment (2005, p. 66) report: "The expansion of agriculture will continue to be a major driver of wetland loss. In regions where agricultural expansion continues to be a large threat to wetlands, the development, assessment, and diffusion of technologies that could increase the production of food per unit area sustainably, without harmful trade-offs related to excessive consumption of water or use of nutrients or pesticides, would significantly lessen pressure on wetlands".

The case is not different in the Kilombero valley floodplain wetland in Tanzania.

As one of Africa's most extensive freshwater wetlands, the KVF is facing pressure both from Macro and local drivers (Sulle, 2020). Although these drivers prevail and interact at different hierarchical scales, two macro drivers are evident in KVF; Economic development and in-migration, which are directly associated with the perceived availability of land, government policy, a local institution, and market (ERM, 2012; KILORWEMP, 2017; Msofe et al., 2019).

Being the hotspot for agricultural production and conservation, the KVF is subject to intervention from government and non-governmental organizations (Bergius, 2014; Sulle, 2020). For instance, the planned growth corridor (SACGOT) includes KVF as one of the clusters to boost agricultural production by linking farmers to the market and supply chain. More so, both the state and non-state actors with interest in the valley have pledged to boost the infrastructure and accessibility of the valley (ERM, 2012). The recently completed "Magufuli bridge" on the Kilombero river that connects Ulanga, Malinyi, and Kilombero districts and the ongoing rehabilitation of Ifakara -Kidatu road are examples of such road infrastructure investments in the valley.

On the other hand, the perceived availability of abundant land has also attracted a significant number of immigrants from arid and semi-arid areas of the country into the valley. The valley has a long history of in-migration starting from colonial agricultural programs, which encouraged tsetse fly-infested areas to be converted to cropland (Blache, 2019). The post-independence resettlement policy of Ujamaa Vijijini also contributed to state lead migration to the valley (Bergius, Benjaminsen, Maganga, & Buhaug, 2020). And the expansion of conservation areas across Tanzania has also lead migrants to look for productive land in sparsely populated area, including Kilombero valley (Salerno, Mwalyoyo, Caro, Fitzherbert, & Mulder, 2017).

The extent of immigration to the KVF can also be seen from the current population composition in the valley. The result from our survey shows out 447 randomly selected households; almost 60 percent of the families identify themselves as a migrant to the valley. Households started to migrate to the valley from 1950, with the majority arriving in early 2000. Eighty percent of the migrated households also mention agriculture as the main reason for their decision to migrate. Ndamba, Mbunga, and Pogoro are generally considered as the natives of the valley. Barbaigs had

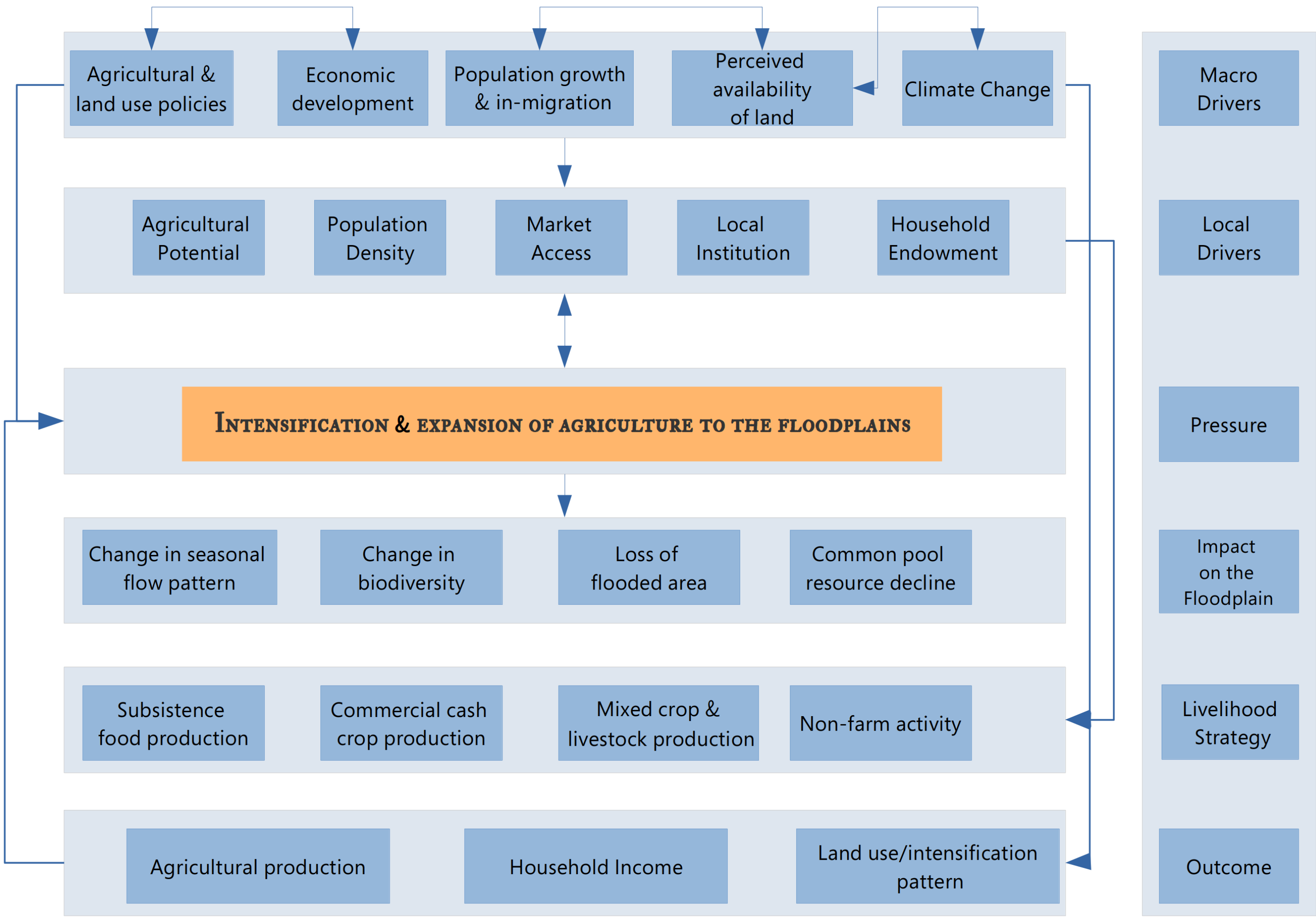


Figure 5.1: DPSIR model of wetland and agriculture interaction

Modified based on Wood and van Halsema (2008) mainly come from Manyara, Dodoma, Mbeya, and Songea Regions; Maasai from Mbeya, Iringa and Arusha Regions, and Sukumas from Shinyanga, Mwanza, Tabora, Sumbawanga, and Mbeya Regions (ERM, 2012; KILORWEMP, 2017).

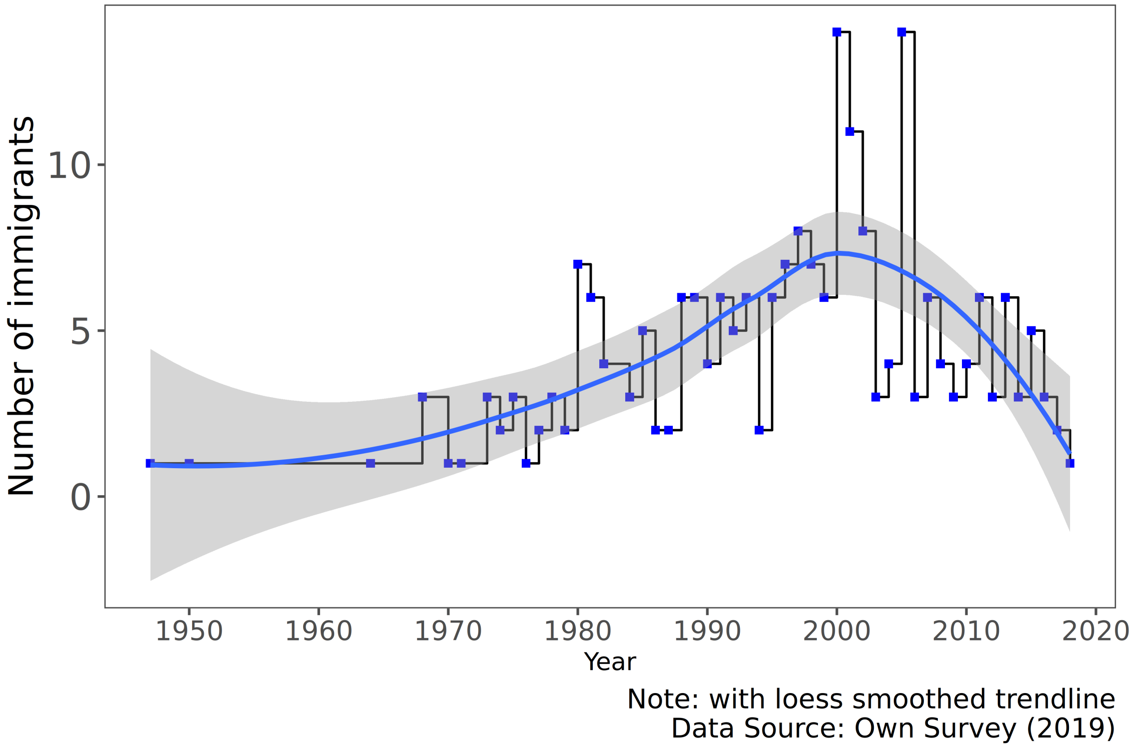


FIGURE 5.2: Households year of migration to KVF

With loess smoothed trendline

Immigration is frequently associated with localized population growth and leads to increasing unsustainable land use practice, and expansion into marginal lands (Hunter & Nawrotzki, 2016; Reardon, Barrett, Kelly, & Savadogo, 1999). As generally observed by Geist and Lambin (2001, p. 63): "... population-driven expansion of permanently cropped land, more cases of subsistence farming in Africa tend to be driven by in-migration and local population growth than cases in other regions."

This is also evident in that the most recent residents of the valley tend to live further from the villages, often on land that is officially gazetted as a protected area. Migrant families have been generally found to be the primary cause of cropland conversion, especially in the wetland edges(Bamford, Ferrol-Schulte, & Smith, 2010). However, as the number of migrants increases, the area of land available for clearing has also diminished significantly in recent years, and the carrying capacity of the valley has reached its limit(ERM, 2012). This even the cause of the recent increase in conflict and litigation between farmers, the migrants, and the government.

The increase in population density, coupled with the government's effort to boost production in the valley, will have the potential to transform the intensification status and thereby the aggregate production system in the valley (Otsuka & Place, 2013; Smith et al., 2010; Vanlauwe et al., 2014).

Although several studies acknowledge the complex and dynamic drivers and changes taking place in the valley, the potential effect of such trends and interactions are under-researched. Hence, there is a clear need for a study that explores the relationship between migration, better access to the market, land use, and intensification for KVF. This chapter examines the potential effect of immigration and access to infrastructure on the dynamics of intensification, land use, and crop production in KVF. Such exogenous activities exert pressure on this complex environment. Their aggregate effect on the population and the landscape varies with the pressure's intensity, spatial, and temporal scale. The study of any impact of the management and intervention requires an in-depth understanding of the interplay between the intervention and the structure, function, and resilience of the floodplain at a different scale.

To this end, we built an agent-based model called WetABM that captures these complex relationships between immigration, land use, intensification, and market. The WetABM is a spatially and temporally explicit model for KVF, which is parameterized with empirical data collected through a survey and geospatial data.

The chapter is structured as follows. The next section provides the specific elements and structure of the WetABM model using ODD+D protocol. The third section describes our scenarios in relation to the model elements. The fourth section will provide the result and discussion of our simulation exercise. The last part will conclude the chapter.

5.2 WetABM: ODD+D protocol

In this section, we will briefly explain the details and principal components of WetABM: An agent-based model of land use and intensification in Kilombero Valley using the ODD+D protocol. A more detailed description of the model and its parameters are provided in a separate document [1]. Here we will concentrate on the principal elements and migration sub-component of the Model.

5.2.1 Overview

5.2.1.1 *Purpose*

WetABM is designed to investigate the impact of alternative policy intervention and exogenous factors on the changing land use and farmer's adoption of intensification options intended to increase the production and income of farm households. It aims to depict as realistically as possible around 38 000 farms within the KVF in all their heterogeneity in terms of production structure, farm as well as modes of social behavior. WetABM can be used to estimate the repercussions of internal and external exogenous influences, and the effects of the heterogeneous site conditions specific to KVF on income trends, structural change, and land management.

5.2.1.2 *Entities, state variables, and scales*

There are different types of entities in the model. A farm household is a primary agent being composed of household members modeled individually. Agents interact via the spatial landscape consisting of Region, Ward, Parcel, and Land Cell and via the Market. Figure 5.3 shows the UML representation of the main agents with WetABM.

Individual Agents (IA): Individual agents in the model are a rather simplistic representation of a single individual represented by their age, gender, and if she is a household head and household id. Individual decision making is not considered.

[1] The full ODD is available at https://bsrthyle.github.io/ODD-DforWetABM/. The detailed ODD+D contains, in addition, the parameters and the regression coefficients for the production function and the spatial input data

Farmer Agents(FA): The farmer agent (FA) is an aggregation of individual agents. Being the primary agent in the model, FA is characterized by several static and dynamic attributes and decision-making routines. Main attributes include age, gender, location of homestead within a spatial landscape, farmer type, plots, commercialization index, income, minimum food requirement, expenditure, savings, dynamic income aspiration threshold, a memory of past expected and realized prices, available labor, proximity to market, proximity to the road and average proximity from their homestead to farm plots. FA are aggregated in a hierarchy of three organizational levels for analytical purposes of simulation results: farmer agent, group of farmers (Mono crop rice producers, diversifier and agropastoralists) [2] and population.

