Center for Development Research (ZEF-Bonn)

The (Ir)relevance of the Crop Yield Gap Concept to Food Security in Developing Countries

With an Application of Multi Agent Modeling to Farming Systems in Uganda

Inaugural - Dissertation

zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der

Hohen Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

Vorgelegt am

5 Dezember 2005

von

Pepijn Schreinemachers

aus

Weert, Die Niederlande

Examiner:	Prof. Dr. Joachim von Braun (IFPRI)
Co-examiner:	Prof. Dr. Ernst Berg (University of Bonn)
Date of oral examination:	10 April 2006
Year of publication:	2006

D 98 (Deutsche Zentralbibliothek für Landbauwissenschaften)

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert

The (Ir)relevance of the Crop Yield Gap Concept to Food Security in Developing Countries

With an Application of Multi Agent Modeling

to Farming Systems in Uganda

Table of contents (brief)

Table of contents (detailed)

Abstract

This thesis scrutinizes the relationship between the width of the crop yield gap and farm household food security. Many researchers have argued that an exploitable gap between average crop yields and the genetic yield potential contributes to food security and that this potential should therefore be improved. Yet, crop yield gaps in developing countries are mostly wide, which is prima facie evidence that factors other than the yield potential are most constraining. A significant negative correlation between the width of the rice yield gap and food security for 19 Indian states confirms this.

The concept and pitfalls of the crop yield gaps are further analyzed at the farm household level for the case of improved maize in two village communities in southeast Uganda. Multi-agent systems are used to model the heterogeneity in socioeconomic and biophysical conditions. The model integrates three components: (1) whole farm mathematical programming models representing human decision-making; (2) spatial layers of different soil properties representing the physical landscape; and (3) a biophysical model simulating crop yields and soil property dynamics. The thesis contributes to methodology in four ways: First, it is shown that MAS can be parameterized empirically from farm survey data. Second, it develops a non-separable three-stage decision model of investment, production, and consumption to capture economic trade-offs in the allocation of scarce resources over time. Third, a three-step budgeting system, including an Almost Ideal Demand System, is used to simulate poverty dynamics. Fourth, coping strategies to food insecurity are included.

Simulation results show that neither the width of the yield gap nor the change in its width over time relate to food security at the farm household level. The maize yield gap is decomposed in both proximate and underlying factors. It is shown that the existence of maize yield gaps does not signal inefficiencies but poverty can be reduced substantially by addressing the underlying constraints such as access to innovations and credit. Improvements in labor productivity are crucial and are a much better indicator of development than crop yields and yield gaps. The results suggest that a strong focus on crop yields and yield gaps might not only be inefficient but even counterproductive to development.

Kurzfassung

Die vorliegende Dissertation untersucht die Beziehung zwischen der Grösse des Crop Yield Gap und der Ernährungssicherung von landwirtschaftlichen Betriebs-Haushalten. Ein grosser Teil der damit befassten Wissenschaftler vertritt die These, dass eine Ertragslücke zwischen dem durchschnittlichen erzielten und dem genetisch bedingten Ertragspotential besteht. Durch eine Verbesserung des genetischen Ertragspotentials könne deshalb ein wichtiger Beitrag zur Ernährungssicherung geleistet werden. Tatsächlich ist der Crop Yield Gap in Entwicklungsländern meist gross, was zu der Annahme führt, dass andere Faktoren als das genetische Ertragspotential von wesentlicher Bedeutung sind. Dies bestätigt eine Studie für 19 indische Provinzen, die eine signifikant negative Korrelation zwischen der Grosse des Crop Yield Gap und der Ernährungssicherung zeigt.

Das Crop Yield Gap Konzept wird in der vorliegenden Arbeit auf der Ebene ländlicher Haushalte für verbesserte Maissorten in zwei Dörfern im Südosten Ugandas empirisch analysiert. Zur Modelierung der Heterogenität der sozioökonomischen und biophysikalischen Ausgangsbedingungen wurde ein Multiagentenmodell verwendet. Das Modell integriert drei Komponenten: (1) Mathematische Programmierungsmethoden zur Modeliierung des Entscheidungsverhaltens auf der Ebene der Betriebs-Haushalte; (2) Ein räumliches Modell zur Erfassung und Analyse der Agrarökosysteme; (3) Ein biophysikalisches Modell zur Erfassung und Simulation von Ernteerträgen und Agrarökosystem-Dynamiken. Die vorliegende Arbeit leistet die folgenden methodischen Beiträge: Sie zeigt erstens, dass Multiagentenmodelle empirisch auf Grundlage von landwirtschaftlichen Haushaltsdaten parametrisiert werden können. Zweitens entwickelt sie ein nicht separables, dreistufiges ökonomisches Entscheidungsmodell für Investition, Produktion und Konsum zur Erfassung von Verteilungsprozessen knapper Ressourcen. Drittens verwendet die vorliegende Arbeit ein dreistufiges Budgetsystem zur Analyse von Armutsdynamiken, welches ein Almost Ideal Demand System integriert. Viertens werden unterschiedliche Strategien zur Ernährungssicherung integriert.

Die Ergebnisse der Simulationsexperimente zeigen, dass weder die Grösse des Crop Yield Gap noch die Änderung im Zeitablauf mit der Ernährungssicherung auf der Betriebs-Haushaltsebene verbunden sind. Die Mais-Ertragslücke lässt sich in mittelbare und grundlegende Faktoren aufteilen, wobei die Simulationsergebnisse

verdeutlichen, dass die Existenz der Mais-Ertrӓgslucke kein Zeichen für Ineffizienzen sind, und eine wirksame Armutsbekämpfung von grundlegenden Faktoren wie Zugang zu Innovationen oder Krediten entscheidend beeinflusst wird. Die Verbesserung der Arbeitsproduktivität ist von herausragender Bedeutung und ein wesentlich besserer Entwicklungsindikator als Ernteerträge oder Ernteertragslücken. Die Ergebnisse dieser Arbeit legen nahe, dass eine zu starke Fokussierung auf Ernteerträge oder Ertragslücken nicht nur ineffizient ist, sondern sogar kontraproduktiv für Entwicklung sein können.

Acknowledgements

This study would have been a lot more difficult, if not impossible, without the tremendous support of many colleagues and friends. Some of whom I would like to mention here. I acknowledge the support of my supervisor Prof. Joachim von Braun for awakening my interest in crop yield gaps and challenging me in my thinking. I also thank Prof. Ernst Berg for taking up the second supervision. I am most thankful to Thomas Berger for teaching me the ins and outs of the multi-agent system, and everything what happens with the data in between. I furthermore thank Jens Aune, Soojin Park, and Hosangh Rhew for their collaboration on the ecology and landscape side and acknowledge that this study would be a lot less interesting without their inputs. I thank Johannes Woelcke for his initial data collection in Uganda and modeling work, which gave me a fundament to build on.

My stay in Uganda was highly pleasant, which must be attributed to Ephraim Nkonya, Aggrey Bagiire, Albert and Diana Mudhugumbya, Musinguzi, Richard Oyare, and Sarah Sanyu. I also thank Kaizzi Cramer Kayuki, Almut Brunner, Justus Imanywoha, and James Sessanga for their invaluable support.

Good friends made my time at ZEF a pleasant one. Among these, I thank Denis Aviles, Quang Bao Le, Cristina Carambas, Arisbe Mendoza, Kavita Rai, Daniela Lohlein (also for the great editing), Puja Sawney, and Charlotte van der Schaaf. I also realize that without Günther Manske, Hanna Peters, and Rosemarie Zabel at ZEF and Gisela Holstein in Hohenheim, some things would never have been accomplished or at least not as smooth. Finally, I thank my dear family: mom and dad, Maurice and Dominique, and of course … Paan for their company and support in the past, present, and future.

Pepijn Schreinemachers

List of abbreviations

CIMMYT	International Maize and Wheat Improvement Center
FAO	Food and Agriculture Organization of the United Nations
IFPRI	International Food Policy Research Institute
IRRI	International Rice Research Institute
MAS	Multi-agent systems
MILP	Mixed integer linear programming
MP	Mathematical programming
MP-MAS	Mathemetical programming based multi-agent systems
SD	Standard deviation of the mean
TSPC	Tropical Soil Productivity Calculator
UNHS	Uganda National Household survey (conducted in 1999-2000)
USDA	United States Department of Agriculture
Ush	Ugandan shilling (1,000 Ush ≈ 0.63 Euro on 01.01.2001)
ZEF	Center for Development Research in Bonn

1 Introduction

1.1 Introduction

This thesis discusses the relevance of the concept of crop yield gaps with respect to food security in developing countries. It applies a novel methodology based on multi-agent systems (MAS) to decompose and simulate crop yield gaps while simultaneously measuring the economic well-being and food security of farm households in a developing country context. This first chapter introduces the crop yield concept and methods used to analyze it. The chapter is organized in six sections. Section 1.2 describes the problem background and introduces the concept of crop yield gaps; Section 1.3 defines the objectives of the study, while Section 1.4 introduces the methodological approach and Section 1.5 outlines how the remainder of the thesis is organized.

1.2 Problem background

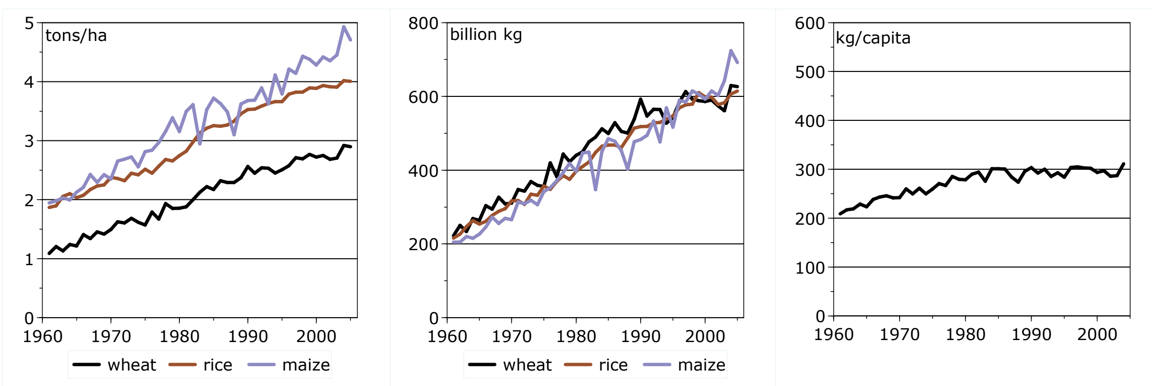
1.2.1 The crop yield gap and food security

A recent decline in the global growth rate of cereal production, production per capita, and cereal yield (see **Figure 1.1**) has intensified concerns about food sufficiency and food security. Cereal yields, many scientists have argued, need to be boosted to supply the growing human population with sufficient amounts of food (*e.g.*, Lampe 1995; Khush and Peng 1996; Pingali and Heisey 1999; Timsina and Connor 2001). An increase in yields is necessary because the possibilities to further expand the agricultural land area are being exhausted at a global level, and current land is rapidly being degraded and lost to expanding urban areas.

It is often written that growth in cereal yields is constrained by insufficient genetic gains in the yield potential and a subsequent narrowness of the yield gap (Peng *et al.* 1999; Reynolds *et al.* 1999; Timsina and Connor 2001). Technologies with a higher yield potential would therefore be required, especially in irrigated areas, to meet the increasing demand for food (*e.g.*, Reynolds *et al.* 1999).

The concern about yield gaps in relation to food security can be judged from the fact that much of the literature on the issue of crop yield potentials starts by summing up global population statistics (*e.g.*, Lampe 1995; Kush *et al.* 1996: 38; Reynolds *et al.* 1996: 1; Duvick 1999; Peng *et al.* 1999: 1552; Pingali and Rajaram 1999: 1; Rejesus *et al.* 1999: 1; Reynolds *et al.* 1999: 1611; Pingali and Pandey 2001: 1; Fischer *et al.* 2002: 1; Tiongco *et al.* 2002: 897). Several authors have called for more sustained efforts in 'beaking the yield barrier' (Cassman 1994; Reynolds *et al.* 1996). Raising the yield potential, in this respect, is implicitly assumed to increase actual cereal supply (*e.g.*, Peng *et al.* 1999; Reynolds *et al.* 1999). A reduction of the difference between yield potential and actual yield, often referred to as the narrowing of the yield gap, is interpreted as a worrying sign for long-term food security as farmers have less technological potential to exploit.

Figure 1.1: Global cereal yield trends and per capita availability, 1961-2005



Source: FAO 2006

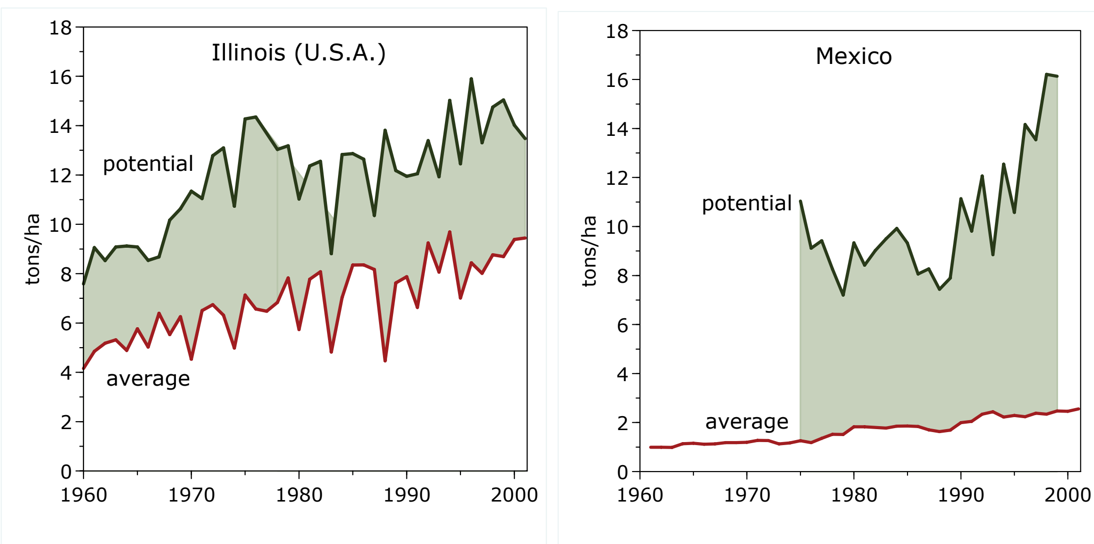
1.2.2 The crop yield potential

The yield gap is commonly defined as yield potential minus average yields. This yield potential refers to the genetic maximum yield of a crop. Evans (1996: 292) defines this yield potential as "the yield of a cultivar when grown in environments to which it is adapted, with nutrients and water non-limiting and with pests, diseases, weeds, lodging and other stresses effectively controlled".

Figure 1.2 shows yield gaps for maize grown in Illinois (left pane) and Mexico (right pane). The yield potential is quantified as the average of the three highest yielding experiments in a particular year. This figure shows that the average maize yield in

Illinois has closely followed the growth in yield potential at the state's experiment stations. Not only are the trends the same but also the variations around the trends resemble one another. Average yields in the beginning of the 1960s reached 4 tons but doubled to 8 tons by 2000 with the yield gap being—a more or less permanent—2 tons/ha. The picture for Mexico strongly contrasts that of Illinois. Mexican average yields also doubled in the same period but remain at a low average of about 2.5 tons/ha. The yield gap has, however, widened considerably since the early 1990s from about 6 tons/ha to more than 12 tons/ha.

Figure 1.2: Maize yield gaps for Illinois and Mexico



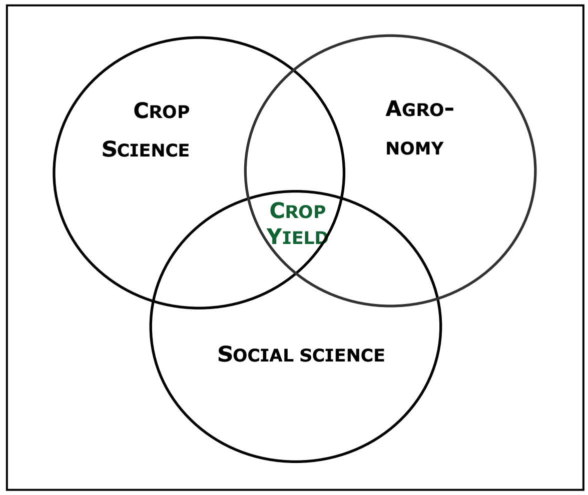
Sources/notes: The Illinois yield gap is based on the maximum over three trial locations: DeKalb, Urbana, and Brownstown, and the average state yield (which is slightly above the United States average maize yield) (Illinois Experiment Station 1960-2001, USDA 2002). Similarly, the Mexico yield gap is based on the three best yielding CIMMYT cultivars and the corresponding national average yield (CIMMYT 2002; FAO 2006).