Land Cell (LC): is the smallest spatial scale of the biophysical space. It represented by grid cells of each 1 ha (100m by 100m). Its characterized by several features relevant to the model. Each attribute is organized as a spatial raster layer storing the location of plots and homestead, current land use (both aggregate and crop), distance from the edge of the Wetland, distance from the river, topographic wetness index, slope, elevation, and ownership layers. The primary state variable of the LCA is the Topographic wetness index (TWI). TWI is an index derived from surface elevation data (DEM) and estimates the relative wetness within a catchment (Nystrom & Burns, 2011). It approximates the spatial soil moisture patterns and distribution of groundwater levels (Qin et al., 2011; Sörensen, Zinko, & Seibert, 2006). The measure is widely used in many applications, including precision agriculture (Qin et al., 2011). It is defined as

$$TWI = ln(\frac{\alpha}{\tan\beta}) \tag{5.1}$$

Where TWI is the topographic wetness index, ln is the natural logarithm, $\propto$ is the upslope contributing area per unit of contour length, in meters, $\tan\beta$ is the local slope.

Parcel Agent(PA) is an abstract representation of a farm. A collection of plots for a single farmer agent forms parcel. The attributes of the parcel agents are merely an aggregation of the plots it contains.

[2] See Chapter 3 three of the dissertation for details of farmer groups and their characteristics, and for the empirical characterization and the methodology see `https://bsrthyle.github.io/FarmTypolgyV5/`

Ward Agent (WA) is the third hierarchical scale in the landscape representation. It represents the official administrative organization of the valley. There are 11 Ward agents, each representing their official administrative counterparts. It is represented by many aggregations from plots, parcels, and farmer agents. More so, they are represented by the cost of transportation from the centers of the ward to the major market in the region, the annual average price for maize and rice, total output for rice and maize, income distribution,

Region Agent (RA) encapsulates all the wards and different aggregate attributes from the ward's agents. Thus, overall patterns of the entire landscape are the result of the aggregated and accumulated impacts. As the spatial system is self-organized, emergence properties may be observed.

Market Agent (MA) is structurally different from the other agents in the model. MA is an abstract market that coordinates demand and supply of rice and maize in the region. It does have only one purpose to collect the supply of rice and maize from all the wards and provide endogenous market prices(details are in the market submodule section).

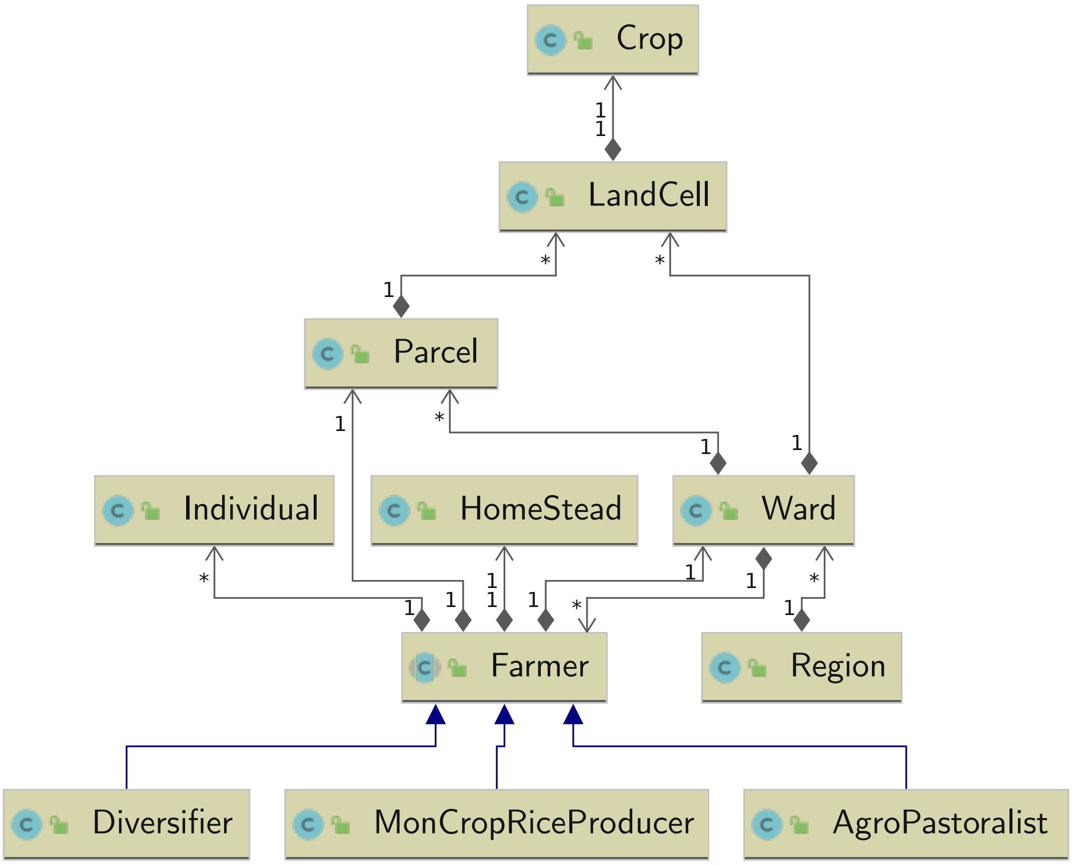


FIGURE 5.3: UML representation of agents in WetABM

5.2.1.3 *Process overview and scheduling*

The diagram in Figure 5.4 gives the process overview of the model. In the current version, each simulation run represents one year and involves the following primary sequential scheduling for farmer agents:-

- ➔ Evaluate their objective and decide whether to change their land management behavior or exit farming (farmers exit farming based on two conditions:
- ➔ when their objective is not satisfied for five consecutive years and when the age of household head is greater than 70 with no heir in the household)
- ➔ Form price expectation for rice and maize
- ➔ Evaluate crops to plant and intensification strategy using BBN

➔ update the attributes of their farm plot if there is a change in land management

➔ Evaluate their annual yield using a production function depending on their production factors and the topography of their farm

➔ Decide on their market participation (exogenous in this version of the model, the proportion of output sold in the market varies by farm types)

➔ The sold output is collected from the farmer and sent to the market agent, and actual prices will be updated then

➔ Farmers evaluate their farm accounting and financial indicators such as income, saving, consumption

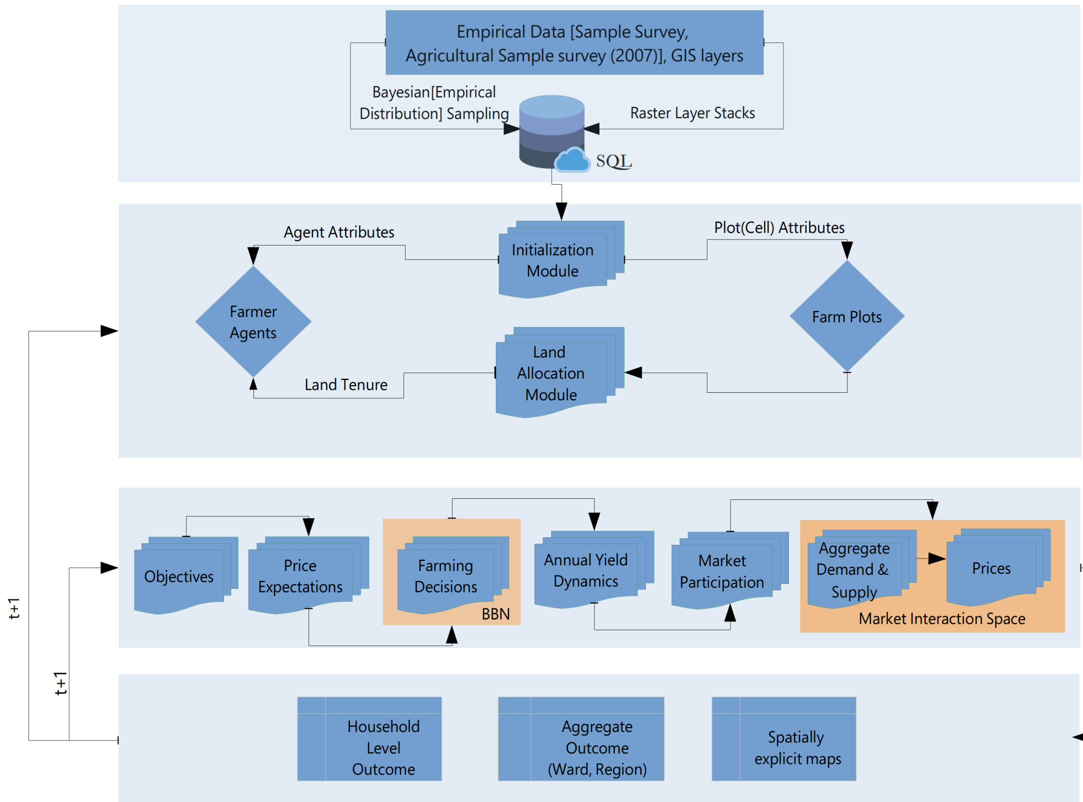


FIGURE 5.4: Process overview for WetABM

5.2.2 Design concepts

5.2.2.1 *Theoretical and empirical background*

WetABM is built based on the general conceptual framework of the coupled human-environment system (An, 2012; Murray-Rust et al., 2011; Turner et al., 2003). Land use and intensification is an emergent property that evolves from the interactions among various components of the entire humanenvironment system, which themselves feedback to influence the subsequent development of those interactions (Lambin, Geist, & Lepers, 2003; Le, Park, & Vlek, 2010). According to Parker (2003), by definition, emergence phenomena cannot be reduced to the system's parts: the whole is more than the sum of its parts because of interactions among the parts, and they are directly related to the nested hierarchies and interdependence that characterize the complex system. At the lower scale of the system's constituent units (e.g., household and land plot), many small changes in land allocation and localized changes are the results of multiple decisions made by agents, who act under certain specific conditions, anticipate future outcomes of their decisions, and adapt their behavior to changes in their external and internal conditions (Lambin et al., 2003). Temporal accumulations of these short term changes and spatial aggregations of these localized changes generate continuously emergent patterns of both land use at the landscape scale and socio economic dynamics at the population scale (e.g., Ward) (Le et al., 2010).

In this context, WetABM combines the socio-economic and biophysical aspects of the KVF in a single coherent modeling approach. In the current version of the model, the biophysical system is roughly modeled. It only embodies a static representation of the topographic feature of the valley. Details of the relation between the two systems are presented in the subsequent sections.

To empirically parametrize different components of the model both primary and secondary data, was collected. The core data source is a household survey in 21 villages in two districts of the KVF, Ulanga, and Kilombero. In total, 304 farm households were interviewed to provide information on the farming systems,intensification choices, resource use and management as well as their relevance for the livelihoods of the households.

The household selection was based on a multi-stage sampling strategy. First, 11 wards were purposively selected based on the occurrence of floodplain farming. In the second stage, 21 villages were randomly selected using probability to population size within the wards. In the final stage, households were randomly selected from the list provided by village leaders.

The core geospatial data for WetABM are land use map for 2014, Digital Elevation Model (DEM) at 90m resolution and proximity raster maps. The DEM was the basis for generating other raster layers, including the Topographic wetness index and elevation. Administrative ward boundaries were obtained from the Kilombero district land and settlement office. Proximity raster (to road, market, and river) are based on the open-source database OpenStreetMap. The location of homestead areas is also obtained from OpenStreetMap and pre-processed using QGIS (see initialization section below for details). All the data used for parameterization of WetABM are documented with the source code on a private repository on GitHub [3].

5.2.2.2 *Individual decision making*

FA make several sequential decisions during a specific simulation period. The primary decision-making model is an intensification decision. After evaluating their objective and if the intensification decision is triggered, FA decide on which intensification option to choose based on the information they have at hand and on their previous experiences. The choice of intensification option is determined by a probabilistic rule using the Bayesian belief network. The details of the Bayesian belief network in terms of the state variables, structure, and parameter learning are provided in the third chapter of this thesis. Figure 5.5 shows the BBN used within WetABM.

FA do not explicitly optimize and use rule-based criteria to evaluate their decision. There are two main such criteria (1) if the household can provide the minimum food requirement for all the household members (2) the relative position of the farm households in terms of income compared to his peers in a particular ward. If the per capita income is in the bottom 25 percent of the cumulative distribution of income in the ward (see intensification and crop choice decision submodel).