The stark contrast between the two pictures is the only reason for showing them here. A multitude of factors determines the width of a yield gap. Farmers in Illinois rapidly adopt higher yielding varieties, yet the situation in Mexico seems to be much more complex. A weak linkage between yields at experiment stations and yields in farmers' fields can result from a lack of agricultural service provision, lack of knowledge among farmers, insufficient adaptation of crop varieties to farmers' conditions, missing or incomplete input markets including credit, high levels of risk impeding adoption, or a misalignment of researchers' and farmers' objectives. It is, however, not the intention to go into much detail at this stage. Yet, one hypothesis would be that crop yield gap dynamics for most developing countries come closer to the Mexican than to the Illinoisan picture.

1.2.3 Need for integrated approaches

The concept of crop yield is situated at the fault lines between three scientific disciplines: crop science, agronomy, and social science. Each of these disciplines has a strong interest in crop yields but from a different point of view. That is not to say that these scientific disciplines can be delineated neatly; they are more like a Venn diagram, as in **Figure 1.3**, with crop yield at its center.

Figure 1.3: Positioning the yield gap



The debate on yield gaps can largely be brought back to a difference in scientific perspectives on the factors determining crop yield. Biophysical sciences tend to focus on proximate factors—such as genes, soil nutrients, and energy—while social sciences tend to focus on underlying determinants—such as markets and institutions. The figures below illustrate these three contrasting perspectives.

First, **Figure 1.4** illustrates the determinants of crop yield from a crop science perspective (*i.e.*, crop physiology). Crop yield, in this view, is a function of total biomass and harvest index. Crop breeders generally concentrate on the absolute size of the yield difference between a new variety and farmers' varieties (Sanders and Lynam, 1982: 99). This yield difference can be widened either by an increase in total biomass—*i.e.*, increasing the size of all parts of the plant, or by an increase in harvest index—*i.e.*, increasing the proportion of grain in the total biomass. This perspective focuses on the level of the individual crop and the increase in crop yield is very much an objective in itself.

Figure 1.4: Crop yield as studied in crop physiology

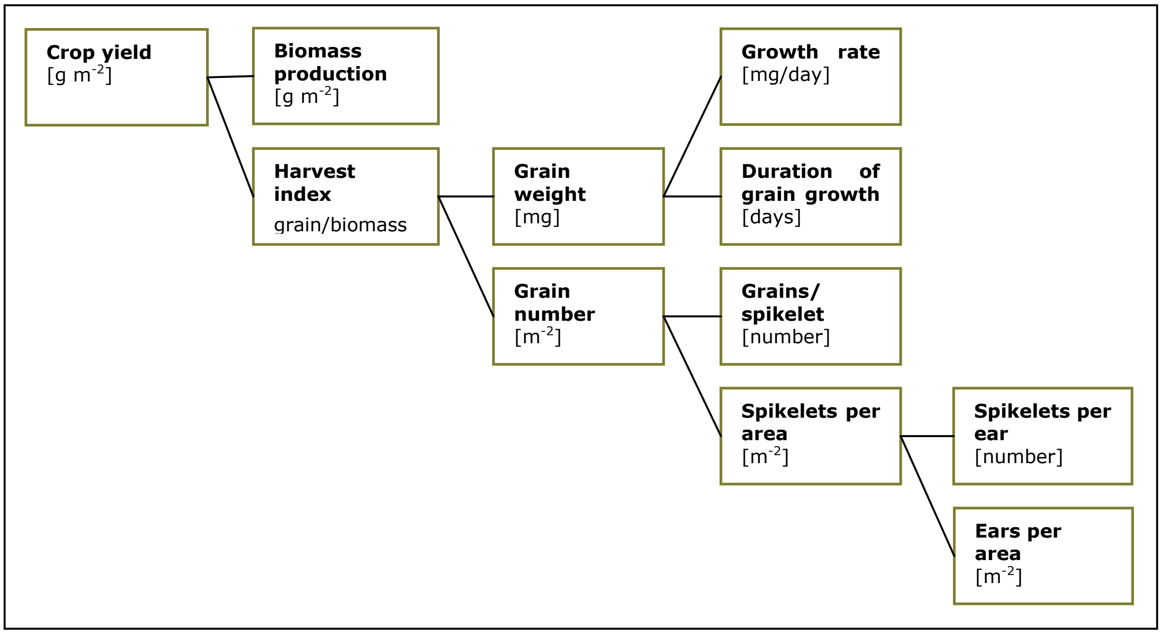
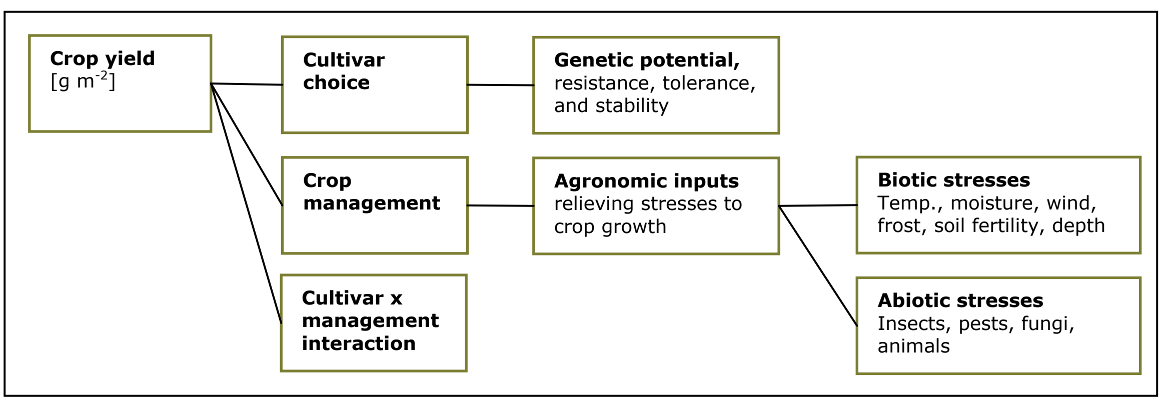


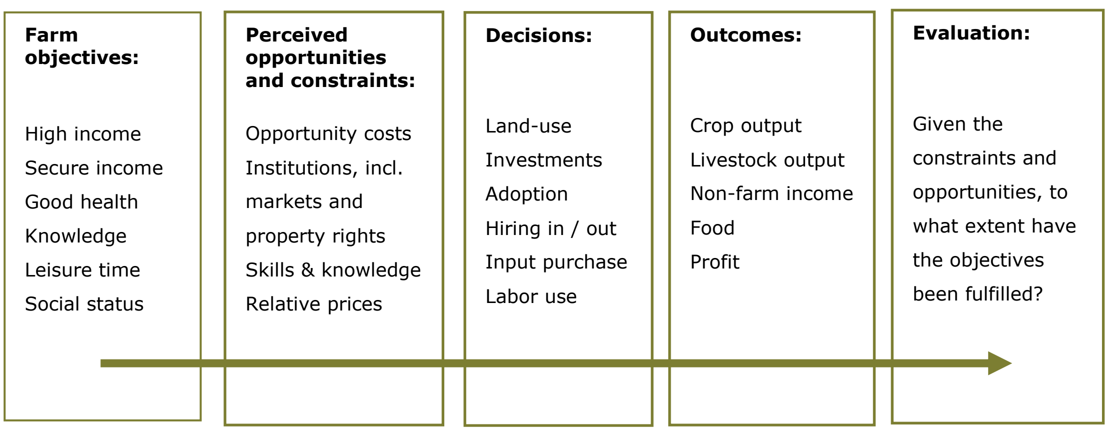
Figure 1.5 shows an agronomist's perspective. Agronomists focus on the field rather than the plant level. The yield of a crop can be increased by using higher yielding cultivars, improving crop management, or improving the interaction between these two (Evans and Fischer 1999). Similar to crop physiology, increasing crop yield and maximizing agronomic response is an objective in itself.

Figure 1.5: Crop yield as studied in agronomy



In the socioeconomic perspective, farm households—unlike crop scientists—do not usually attempt to obtain maximum crop yield. Farm households maximizing crop yields are destined to bankruptcy in any functioning market economy. Farm households have different objectives, such as meeting the food needs of the household, attaining a high level of income, having a stable income over time, increasing their knowledge, and having leisure time. **Figure 1.6** conceptualizes the socioeconomic perspective on the farm household. It shows that crop yield is one particular outcome of farm decision-making, rather than an objective in itself. In their decision-making, farm households are guided by their objectives and their perceptions of the environment, such as the availability and price of inputs, the sale of output, the security of their land tenure, the amount and distribution of rainfall, and the fertility of their soils. When evaluating their decisions, farm households will assess the extent to which their expectations with respect to objectives have been met and compare their performance with other farms.

Figure 1.6: Socioeconomic view on crop yield



Household objectives seldom overlap with attaining maximum crop yields, yet they may come close under certain conditions: (a) if land is the scarcest factor of production and land rents are therefore high; (b) if labor or mechanization is in ample supply; (c) if yield risks and price risks are low or covered by insurance; (d) if variable inputs such as fertilizers and agrochemicals are relatively cheap, supply is certain, and credit is available; and (e) if farmers are well-informed about the characteristics of improved varieties. These conditions apply more to agriculture in Illinois than to agriculture in most developing countries.

The three disciplinary perspectives on crop yield complement rather than substitute each other. Each perspective focuses on a different scale, from the plant, to the plot, to the farm level. Though the above contrast between disciplines is rather simple and incomplete, it helps to highlight two issues. First, caution is needed when linking determinants of crop yield at the plant level to factors at higher levels, such as the link between increasing the harvest index of wheat on the one hand and the food security of farm households on the other hand. Second, the understanding of crop yields and the relevance of crop yield gaps requires an integrated approach. Neither economic nor biophysical models alone can explain the level of and variation in average crop yields.

1.3 Objectives

The general objective of this thesis is to scrutinize the concept and pitfalls of crop yield gaps with respect to developing country agriculture. More specifically, the objectives are:

1. To review the linkages between a higher crop yield potential on the one hand and an increase in average yields and food security on the other hand.
2. To build a dynamic simulation model that integrates the biophysical and socioeconomic factors driving the width of the crop yield gap, and use this model for three purposes: (a) to quantify yield gaps and yield gap dynamics at the farm household level and to decompose them in proximate and underlying factors; (b) to assess the relationship between the width of the crop yield gap on the one hand and farm household well-being and food security on the other hand; and (c) to analyze how improved varieties with a higher yield potential affect incomes and food security at the farm household level.

1.4 Approach

After an in-depth discussion on the (ir)relevance of crop yield gaps for developing country agriculture based on a review of literature in **Chapter 2**, the concept is analyzed at the farm household level in the remaining chapters. For this, a multi-agent system (MAS) is calibrated to two villages in southeast Uganda. The MAS is used as a framework for integrating three main model components: an agent component representing farm household decision-making, a landscape component, and a biophysical component simulating crop yields and soil property dynamics.

1.4.1 Main methodological contributions

The thesis makes the following four contributions to the methods of farm household modeling and MAS:

First, this thesis shows that it is possible to empirically parameterize multi-agent systems from farm household survey data by using Monte-Carlo techniques to extrapolate from survey estimates.

Second, the thesis describes a novel approach to simulate farm household decision-making with mathematical programming by sequentially simulating investment, production, and consumption decisions while treating consumption and production as non-separable. This three-stage sequence of decisions is a realistic way of representing farm household decision-making and is well able to capture economic trade-offs in the allocation of scarce resources over time.

Third, the consumption side is modeled using a three-step budgeting process involving savings, food expenditures, and expenditures on specific categories of food. A linear approximation of the Almost Ideal Demand System (LA/AIDS) is included in the third step. The inclusion of a complete and flexible expenditure system in MAS opens new opportunities for applying MAS to the analysis of poverty, food security, and inequality.

Fourth, coping strategies to food insecurity are included. Agents can choose to spend their monetary savings or sell off livestock if food consumption falls short of their needs. The inclusion of coping strategies in MAS gives a realistic representation of the strategies of food insecure farm households in developing countries.

1.4.2 Main collaborations

Thomas Berger (University of Hohenheim) wrote the source code for the multi-agent model. Jens B. Aune (Norwegian University of Life Sciences) calibrated the Tropical Soil Productivity Calculator (TSPC) for soil conditions and 11 crops in Uganda. The TSPC was adjusted and integrated into the MAS by the author together with Thomas Berger. Hosahng Rhew and Soojin Park (both from the University of Seoul) estimated continuous soil maps from soil samples that were collected by the author and Gerd Ruecker (ZEF/ German Aerospace Center, DLR). Johannes Woelcke (The World Bank) developed the first version of the mathematical programming matrix that served as a basis for the matrix developed for this thesis. Thorsten Arnold (University of Hohenheim) wrote the MatLab routines that collected the MAS output and compiled it into single data files, which were used for statistical analysis.

1.5 Outline of the thesis

The thesis consists of 10 chapters. **Chapter 2** introduces the yield gap debate and highlights four important misconceptions commonly voiced in this debate. These misconceptions concern the assumed linkages between an improvement in yield potential and an increase in average yields, food availability, and food security. The chapter will point to the microeconomic factors affecting the yield gap. For analyze these, the focus turns to the farm household level in the following chapters. A novel methodology is developed based on multi-agent systems to integrate dynamic models of biophysical processes and farm household behavior at a very fine spatial resolution. **Chapter 3** describes the conceptual frame of the study. The general methodology is outlined in **Chapter 4**. Four subsequent chapters describe the calibration of the main model components. These are respectively, the landscape component in **Chapter 5** and the biophysical component in **Chapter 6**. The agent decision component is split into two with the production part outlined in **Chapter 7** and the consumption part outlined in **Chapter 8**. Results of the study are presented in **Chapter 9**. Finally, **Chapter 10** highlights the strengths and limitations of the applied methodology.

2 The (ir)relevance of crop yield gaps in developing countries

2.1 Introduction

Much has been written about the need to increase the crop yield potential of cereals for developing countries (*e.g.*, Cassman 1994; Khush 1995b; Reynolds *et al.* 1996). The commonly espoused argument is that the gap between average yields and the yield potential is too narrow to meet the increasing demand for food (*e.g.*, Khush 1995b; Peng *et al.* 1999; Reynolds *et al.* 1999) and that this has slowed down growth in average yields (*e.g.*, Cassman 1999; Timsina and Connor 2001). It is the objective of this chapter to show major shortcomings in this line of argumentation. In doing so, the irrelevance of the crop yield gap concept for developing countries is shown.

The chapter is structured as follows. Section 2.2 introduces the yield gap concept and gives an overview of the debate surrounding it. In Section 2.3, four major misconceptions about yield gaps commonly found in the literature are listed. Section 2.4 highlights two aspects of the debate that have received too little attention. The chapter ends with a summary in Section 2.5.

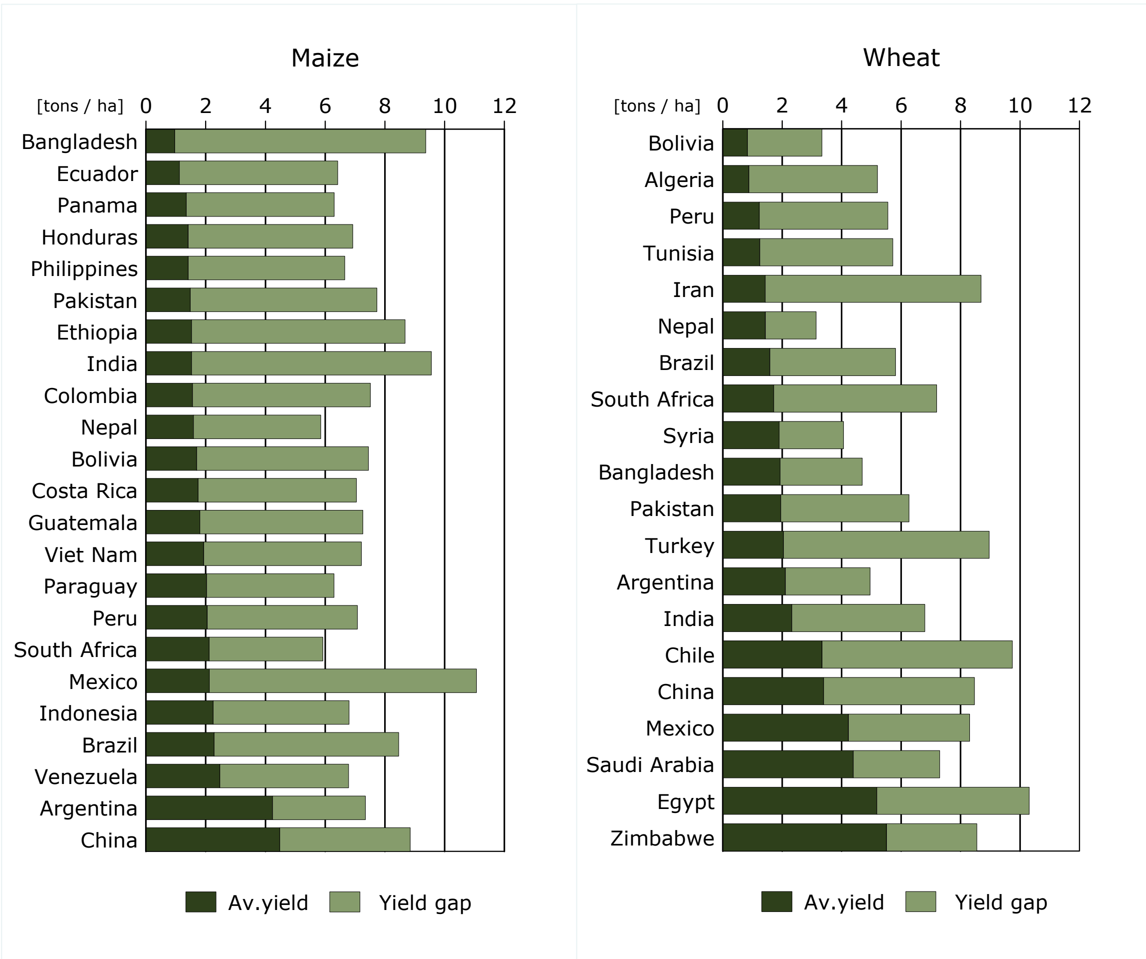
2.2 The crop yield gap

2.2.1 The yield gap concept

The yield gap is defined as the yield potential minus the average yield level, with the first being the genetic potential of a cultivar, achieved at experiment trials where temperature and radiation are the only factors uncontrolled (de Wit 1958, as in Rockström and Falkenland 2000: 321).

Yield gaps are wide for most developing countries. **Figure 2.1** quantifies the yield gap for a few developing countries for maize (left pane) and wheat (right pane). The wheat data come from CIMMYT's International and Elite Spring Wheat Trials (ISWYN and ESWYN), while the maize data come from CIMMYT's International Maize Testing Unit (IMTU). As in **Figure 1.2** in the previous chapter, the yield potential is quantified as the average of the three highest yielding experiments in a particular year, and because yield gaps tend to vary considerably between years, a 15-year average is used for **Figure 2.1**. The picture shows a wide yield gap for nearly all countries, with the exception of some wheat growing countries where the use of irrigation is widespread.

Figure 2.1: Maize and wheat yield gaps for a selection of developing countries



Sources: FAO 2006 (national averages); Payne 2002 (wheat); CIMMYT 2002 (maize); Lantican *et al.* 2003 (wheat)

Notes: Yield potential is the average over top three yields in a year. Average 1985-2000. Only countries with data for at least 10 years and where the contribution of the crop to total cereal production is more than 5 percent were included.

2.2.2 Background

The interest in yield gaps can be traced back to research at the International Rice Research Institute (IRRI), or possibly the Indian rice research institute CUTTACT, which was a predecessor of IRRI (van Tran, personal communication 2001). The research at IRRI focused on Philippine rice farmers whose average yields of less than 1 ton stood in stark contrast to the 5 to 10 tons researchers were achieving. Gomez (1977) and De Datta (1981) are the earliest references found in the literature who try to explain this difference by decomposing the yield gap. Gomez (1977) developed a series of on-farm experiments to decompose the yield gap into factors such as variety choice, fertilizer use, and pest control (see Sall *et al.* 1998 for a recent application).

Herdt and Mandac (1981) extended this methodology to include farm survey data. They decomposed the yield gap into three parts: 1) profit-seeking behavior, as farmers do not maximize yields but profits; 2) allocative inefficiency, as farmers misallocate production factors; and 3) technical inefficiency, as farmers do not use the production factors correctly up to their optimum. Herdt and Mandac (1981: 379) wrote: "the yield gap due to profit-seeking behavior should never be eliminated because it indicates a socially efficient economy. But if artificially distorted prices can be shown to significantly increase the portion of the yield gap due to profit-seeking behavior, it may be possible to recommend that the government remove such distortions." In their application to Philippine rice farmers, Herdt and Mandac found a yield gap of about 1 ton/ha, the majority of which they attributed to technical inefficiency.

2.2.3 The yield gap debate

The main topic in the yield gap debate is whether a higher yield potential—that is a widening of the yield gap—is needed to increase average cereal yields. Two camps can broadly be identified in the debate; hereafter called 'yield gap pessimists' and 'yield gap optimists'.

The yield gap pessimists argue that yield gaps are narrow and the yield potential needs to increase for average yields to continue growing (*e.g.*, Cassman 1994; Reynolds *et al.* 1996). These pessimists find support from an empirically observed slowdown in the growth rate of cereal yields at global and regional levels and a simultaneous reduction of investments in agricultural R&D (*e.g.*, Lampe 1995; Khush and Peng 1996; Pingali *et al.* 1999: 1; Pingali and Heisey 1999; Reynolds *et al.*

1999: 1611; Heisey *et al.* 2002: 47). Increasing the yield potential, in this respect, is implicitly assumed to increase cereal supply (*e.g.*, Peng *et al.* 1999; Reynolds *et al.* 1999). The pessimists find further evidence in the fact that the yield potential of some rice cultivars (notably IR8, released in 1967 by IRRI in the Philippines) in some long-term experiments in India and the Philippines has declined (Flinn *et al.* 1982; Flinn and De Datta 1984; Cassman *et al.* 1995; Dawe *et al.* 2000).

The yield gap optimists, on the other hand, argue that yield gaps are still sizable, and average yields can continue to grow if available technologies are more fully exploited. Most developing countries have very low average cereal yields, and the yield gaps in rice too, are sizeable, ranging from 10 percent to 60 percent, across ecologies and crop seasons in all rice-growing countries in the Asia-Pacific region (FAO 1999). These optimists argue that not a low yield potential but poor crop management and problems of institutional support account for the large variability in average rice yields of irrigated rice between countries (Duwayri *et al.* 2001).

Furthermore, these optimists point out that the growth in yield potential has not slowed down but has risen linearly over time (Evans and Fischer 1999). Although the yield of IR8 might have stagnated, other higher yielding rice varieties are available or are in the pipeline. China, for instance, released hybrid rice varieties in 1976 and this has shown a yield premium of about 20 percent compared to other improved varieties (Lin 1994; Yuan 1998). Evidence also abounds that growth in the yield potential of wheat and maize has neither slowed down (Duvick 1992; Canevara *et al.* 1994; Eyhérabide *et al.* 1994; Austin 1999; Evans and Fischer 1999).

Further support for the optimists' claim comes from the many success stories of countries rapidly increasing their average yields in spite of slow advances in yield potential. The experience of 'Ricecheck' in Australia is worth mentioning in this respect. Ricecheck is a collaborative learning system of farmers, researchers and extension services and was introduced in Australia in 1986 (Lacy *et al.* 2000; Clampett 2001). The Ricecheck approach tries to find the answer for high yields in high yielding fields of farmers rather than from research plots. Seven key recommendations (*e.g.*, field layout, plant density, and timing of input use – to name a few) linked to high yields are identified in farmers' fields through a continuous cycle of monitoring, recording, data exchange, and feedback. The aim is to educate farmers to improve their learning and performance. Farmers are encouraged to monitor their crop, compare it with the key check recommendations, and to record their findings. Extension agents give farmers individual feedback, based on statistical analysis of these records; this feedback shows how their performance compares with the key checks as well as with other farmers in the same district. The Ricecheck approach resulted in a significant increase in farmers' yields over the last 15 years, although the yield potential did not increase during this period (*ibid.*). The Ricecheck approach is now promoted by FAO and IRRI for other countries as well.

2.3 Misconceptions about crop yield gaps

It is not the purpose of this thesis to take sides in the yield gap debate. Most commonly, pessimists and optimists find some agreement in that both a higher yield potential and improved management are needed, with a higher yield potential to be emphasized in the irrigated areas and improved management to be emphasized in the rainfed areas.

Instead, the purpose of this paper is to show that some of the arguments, of both pessimists and optimists, are unjustified on socioeconomic grounds. To do this, the chapter focuses on four common and pervasive misconceptions surrounding the debate. The first misconception is that farmers want a higher yield potential. The second is that a higher yield potential is needed to meet the future demands for food. The third is that a higher yield potential will improve food security. The fourth is that a higher yield potential is needed to keep cereal prices low.

2.3.1 'Farmers want a higher yield potential'

There is a persistent, but wrong, belief among researchers that farmers in developing countries, like farmers in many developed countries, adopt technologies only when these increase yields. Furthermore, Evans and Fischer (1999: 1547) noted that it is often erroneously assumed that progress in yield potential automatically translates into progress in farmers' yields. The fixation on higher yields is likely to be a Western bias, as land is usually the scarcest factor of production in high-income countries, with all other factors such as labor, credit, fertilizers, pesticides and herbicides as well as crop insurance in ample supply. This is, however, not usually the case in most developing countries. Labor, capital, fertilizers, pesticides and herbicides, and increasingly water are often equally scarce factors as land, and not the size of the plots but the security of property rights over them constrain production.

Yield enhancing technologies compete for scarce resources at the farm level. This relative scarcity means that production resources have an implicit price attached, called opportunity cost. The opportunity cost is the value of the best alternative choice that is foregone as a result of a decision (Coleman and Young 1989: 17). The level of opportunity costs plays a crucial role in the adoption of technologies. Farm households will prefer a technology that substitutes the production resource with the highest opportunity cost, as freeing one unit of this resource will give the highest additional return.

Because resources are scarce, an improved variety can increase the yield of one crop but simultaneously lower the yield of other crops. The true return of an improved variety can therefore not be inferred from the yield premium it gives. For instance, von Braun (1988), in his study on rice technology adoption in The Gambia, calculated that for any additional ton of rice produced, 390 kilograms of other cereals and 400 grams of groundnut are foregone. Expressed in monetary terms, this means that for each additional dollar earned in rice, 71 cents are foregone in the cultivation of alternative crops. This shows that the markup in farm earnings is much lower than the markup in yield of the crop that is improved.

Higher yields can even lead to lower farm earnings. For example, Sanders and Lynam (1982) described the introduction of an improved cassava variety together with improved management, which gave a yield premium of 108 percent over farmers' varieties and farmers' management in Colombia. Yet, the lower starch contents of the new variety resulted in 40 to 60 percent lower prices, making the high yielding variety less profitable than the traditional one.

Equally, lower crop yields are compatible with greater farm earnings. Byerlee and Siddiq (1994: 1354) for instance observed that farmers in Pakistan's Punjab postpone the date of planting wheat in order to extend the cultivation of cash crops, although this leads to lower wheat yields. One increasingly frequent observation is that the rising opportunity cost of labor constrains technology adoption, and that the opportunity cost of labor exceeds that of land, which increases the importance of labor saving technologies relative to that of land saving (yield enhancing) technologies. Moser and Barrett (2003), for instance, showed that a rice yield increasing technology is not widely adopted by farmers in Madagascar because it has high seasonal labor demands. Another example is green manure; this technology has a long-term positive effect on cereal yields, especially on poor soils. Nevertheless, green manure has not been widely adopted and has even largely disappeared from the rice systems of Asia (Ali 1999). The reason is that green manure is land and labor intensive and the relative prices of these factors have increased over the last decades, making mineral fertilizers more cost-effective (*ibid.*).

Risk is another important consideration, especially for agriculture where the time between input decisions and outcomes is long and the outcome much depends on the vagaries in weather, pests, diseases, and market prices. If the variability in returns is high, because of a fluctuating climate or variable market prices, then farm households rationally lower their expectations below the average returns to shield themselves from disaster. For instance, de Rouw (2004) in her study on pearl millet in the African Sahel of Niger, showed that farmers did not adopt a high yielding technology package (consisting of short-cycle varieties, a high planting density, and mineral fertilizer) as this technology did not reduce yield variability caused by unreliable rainfall. She found that farmers' priority is risk reduction, *i.e.* obtaining at least a minimum yield in the worst year, rather than obtaining a high yield in the average year (*ibid.*).

In other cases, farmers do not adopt because they have strong cultural preferences concerning the quality of a crop, especially when it is indigenous to a country (*e.g.*, Adesina and Baidu-Forson 1995; Bellon and Risopoulos 2001). Bellon and Risopoulos (2001) showed how Mexican farmers only partially adopt high yielding maize varieties and actively mix these with traditional varieties to combine desirable properties while compromising on average maize yields—a process they called 'creolization' of maize.

What these seven examples show is that a higher yield is neither a sufficient nor a necessary condition for farmers to increase their productivity and to improve their well-being. Farming systems in developing countries are diverse as well as complex. Higher yielding varieties need to be tailored to farmers' objectives, preferences and constraints.

2.3.2 'A higher yield potential is needed to meet future demands'

Much of the literature on crop breeding in developing countries compares human population growth with progress in crop breeding (*e.g.*, Reynolds *et al.* 1999; Slafer *et al.* 1999; Peng *et al.* 2000). Many studies begin with a summary of human population growth in the first paragraph of the treatise (*e.g.*, Poehlman and Quick 1980: 1; Khush 1995b: 329; Kush *et al.* 1996: 38; Duvick 1999; Peng *et al.* 1999: 1552; Reynolds *et al.* 1999: 1611; Slafer *et al.* 1999:379; Fischer *et al.* 2002: 1).

Some of the arguments most commonly advanced are put together in **Box 2.1**. Most of these assume a direct linkage between crop breeding and the feeding of a country's population.

Box 2.1: Selection of quotes relating yield potential with food demand and supply

"Future genetic gains in grain yield must be attained at the same pace as before, or even accelerated, to meet the increased demand for food from an increasing population, estimated to be 6 billion by 2010" (Slafer *et al.* 1999: 379)

"Global demand for wheat (*Triticum aestivum L.*) is growing faster than gains in genetic yield potential are being realized, currently under 1% per year in most regions." … "This means that current trends in the improvement of genetic yield potential are too low to keep pace with future demand." (Reynolds *et al.* 1999: 1611, 1617).

"Given these constraints on the availability of arable land, crop yield potential will be a primary factor governing the nature of agricultural systems in the next century" (Duvick and Cassman 1999: 1622).

"Rice-wheat is the most common cropping system in the region [Indo-Gangetic plains]. Understanding and increasing its yield potential is essential to meet the growing food demand" (Aggarwal *et al.* 2000).

"Scientists at IRRI […] feel a responsibility to be prepared to develop and provide the technologies and the knowledge that will allow the world's rice farmers to produce enough rice to meet the population growth that is realistically expected in the next century" (Lampe 1995: 256)

"Over the next 30 years, Asia must increase its rice production by at least 60% to meet the needs of population growth […]. To achieve this goal, our best option is to develop rice cultivars with higher yield potential through crop improvement." (Peng *et al.* 2000: 307)

Farm households and governments constitute the missing links in these arguments. Farm households are the first missing link. It is in the interest of society as a whole that cereals are available at reasonable prices; yet, the interest of the farmers is to feed their own household and to improve its well-being. Trade-offs exist between the 'private' goal of farmers' well-being and the 'social' goal of food availability. For instance, a high level of cereal production does not necessarily correlate with a high level of income at the farm level. De Datta (1981) noted that in villages in Asia, incomes are greater where much land is planted to crops other than rice, than in villages where only rice is grown (IRRI 1978, as in De Datta 1981: 553). That private and social objectives should not be confused, was eloquently stated by Adam Smith in 1776: "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages" (Smith [1776] 1937: 14).