[3] https://bsrthyle.github.io/WetABMv1.0/

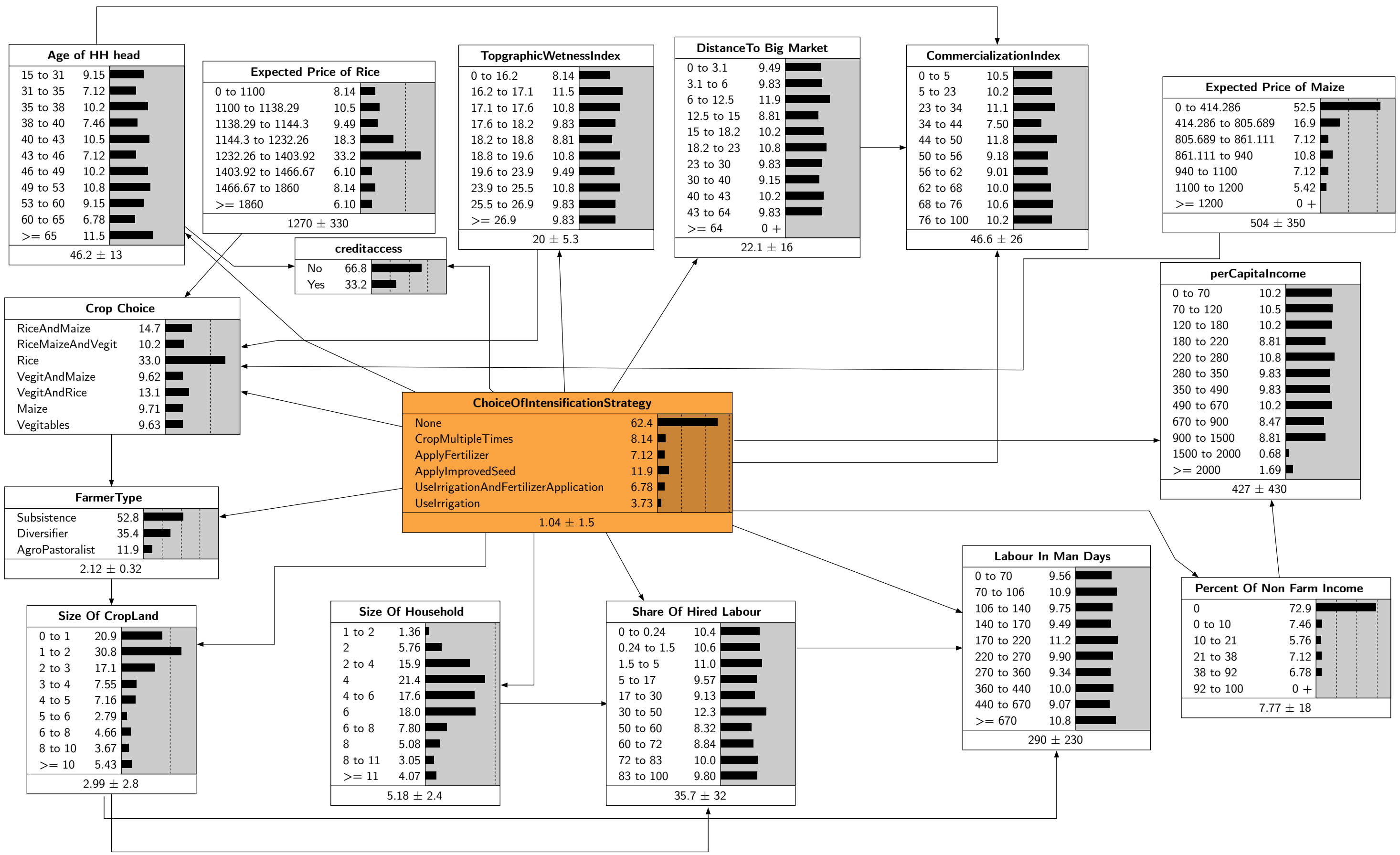


FIGURE 5.5: A Bayesian belief network of intensification decision in KVF

Note: The BBN used within WetABM is slightly modified from the one presented the fourth chapter. Here the state intervales are increased to the point where its not computationally expensive.

Besides, the intensification decision, FA form price expectation, choose crops to plant, decide on land expansion, labor allocation, and on exiting farming. Figure 5.6 shows the UML representation of the decision models within WetABM.The submodel section provides the details of these decision-making routines.

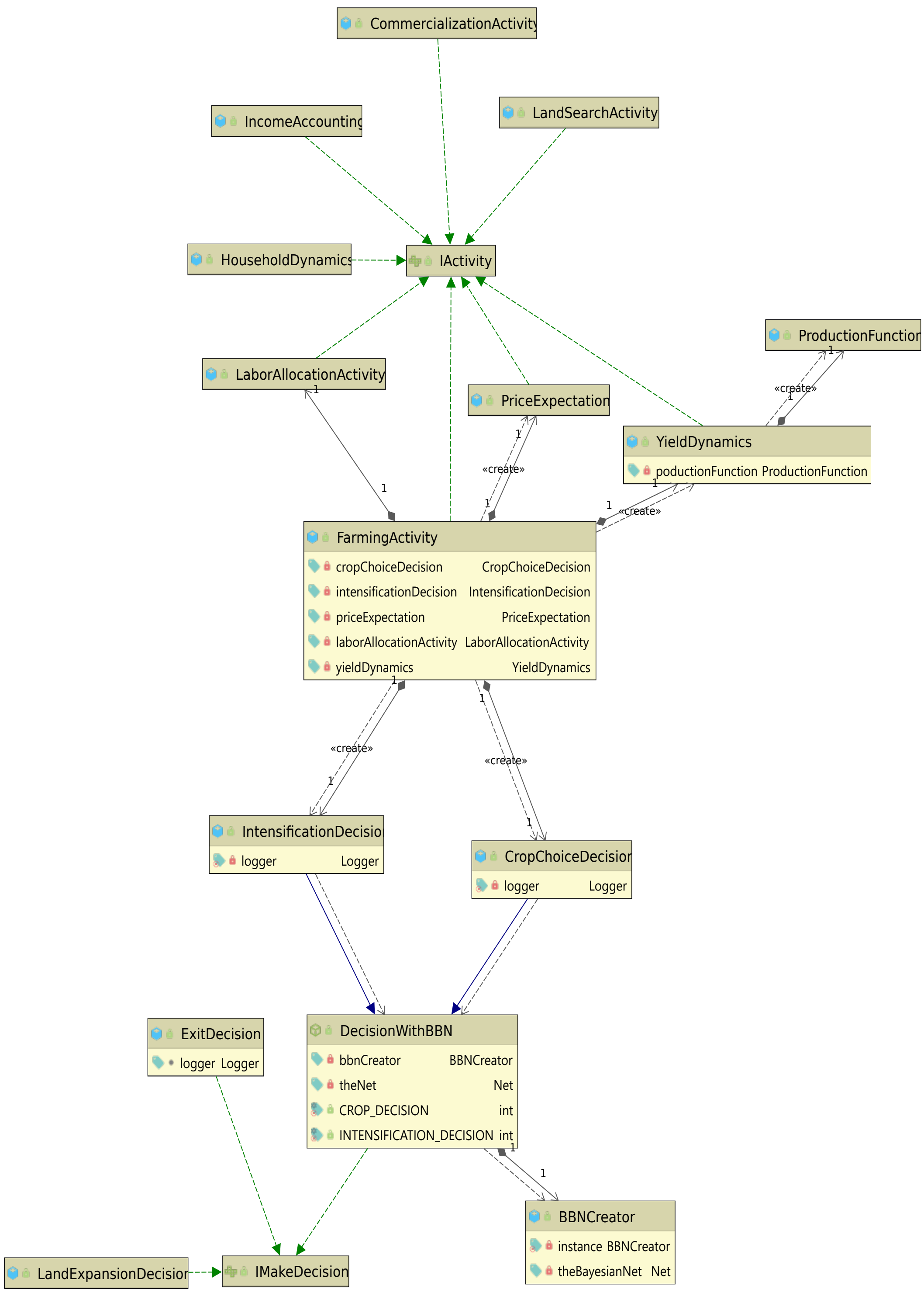


FIGURE 5.6: UML representation of the decision module in WetABM

5.2.2.3 *Learning*

In the current version of the model, there is no learning mechanism for the agents.

5.2.2.4 *Individual sensing*

Farm household agents are assumed to have access to all their attributes and biophysical characteristics of all the plots they make decisions for.

5.2.2.5 *Individual prediction*

Farmers have different expectation formation heuristics for crop prices by using their past price memory. Furthermore, use these forecasts for their production and adoption decisions. FA also predicts their vectors of probability for each crop they plan to plant and which intensification options they plan to adopt.

5.2.2.6 *Interaction*

WetABM captures multi-scale interactions within and between different entities of the model. Interactions represented in the model are household-to-household and household-to-environment interactions. A household-household interaction represents the indirect interaction between different households via markets through the endogenous formation of prices. Household agents also interact with the environment itself when they engage in farming. Households can change the states of the land as they convert it from one land use and land cover to another.

5.2.2.7 *Collectives*

In the current version of the model, there is no collective for the agents in WetABM.

5.2.2.8 *Heterogeneity*

Within the model, households are heterogeneous in their attributes, livelihood choices, and characteristics of their farm. Besides, some FA are heterogeneous on some of the state variables depending on their farm type. FA belongs to one of the three farm types defined during the initialization. The type they belong to affect some of the state variables, including Off-farm income, access to credit, and commercialization index. In the current version of the model, farm types are imposed during the initialization, and FA does not change their type during the simulation period. In the later versions, it possible to allow livelihood transition between farm types depending on their characteristics, at least between diversifier and mono-crop rice producers.

5.2.2.9 *Stochasticity*

The model has stochasticity built in several ways. The choice of intensification strategy is stochastic, as the farmers are most likely to select an option with a certain conditional probability. However, as the farmers are not modeled as purely rational decision-makers, the highest-ranking strategy is not always chosen. Moreover, to better represent the decision environment, the BBN is fully stochastic to represent the uncertainty and variability observed in nature. Moreover, crop output for rice and maize are stochastic as they vary between the prediction interval from the regression model (see the crop yield sub module section below).

5.2.2.10 *Observation*

Different farm and landscape related indicators are observed at the end of the simulation. These include household and regional indicators (income, agricultural production, number of farmers adopting specific intensification strategy at ward and regional level). Additionally, landscape-related indicators are tracked (maps for land use and application of intensification strategy). WetABM displays several simulated data panels of the Graphical User interface (GUI). The GUI displays different types of statistical outputs at runtime, including a spatially explicit map.

5.2.3 Details

5.2.3.1 *Implementation Details*

WetABM is written in java using the Repast Simphony framework (North et al., 2013). Repast provides a tool kit that is advanced and flexible and has the advantage of all the tools and packages provided by java. The Bayesian network is similarly written in Java using a commercial library called NETICA (NorsysSoftwareCorp, 2016). Netica provides the Java API for learning the structure and parameters of BBN and also provide efficient algorithms for inferences. The data for initialization is provided through a cloud-based relational database PostgreSQL hosted on Amazon

Web Services (AWS). The relational database management has the advantage of being robust, fast, and provides easy accessibility.

Given a relatively large number of agents (38,000), high-resolution landscape representation (1 ha) for an area of 5,200km^2 , and agent-specific Bayesian inference, WetABM is computationally expensive. To increase the computational efficiency of the model run, WetABM leverages on a parallel feature introduced in java 1.8.

Also, simulation runs were performed in parallel on high computational power computers hosted on the Google Cloud Platform (GCP), and local personal computer.

Once the model is initialized, and experiments are conducted, the results are analyzed using an open-source software R through direct linking WetABM and R and R-markdown (R Core Team, 2018). Analyzing the results through R and Markdown provides us with more flexibility, transparency, and reproducible outputs from the model. The code for WetABM is under version control on a private repository in Github.

5.2.3.2 *Initialization*

The model initialization of WetABM involves several steps before the model is ready for simulation and experimental analysis. There are three sub-modules for the initialization of the model (Figure 5.7).

The Agent initialization module creates the farmer and individual agents using the number of farmer households and household sizes defined for each ward. WetABM provides multiple options to upscale and generates farmer attributes from a sample. (1) Sampling based on the empirical distribution of the attributes from the survey (Berger & Schreinemachers, 2005). To capture the correlation between attributes, empirical distribution is fitted for each farm type we identified. Then, the sample is drawn from the distributions for the total number of actual farm households in each ward. (2) Attributes of agents can also be created using a Bayesian belief network sampling (specifically forward sampling) to upscale the sampled household data. The Bayesian belief network sampling will allow as to generate the synthetic data based on the marginal distribution and the conditional dependence of the variables (Young, Graham, & Penny, 2009).

Land space initialization, on the other hand, creates the virtual landscape of the model. A grid-based representation of the KVF is created using raster data from different sources. The landscape is represented by 100m by 100m (1 hectare) resolution, and each cell is representing an object in Object-Oriented Programming language that stores different states of the landscape. The main attribute is a current land use state of the cell, which is initialized based on the land use map created using Landsat satellite images Leemhuis et al. (2017). Besides, the cell also contains attributes or states for distance from the nearest river, distance from the road and market, topographic wetness index, and the code of the ward the cell belongs. Each attribute was created based on the pre-processing of the vector data (rasterization and changing the resolution of the raster data) in QGIS (see Figure 10; Figure 11; Figure 12; Figure 13; Figure 16; Figure 14; Figure 15 in WetABM online documentation) .

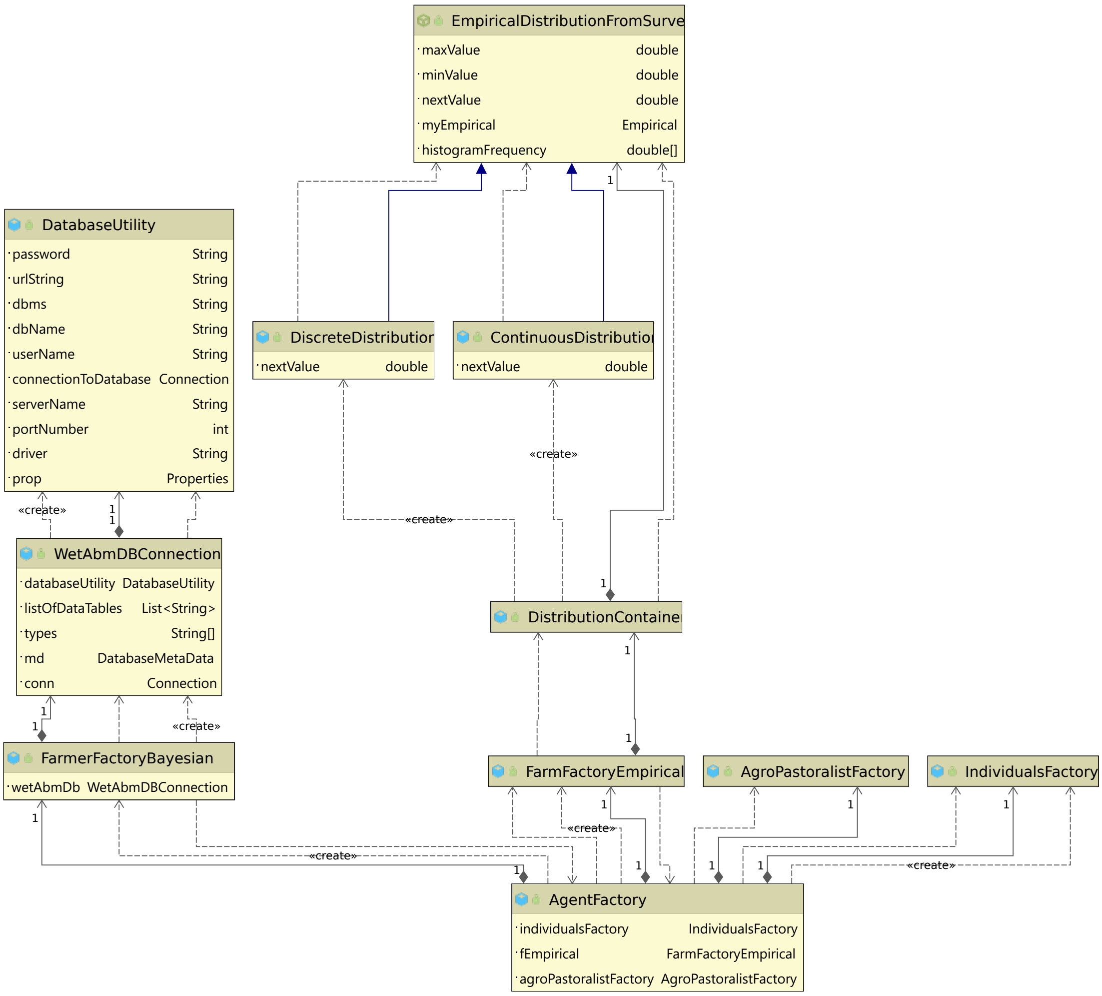


FIGURE 5.7: UML representation of Initalization in WetABM

The land allocation module creates the relationship between the farm agent and the landscape through “land tenure” relation. Since cadastral maps and farm boundaries are not available for the study region, a simple neighborhood-based allocation algorithm was created. The land allocation algorithm works in the following sequence of routine:

➔ Randomly choose homestead

- Each ward contains plots that are designated as built-up areas or "Village land". The areas are classified based on OpenStreetMap village center areas and GPS points collected during the household survey.