Governments are the second missing link. If private objectives of farm households do not coincide with the society's objective of producing more cereals, then governments may want to intervene to give farm households more incentive to produce cereals. Keeping people on the farm at times of a relative decline of agricultural sector's contribution to the economy might be another social objective. It is a general feature that in the process of economic development, the opportunity cost of farm labor rises as agricultural labor productivity declines in relative terms to the other sectors in the economy (Martin and Warr 1994; Pingali and Rosegrant 1995). As a result, people move out of agriculture to seek better-paid jobs. Papademetriou, for example, noted that the number of rice farmers in Asia decreases proportionally to the rate of industrialization, while the age of the remaining farmers also increases proportionally to it (Papademetriou 2000, 2001). Many studies have shown a decline in the profitability of agriculture in general and rice production in particular. Tiongco and Dawe (2002) and Pingali and Heisey (1999) showed it for Philippine rice farmers by estimating the change in total factor productivity from panel data. Estudillo *et al.* (1999) showed it for the same country by calculating domestic resource costs for assessing the comparative advantage. In the light of this declining profitability, Pingali and Rosegrant (1998: 956) pointed out that "for wheat farming to remain profitable, technological change must ensure that production costs per ton of wheat fall at the same rate as the real price of a ton of wheat". The experience of most high-income countries is that technological change alone is not enough, and governments step in to support farm production by upholding its relative profitability. Government support to agriculture can include price and income support, but can also be in the form of investments in R&D. To argue for a higher yield potential for the irrigated rice areas of Asia means subscribing to one aspect of the latter. However, it is important to recognize that alternative combinations of policy intervention are available.

2.3.3 'A higher yield potential increases food security'

Related to the argument that a higher yield potential increases food availability, is the argument that it enhances food security (*e.g.*, Lampe 1995; Cooper 1999). Lampe, for instance, claimed that "[crop] breeding remains one of the most powerful tools to eliminate hunger" (Lampe 1995: 258). This argument assumes that crop varieties with a higher yield potential ultimately end up in the mouths of the hungry. This is a misconception for mainly four reasons.

First, the linkages between technological change and food security are complex and cannot be assumed as linear. Among other things, these linkages depend on the diffusion process of a technology, price and income effects including multiplier effects, and the functioning of markets. The remaining chapters of this thesis will scrutinize these linkages using agent-based modeling on a case study of Ugandan farm households. This chapter merely stresses that non-market and non-technology sources of food insecurity, such as violent conflict, bad governance, discrimination, and natural disasters are important, if not the main, factors behind food insecurity (von Braun *et al.* 1998; Paarlberg 2000).

Second, the definition of yield potential, as the unconstrained yield, contrasts to the fact that food security is concentrated in those areas with a relative abundance of constraints – *i.e.*, the less-favored areas. A wide array of biophysical as well as socioeconomic factors constrain agricultural growth in the less-favored rural areas of developing countries: rainfall is uncertain, soil fertility is poor, slopes are steep, irrigation is lacking, the physical infrastructure is poor, transaction costs are high, and markets are either imperfect or completely missing (Wade *et al.* 1999; Kuyvenhoven 2004; Ruben and Pender 2004). Yet, the yield potential is the unconstrained yield and hence assumes the absence of all of these constraints as well as yield maximizing labor use by farm households. Under such conditions, the yield potential becomes irrelevant, as it does not represent realistic opportunities for farm households to exploit.

Yet, the number of poor people living in these less-favored areas is vastly larger than the number of poor people in the favored areas (von Braun 2003). Kuyvenhoven *et al.* (2004) estimated that roughly 40 percent of people in developing countries live in the less-favored areas. According to Mackill *et al.* (1996), only 8 percent of the major rice areas in South and Southeast Asia are favorable (Mackill *et al.* 1996, as in Wade *et al.* 1999: 5). Hence, for an increase in yield potential to be beneficial for those who are food insecure, current constraints need to be addressed simultaneously, if not primarily. Signs of an increase in yield potential for the less-favored areas (*e.g.*, Lantican *et al.* 2003) need to be treated with cautious optimism, as an increase in yield gap is an irrelevant indicator for an increase in real opportunities for these areas.

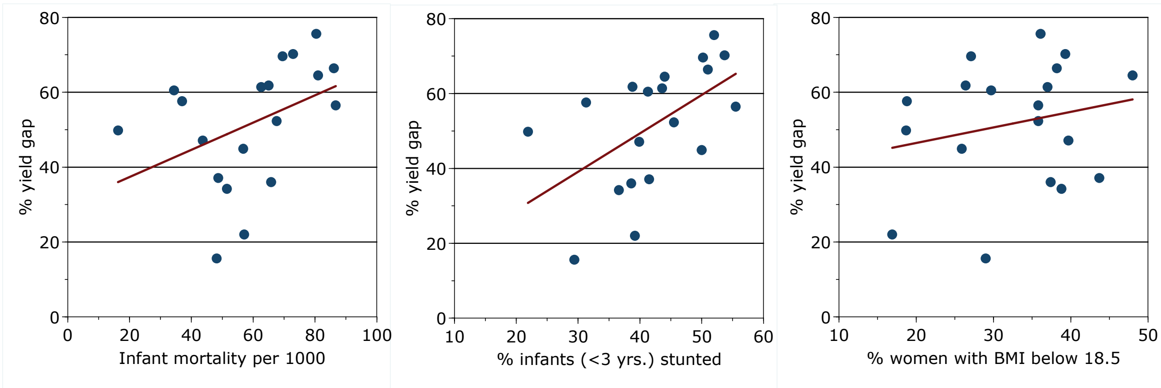
Third, technological change is often location-specific, which makes linkages between locations important in assessing the impact of technological change on food security. Consumers and early adopting farm households generally benefit from improved technologies, but non-adopting farm households experience real losses from deteriorating terms of trade if farmgate prices fall (Cochrane 1958; Renkow 1994; Hazell and Haddad 2001; Evenson and Gollin 2003). The problem is that the non-adopters tend to be concentrated in the less-favored areas and are more likely to be food insecure than the adopters. The introduction of a variety with a higher yield potential might therefore worsen the food security situation of poor farm households whose adoption lags behind.

Fourth, a higher yield potential might be the least binding of all constraints for those areas where food insecurity is most severe. To support this claim, attention is turned to India as much of the debate on decreasing yield gaps has focused on this country. Data are available on the yield gap in rice for 19 Indian states, including both favored and less-favored areas. These data come from Siddiq *et al.* (2001). Yield potentials are seven-year averages of the highest yielding entries at test locations of the All-India Coordinated Rice Improvement Program (AICRIP) in each state. Average rice yields are state-level averages calculated over the same period as the yield potential. The state-level yield gap is quantified as the difference between yield potential and average yield, expressed as a percentage of the average.

Three outcome indicators for nutrition and health are used as indicators of food security: the infant mortality rate, the percentage of stunted children under the age of three, and the percentage of women with a body mass index (BMI) below 18.5 kg/m^2. State-level data come from the National Family and Health Survey 1998-99 (IndiaStat 2004). These indicators are plotted against the state-level yield gap in **Figure 2.2**.

The figure shows a clear and positive correlation between the width of the yield gap and the three indicators of food insecurity. The correlation is strongest for infant mortality and stunted children. These figures show that the gap between the yield potential and average yields is especially large in those Indian states with a high level of food insecurity. This might not point to deficiencies in technology but, more likely, to a failure of agricultural services and health services in these states. These and other non-technology related factors constrain average yields and impede the food security and well-being of people.

Figure 2.2: Correlation between the rice yield gap and three food security (outcome) indicators for 19 Indian states, 1998-1999



Sources: Siddiq *et al.* 2001 (yield gap); IndiaStat 2004 (food security indicators)

These results suggest that food insecurity is highest in those Indian states where the rice yield potential is least constraining, and that a higher yield potential is no guarantee for food security. The results also suggest that there is a large technological potential for states to exploit. Fan and Hazell (1999), for instance, showed that returns to public investments in India, like infrastructure, are currently greater for the less-favored than for the favored areas.

2.3.4 'A higher yield potential is needed to keep prices low'

Many authors have claimed that varieties with a high yield potential are needed to keep food prices down and that this enhances food security. For instance, the IRRI 2000 Annual Report states that "The price of rice must be kept down so that the position of the worlds' poor and hungry does not deteriorate" (IRRI 2000: 23). Yet, the idea that, as a rule, the food security situation improves with lower cereal prices is a misconception. The linkage is complex, location and time specific, and can go either way (Pinstrup-Andersen 1988; Renkow 1994).

Lower real cereal prices increase the purchasing power of consumers who can consume more for the same amount of money. Lower cereal prices especially augment the income of the poor as they spend a large share of their budget on cereals. However, lower cereal prices at the farm gate decrease the revenues of the producers, weakening the incentive to produce cereals. Too low cereal prices can impede on the food availability and worsen the food insecurity if trade is restricted. Too high cereal prices, on the other hand, can also worsen food insecurity by limiting the access to food. A positive linkage between low cereal prices and food security can therefore not be assumed but requires careful analysis taking into account both spatial and temporal dimensions. Six main factors are to be taken into account:

(a) whether the food insecure are net buyers or net sellers of cereals;

(b) price and cross-price elasticities of expenditures on cereals by the food insecure;

(c) multiplier effects, such as a possible reduction in the wages for the rural landless resulting from lower cereal prices (*e.g.*, Barnum and Squire 1979);

(d) the opportunity cost of cereal production and its marginal rate of substitution;

(e) the terms of trade in agriculture, such as the price of inputs relative to the price of output; and

(f) the openness of the economy and functioning of markets.

The effect of low cereal prices on food security is therefore ambiguous. The objective of the current chapter is merely to highlight the complexity, while the second part of the thesis will analyze this complexity using simulation modeling on a case study in Uganda.

2.4 More than genes

A higher yield potential alone will not provide the virtues augured by crop breeders. The creation of an enabling environment is a precondition for improved seeds to have their desired impact. Yet, this impact should not be judged in terms of crop yields only, and especially not in terms of average yields. The issues of enabling environments and average yields are discussed in the following.

2.4.1 The importance of creating an enabling environment

Improved varieties with a higher yield potential have undoubtedly contributed to growth in average yields. Duvick (1992) estimated that the genetic improvement in maize has accounted for 56 percent of annual average yield growth in Iowa maize yields between 1930 and 1989. Similarly, Byerlee *et al.* (2000, as in Wiebe 2003: 10) estimated that about half of the increase in global crop yields in recent decades has come from genetic improvement, the remainder coming from improved management. Yet, does this mean that to raise crop yields, we merely need to improve the genetic potential? Not really. The example of hybrid maize below is used to underpin this claim.

Hybrid maize was developed and first introduced in the United States in the late 1930s. These hybrids spread at an enormous pace, which resulted in 95 percent of the United States Corn Belt covered by it within 15 years of its release (Griliches 1957; Duvick 1996). The success of this hybrid is still one of the reasons for a very strong advocacy for hybrid maize in developing countries.

Though hybridization lay at the core of this success story, explaining the success of hybrid maize only from the advance in hybridization does not do historical justice to the socioeconomic context in which it was achieved. One might apply similar arguments to the Green Revolution in Mexico and South and Southeast Asia (*e.g.*, Herdt 1987; Hazell and Ramasamy 1991). Though nothing new, it is important to keep in mind the non-gene related factors that spurred the success of, in our example, hybrid maize in the United States. These factors can be divided into basic structures and incentives. Basic structures include the physical and institutional capital, which affect the functioning of the economic system, and which facilitate, but not immediately influence, farm household decision-making. Incentives, on the other hand, directly influence economic decision-making. Three such basic structures and five incentives that played a crucial role in the success of hybrid maize in the United States are listed below.

(1) Supportive basic structures:

- A wide network of experimental stations at so-called land-grant universities facilitated the local improvement and adaptation of hybrid maize (Goldman 1998).
- Early institutions of intellectual property rights protection safeguarded large private investments in R&D of hybrid maize. High profits from sterile seeds and relatively inexpensive seed production shifted maize breeding from the public to the private sector.
- Substantial federal investments in institutional and physical infrastructure in the 1930s facilitated the development of an input service and food processing industry (Cochrane 1979).

(2) The right incentives:

- Hybrid maize had a very large yield premium of 15 percent over the best open-pollinated varieties (Duvick 1997: 6). Drought-resistant and disease-resistant plant varieties were developed at the same time.

- The United States Department of Agriculture actively supported the adoption of hybrid maize (Goldman 1998).
- Farm-gate prices from the late 1930s to the late 1940s were high because of wartime demands and government payments to farmers, which raised farm incomes substantially and allowed farmers to purchase technologies (Schultz 1945; Cochrane 1979).
- There was strong demand for land and labor saving technologies because of rising opportunity costs of labor, strong rural-to-urban migration, and an increase in farm sizes (Schultz 1945; Cochrane 1979).
- The introduction of hybrid maize largely coincided with the introduction of three other technologies in United States agriculture to which the crop was highly responsive: (1) mineral fertilizer; (2) agrochemicals to control pest, insects, and weeds; and (3) farm mechanization (Cochrane 1979).

2.4.2 The limited relevance of national average yields

In the current debate on food production, the yields of cereal varieties get too much attention, while the conditions that enable farmers to produce such yields get too little attention. Even more so, the discussion about cereal yields focuses almost exclusively on average national yields. Yet, national average cereal yields are not good indicators of agricultural development for at least six reasons:

a. One output focus

Cereals are only one crop in a whole range of crop, livestock, and non-farm enterprises performed by farm households. High productivity of non-cereal enterprises might easily compensate low productivity in cereal production.

b. Economically void

Low cereal yields do not signal inefficiencies but can well be an economically optimal allocation of scarce inputs. A large amount of empirical literature shows that farm size negatively relates to land productivity but does not show that the lower land productivity of larger farms is associated with inefficiencies (Cornia 1985; Singh 1988; Heltberg 1998; Dorward 1999; Lerman and Schreinemachers 2005).

c. Partial indicator

The importance of the yield indicator hinges on the assumption that land is the scarcest production factor for farm households so that land saving technological change is a good indicator of overall productivity growth. Often this is not the case.

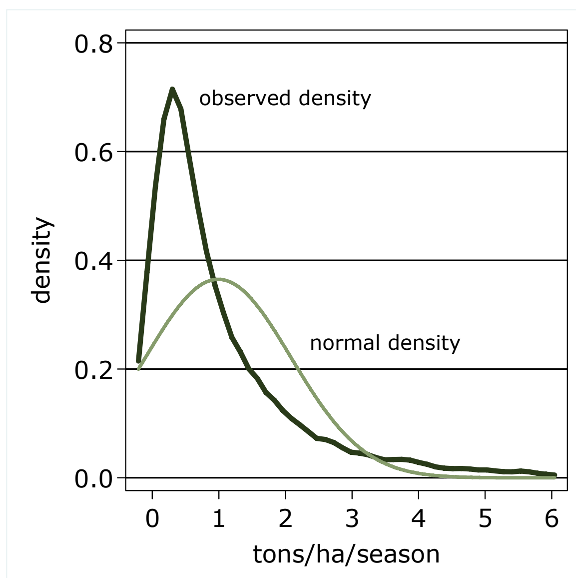
In many parts of the world, the relative scarcity of fertilizers, pesticides and herbicides, and credit is more constraining than the availability of land. In addition, labor and water are increasingly constraining agricultural productivity in developing countries. If this is the case, labor productivity, fertilizer productivity or water productivity offer more guidance to farm households and policy makers in their allocation of scarce resources.

d. The arithmetic mean

Cereal yield is an average at all levels of aggregation (Just and Weninger 1999). At the plot level, yield is an average over a plot. At the farm level, yield is an average over a farm's area under one crop. At the national level, yield is an average over a country's cultivated area, and so forth. An average is meaningful if the underlying distribution is statistically normal and not too skewed in either direction. Yet, farm household surveys in developing countries often find farm-level crop with a strong negative skew; that is, many farm households attain very low yields while only few attain very high yields. If this is the case then the arithmetic mean is not a good indicator.

Figure 2.3 illustrates this with a kernel density estimate for maize yields in Uganda, as estimated from household survey data. The arithmetic mean maize yield is 0.96 tons/ha/season, which is the yield at the maximum density of the overlaid normal distribution. Yet, the actual distribution of maize yields is far from normal and the arithmetic mean hence badly represents what most farm households obtain from their plots. Under such circumstances, median values or trimmed arithmetic means usually give a more realistic idea of crop yields in a country.

Figure 2.3: Kernel density distribution of maize yields in Uganda



Source: Estimated from 2000-01 IFPRI survey
Notes: Epanechnikov kernel used. Zero yields included.

e. Bias toward monoculture

Where intercropping is common, the concept of crop yield, defined as production of a single crop per hectare per season, is unsuitable. If two crops are grown in combination on the same hectare, then the crop yield indicator grossly underestimates the true land productivity. Individual crop yields are usually lower because of increased competition between the crops, but total land productivity is often greater than in monoculture (see **Chapter 7**). The possible advantages of monoculture mainly stem from mechanization, which is still not common in many developing countries, especially in sub-Saharan Africa. One can therefore discern a Western bias in the importance attached to average yields, because intercropping is common in rainfed agriculture in the tropics, whereas monoculture dominates agriculture in the West (Hildebrand 1976).

f. Conditionality

Trends in average national yields might not reflect trends in farmers' fields, if the underlying resource conditions change. This implies that for growth in the average national cereal yield to be a meaningful indicator of growth in average land productivity, it needs to be assumed that: (1) the area under cereals remains constant and there is neither substitution between crops nor a shift to alternative on-farm or off-farm activities; and (2) the number and composition of farm households remains constant. The literature contains an abundance of examples, especially from Southeast Asia, that these conditions are not met.