➔ Choose plot with the boundaries of the ward

- Once the farmer agent chooses his homesteads, he will choose the first farm plot within the boundaries of his ward given the plot is not occupied and the current land use is cropland

➔ If a land size is greater than one, allocate plot in the neighborhood of the first plot

- Assign ownership of the cell to the farmer through ID

➔ One the plot allocation is completed, a relationship between the plots and the FA is created through ownership and tenure.

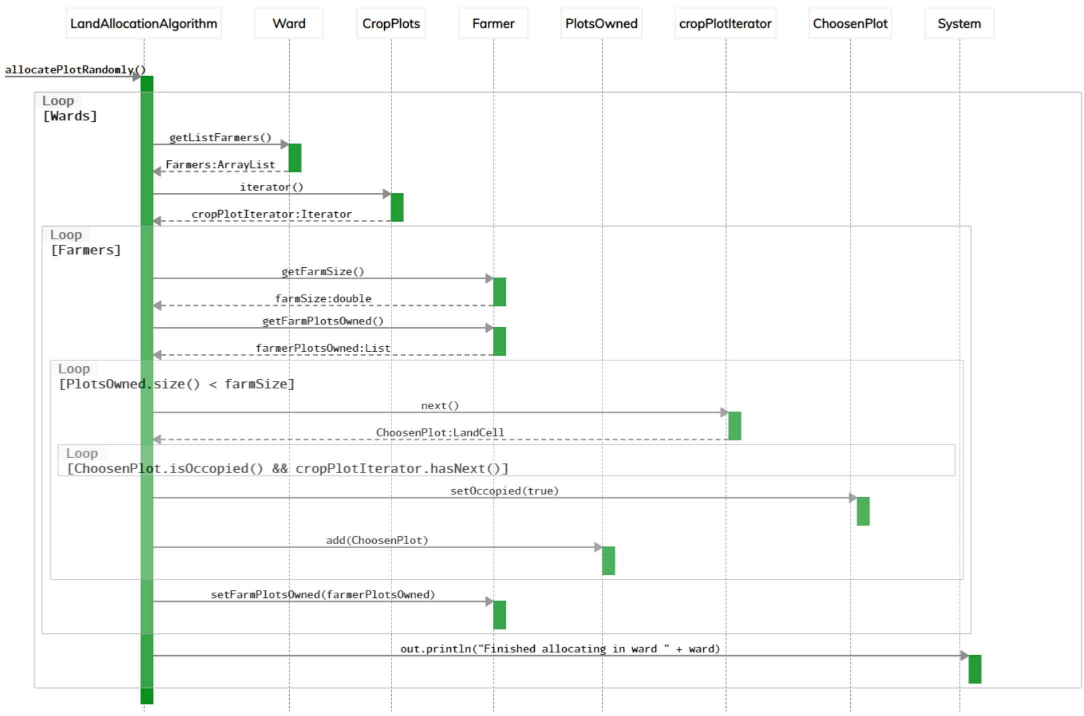


FIGURE 5.8: Sequence diagram of land allocation algorithm

5.2.4 Sub Models

5.2.4.1 *Household dynamics*

The household dynamics model controls the FA and household demographics.. In every simulation period, age progresses for the head and the individual household members. Depending on the probability of birth, a new individual agent will be created and added to the household, which in turn will increase the household size and the adult equivalent of the household. Similarly, depending on the probability of death, each individual has the probability of dying. In the case of the death of the head, the household will have a new head selected from its members provided that there is a hire. Once aging and the dynamics are calculated, the model will update the adult equivalent for the household. An adult equivalent is calculated based on the OECD scale (Haughton & Khandker, 2009).

$$AE = 1 + 0.7(N_{\text{adults}} - 1) + 0.5N_{\text{children}} \tag{5.2}$$

where AE is adult equivalent,$N_{\text{adults}}$is the number of individuals between the age of $15 - 65$ and N_{children} is between the age of $0 - 15$. Within WetABM, the adult equivalent is essential for calculating minimum food requirement and amount of labor availability.

5.2.4.2 *Land expansion*

Farm agent also has the probability of expanding their land by clearing new land within the boundaries of their respective ward. The result from our household survey shows that 8 percent of the households have cleared new land with an average of 5 hectares. Clearing new land also depends on the protection effectiveness of a particular land cell. The probability of protection is assumed to have only two values: There are protected areas that are adequately protected, for example, the Udzungwa mountains and the river banks. On the other hand, the Ramser site has only 39 percent of being protected effectivelly (Munishi, Chuwa, Kilungu, Moe, & Temu, 2012).

5.2.4.3 *Exit decision*

FA will exit farming as a result of two conditions. First, if the age of the HH head is more than 70 years, and there is no hire in the household, then the household will exit farming, and its parcel will be unoccupied. Second, if FA is not able to satisfy the minimum food requirement of its members for three consecutive years, then the FA will exit farming.

5.2.4.4 *Labor allocation decision*

The labor allocation submodule is responsible for monitoring the total amount of labor availability both from family and hired labor and allocating to crop production. The amount of family labor supply per man-days is modeled as a linear function of the adult equivalent of the FA. Each adult equivalent is assumed to provide 47 man-days per year. The share of hired labor is a constant variable that varies between farm types. Here we make a strong assumption that hired labor is a variable when needed. However, our observation in the KVF was that there is enough labor during the land preparation and harvesting periods due to high seasonal migration to the region.

5.2.4.5 *Price expectation*

Farm agents follow different expectations heuristics which consider both the expectation p_t^e and the effective realization of the actual price p_{t-1}. The expected prices are used for crop choice and intensification decision during the planning period. And actual prices are calculated at the end of the production activity based on total production of rice and maize and commercialization activity. Within WetABM the expected price in period $t+1$ can be determined based on the following three expectation heuristics developed by (Brock & Hommes, 1997; Caiani, Russo, Palestrini, & Gallegati, 2016, p.17-18)

1. Adaptive Heuristic
$$p_{t+1}^e = p_{t-1} + \omega(p_{t-1} - p_t^e) \tag{5.3}$$

   If $\omega = 1$, the equation will give us the "naïve" expectation $p_{t+1}^e = p_{t-1}$ where agents expectation is equal to the previous realized price.

2. Trend-Following Heuristic
$$p_{t+1}^e = p_{t-1} + \gamma(p_{t-1} - p_{t-2}) \tag{5.4}$$

   With $\gamma > 1$, the higher γ, the stronger the impact on the trends on an expectation

3. Anchoring and Adjustment Heuristic

$$p^e_{t+1} = 0.5 * (p_f + p_{t-1}) + (p_{t-1} - p_{t-2}) \tag{5.5}$$

Where p^f the fundamental level of price used as an anchor. In WetABM, it defined as the average of past realization of prices.

$$p_f = \frac{1}{t}\sum_{i=0}^{t-1} P_i \tag{5.6}$$

5.2.4.6 *Intensification and crop choice decision*

The module represents the mechanisms of the crop choice process and choice of intensification option of the farm household agent. Farm agents use the Bayesian belief network to decide on the type of crops to plant for a specific plot and period. The BBN acts as an internal decision-making routine encapsulated into the blueprints of the agent. In other words, farmers use the BBN as a mental model to make inferences about which crop to plant and which intensification option to choose conditioned on their characteristics, characteristics of their plot, and the market prices of each crop. The crop choice routine is formally expressed as:

$$[P_A(c|X_1...X_i)]_{c\varepsilon C} \tag{5.7}$$

$$Acceptance\, of\, c^* = \begin{cases} true, & if\, \mu[0,1] \leq P_a(c) \\ false, & otherwise \end{cases} \tag{5.8}$$

And the intensification routine is formally expressed as:

$$[P_A(s|X_1...X_i)]_{s\varepsilon S} \tag{5.9}$$

$$Acceptance\, of\, s^* = \begin{cases} true, & if\, \mu[0,1] \leq P_a(s) \\ false, & otherwise \end{cases} \tag{5.10}$$

where $\mu\,[0,1]$ is a random number from a uniform distribution

For each farm agent, the BBN returns a vector of probability for each option. Based on these conditional probabilities, a strategy is chosen that satisfies the above random choice condition.

However, contrary to the crop choice sub-module where farm agents make crop choices every production year, intensification decision is triggered by different conditions. There are two primary intensification triggering mechanisms in WetABM: first, the farm agent will consider to intensify his production if the income from the last production period was not sufficient to meet the minimum food requirement of the household. The second triggering mechanism, on the other hand, uses the relative position of the farm households in terms of income compared to his peers in a particular ward. If his income is in the bottom 25 percent of the cumulative distribution of income in the ward, then the agent will consider intensifying his production. As a result, WetABM contains an implicit dynamic aspiration of farm households.

5.2.4.7 *Crop yield*

The A crop yield sub-model is a module for performing the dynamics of Crop yield in response to variations in natural conditions and management practice. Following (Le et al., 2010) and (Julia Schindler, 2010), WetABM uses a production function approach to estimate the yield of a particular crop given the factors of production and characteristics of the farm plots. A quadratic production function is empirically estimated for rice and maize outside of the WetABM, and the parameters protocol in the crop yield sub-model. Yield for rice and maize for each agent is given by

$$Y_{AC} = \alpha + \sum_i \beta_i x_i + \sum_i \sum_j \delta_{ij} x_i x_j \tag{5.11}$$

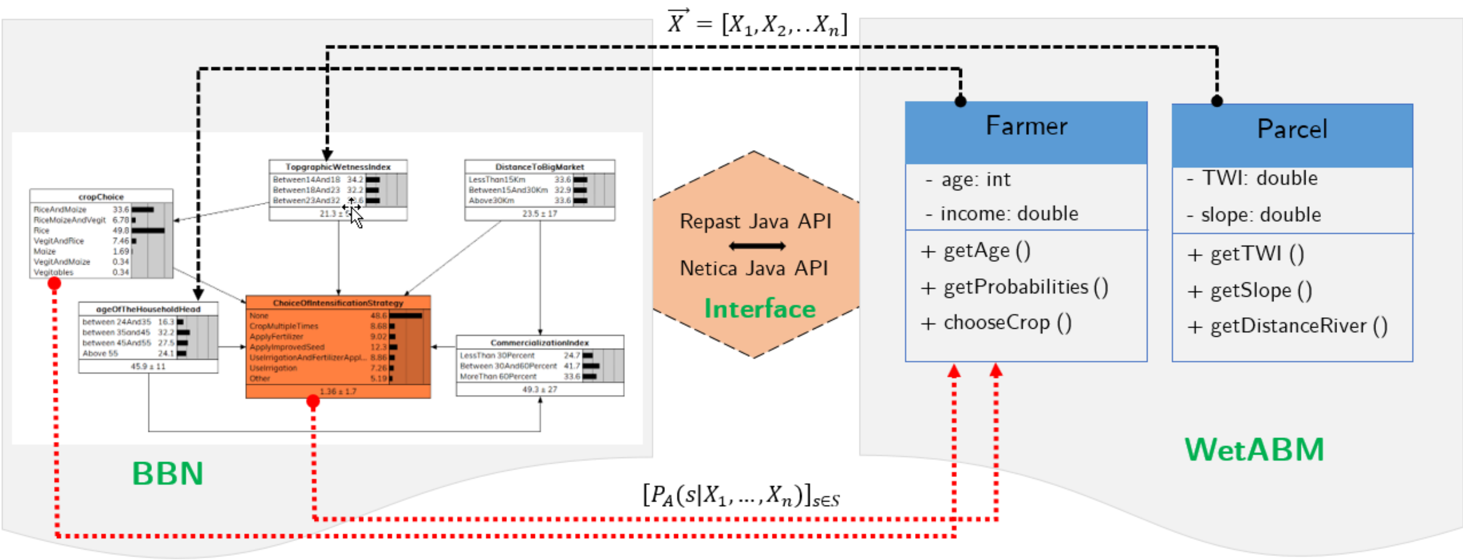
Where y is Y_{Ac} is the yield for agent A crop c, β_i and δ_{ij} and x_i are

- Land size (ha)

- Capital (Tsh)

- TWI (index)

- Labor (man-days)



$\overrightarrow{X} = [X_1, X_2, ..X_n]$: Vectors of Attributes

$[P_s(s|X_1 ... X_i)]_{s \in S}$: Vectors of Conditional probabilities for each state

: Output

: Input

FIGURE 5.9: Illustration of Bayesian inference within WetABM

- Fertilizer application (yes=1, No=0)

- Irrigation application (yes=1, No=0)

- Multi-Season farming (yes=1, No=0)

- Improved seed (yes=1, No=0)

In addition, the yield is stochastic for each agent where the estimates yield varies within the prediction interval

$$\hat{y}_{ac} \pm t_{(\alpha/2,n-p)} \times \sqrt{MSE + (se(\hat{y}_{ac}))^2} \tag{5.12}$$

Where $\hat{y}_{\mathrm{ac}}$ is predicted output for crop c and agent a and $\sqrt{MSE + \ (se(\hat{y}_{\mathrm{ac}}))^2}$ is the standard error of prediction and $t_{(\frac{\alpha}{2},n-p)}$ is "t-multiplier" with $n-p$ degrees of freedom.