First, when cereal production expands into areas of lower land quality, where yields are below the average national yield, then this average yield will go down although no individual farmer needs to have observed any yield decline. David Ricardo already identified this process (Ricardo [1818] 1957: 35). The opposite can also be true. For example, Duwayri *et al.* (2000) observed that high growth rates of national average rice yields in Latin America during the 1980s can mainly be attributed to reductions in the area devoted to low yielding upland rice in Central America and central Brazil. Dawe (2002: 361) also pointed to a 25 percent absolute decline in the area of deepwater and upland rice in Asia, which increased the share of irrigated area in the total area and increased average rice yields.

Second, the same is true for differences in labor quality. If young and highly productive farmers decide to leave agriculture, for instance to take up employment in the cities, then the national average yield will go down although none of the remaining farmers might observe any decline in their yields. Papademetriou (2000), for example, observed that the younger generation of farmers in Southeast Asia is moving away from agriculture and from rice farming in particular with only the old generation staying behind.

g. Bifurcation

National statistics do not usually produce a measure of the regional variation around the average, despite the fact that averages without standard deviations are not particularly meaningful. The reason is that the national average yield is commonly estimated from total production divided by total area (*i.e.*, from the sums of output and area, rather than the average of farm-level average yields).

Sub-national yield estimates, where available, can be used to get an impression of how standard deviations have developed over time. Our interest is the direction of these trends, and not the comparison of average yields and standard deviations between countries (the data come from different sources, are compiled using different methods, and sub-national units are of different size, number, and quality).

Data were available for 34 Indian states (rice and wheat, 1970-2000), 29 Chinese provinces (rice, wheat, and maize 1979-1995), 24 Bangladeshi districts (rice and wheat, 1979-2000), 48 American states (wheat and maize, 1970-2001), 32 Mexican states (wheat and maize, 1980-2002), and 27 Brazilian states (rice and maize, 1990-2002). Ideally, one would like to use farm-level yield data, but these typically have no national coverage, and are certainly not available in comparable time series.

Table 2.1:Growth rates in national average cereal yields and sub-national variations around the average for a selection of countries, in % per annum

Country	Period	Rice		Wheat		Maize	
		average	SD	average	SD	average	SD
Bangladesh	1979-2000	1.75**	5.14**	0.57	-1.10	-	-
Brazil	1990-2002	4.57**	2.31**	-	-	4.13**	3.02**
China	1979-1995	2.13**	1.70**	3.37**	2.65**	3.41**	4.23**
India	1970-2000	2.28**	2.37**	2.91**	2.32**	-	-
Mexico	1980-2002	-	-	0.77**	1.80**	2.05**	2.90**
United States	1970-2001	-	-	0.81**	1.91**	1.91**	0.34

Sources: CIMMYT Bangladesh Regional Office (2004); Instituto Brasileiro De Geograpia e Estadística (IBGE) (2004); International Rice Research Institute (IRRI) (2004) (China and India); Servicio de Información y Estadística Agroalimentaria y Pesquera (SIAP) (2004); United States Department of Agriculture (2002)

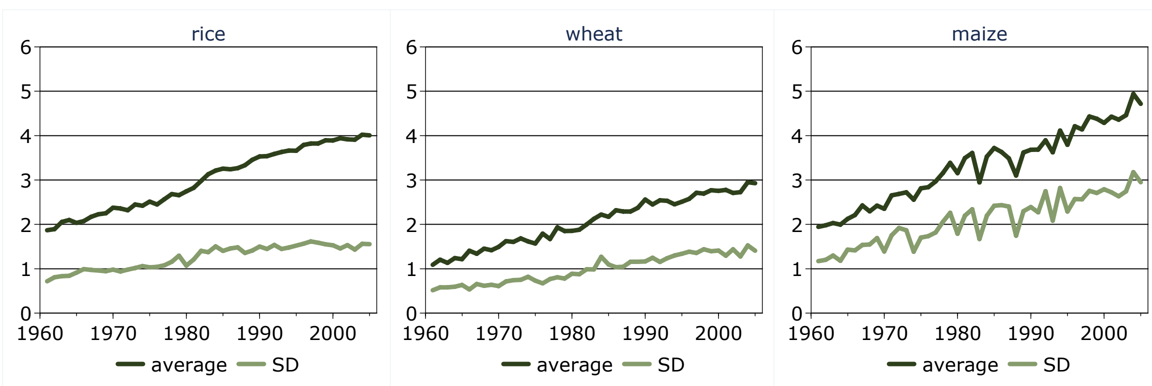
Notes: ** significant at 0.05. Figures not comparable across countries. SD is standard deviation of the mean. Linear growth model is ln(y)=a+bx. - means data not available.

For each year, the national average yield was estimated as the average of sub-national yields using acreage-weights; this is the same as the ratio of total production over total area, but has the advantage that it produces a standard deviation. A linear growth model was fitted to the trends in both yield and standard deviation. The results are shown in **Table 2.1**.

For all countries and all crops in our sample, the standard deviation of the average has increased over time. All growth rates in national average yield and standard deviation are significant except for Bangladeshi wheat. For a number of countries, notably India, Bangladesh, and Mexico, the growth in standard deviation even outpaces the growth in average yield. Sub-national yield averages are clearly bifurcating.

The observed bifurcations of subnational average yields suggest that crop yields grow unequally within countries. It can be interpreted as realized opportunities to increase yields in some regions. Whether this means that some farmers are losing out from technological change depends, among other factors, on the opportunity costs of cereal production. Sub-national bifurcation is not necessarily bad as it can also mean efficiency improvements resulting from regional specialization.

Figure 2.4: Global development of average cereal yields and standard deviations



Source: FAO 2006. **Note:** In tons/ha/season

Out of curiosity, the exercise was repeated at the global level using countries instead of districts and states (see **Figure 2.4**). Though national figures are not necessarily comparable between countries, the pattern that emerges is sufficiently robust to accommodate severe statistical errors. A much more dramatic pattern emerges at a global level, with the standard deviations closely following the trend in average yields, especially for maize and wheat. Again, the growth in average yield of farmers in some countries is obviously lagging behind yield growth achieved in other countries.

2.5 Summary

A direct relationship between an increase in the yield potential and an increase in food security is often claimed; yet, this chapter pointed to four major misconceptions that underlay this claim. It is not true that farmers generally want a higher yield potential, that meeting the future demand is a question of production only, that extra production ends up in the mouths of the hungry, or that low cereal prices help the poor. Yield gaps for most developing countries are wide, which is *prima facie* evidence that there is no straightforward relationship between the level of yield potential and food security. This was further confirmed by a positive correlation between the width of the rice yield gap and food insecurity for 19 Indian states. Although seeds with a higher yield potential can contribute to increased productivity under certain conditions, the situation in many developing countries points to the fact that average yields are mostly constrained by non-technology factors; although the concept of average yield itself needs to be treated with caution.

3 Conceptual frame and analytical approach

3.1 Introduction

Despite the simplicity of its definition, the concept of crop yield gaps is complex as a multitude of factors determines its width. The previous chapter showed that economic factors are often missing in arguments that connect the progress in crop breeding for a higher yield potential with meeting the food requirements of the poor. One important reason was a lack of understanding about the impact of improved varieties on a heterogeneous population of farm households. The remaining chapters of this thesis scrutinize the concept and pitfalls of crop yield gaps at the farm household level. An integrated model that captures soils, crops, and households is developed and calibrated to two village communities in southeast Uganda. **Chapter 4** describes the model components in general terms; four subsequent chapters give details on each component, followed by the simulation results in **Chapter 9**.

The type of model is a multi-agent system, which is a suitable methodology for capturing biophysical and socioeconomic complexity and for spatially integrating different models of farm household decision-making and biophysical dynamics. Such spatial integration is useful for analyzing empirical problems of an integrated nature of which crop yield gaps are a good example. The developed model has wider application than for analyzing yield gaps; other potential applications include soil fertility decline, technological change in agriculture, and the impact of HIV/Aids on agriculture. Without moving focus away from the yield gap, the following chapters will touch on all these topics, as these are relevant for Ugandan agriculture today and impact on the crop yield gap as well.

The present chapter is structured as follows. Section 3.2 describes how maize yield gaps are quantified and decomposed in both proximate and underlying factors. Section 3.3 adds socioeconomic dimensions to the yield gap and Section 3.4

introduces the research area where the method is applied. The chapter ends with a summary.

3.2 Decomposing crop yield gaps

Causes of crop yield gaps can be divided into proximate and underlying causes. Proximate causes for instance include variety choice, and levels of fertilizer and labor use. Underlying causes, on the other hand, explain the reasons why farm households do not grow improved varieties and do not apply maximum levels of input.

3.2.1 Proximate factors

In a first stage, the yield gap is decomposed based on the standard methodology developed by Gomez (1977). In its original form, this method involves a series of on-farm experiments to decompose the yield gap into factors such as variety choice, fertilizer use, and pest control. Each input is applied at two levels, of which one is the input use by the farm household and the other is the recommended, or optimal, level. The experiments exhaustively combine each combination of input levels to separate the contribution of each factor statistically.

The method applied here is similar, yet with the crucial difference that computer experiments replace the on-farm experiments. These computer experiments use empirically based crop yield functions that represent the relationship between kilograms of harvested crops and input decisions (crop choice, land allocation, fertilizers, management). This might not be as precise as on-farm experiments for a particular level of input use. Yet, it has the great advantage, apart from low research costs, that the yield gap can be decomposed not only for one input level for one crop and for an average farm household, but for every input level, for all crops and for every 'agent'. The term agent is used instead of farm household to indicate that the method uses a model representation of a real-world farm household.

The present decomposition analysis uses four factors: variety choice, fertilizer use, labor use, and intercropping. Labor use and intercropping are not commonly used in this type of analysis but are included here because these are important for agriculture in southeast Uganda as few farmers have access to animal or mechanical traction and most crops are grown in combination.

Table 3.1 shows the factorial design of the experiments, with an empty square indicating the simulated actual input use of the farm agent, and the full square the maximum level of input use. For the fourth factor called 'pure stand', the empty squares stand for intercropping and the full squares for cropping in pure stand.

Each component of the yield gap can be identified using the following formulas:

$$YGAP_{var.} = \frac{Y_2 + Y_6 + Y_7 + Y_8 + Y_{12} + Y_{13} + Y_{14} + Y_{16}}{8} - \frac{Y_1 + Y_3 + Y_4 + Y_5 + Y_9 + Y_{10} + Y_{11} + Y_{15}}{8}$$

$$YGAP_{fert.} = \frac{Y_3 + Y_6 + Y_9 + Y_{10} + Y_{12} + Y_{13} + Y_{15} + Y_{16}}{8} - \frac{Y_1 + Y_2 + Y_4 + Y_5 + Y_7 + Y_8 + Y_{11} + Y_{14}}{8}$$

$$YGAP_{lab.} = \frac{Y_4 + Y_7 + Y_9 + Y_{11} + Y_{12} + Y_{14} + Y_{15} + Y_{16}}{8} - \frac{Y_1 + Y_2 + Y_3 + Y_5 + Y_6 + Y_8 + Y_{10} + Y_{13}}{8}$$

$$YGAP_{stnd} = \frac{Y_5 + Y_8 + Y_{10} + Y_{11} + Y_{13} + Y_{14} + Y_{15} + Y_{16}}{8} - \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_6 + Y_7 + Y_9 + Y_{12}}{8}$$

Table 3.1: Experiments for decomposing the yield gap in proximate factors

Experiment	Variety	Fertilizer	Labor	Pure stand	Other
1 (baseline)	□	□	□	□	□
2	■	□	□	□	□
3	□	■	□	□	□
4	□	□	■	□	□
5	□	□	□	■	□
6	■	■	□	□	□
7	■	□	■	□	□
8	■	□	□	■	□
9	□	■	■	□	□
10	□	■	□	■	□
11	□	□	■	■	□
12	■	■	■	□	□
13	■	■	□	■	□
14	■	□	■	■	□
15	□	■	■	■	□
16 (max. yield)	■	■	■	■	□

Note: □ = agent level of input use ■ = maximum level of input use

3.2.2 Underlying factors

The existence of yield gaps does not immediately imply economic inefficiencies. Low levels of yield can still be optimal if resources are limited and farm households therefore face tradeoffs in their input allocation. Performance gaps instead of crop yield gaps are then more meaningful. Herdt and Mandac (1981), for instance, measure performance in terms of efficiency.

The present study decomposes performance gaps in three underlying factors following the same factorial design as for the proximate factors. Three underlying factors are of interest: knowledge about innovations, short-term credit, and hiring in temporary labor and/or leasing a tractor. These underlying factors do not relate to crop yield in the same straightforward way as inputs enter a crop yield function. Moreover, the interest now goes to the overall performance of the farm rather than the yield of a single crop, and trade offs in land allocation and input use then become important. The above computational experiments based on the crop yield function cannot be used because of dynamic feed back effects: the performance in one year influences the performance in subsequent years. Dynamic simulation experiments are used instead. This involves a series of simulation experiments in which constraints on each factor and combination of factors are relieved in turns. **Table 3.2** shows the factorial design of these simulation experiments. The crop yield gap can be decomposed using eight computational experiments and the following formulas:

$$PGAP_{knowledge} = \frac{P_2 + P_5 + P_6 + P_8}{4} - \frac{P_1 + P_3 + P_4 + P_7}{4}$$

$$PGAP_{credit} = \frac{P_3 + P_5 + P_7 + P_8}{4} - \frac{P_1 + P_2 + P_4 + P_6}{4}$$

$$PGAP_{LaborHiring} = \frac{P_4 + P_6 + P_7 + P_8}{4} - \frac{P_1 + P_2 + P_3 + P_5}{4}$$

Table 3.2: Experiments for decomposing the yield gap in underlying factors

Experiment	Knowledge about innovations	Short-term credit	Hiring in labor / leasing tractor	Other
1 (baseline)	□	□	□	□
2	■	□	□	□
3	□	■	□	□
4	□	□	■	□
5	■	■	□	□
6	■	□	■	□
7	□	■	■	□
8 (max. yield)	■	■	■	□

Note: □ = constrained level ■ = unconstrained level

3.3 Socioeconomic dimensions of the yield gap

What are good performance indicators of the farming system? A good indicator would be the extent to which the system satisfies farm households' objectives. These objectives may include income, health, education, social status, security, leisure time, and savings to smooth consumption in the face of shortfalls. These are private objectives because they do not necessarily coincide with the objectives of society as a whole. A principal social objective, which many governments subscribe to, is national self-sufficiency in cereals. Other social objectives, flowing from the current development paradigm, are poverty reduction, food security, sustainability, and equality. Social and private objectives are compatible in principle, as the well-being of society is the sum of the well-being of its members. Yet, social objectives not only relate to the level of well-being but also to its distribution; it is here that social and private objectives diverge. **Table 3.3** summarizes the private and social objectives and the way they are quantified in this thesis, each is detailed in the following subsections.

Table 3.3: Measurement of private and social objectives

Private objectives		Social objectives		
Economic well-being	Food security	Food sufficiency	Equality	Ecological sustainability
Food energy consumption	Savings, assets and crop diversity	Surplus food production	Gini inequality index	Nutrient stocks

3.3.1 Private objectives

Private objectives of farm households have multiple dimensions that can be captured by a utility function, which will be specified in **Chapter 4**. The utility function is specified at the farm household level assuming that utility is equally spread over all household members, although this might not be the case in reality. The fulfillment of these private objectives is measured in the light of the current development paradigm, including poverty reduction and food security. Equality and sustainability are additional aspects of this paradigm, but these are treated as social objectives. It is useful to note here that private and social objectives are normative and that it is for policy makers to decide on 'acceptable' levels of each of these indicators.

3.3.1.1 Economic well being

Economic well being, or its reverse poverty, concerns the command over commodities deemed essential to constitute a reasonable standard of living in society. The analysis is restricted to indicators of economic welfare rather than the broader concept of well-being, which would include aspects such as health and education.

In a developing country context, the concepts of food security and poverty are very much alike. Although poverty is most often quantified as a dichotomous variable, the best poverty measures are scalars and the best scalar is food energy consumption (Lipton 1983; Coudouel *et al.* 2002). Food consumption can be quantified in terms of food energy in joule.

The food energy intake is quantified on a per capita basis in male adult equivalents to control for differences in size and composition of households. The adult equivalent is set to an average annual energy requirement of an adult male (18-62 years old) in Uganda of 3.259 billion joule. Because food energy needs depend on the size and composition of the household, the poverty line is household specific. Food energy consumption is estimated from a complete demand system estimated in **Chapter 8**. Such a system allows for substitution between food products, which makes the poverty line more flexible as the relation between food expenditures and energy consumption is non-linear. Rich households, for instance, consume more protein-rich than calorie-rich foods (*e.g.*, more meat, fewer cereals) and hence get less food energy for the same amount of money.

3.3.1.2 Food security

In addition to food energy consumption, food security also refers to a stable command over food over time. Two indicators measure this type of security: first, the value of farm assets plus savings; and second, crop diversification. Farm households with a diversified income and a high value of farm assets spread risk in two dimensions: diversification spreads risk within a year, while farm assets spread risk over several years.

Ownership of farm assets spreads the risk of income variation as it allows smoothing of consumption in the face of shortage. Investments in livestock are a good example commonly found in developing countries (Kristjanson *et al.* 2004). Farm households invest in livestock in years when income is ample and sell when income falls short; **Chapter 8** presents a quantitative model for this. Farm assets are quantified in value terms at the end of the cropping season. The value farm assets plus savings is expressed in per capita terms.

The crop diversification index (CDI) was used to quantify the spread of risk over different crops. The CDI is calculated as one minus the Herfindahl concentration index, which is calculated as the sum of squares of all proportions under each crop. The CDI ranges from 0 to 1, with larger values indicating more diversification.

3.3.2 Social objectives

Social objectives encompass the private objectives but add three more dimensions: (1) a high level of national self-sufficiency in cereal production; (2) an acceptable level of inequality; and (3) ecological sustainability.

3.3.2.1 Self-sufficiency in cereal production

Many national governments adhere to a high level of self-sufficiency in cereal production. Assessing the level of national autarky would need a different modeling approach, such as a computable general equilibrium model, to capture the economic linkages between farm and non-farm sectors. Yet, at the village level, where non-farm activity is limited, surplus production is a good indicator of self-sufficiency. Surplus production is quantified as the value of cereals produced minus what is consumed within the community.

3.3.2.2 Inequality

Inequality concerns the dispersion in the distribution of income or assets. It is important as it is closely associated with both poverty and food insecurity. The Gini coefficient, or Gini concentration ratio, is a popular measure of inequality (Gini 1912). It is scaled from zero (perfect equality) to one (perfect inequality). Let y_i be the per capita income of household *i* and y_j the per capita income of household *j* in a population of *N* households. Then the Gini coefficient is defined as the mean of the difference between every possible pair of individuals *i* and *j*, divided by the arithmetic mean income $\bar{y}$ of the total population (Litchfield 1999):

(3-1) $$\text{Gini} = \frac{1}{2N^2\bar{y}} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| y_i - y_j \right|$$

3.3.2.3 Ecological sustainability

Ecological sustainability here refers to the long-term biophysical capacity of the soil to sustain the same level of crop yields over time. Ecological sustainability is both a private and a social objective. It is a private objective because soil fertility decline reduces crop yields of farm households, who hence have an incentive to produce in a sustainable way. It is a social objective as farm households might not immediately notice the impact of soil fertility decline on crop yields, or might overexploit their soils to increase crop yields at the cost of future generations. **Chapter 6** describes how crop yield and the level of soil nutrients are endogenous in the model. Ecological sustainability is quantified using soil nutrient balances for nitrogen, phosphorus, and potassium.

3.4 Application to Uganda

Uganda was chosen for two reasons. First, crop yield gaps in this country are notoriously large, while average yields are reported to be stagnating. Second, for the practical reason that at the time this research began farm household and soil data had been collected in the framework of the research project *Policies for Improved Land Management in Uganda* (Pender *et al.* 2001; Kaizzi 2002; Nkonya 2002; Brunner 2003; Pender *et al.* 2004; Woelcke 2004, 2006; and Ruecker 2005). The International Food Policy Institute (IFPRI) led this project in collaboration with several national agricultural research organizations in Uganda and the Center for Development Research in Bonn.

Main findings of the project concern the determinants of land degradation in Uganda. Increased sales of agricultural production, better infrastructure, and better access to output markets have stimulated the outflow of nutrients from agriculture. At the same time, these nutrients have not been sufficiently replenished by increased input use, as input markets have remained poorly developed (Pender *et al.* 2001, 2004). The research stressed the importance of assessing the profitability of technologies to reduce soil degradation, as well as the importance of better extension services and market information for farmers to enhance yields and reduce land degradation (*ibid.*). The present study builds on these findings by explicitly including the identified constraints and opportunities into the simulation model.

3.4.1 Southeast Uganda

The present study focuses on two village communities in southeast Uganda. Data were collected in 1999–2000 and were used to build a bio-economic model (see Woelcke 2004, 2006). This model included mixed integer linear programming models at the farm household level and artificial neural networks used as yield estimators (*ibid.*). The model was used to assess the financial and technical feasibility of sustainable farming practices. The research found that under current conditions, most of these farming practices are not profitable, even if credit were provided (*ibid.*). Woelcke's model was adapted and extended for the present purpose using various sources of data listed in **Appendix A**.

It was not the primary objective of this study to give tailored policy advice to the government of Uganda concerning technological change in agriculture. Yet, the model was carefully calibrated to local conditions and relevant policy issues are taken into account, such as: (1) the impact of HIV/Aids on agriculture; (2) population growth; (3) soil fertility decline; and (4) declining crop yields.

3.4.2 Maize in Uganda

This thesis focuses on the adoption of maize varieties with a higher yield potential, and particularly on hybrid maize. The focus is on maize for three reasons: (1) together with roots and tubers, maize is the most important food staple in Uganda; (2) maize has received most attention in research, government agencies and NGOs heavily promote its cultivation, and the crop has a main role in food security strategies; and (3) some organizations in Uganda have made the increase in yield potential a research priority for maize (*e.g.*, ASARECA 2001).

There is wide consensus that maize yields, and crop yields in general, in Uganda have declined or, at best, are stagnant (Deininger and Okidi 2001 as in Nkonya *et al.* 2003; Sserunkuuma *et al.* 2001; Pender *et al.* 2001). The statistical evidence is, however, weak. Reporting results from the 2000-2001 IFPRI Farm Household Survey, Pender *et al.* (2001) wrote that farmers reported declining yields of all crops in all zones of the country. They furthermore reported that maize yields have declined more in those areas most suitable for cereal expansion, which they attributed to the expansion of production into less suitable areas and a decline in soil fertility (*ibid.*)

In contrast, FAO data show a significant increase in maize yields since the mid-1980s. These data are, however, not very reliable as Uganda has no sound system of annual crop yield data collection. Yield estimates are conducted by the Ministry of Agriculture and rely on expert opinion rather than crop cutting and sample surveys. There are no reliable estimates of current crop yields at a national level and neither of time series. The 1999-2000 Uganda National Household Survey (UNHS), which would normally be a good data source for estimating farm level yields, is not suitable for it recorded plot sizes but not the area under each crop.

3.5 Summary

This chapter outlined the conceptual and analytical frame for the remainder of the thesis. The yield gap can be decomposed in proximate factors based on the crop yield equation or in underlying factors using scenario experiments. Agent-based modeling is a suitable tool for doing the analyses for it can capture much of the real-world complexity that derives from heterogeneity, interaction, and system dynamics. In addition, agent-based modeling allows the decomposition of crop yield gaps not just for a few plots on a few selected farms as in traditional experiments but virtually, for all agents and for all plots. The following chapters calibrate the model to two village communities in southeast Uganda and use it to analyze the yield gap and simulate the diffusion of a high yielding maize variety.

4 General methodology

4.1 Introduction*

Crop yield gap dynamics are simulated at the farm household level using multi-agent systems (MAS). MAS serve as a framework for integrating 'conventional' models of biophysical processes and human behavior. The word conventional is used to underline that these separate models do not break with well-established biophysical or economic theories. The novelty of the methodology is the way these conventional models are integrated. To emphasize the aspect of integration in MAS, the separate models are referred to as 'system components'. The technical details of these system components are given in **Chapters 5**, **6**, **7** and **8**, while this chapter describes the MAS approach in general terms.

The chapter is structured as follows. Section 4.2 justifies the use of MAS. Section 4.3 introduces the three main system components. Section 4.4 discusses how the taken approach meets the challenges of capturing heterogeneity, interaction, and dynamics. Section 4.5 describes the use of mathematical programming (MP) to simulate farm decision-making, and Section 4.6 expands the use of MP with a three stage non-separable decision process to capture economic trade-offs over time. The chapter ends with a summary.

4.2 Methodological approach

4.2.1 Heterogeneity

Development research has come to focus on poverty reduction, inequality, and ecosystem sustainability, all of which require an enhanced understanding of heterogeneity at levels of farm households. The importance of heterogeneity is also

---

* Parts of this chapter were used for SCHREINEMACHERS, P. and T. BERGER 2006. Land-use decisions in developing countries and their representation in multi-agent systems, *Journal of Land Use Science*, 1(1): 1-16.

realized in agricultural research and extension services as technological change is observed to occur unevenly across and within different production environments (*e.g.*, Renkow 1994). Blanket input recommendations have proven unsatisfactory because these do often not correspond to farmers' financial optima and strategies that strengthen the learning abilities of farmers to optimize their individual input use have proven more fruitful instead (*e.g.*, Lacy *et al.* 2000; Snapp *et al.* 2003).

Analytical approaches addressing this heterogeneity in agricultural economics have remained weak. Statistical models thrive on heterogeneity for estimating parameters but are commonly used for deriving averages only (*e.g.*, average income elasticity, average producer response). Farming systems research has long realized that agriculture in developing countries typically consists of large numbers of farm households that are heterogeneous in terms of opportunities and constraints (Dillon and Hardaker 1993). Yet, most model approaches have simplified this heterogeneity by *ex ante* categorizing farm households into more homogenous groups (typically four) and then specifying a mathematical programming model for each (*e.g.*, Kuyvenhoven *et al.* 1998; Kebbeh and Miezam 2003; Holden *et al.* 2004; Woelcke 2006). This categorization can be rather arbitrary as based *a priori* knowledge or statistical analysis, while whole farm mathematical programming models tend to be sensitive to small changes in resource endowments or constraints.

4.2.2 Mathematical programming-based multi-agent systems (MP-MAS)

It is with respect to heterogeneity that MAS based on mathematical programming (MP-MAS) distinguish themselves most clearly from approaches based on representative farm household models. Using MP-MAS the *ex ante* categorization into representative farm households can be avoided and the grouping can be done *ex post* for the purpose of analysis or for summarizing results.

MAS are a methodology originating from computer science, which is now widely applied to the analysis of complex systems (Berger 2001). MAS are also applied to land-use/cover change, but still little to farming systems research. Most MAS of land-use/cover change are based on relatively simple behavioral rules that have only little in common with microeconomic theory. However, Balmann (1997) showed that MAS can be successfully built on farming systems research by simulating agent decision-making using whole farm mathematical programming. The pioneering work of Balmann (1997) was followed by empirical applications of Berger (2001) and Happe (2004).

MAS contribute to farming systems research by more fully capturing the heterogeneity in farming systems and allowing farm household models to interact directly in the exchange of resources or the communication of knowledge (Balmann 1997; Berger 2001; Happe 2004). The use of mathematical programming has furthermore the advantage that in their process (or activity) approach to agricultural production they provide an appropriate format for integrating information from different disciplines (Hazell and Norton 1986). **Chapter 10** will return to a comparison of alternative approaches.

4.3 Introduction of system components

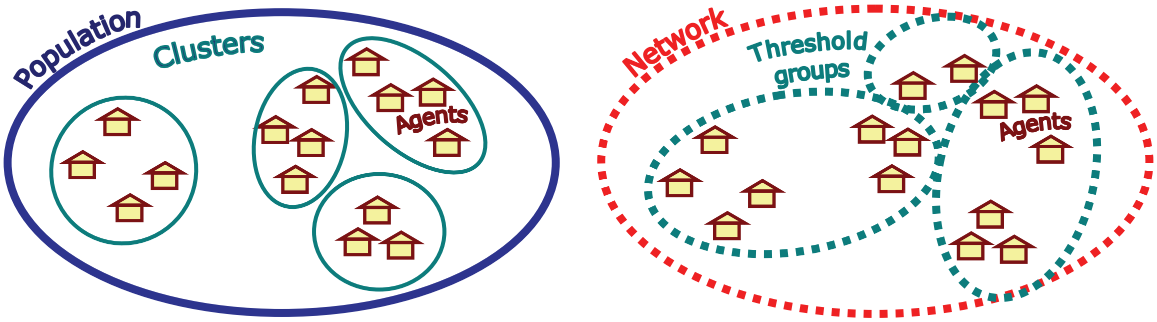
The present MAS approach integrates three main system components: farm agents, a landscape, and biophysics. The following four chapters describe each in detail but here they are briefly introduced and their interaction is made explicit.

4.3.1 Farm agents

The term agent refers to a computer model representing a real-world farm household. Each agent represents a single real-world farm household, *i.e.*, there is a one-to-one correspondence between farm households and model agents. In the empirical application of this study, there are 520 agents representing all farm households in two villages in southeast Uganda. The decisions of model agents are simulated using whole farm mathematical programming.

Figure 4.1 explains some basic terminology used with respect to the agents; with the agents represented by farmstead icons. The left diagram shows that all agents together form the agent population. Agents belong to clusters, which are subgroups of the population characterized by similar resource endowments. These clusters are used to generate the agent population from survey data, which will be explained in the next chapter. The right diagram shows that every agent also belong to a network; networks define the possibilities for communication of innovations and are discussed in **Chapter 7**. Just as the population is composed of clusters, the network is composed of threshold groups. Membership of a threshold group determines the availability of, and access to, innovations. The figure furthermore shows that clusters and threshold groups do not necessarily overlap.

Figure 4.1: The concepts of agent populations and networks



4.3.2 Landscape

The landscape consists of two villages in the Lake Victoria Crescent of southeast Uganda. Geographical information about these villages is organized in layers. This information includes soil properties as well as locations of agents and farm plots; **Chapter 5** describes each layer in detail. Layers are composed of grid cells with each cell representing a small unit of land about the size of the smallest plot.

4.3.3 Biophysics

Whereas the landscape component contains the spatial data on soil properties, the biophysical component simulates changes in these properties resulting from land-use decisions and natural processes such as decomposition of organic matter and atmospheric deposition. How soil properties and land-use decisions influence crop yield is modeled using crop yield response functions. The yield response is crop, plot, and agent-specific. The Tropical Soil Productivity Calculator (TSPC) was used for this purpose (Aune and Lal 1995, 1997). The TSPC has non-linear multiplicative yield functions that resemble the theory of Mitscherlich because factors are assumed complementary while interaction between factors is allowed.

4.4 Heterogeneity, interaction, and dynamics

Complexity in agro-ecosystems can be conceptualized as deriving from three aspects: heterogeneity, interaction, and complex system dynamics. The following three sub-sections describe how this complexity is represented in the model.

4.4.1 Heterogeneity

A distinctive feature of MAS is the existence of many individual agents. In the present MAS model, heterogeneity among agents is introduced by varying the following five groups of variables:

1. land quantity (hectares) and quality (soil physical and chemical properties, and land-use history);
2. labor quantity (household size) and quality (sex and age composition);
3. livestock quantity (number of animals) and quality (species and age);
4. quantity of permanent crops (ha of coffee) and quality (age of plantation); and
5. membership to threshold groups determining the access to innovations.

There are other sources of variation among agents one can think of, such as variation in price expectations, the incidence of crop pests, human sickness, skills, and educational level, which are not considered in the model. These could however, be captured within the same framework, if data were available. Heterogeneity in the landscape is captured by modeling each separate plot. All soil properties are scalars and changes in properties are determined endogenously in the model.