5.2.4.8 *Market*

WetABM also has a virtual market agent that coordinates output transactions between farmers and the outside world. In the present version, only the output market component is modeled. The market agent is spatially fixed at the biggest market in the region and collects Rice and maize output from farmers and provides disaggregated prices based on distance and transaction costs at the ward level. Following Happe, Balmann, and Kellermann (2004) and Kruseman (2000), the market determines the price of rice and maize for a particular production period using neither fully elastic nor fully static demand (endogenous price formation). Aggregate price for crop c at time t is given by the equation:

$$P_{C,t} = P_{C,0} \ \ .\tau_c^{-(t+1)} \ \ . \left(\frac{\sum_A Q_{AC,t}}{\sum_A Z_A}\right)^{-\theta_c} \tag{5.13}$$

where $P_{c,0}$ is price at the initialization for crop C,τ_c controls for price trend overtime, θ_c is inverse demand elasticity for crop C

Once the aggregate prices are estimated in the central market, the virtual agent then allocates a heterogeneous price for each ward, depending on the distance and transportation cost between the ward centers and the central market. Although prices are heterogeneous among the wards, farm agents with in the same wards will receive the same farm get price.

5.2.4.9 *Income accounting*

The income accounting model contains different routines that calculate the financial status of the household. The following HH balances are calculated by income accounting submodel. Total revenue is the amount of rice or maize sold to the market multiplied by the market price. Total farm income is the difference between total revenue and total agricultural expenditure. Agriculture expenditure contains overhead cost, cost of hired labor, cost of Intensification (if adopted) and transportation cost from the parcel to the homestead.

Total household income is calculated as the sum of agriculture income and non-farm income. Non-farm income is a parameter that varies between farm types.

Total household saving is calculated as the difference between total income and household expenditure. Household expenditure is composed of food and basic needs expenditure that depends on the adult equivalent of the household.

5.2.4.10 *Migration*

Within WetABM, the immigration module controls the exogenous increase of agropastoralists to the valley. Here we assume that the agent has decided to migrate to the valley, and we do not explicitly consider his/her decision to migrate. As outlined earlier, mainly agropastoralists are migrating into the valley. They model adds new agropastoralists in every period depending on the migration rate being an exogenous parameter. The migrating agent will first scoop across wards for potential migration areas depending on population density, potentially available land, and the proportion of agropastoralist within the ward.

Once he decides on the ward to which he will migrate, he will look for land to clear, and if successful, he will clear land depending on a pre-determined farm size for that particular agent. Similar to the initialization, the attributes of migrating farmers are created based on the empirical distribution of agropastoralist farmers. The size of land to be cleared is rather an aspiration of the migrating farmer, and it is not based on the availability of land. In order to capture the complex interplay of land tenure and conservation in the valley, each plot land has a probability of being protected and enforcement of protection. One particular characteristic of the Kilombero valley in terms of the protected area is the variation in the enforcement of protection zones. Some of the protected areas are relatively highly protected either due to biophysical constraints (Uduzungua Mountain and the river banks) or based on the priority given to areas by the government and conservation groups. For example, the source of income for the regional and national government in the Selous game reserve area is highly protected compared to the neighboring Ramser site and Kilombero Game controlled area. Hence the highly protected area is masked out from land conversion, and the remaining protected areas will have a constant probability of 39% of being protected (Munishi et al., 2012).

Besides, there is a competition between the native farmers who decide to expand their land and the migrant farmers for a specific plot. Assuming the native farmers having an advantage in terms of better information on the availability of land within his ward, he will have the priority of getting the land in case both farmers planned to clear the same plot of land.

5.3 Validation and verification of WetABM

Before presenting model results, we will briefly describe our approach to verification and validation of the WetABM. Verification is the process of "determining that a computational software implementation correctly represents a model of a process" (Ormerod & Rosewell, 2009, .p 131). Validation, on the other hand, is the process of "assessing the degree to which a computer model is an accurate representation of the real world from the perspective of the model's intended applications" (Ormerod & Rosewell, 2009, .p 131).

Validation and verification of ABM is often challenging, especially when data is limited to perform result validation through comparison of outputs of a simulation run against the real world (Olsen & Raunak, 2016).

Three different approaches were used in terms of verification and validation. The first approach was to validate the critical component of the model, the Bayesian belief network, using a cross-validation technique (refer to the third part of the dissertation for more details); besides, we validated the agent typology based on secondary data. This is very important as it determines key parameters and attributes that are used to initialize the agents within the model.

The second approach is related to the verification of the models. With such a large-scale nature of the WetABM, software implementations are prone to errors and its easy to overlook coding mistakes. In building WetABM, we followed a step by step upscaling approach (Figure 5.10). A smaller prototype model with random parameters and the landscape was built in the first step. In the next step, we built an empirical ABM for one particular ward with all the details, except the market interaction, we planned for WetABM. Once we are confident the Ward model was performing as intended, we upscaled the software implementation to the other ten wards, and the market interaction was implemented thereafter. Our third attempt at verification was code walkthrough, debugging, and unit testing. When appropriate, we used unit testing within the java framework to test if methods are performing correctly under different plausible conditions.

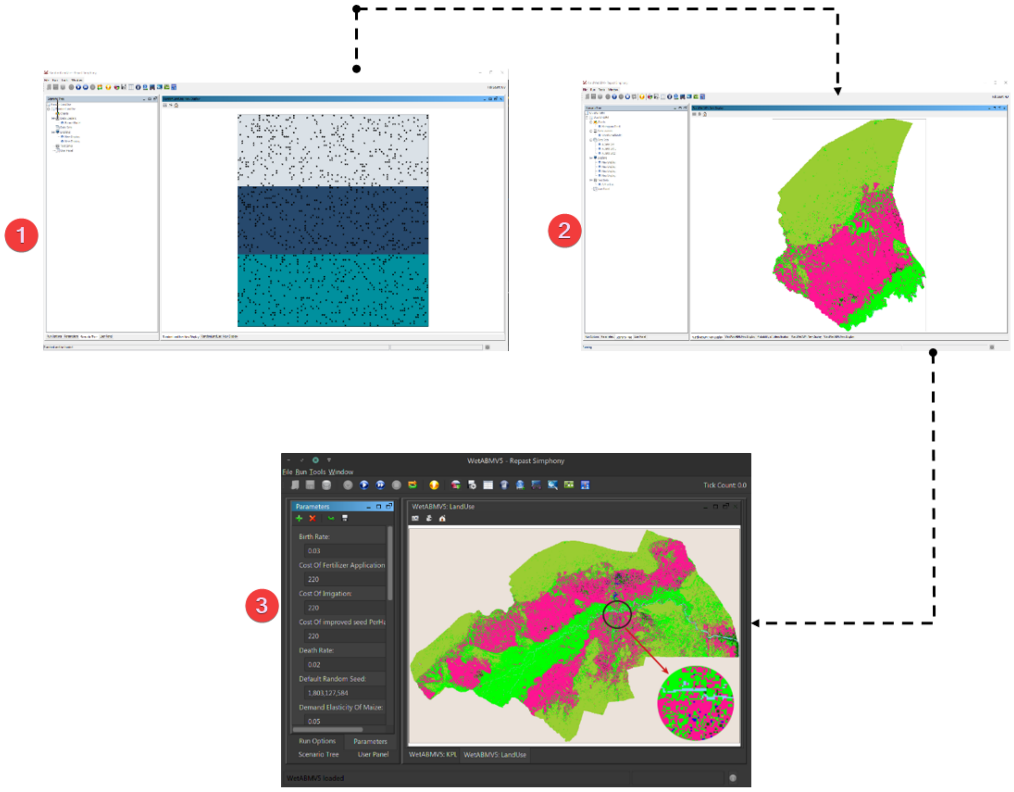


FIGURE 5.10: Upscaling steps in building verified WetABM

5.4 Scenarios

BAU The business-as-usual (BAU) scenario is simulated to estimate projected baseline changes in the Kilombero basin. This scenario assumes a continuation of historical population growth and land conversion for agriculture. It implements all aforementioned submodels of household dynamics, crop choice, and intensification decisions

Road Infrastructure (RI)

Without going into the details of the effect of road infrastructure on the functioning of the agriculture sector, we assume the planned construction and rehabilitation of existing road network will reduce 20 percent of the baseline transportation cost for output into the market. In other words, under the RI scenario, we assume the difference between the farm get price and the aggregate price at the central market will reduce by 20 percent.

Immigration (IM) There is no official number of immigrants in the valley. This is due to the problematic relation between the village offices and the majority of immigrants not being officially registered in the village when they arrive. From our survey and field visit, it is evident that migration rates are not linear, and the number of migrants varies between years. Social networks play a significant role in pulling family members and relatives after the first migrant is settled into the valley. The immigration scenario simulates a potential 3 percent increase in the number of households due to immigration. Two different aspects differentiate between localized population growth and migration in the model:

1) Compared to localized population growth, immigration leads to an increase in agropastoralists. Furthermore, (2) in the valley, migration is associated with land expansion to marginal lands rather than buying existing farmland.

5.5 Result and discussion

Next, we present the results of our simulation experiment using the aforementioned scenarios by looking at the effect on the pattern of intensification, agricultural production, and land use. To take the uncertainties associated with random numbers into account, we ran a set of 30 simulations for each scenario, and the results presented is the smoothed mean trend of these model runs. Each iteration step represents one year, and the model runs for a total of 25 years.

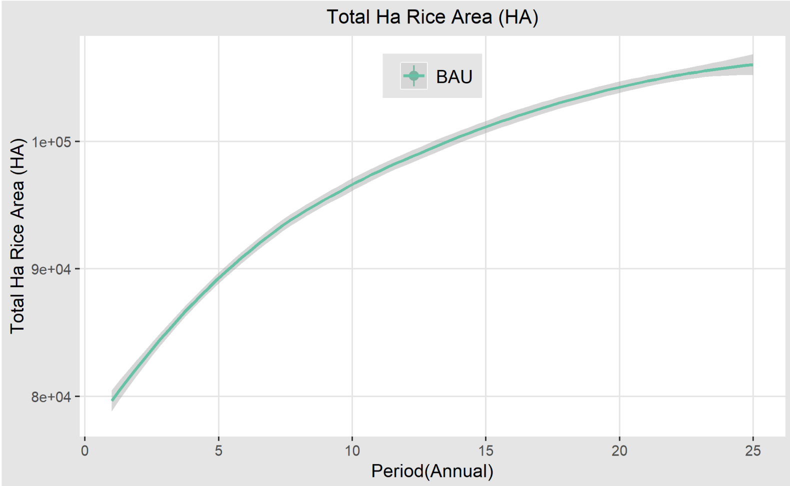
Under the "baseline scenario," farm households conduct their annual production routine by starting to form price expectations, cropping, and intensification decision, production, and market participation. The state variables under the baseline scenario are household composition (both household size and age of members), labor availability, household food requirement, expected prices for rice and maize, and per capita income. Figure 5.11 presents trends in agricultural output and land use over a period of 25 years. Under the BAU scenario, aggregate rice output increases by 14% and reached 378, 000 metric tones over the simulation period (Figure 5.11b). On the other hand, the total maize output also increased by 22% percent to 72 000 tones (Figure 5.11d). The increase in the production for both crops is attributed to more land allocated to the crops. The land area allocated for rice increased significantly to 106,044 ha, which is an increase from 42% to 56% of the total cropland in the base year. The total hectares of land allocated to maize also increased steadily to 22,269 ha, which rise from 9 % to 12% of the total cropland in the base year. The additional cropland allocated to the two crops is the result of land expansion by farmers and reallocation from land to other crops.

Looking at the trend of intensification measured by the number of farm households using one or more of the four intensification options, the result shows a mixed trend. While a number of improved seed and irrigation users stabilizes after the 5th simulation year (Figure 5.12a and Figure 5.12b), the trend for fertilizer users and farmers with multi-season farming decline over time. One possible explanation for the trend is that previous year per capita income is affecting the adoption of the intensification choice and farm households might be in a income trap where they can not get enough per capita income to jump from one income state to the next. Thus not enough capital is available for adopting one of the options.

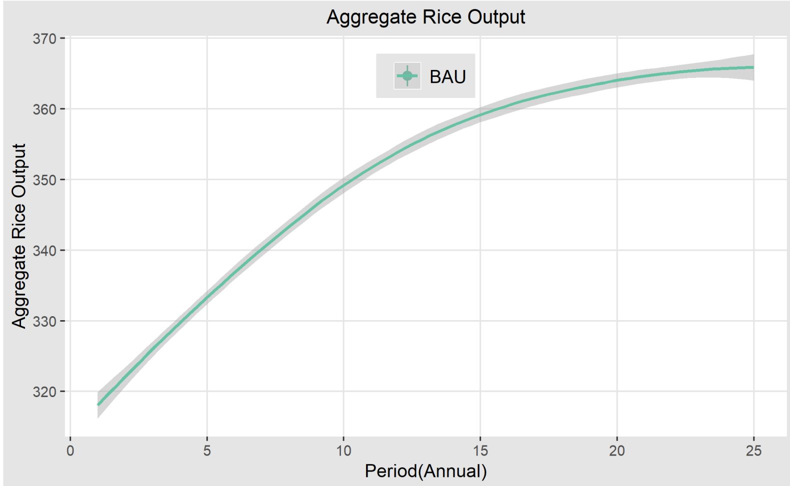
Under the migration scenario, new farm households are introduced in each simulation period which increases the number of farm households by 3% percent annually. As expected, the total output for rice and maize rises over time, and the increase is higher than under the baseline scenario (Figure 5.13a and Figure 5.13c). The increase in aggregate output is attributed to both an additional cropland area that is brought to cultivation by the new farm households and an increase in land area allocation for the two crops. Since all migrating farmers are assumed to be agropastoralists who are characterized by large farm sizes, the total new cropland in the study area increased by 37% at the end of the simulation period. One of the interesting results under the migration scenario is the total ha of land allocated to maize also increases relatively higher than under the baseline scenario.

Figure 5.14 shows the trend in intensification over the simulation period under the immigration scenario. Although the number of households applying fertilizer shows the same declining trend as the BAU scenario over time, the number of households using improved seed variety (Figure 5.14a) and crop multiple times(Figure 5.14b) indicates an increasing trend over time. After an initial decline in the number of households who use small scale irrigation(Figure 5.14c), the number of adopters stays stable and constant of the 10th simulation period.

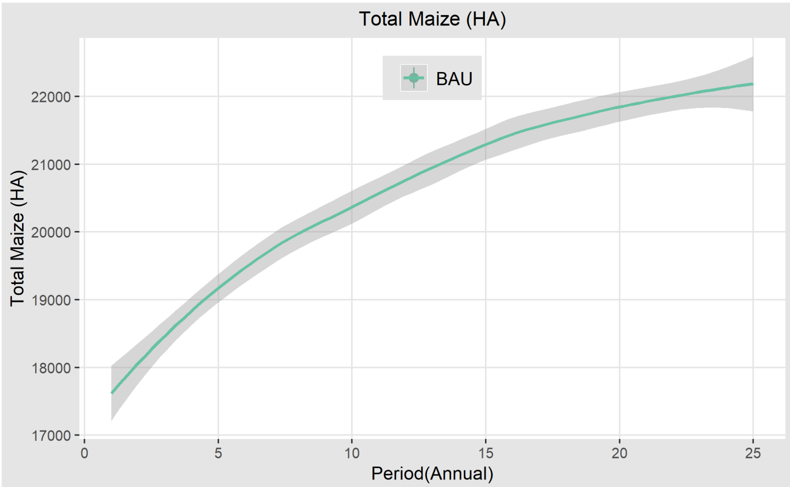
Under the road infrastructure scenario, we assume rehabilitation of rural road from the central market to the ward centers reduces the transportation cost, which will have a direct effect on the price difference between the primary market and farm get prices. The result from our simulation shows there is no significant effect on both production and intensification trend (Figure 5.15). While total rice production remains similar to the BAU scenario, the total maize production is higher under the RI scenario. When it comes to the trend of intensification, it remains the same as the BAU scenario with a small variation for a number of improved seed users and the number of fertilizer users. One plausible explanation for the negligible effect of the RI scenario is the reduction in transport cost didn't introduce a significant decrease in price differences between the ward and the central market. Hence no change in their crop production and intensification decision.



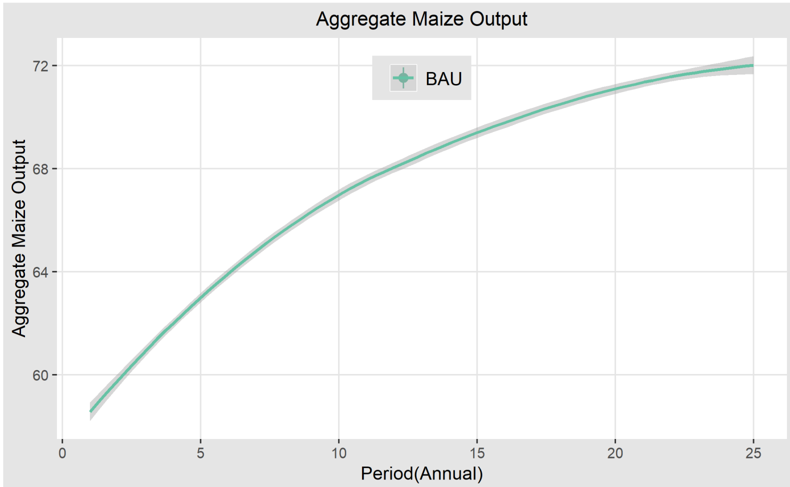
(A) Total area of land in hectares allocated to rice



(B) Total regional rice output in 1000 metric tone

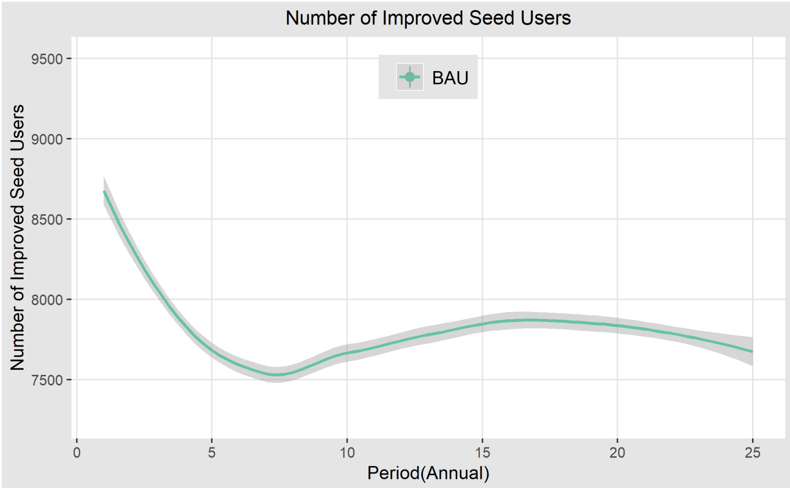


(C) Total area of land in hectares allocated to maize

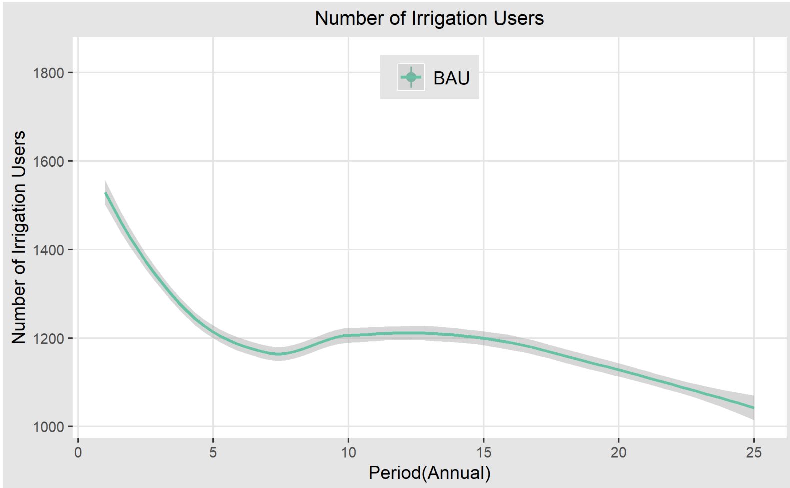


(D) Total regional maize output in 1000 metric

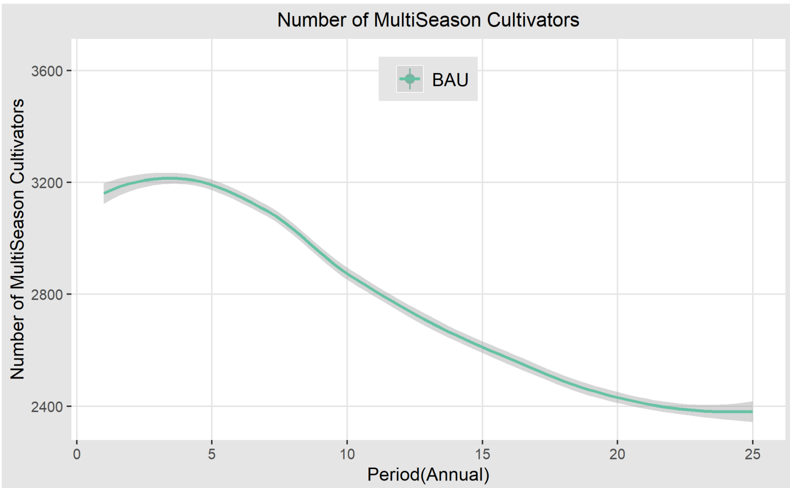
FIGURE 5.11: Crop production under baseline scenario



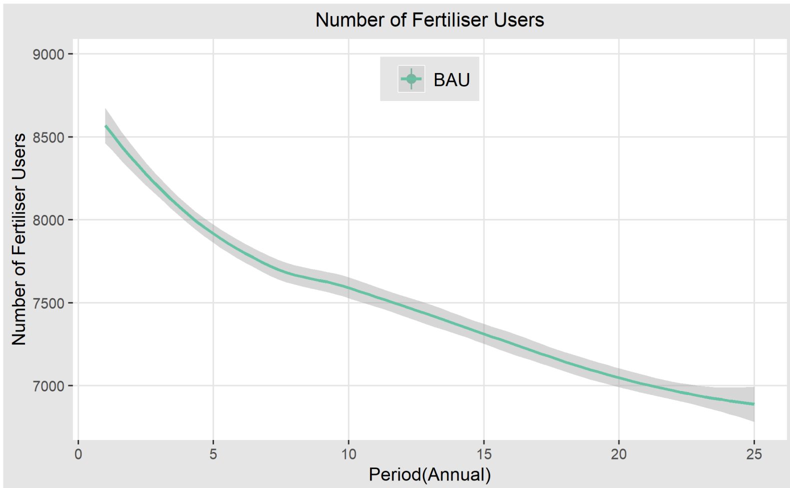
(A) Number of households with improved seed users



(B) Number of households using small scale irrigation

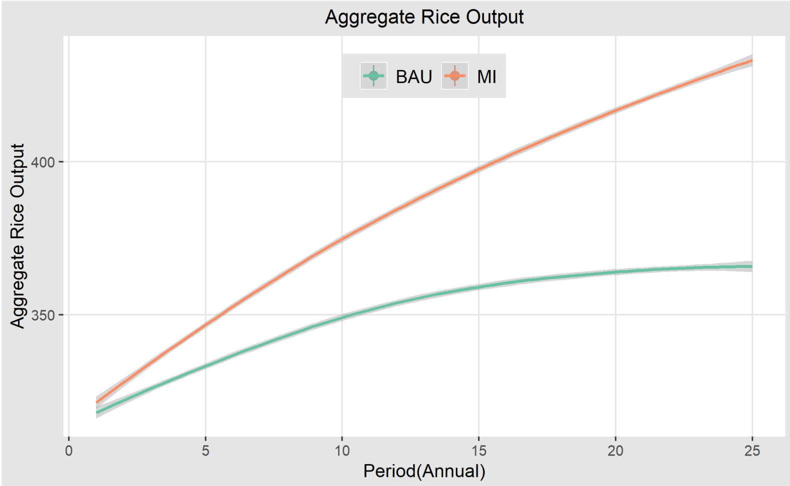


(C) Number of households cropping multiple seasons during the year

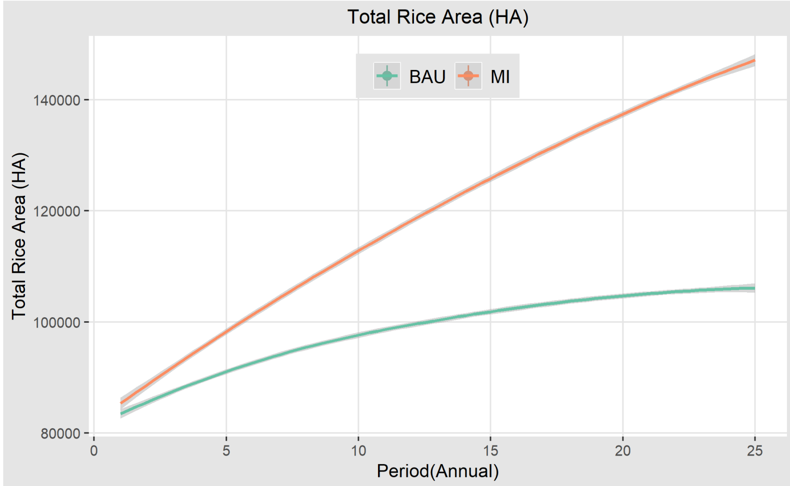


(D) Number of households using fertilizer

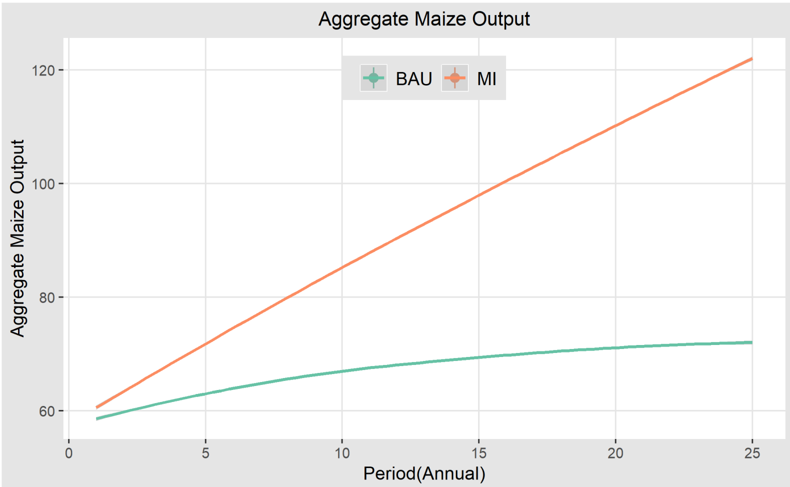
FIGURE 5.12: Number of Households by intensification option under the base line scenario