4.4.2 Interaction

The word 'system' suggests that there are linkages between things or people. Three types of interaction are distinguished:

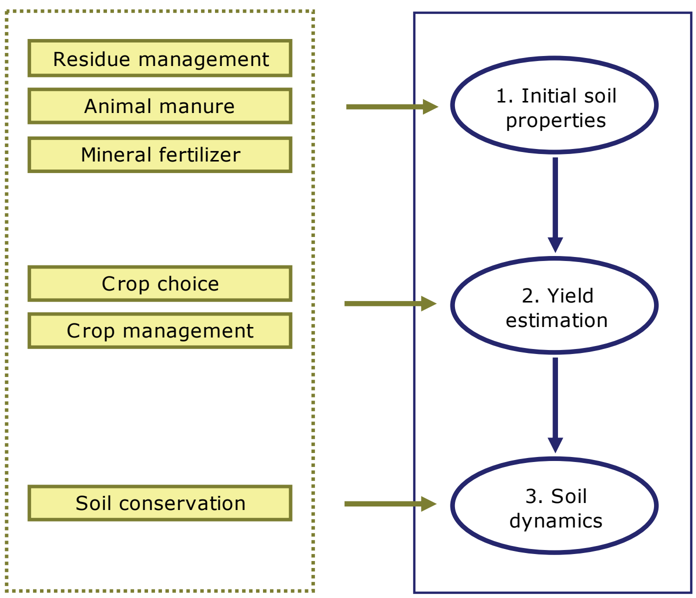
1. Agent-agent interaction. Interaction among agents takes place in the communication of innovations. Quantitative as well as qualitative information can be used to specify how innovations are communicated among agents. Depending on this, an innovation becomes accessible to some but not all agents.

2. Agent-environment interaction. Agricultural production is the main such type of interaction. **Figure 4.2** shows how management decisions influence soil properties and crop yield. The three vertically ordered ovals represent the biophysical process. Agents affect the environment at three levels. First, residue management, livestock (manure) and mineral fertilizer affect crop yield indirectly by altering the properties of the soil. Second, crop choice and crop management (*e.g.*, weeding, harvesting) affect crop yield directly. Third, conservation measures, such as erosion control, have an impact on soil property dynamics and can have a long-term impact on the soil properties. Yet, erosion control is neither

a common practice in the study area nor is erosion very severe (Brunner *et al.* 2004); erosion control is therefore not included as a management option in the model.

3. Environment-environment interaction. Soil erosion, deposition, leaching, and volatilization are spatially explicit soil processes, meaning that the process in one location is a function of processes in surrounding locations. The grid-cell-based landscape component can capture such interaction, but in the current model version only erosion is spatially explicit by the inclusion of a slope length factor (see **Chapter 5**).

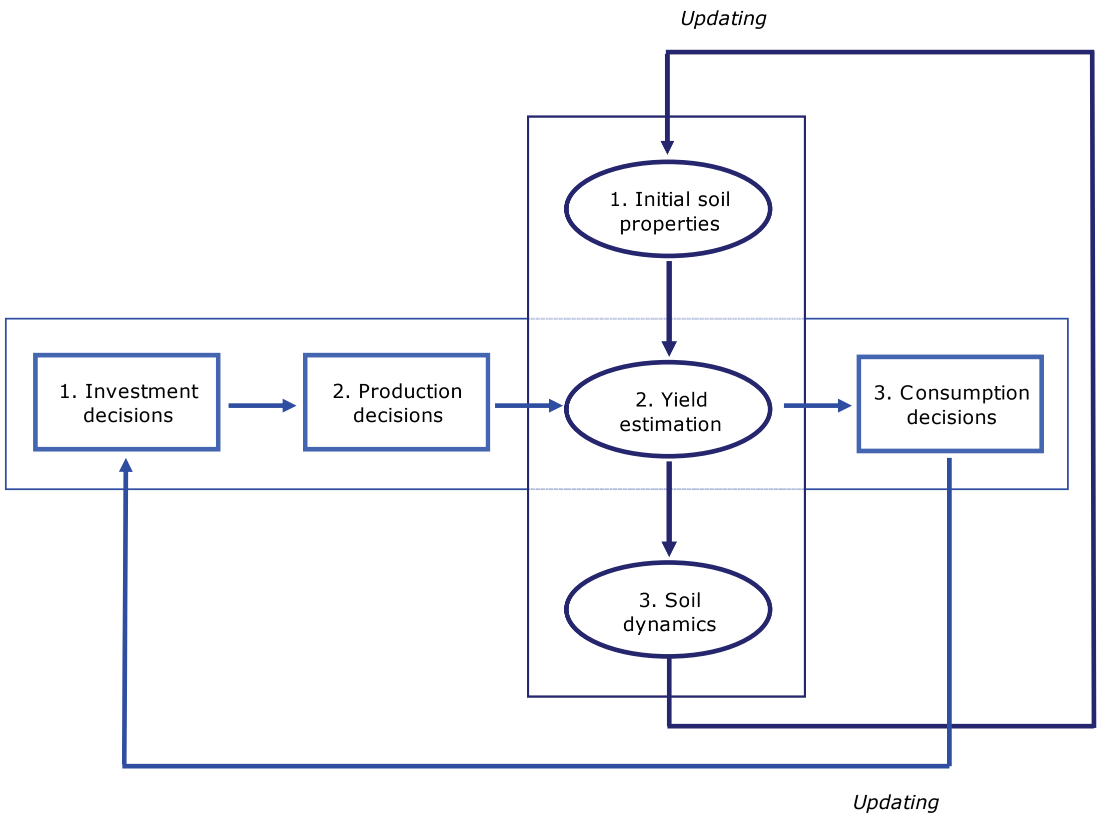
Figure 4.2: Three levels of agent-environment interaction



4.4.3 Dynamics

The basic time step of the current MAS is one year with a complete simulation run spanning 16 years. The outcome of one year is the basis for the following year. This is true for both agents and the soils they cultivate. **Figure 4.3** shows the two interacting processes; the three vertically ordered ovals are the same as in **Figure 4.3** and the three horizontally ordered rectangles represent the farm decision-making process. More details about this three-stage decision-making procedure follow.

Figure 4.3: Dynamics and interaction of soil processes and farm decision-making



4.5 Mixed integer linear programming (MILP)

Agent decision-making is modeled using mathematical programming (MP). The essence of this methodology is optimization of decisions under constraints. As some of the decision variables, such as livestock production, can only take integer values, the type of mathematical program used is a mixed integer linear program (MILP).

4.5.1 Non-separable farm decision-making

Maximization of expected utility is the objective that guides agents in their farm decision-making. The utility function has three dimensions: (1) cash income from sales and off-farm labor; (2) in-kind income from auto consumption of crop and livestock products; and (3) the annuity of future expected income from investments.

Cash income and in-kind consumption objectives are included separately as these objectives are not identical in the case of market imperfections and risk. When markets are imperfect, the resource allocation that optimizes the level of income does not necessarily optimize consumption (Sadoulet and de Janvry 1995). In microeconomic literature, this is known as the non-separability of consumption and production decisions. This non-separability means that consumption decisions need to be taken into account when optimizing production decisions. Production and consumption decisions can, in recursive fashion, be integrated in a mathematical programming model (Dillon and Hardaker 1993; Ruben *et al.* 1994).

4.5.2 Concise theoretical model

The theoretical model is here presented in concise mathematical form. **Chapters 6** and **7** will give more details about the separate production and consumption parts.

Let OUT be the total output sold and *a* the farm gate price, and let INP be the matrix of variable input requirements purchased at price *b*. The difference between these two factors (*a**OUT – *b**INP) is then the current cash gross margin. Furthermore, let AUT be the quantity of production that is auto-consumed and *c* its internal price, FUT the future output from investments expressed as an annuity at price *e*, and FIX be the fixed costs. The objective function is then specified as:

(4-1) $$\text{Maximize} \quad a * OUT - b * INP + c * AUT + e * FUT - FIX$$

Let PUR be the quantity of purchased food products and *d* its market price. PUR does not appear in the objective function because its internal valuation by the agent equals its market price and hence the terms cancel out as agents consume all what is purchased. Nevertheless, the term appears as a constraint as cash expenses cannot exceed cash revenues:

(4-2) $$a * OUT - b * INP - d * PUR - FIX \geq 0$$

Furthermore, let ACT be a set of alternative production activities, and YLD a matrix of per unit output levels. The total quantity sold (OUT) and auto-consumed (AUT) cannot exceed the production possibilities range as defined by the product of the matrices YLD and ACT:

(4-3) $$OUT + AUT - YLD * ACT \leq 0$$

Moreover, input levels cannot exceed the matrix of available inputs (AVI):

(4-4) $$ACT * INP \leq AVI$$

On the consumption side of the model, let INC be the disposable income, which equals the value of the objective function but omits the annuity of investments (e*FUT):

(4-5) $$INC = a * OUT - b * INP + c * AUT - FIX$$

The expenditure model includes a novel three-step budgeting process that includes the decisions to save, to allocate income to food, and to spend the food budget on specific food categories. A flexible budget system is used to simulate poverty dynamics in terms of food energy levels. **Chapter 8** specifies and estimates this model; here it suffices to state that the value of food consumption (CON) is a function of disposable income (INC) and household size (HHS):

(4-6) $$CON = f(INC, HHS)$$

and that the value of food expenditures is the sum of the value of own produce and food purchased from the market:

(4-7) $$CON = c * AUT + d * PUR$$

Finally, all decision variables are constrained to positive values:

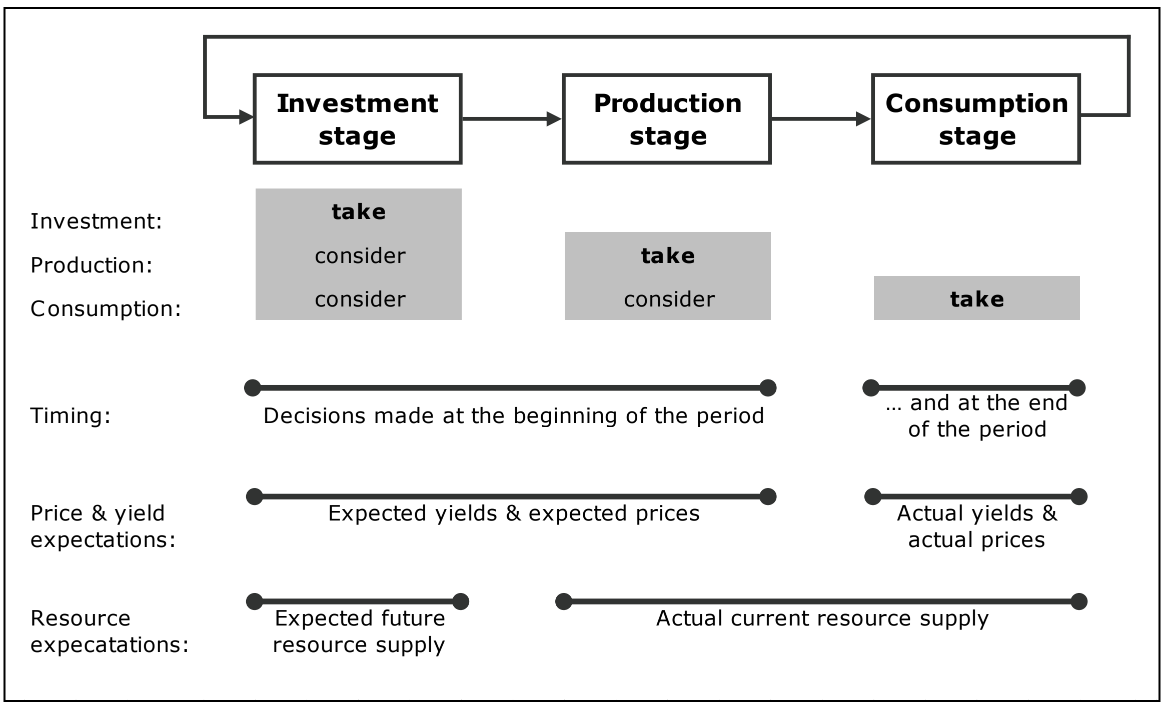
(4-8) $$OUT, AUT, INP, PUR, FUT, FIX, ACT, AVI, CON \geq 0$$

4.6 A three-stage non-separable decision process

Mathematical programming models are suitable to capture trade-offs in the allocation of scarce resources at one point in time, for instance, the trade-off between using the land to grow maize or to grow vegetables. Another type of trade-off is the allocation of scarce resources in time, for instance, between consuming income now and investing it to increase consumption in the future, or between growing a seasonal crop or a perennial crop. These temporal trade-offs can be handled in a multi-period MP that optimizes production and investment decisions over the entire lifespan of an investment. Berger (2001) suggested a two-stage MP as a cost-effective alternative to multi-period programming in MAS and this thesis expands it to a three-stage non-separable MP by including consumption decisions, which makes it more suitable for developing country application.

In each period, agents go through a three-stage process of investing, producing, and consuming as was shown by the three horizontally ordered squares in **Figure 4.3** and is also shown in **Figure 4.4**. Each stage involves the optimization of a unique programming matrix and parts of the solution vectors are transferred between sequential stages. **Appendix B** lists the differences in the programming matrices used at each stage. The matrices differ in internal matrix coefficients (*e.g.*, yields and consumption function coefficients), objective function coefficients (*e.g.*, prices), right-hand-side values (*e.g.*, assets, resource endowments, and liquid means), and in the number of included constraints. **Figure 4.4** summarizes the differences between the stages in a more concise form. The following subsections outline each of these three stages.

Figure 4.4: Conceptual model for a three-stage decision-making process



4.6.1 Investments

Investments are those productive activities with a gestation period between first input use and total output of more than one year. This includes most forms of animal husbandry, perennial crops including forestry, and infrastructure. The difference between investment decisions and current production decisions is the time horizon of the decision-maker.

Agents optimize investment decisions by comparing future and annual costs with future and annual revenues, which are based the agent's yield and price expectations and the long-run expected household labor supply (see **Chapter 7**). Production and consumption decisions are considered in the investment stage by a simultaneous optimization of all three decisions. This means that the agent considers the trade-off between future income and current needs when deciding how much to invest.

The results of the optimal investment plan are then added to the resource endowments of the agent. For instance, if the agent had three cows at the start of the period and invests in two more cows, then the resource endowment is updated to five cows of different vintages while the agent's savings are reduced with the purchasing price of two cows.

4.6.2 Production

After deciding on investments, the agent decides on the current land-use in the production stage. This includes the decisions what crops to grow, on which plots, using which variable inputs, and in what quantities. Like in the investment stage, these decisions are based on the agents' expected yields and prices for the present period. Consumption needs are considered by simultaneously optimizing the production and consumption decisions.

All new investment decisions, such as purchasing new livestock, acquiring more land, or planting additional trees were already taken in the previous stage and cannot be revisited at the production stage. Livestock can, however, be sold and trees be cut down if needed at this stage. For this, previous investments are split into two alternatives: one is the decision to sell or cut at the end of the period based on the present expected costs and benefits, while the other decision values the future expected costs and benefits. If present net benefits exceed the expected future annual net benefits, then the investment will be sold or cut at the end of the present period.

4.6.3 Consumption

In the third and last stage of the decision-making process, the model agents sell and consume products based on actual simulated yields and prices. Investment and production decisions cannot be reversed in the consumption stage. Actual prices replace the expected prices and actual crop yields, as simulated by the biophysical component from actual input levels in the production stage, enter the MP to replace expected yields. The results of the consumption stage can be used to quantify the economic well-being and food security of the agents and the agents' resources are updated to serve as a starting point for the next period.

4.7 Software implementation

There is still no commercially available software for using MAS based on mathematical programming and hence a few paragraphs on the software implementation. The software for the present study was developed in collaboration with Thomas Berger (University of Hohenheim) based on Berger's MP-MAS applied to Chile (Berger 2001) and written in the C++ programming language. This software does not have its own graphical user interface but is controlled through command line functions. Different scenarios are managed through a Microsoft Excel workbook that controls selected parameter values in files that are linked to it (**Figure 4.5**). This is an easy way of setting up many different scenarios and avoids mistakes when having to change parameters in many separate files. The scenario manger uses Visual Basic macros to change parameters and to convert all input files from Microsoft Excel into ASCII format, which is read by the MAS software source code.