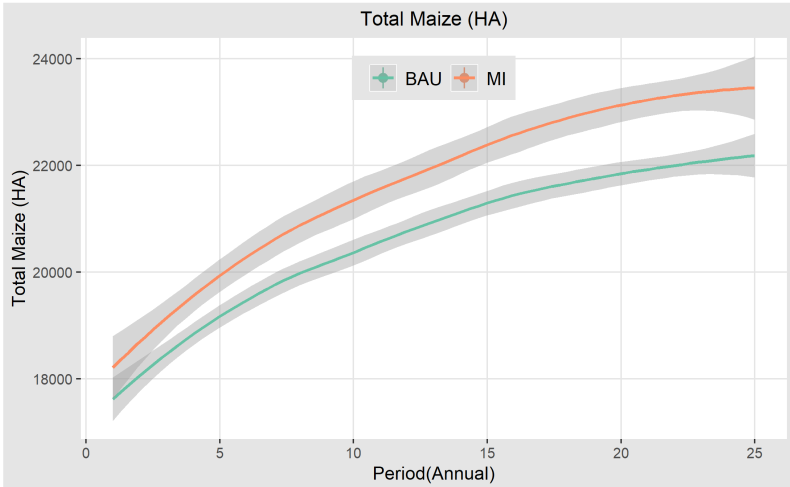
(A) Total regional rice output in 1000 metric tone



(B) Total area of land in hectares allocated to rice

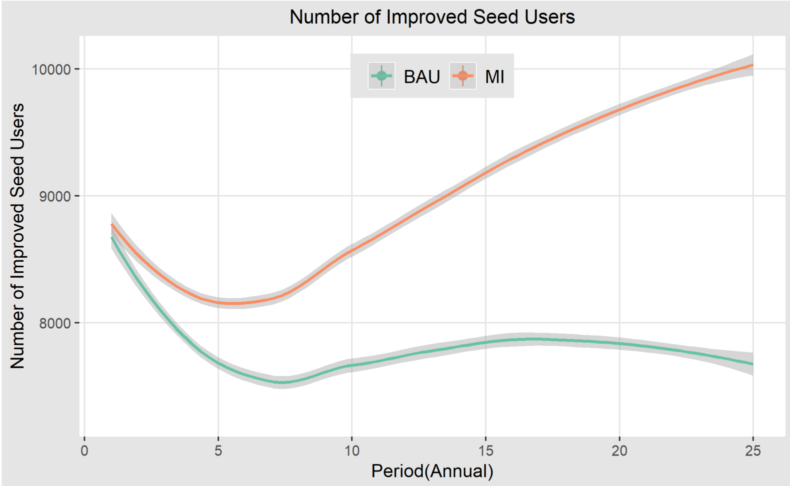


(C) Total regional maize output in 1000 metric tones

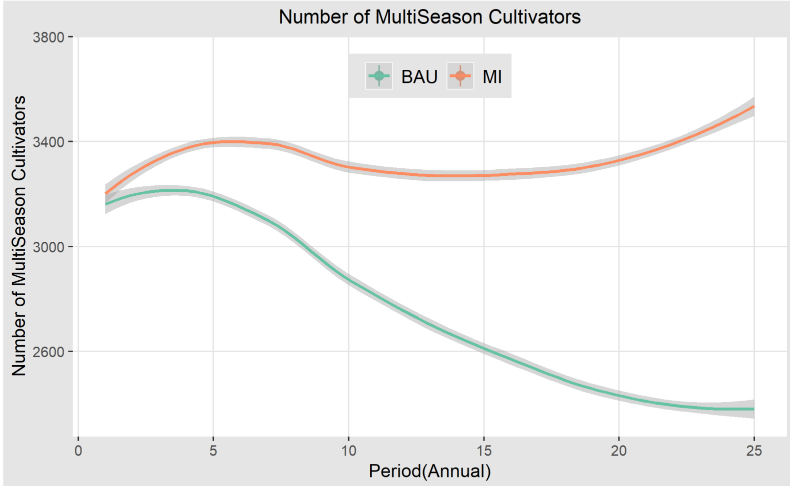


(D) Total area of land in hectares allocated to maize

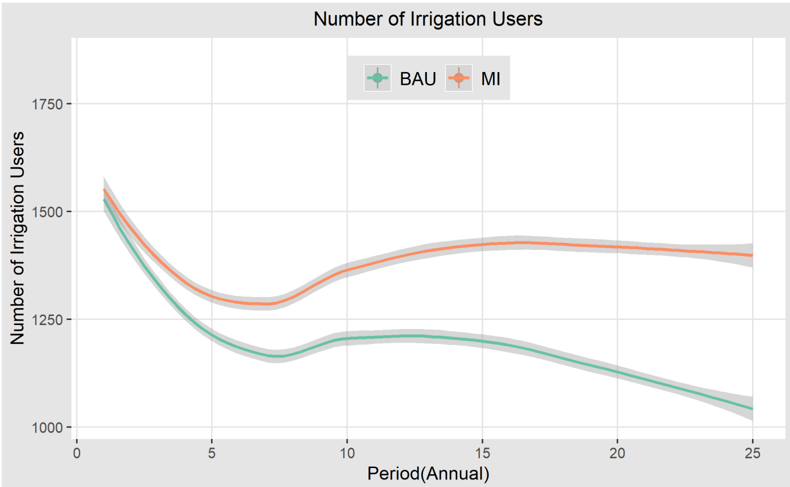
Figure 5.13: Crop production under baseline and in-migration scenario



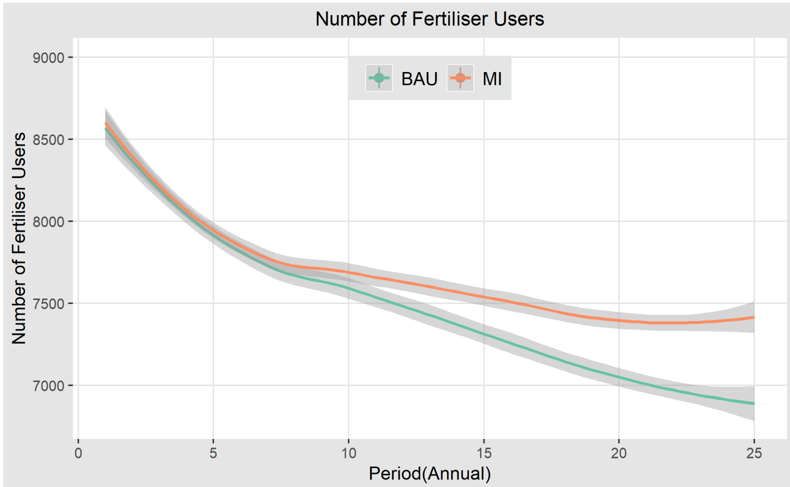
(A) Number of households with improved seed users



(B) Number of households cropping multiple seasons during the year

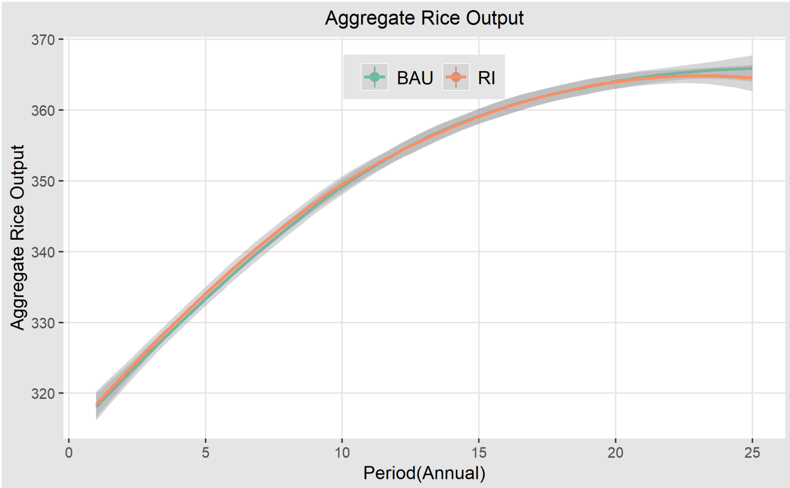


(C) Number of households using small scale irrigation

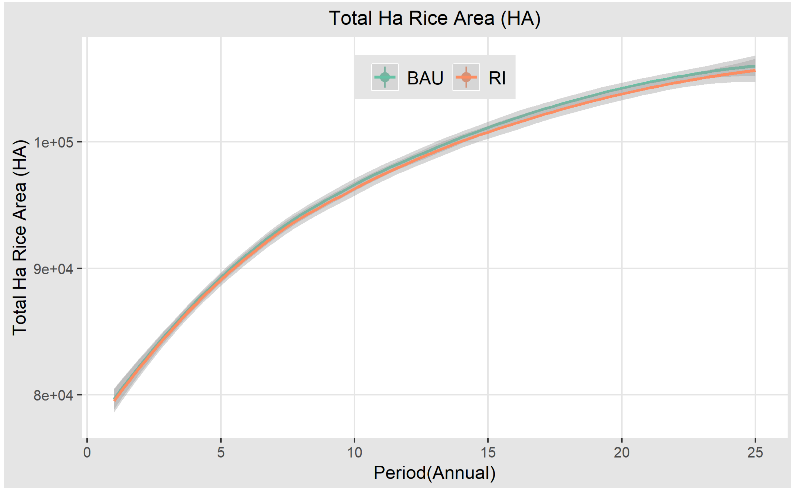


(D) Number of households using fertilizer

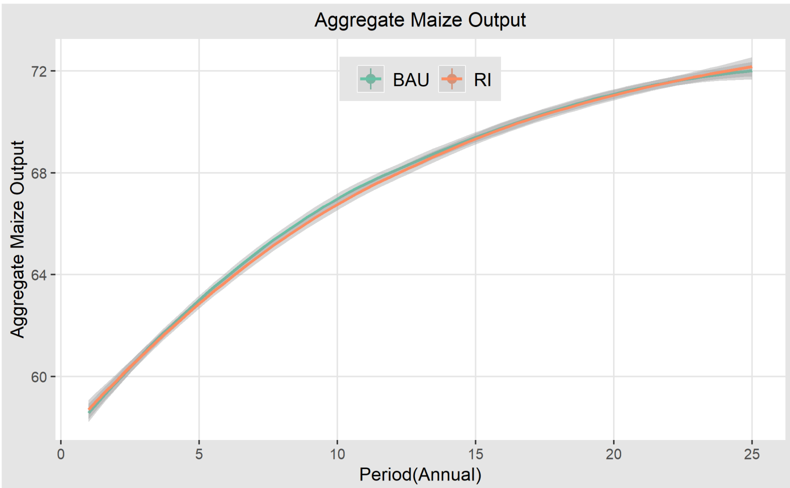
FIGURE 5.14: Number of Households by intensification option under the in-migration scenario



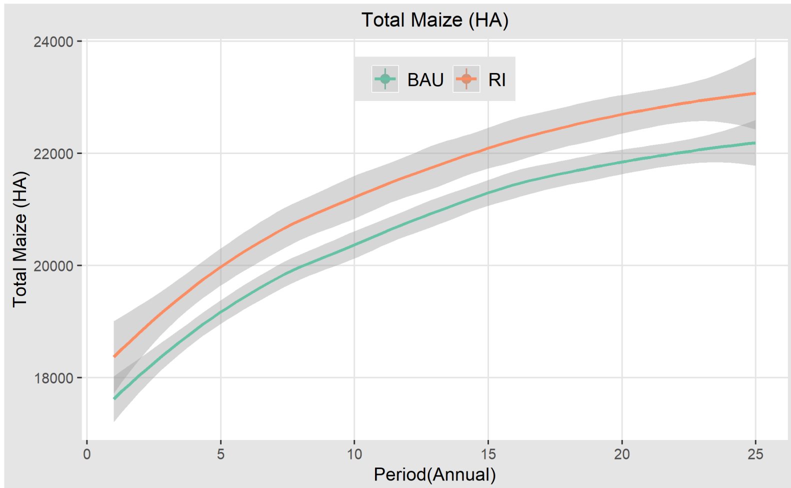
(A) Total regional rice output in 1000 metric tone



(B) Total area of land in hectares allocated to rice

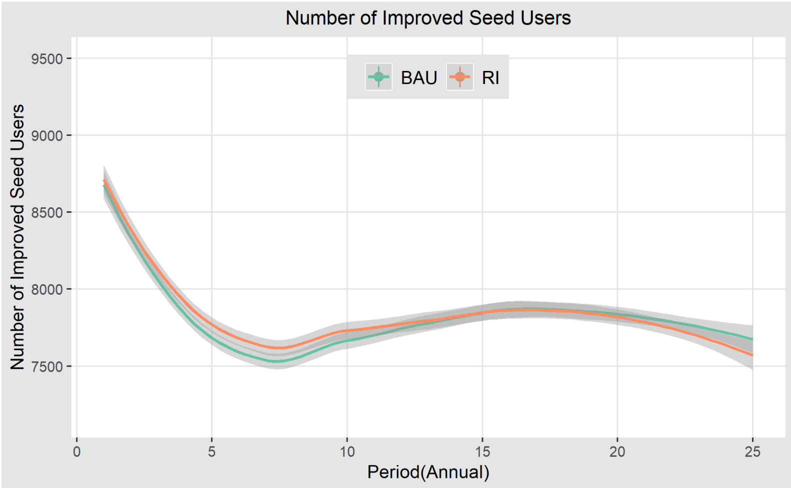


(C) Total regional maize output in 1000 metric tones

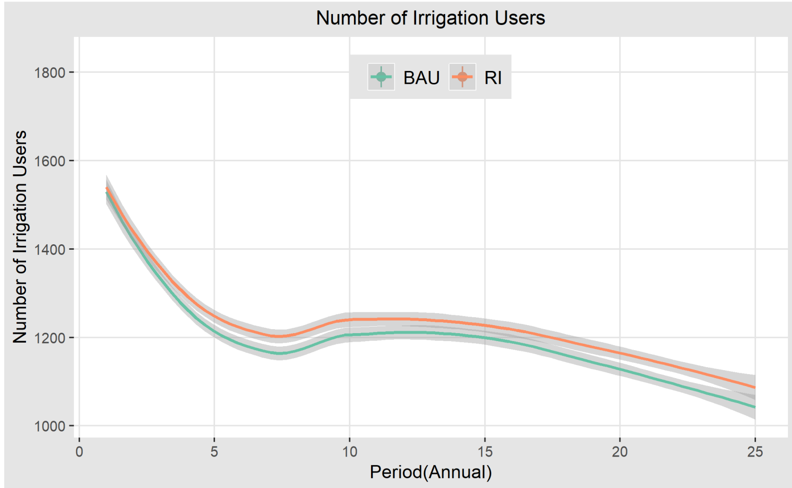


(D) Total area of land in hectares allocated to maize

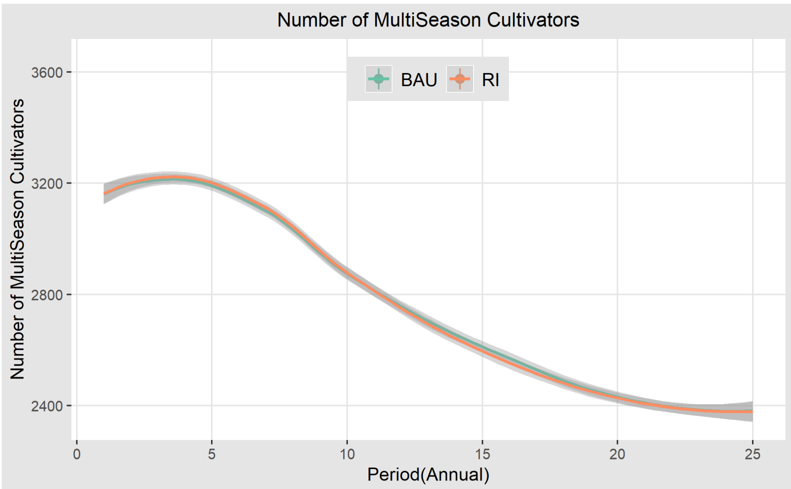
FIGURE 5.15: Crop production under baseline and road infrastructure scenario



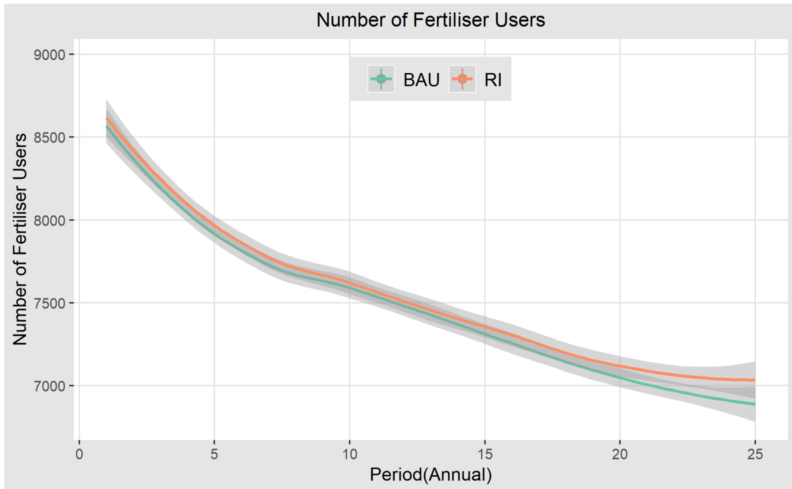
(A) Number of households with improved seed users



(B) Number of households using small scale irrigation



(C) Number of households cropping multiple seasons during the year



(D) Number of households using fertilizer

FIGURE 5.16: Number of Households by intensification option under the road infrastructure scenario

5.6 Conclusion

Our objective in this chapter was to examine the potential effect of sustained immigration and better access to the market on trends of intensification, land use, and agricultural production in KVF. We presented a newly built large scale empirical agent-based model (WetABM) that captured 38,000 - 51,000 farmer agents in all their heterogeneity and a detailed representation of the floodplain landscape. Our modeling approach complements the ongoing efforts in agent-based modeling with respect to scale, alternative modeling of decision making, rigorous empirical foundation, and integration of reach geospatial data. Although the full WetABM was not validated with empirical data (the challenge common to similar ABMs), we have followed multiple procedures to validate and verify different components and software implementation of the model.

Four different scenarios were simulated using the model developed. The baseline scenario shows that with the current production system, the overall intensification (measured by the number of households using one of the four intensification options) will not pick up but will decline in the long runThe production of rice and maize will increase mainly due to an increase in land allocation to the two crops and land expansion.

With a sustained increase of immigration into the valley resulting in increased population density and the current level of protection of wildlife corridors, farmers engage more in land expansion than in intensification. However, we observe a higher number of farm households using improved seed variety, small scale irrigation, and crops multiple seasons compared to the baseline scenario. The use of these three intensification options also stable over the simulation period.

Our simulation result for a reduction in transportation cost as a surrogate for improved road infrastructure shows a negligible effect on both intensification and agricultural production trends. Here we note that the results should be interpreted in reference to our modeling approach and the parameters. Under the baseline scenario, the average transportation cost parameter was 50 Tsh/kg/km(0.02€ /kg/km) . Reducing the transportation cost by 20 percent might not have been significant enough for the farmers to change their crop choice and intensification decision. However this will also require further explorations through sensitivity analysis and different transportation cost for different seasons.

5.7 References

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, *229*, 25–36.

Bamford, A., Ferrol-Schulte, D., & Smith, H. (2010). The status of the Ruipa corridor between the Selous Game Reserve and the Udzungwa Mountains. *Report to Frontier Tanzania, Dar es Salaam, Tanzania*.

Berger, T., & Schreinemachers, P. (2005). Empirical parameterization of multi-agent models in applied development research. In *workshop "empirical techniques for testing agent based models". indiana university, bloomington, usa* (pp. 2–4).

Bergius, M. (2014). *Expanding the corporate food regime: the southern agricultural growth corridor of tanzania. current and potential implications for rural households* (Unpublished master's thesis). Norwegian University of Life Sciences, Ås.

Bergius, M., Benjaminsen, T. A., Maganga, F., & Buhaug, H. (2020). Green economy, degradation narratives, and land-use conflicts in Tanzania. *World Development*, *129*, 104850. Retrieved from https://doi.org/10.1016/j.worlddev.2019.104850 doi: 10.1016/j.worlddev.2019.104850

Blache, A. (2019). Grabbing land, catching votes!. land and the 2015 election campaign in kilombero district, tanzania. *Les Cahiers d'Afrique de l'Est/The East African Review*(53).

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, 1059–1095.

Caiani, A., Russo, A., Palestrini, A., & Gallegati, M. (2016). Economics with Heterogeneous Interacting Agents. *New Economic Windows, Springer Series. DOI*, *10*, 973–978.

ERM. (2012). Southern Agricultural Growth Corridor of Tanzania (SAGCOT): Strategic Regional Environmental and Social Assessment. , 144.

Everard, M., & Wood, A. (2017). Agricultural management and wetlands: an overview. *The Wetland book I: structure and function, management, and methods. Springer, Beijing*.

Geist, H. J., & Lambin, E. F. (2001). *What drives tropical deforestation? A meta-analysis of proximate and underlying causes of deforestation based on subnational case study evidence.* (Vol. 4; Tech. Rep.). Retrieved from http://linkinghub.elsevier.com/retrieve/pii/095937809090005T doi: 10.1098/rsbl.2008.0691

Happe, K., Balmann, A., & Kellermann, K. (2004). *The agricultural policy simulator (AgriPoliS): an agent-based model to study structural change in agriculture (Version 1.0)* (Tech. Rep.).

Haughton, J., & Khandker, S. (2009). *Handbook on Poverty and Inequality*. The World Bank. Retrieved from https://econpapers.repec.org/RePEc:wbk:wbpubs:11985

Hunter, L. M., & Nawrotzki, R. (2016). Migration and the environment. In *International handbook of migration and population distribution* (pp. 465–484). Springer.

Julia Schindler. (2010). *A multi-agent system for simulating land-use and land-cover change in the atankwidi catchment of upper east ghana* (Doctoral dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn). Retrieved from http://hdl.handle.net/20.500.11811/4160

KILORWEMP. (2017). *Technical advisory services for biodiversity conservation and wetland management in Kilombero valley RAMSAR site, Tanzania:PASTORALISM DIAGNOSTIC STUDY* (Tech. Rep.). Retrieved from https://kilomberovalley.files.wordpress.com/2019/02/kilorwemp-2017-scoping{_}imp{_}gmp.pdf

Kruseman, G. (2000). *Bio-economic household modelling for agricultural intensification* (Doctoral dissertation). Retrieved from https://edepot.wur.nl/139103

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. *Annual review of environment and resources*, *28*(1), 205–241.

Le, Q. B., Park, S. J., & Vlek, P. L. G. (2010). Land Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. 2. Scenario-based application for impact assessment of land-use policies. *Ecological Informatics*, *5*(3), 203–221.

Leemhuis, C., Thonfeld, F., Näschen, K., Steinbach, S., Muro, J., Strauch, A., ... Diekkrüger, B. (2017, aug). Sustainability in the Food-Water-Ecosystem Nexus: The Role of Land Use and Land Cover Change for Water Resources and Ecosystems in the Kilombero Wetland, Tanzania. *Sustainability*, *9*(9), 1513. Retrieved from http://www.mdpi.com/2071-1050/9/9/1513 doi: 10.3390/

su9091513

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Wetlands and water synthesis* (Tech. Rep.). World Resources Institute, Washington, DC. Retrieved from http://protosh2o.act.be/VIRTUELE_BIB/Water_in_de_Wereld/MIL-Milieu/W_MIL_E45_ecosystems_human2.pdf

Msofe, N. K., Sheng, L., Lyimo, J., Msofe, N. K., Sheng, L., & Lyimo, J. (2019, jan). Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability*, *11*(2), 505. Retrieved from http://www.mdpi.com/2071-1050/11/2/505 doi: 10.3390/su11020505

Munishi, P. K. T., Chuwa, J. J., Kilungu, H., Moe, S. R., & Temu, R. P. C. (2012). Management Effectiveness and Conservation Initiatives in the Kilombero Valley Flood Plains Ramsar Site, Tanzania. *Tanzania Journal of Forestry and Nature Conservation*, *81*(2), 1–10.

Murray-Rust, D., Dendoncker, N., Dawson, T. P., Acosta-Michlik, L., Karali, E., Guillem, E., & Rounsevell, M. (2011). Conceptualising the analysis of socio-ecological systems through ecosystem services and agent-based modeling. *Journal of Land Use Science*, *6*(2-3), 83–99. doi: 10.1080/1747423X.2011.558600

NorsysSoftwareCorp. (2016). *Netica-J Manual.* Retrieved from http://www.norsys.com/netica-j/docs/NeticaJ{_}Man.pdf

North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M., & Sydelko, P. (2013). Complex adaptive systems modeling with Repast Simphony. *Complex Adaptive Systems Modeling*, *1*(1), 3.

Nystrom, E. A., & Burns, D. A. (2011). *TOPMODEL Simulations of Streamflow and Depth to Water Table in Fishing Brook Watershed, New York, 2007-09.* US Department of the Interior, US Geological Survey.

Olsen, M. M., & Raunak, M. S. (2016). Metamorphic validation for agent-based simulation models. Montreal, Quebec, Canada: Society for Computer Simulation International. Retrieved from https://dl.acm.org/doi/10.5555/3015574.3015607

Ormerod, P., & Rosewell, B. (2009). Validation and verification of agent-based models in the social sciences. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5466 LNAI*, 130–140. doi: 10.1007/978-3-642-01109-2_10

Otsuka, K., & Place, F. (2013). Evolutionary changes in land tenure and agricultural intensification in sub-Saharan Africa. *United Nation University Working Paper*, *2014/051* (November), 47. doi: 10.1093/oxfordhb/9780199687107.013.015

Parker, D. (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.

Qin, C.-Z., Zhu, A.-X., Pei, T., Li, B.-L., Scholten, T., Behrens, T., & Zhou, C.-H. (2011). An approach to computing topographic wetness index based on maximum downslope gradient. *Precision agriculture*, *12* (1), 32–43.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. Retrieved from https://www.r-project.org/

Reardon, T., Barrett, C., Kelly, V., & Savadogo, K. (1999). Policy reforms and sustainable agricultural intensification in africa. *Development policy review*, *17* (4), 375–395.

Rebelo, L. M., McCartney, M. P., & Finlayson, C. M. (2010, oct). Wetlands of Sub-Saharan Africa: Distribution and contribution of agriculture to livelihoods. *Wetlands Ecology and Management*, *18* (5), 557–572. Retrieved from http://link.springer.com/10.1007/s11273-009-9142-x doi: 10.1007/s11273-009-9142-x

Salerno, J., Mwalyoyo, J., Caro, T., Fitzherbert, E., & Mulder, M. B. (2017). The consequences of internal migration in sub-saharan africa: a case study. *BioScience*, *67* (7), 664–671.

Smith, P., Gregory, P. J., Van Vuuren, D., Obersteiner, M., Havlík, P., Rounsevell, M., ... Bellarby, J. (2010). Competition for land. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365* (1554), 2941–2957.

Sörensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions*, *10* (1), 101–112.

Sulle, E. (2020). Bureaucrats, investors and smallholders: contesting land rights and agro-commercialisation in the southern agricultural growth corridor of tanzania. *Journal of Eastern African Studies*, *14* (2), 332–353.

Turner, B. L., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., ... others (2003). Illustrating the coupled human–environment system for vulnerability analysis: three case studies. *Proceedings of the National Academy of Sciences*, *100* (14), 8080–8085.

Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., … Asten, P. V. (2014). ScienceDirect Sustainable intensification and the African smallholder farmer. *Current Opinion in Environmental Sustainability*, *8*, 15–22. Retrieved from http://dx.doi.org/10.1016/j.cosust.2014.06.001 doi: 10.1016/j.cosust.2014.06.001

Villar, C. (2014). Wetlands and agriculture: partners for growth. *Manila Bulletin*, 12. Retrieved 2019-05-12, from https://www.ramsar.org/sites/default/files/wwd14{_}leaflet{_}en.pdf

Wood, A., & van Halsema, G. E. (2008). *Scoping agriculture – wetland interactions* (No. 33). doi: 10.1016/j.biocon.2015.04.016

Young, J., Graham, P., & Penny, R. (2009). Using bayesian networks to create synthetic data. *Journal of Official Statistics*, *25*(4), 549–567.