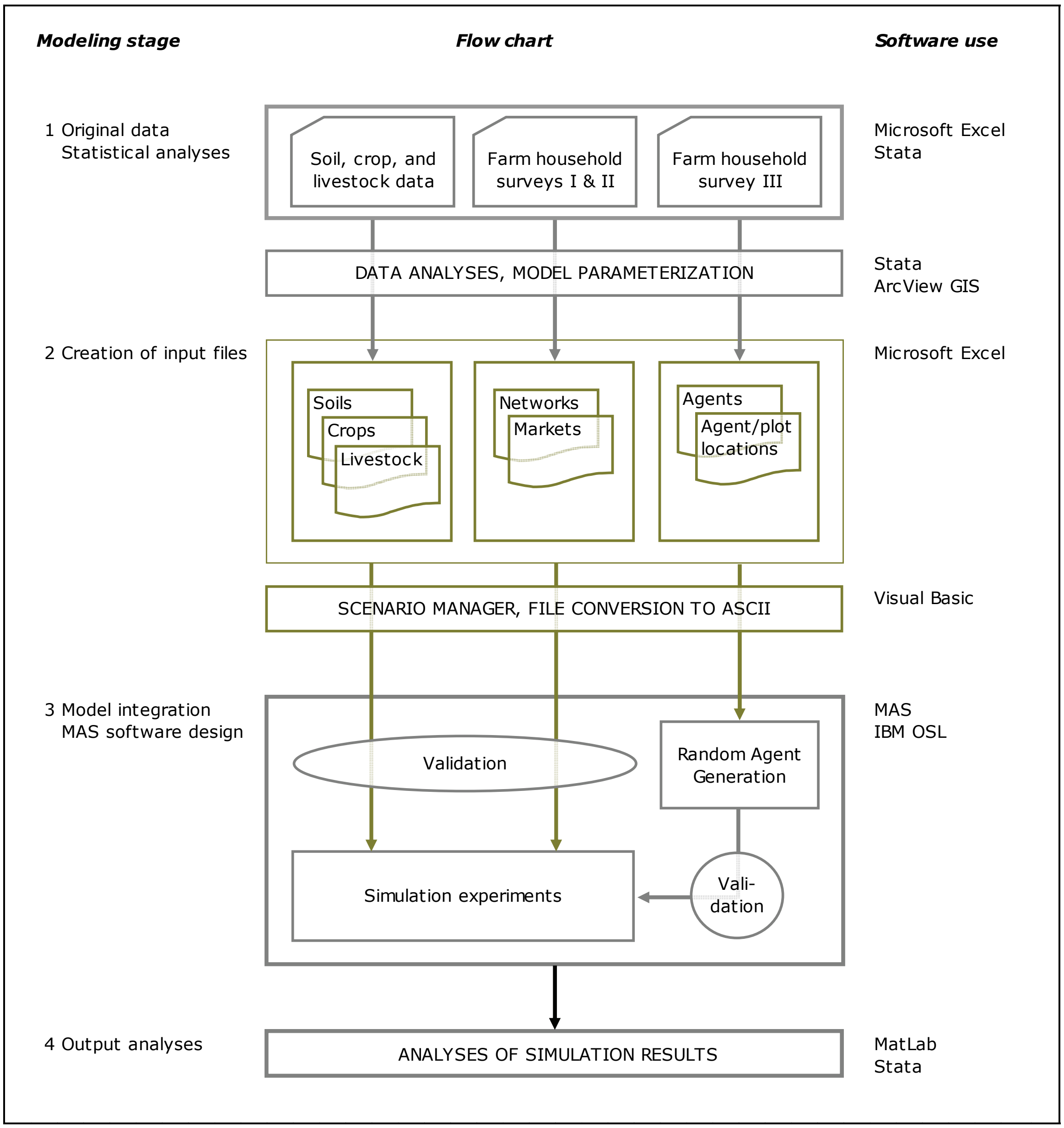
Mathematical programs are solved using the IBM Optimization Subroutine Library (OSL), for which free academic licenses were until recently available from IBM. The OSL is strong in solving large mixed integer problems. After installing the OSL, the model runs on standard personal computers, though the model runs more stable under a UNIX than a Windows operating system.

In its essence, the source code processes input files and returns output files. These output files contain all solution vectors, and changes in soil properties. Output data are analyzed using statistical software packages. For the present study, the output files were pre-processed using routines programmed by Thorsten Arnold (University of Hohenheim) in MatLab and then analyzed using STATA, which is suitable for analyzing large data sets.

4.8 Summary

The multi-agent systems approach is a suitable approach for capturing heterogeneity, interaction, and dynamics of landscape and agents. The combination of this system with mathematical programming offers a powerful tool for simulating farm household decision-making in a changing environment. An innovative three-stage decision process, in which investment, production, and consumption decisions are inseparable but sequentially solved captures temporal and non-temporal economic trade-offs in decision-making.

Figure 4.5: Flow chart of the multi-agent simulation model with software use



5 Generation of landscapes and agent populations

5.1 Introduction[†]

Although multi-agent models have been widely applied in experimental and hypothetical settings, only few studies have tried to build empirically based multi-agent models and the literature on methods of empirical parameterization is therefore limited. This chapter develops and tests a novel methodology based on a Digital Elevation Model (DEM) to generate a landscape from soil samples, and Monte Carlo techniques to generate agent populations from a random sample of farm households. By estimating probability functions from a random sample of farm households and applying random seed values, the method is able to generate large numbers of agents and agent populations that are statistically consistent with the survey population. The alternative populations will later be used to estimate the variation in simulation results from differences in initial conditions.

The chapter is organized as follows. Section 5.2 describes how the landscape of soils and plots was created from a DEM and hand drawn maps of farmsteads, while Section 5.3 describes the creation of agent populations from survey data. The resulting agent populations are validated in Section 5.4. The chapter ends with a summary.

5.2 The landscape

The landscape comprises two village communities in the Lake Victoria Crescent of southeast Uganda (**Figure 5.1** and **Figure 5.2**). The two villages are Magada and

---

[†] Parts of this chapter were used for BERGER, T. and SCHREINEMACHERS, P. 2005. Creating agents and landscapes for multi-agent systems from random samples. *Ecology and Society*, forthcoming.

Buyemba. Both are located in the southern part of the Iganga District, which through redistricting has recently become the Mayuge District.

Figure 5.1: Africa with Uganda marked

Source: ESRI 2003

5.2.1 Data Sources

Data for generating the landscape come from two soil sample surveys. The first sample survey was conducted by Gerd Ruecker (ZEF) in Magada in 2000 (Ruecker 2005), while the second survey of 120 samples was carried out by the author in June 2003 in both Magada (80 samples) and Buyemba (40 samples).

5.2.2 The villages of Magada and Buyemba

The distance from Magada to Buyemba is about 12 km by road. Both villages are connected to the towns of Mayuga and Iganga. Iganga town is about 11 km from Magada and about 23 km from Buyemba. The smaller town of Mayuge is located in between the two villages: 5 km from Buyemba and 7 km from Magada (**Figure 5.2**).

The first farm families moved into Magada in the early 1920s. Settlers mostly followed the paths created by seasonally migrating buffalo. The first families settled at some distance from these paths for enhanced safety, with later arrivals settling more land inward or directly along the path. Buyemba was populated about one generation after Magada, in the 1950s.

Magada counts around 374 households, Buyemba 247. Magada is also more densely populated with 436 people/km^2 compared to Buyemba's 383 people/km^2. High population density has resulted in small average farm sizes, with average farm size being smaller in the more densely populated village of Magada. **Table 5.1** shows some basic statistics for both villages.

Figure 5.2: The Lake Victoria Crescent of southeast Uganda (1:1,000,000)

Note: The villages of Magada ('Magunga') and Buyemba ('Biemba') are encircled.

The Mayuge District is one of the more densely populated districts of Uganda, with a district average of 210 people/km^2 compared to an average population density of 85 people/km^2 for Uganda as a whole (UBOS 2002). It is noted that the population density of the two study villages is substantially above the district average. The already high population density of the district is set to further increase by 3.4 percent annually, according to the latest available population growth rate, estimated over the period 1980-91 (*ibid.*). This rate is above the national average of 2.6 percent per annum (*ibid.*).

Table 5.1: Basic statistics of Magada and Buyemba villages

	Magada village	Buyemba village	Both villages
Area [km^2]	6.75 km^2	5.15 km^2	11.90 km^2
Households in population (number)	373	235	608
Total population estimate (SE)	2888 (339)	1932 (320)	4820 (281)
Population density [persons/km2]	428	375	405
Arable land [ha] / person (SE)	0.39 (0.06)	0.27 (0.04)	0.34 (0.04)
Sampling fraction (households)	0.18	0.17	0.17

Source: Estimated from the 1999-2000 ZEF Survey

Note: Estimates are uncorrected for village boundaries (see text).

5.2.3 Landscape representation

The two villages can be seen as physical landscapes in which farm households reside and earn a living by cultivating crops, raising livestock, and performing off-farm employment. **Figure 5.3** gives a spatial representation of part of the Magada village. The figure shows a landscape, a hillslope in the case of Magada, overlaid by a grid of uniformly sized cells. A grid cell represents a piece of landscape of 71 x 71 meters (0.5 ha) that can be used for agricultural production. Agents' land endowments consist of a variable but discrete number of grid cells. For example, if an agent owns 2.73 ha of agricultural land then this is represented by five grid cells of 0.5 ha, 2.73 is hence rounded down to 2.5 ha. Farmsteads are distributed in the landscape and represent the agents' location. Agents own plots and choose which crops to grow and how much input to apply.

Information about each grid cell is organized in spatial layers; **Table 5.2** summarizes each layer. Layers 1-2 define every agent's location in terms of farmstead and plots, layers 3-5 define the socioeconomic characteristics, layers 6-10 the soil chemical properties, and layers 11-12 define the soil physical properties.

Figure 5.3: Spatial representation of the landscape component

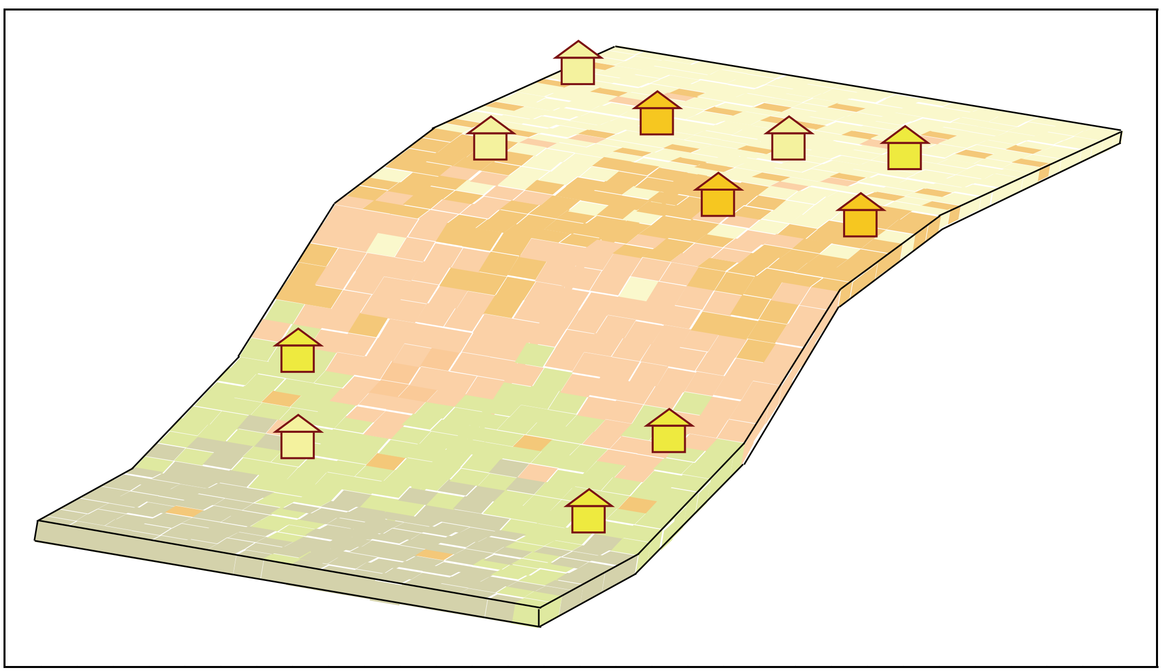


Table 5.2: Twelve spatial layers representing the landscape

Nr.	Layer	Unit	Note
1	Farmsteads	#	The location of farmsteads (agent identification)
2	Plots	#	Ownership over agricultural plots
3	Villages	#	Demarcation of village boundaries
4	Network thresholds	#	Determines the access to innovations
5	Cluster	#	For random generation of socioeconomic characteristics
6	Nitrogen	%	Total nitrogen
7	Phosphorus	ppm	Total phosphorus
8	Potassium	Mg/100 gr.	-
9	Soil organic matter	%	-
10	Acidity	pH	-
11	Slope length	m	-
12	Soil erodibility factors	-	-

5.2.4 Location of agents and farm plots (layers 1-3)

Previous surveys had not recorded the location (latitude/longitude) of farmsteads or plots. Only hand drawn maps were available on which most farmsteads in the random sample had been marked. Based on these maps, about 17 percent of the farmsteads in Magada and Buyemba could be located. Furthermore, using the sampling weights of the survey, the total number of farmsteads and plots in each village could be estimated. The farmsteads and plots with unknown locations were randomly generated using the following procedure.

Figure 5.4 shows the different stages in generating the spatially located farmsteads and farm plots for the example of Magada, with the same procedure applied to Buyemba. The left upper panel (**Figure 5.4 A**) shows the sample points within the village boundary of Magada. Socioeconomic data are available for each of these farm households. The figure shows that the sample farm households are not evenly distributed in the landscape but are clustered around the road network. The figure only shows main roads, but it should be noted that the cluster of dots in the north and center of the village are also close to smaller roads. A random allocation of all remaining agents was used to complete this layer (using ArcView GIS). Yet, a fully random spatial allocation of all remaining agents would be incorrect, as the distribution of sample points does not appear to be random. Two different areas according to population density were therefore first demarcated: the area alongside the road network was designated as of high population density while the remaining areas were of low population density. Since survey households were randomly selected, the geographical distribution of the sample households represents the distribution of the total population. In Magada, for instance, 84 percent of the sample households live in the high-density area, which accounts for 40 percent of the total village area. Of the remaining (non-sample) households, 84 percent were thus allocated in the high-density area and 16 percent in the low-density area (**Figure 5.4 B)**.

This methodology created a problem in the village of Magada in that the estimated area from the survey (see **Section 5.3**) much exceeded the available area within the village boundary, though no such problem arose for Buyemba. The likely explanation is that many farmsteads in Magada, unlike in Buyemba, are located close to the village boundary in the north of the village: these farm households might have plots outsides of the village boundary, which are included in the survey. The data were therefore adjusted by randomly deleting 95 (non-sample) agents from the population so that the estimated and available areas matched. **Figure 5.4 C** shows the agent population after this adjustment. All allocated farmsteads were then converted into grid cells.

Finally, using the estimated sample distribution from the survey, agricultural plots were allocated to agents. The spatial randomizer was not used at this stage, as this would have produced an unrealistically scattered pattern of farm plots. The allocation was therefore done manually (**Figure 5.4 D)**. **Table 5.3** summarizes some characteristics of the generated agent population.

Table 5.3: Characteristics of the agent population

	Magada	Buyemba	Both villages
Agricultural plots (number)	1,337	921	2258
Households	278	242	520
Plots / household	4.8 (2.4 ha)	3.8 (1.9 ha)	4.3 (2.2 ha)

Source: see text.

Figure 5.4: Spatial generation of agent population and agricultural plots from a random sample of farm households

Notes: A. survey sampling points; B. division in areas with high and low population density; C. location of all agents and conversion into grid cells; and D. the distribution of agricultural plots.

5.2.5 The socioeconomic landscape (layers 4-5)

Socioeconomic layers contain information about the agents' resource distributions and their membership to network threshold groups; both were estimated from farm household survey data. The land area cultivated by households appeared to be correlated with household size. Because of this and other correlations, a complete random generation would be unsuitable. Additional constraints were added to the random generation by separating agents into clusters as will be explained in Section 5.3.

5.2.6 Soil chemical properties (layers 6-10)

Soil chemical properties were derived from predictive soil maps for the area. These maps were estimated by a team of researchers, including Hosangh Rhew (University of Seoul), Soojin Park (University of Seoul/ZEF), and Gerd Ruecker (ZEF/German Aerospace Center, DLR).

Soil properties (nitrogen (N), phosphorus (P), potassium (K), organic matter, acidity (pH), and texture) were predicted from terrain parameters using a Digital Elevation Model and multiple regression analyses in which each soil nutrient was regressed on terrain parameters and other soil properties. Terrain parameters included elevation, slope, upslope area, curvature, plan curvature, profile curvature, wetness index, streampower index, and aspect (Rhew *et al.* 2004).

Table 5.4: Predictive soil models

Soil property	Estimated model[1,2]
N	$-1.22 + 0.039(\text{ELEV})^{0.5} - 0.087(\text{SLOP})^{0.5}$
K	$(-21.17 + 0.748(\text{ELEV})^{0.5} - 0.360\text{Log}(\text{STRI}))^{2}$
pH	$6.47 - 1.32(\text{SLOP})^{0.5}$
Organic Matter	$(-8.39+0.30(\text{ELEV})^{0.5} - 0.79(\text{SLOP})^{0.5})^{2}$
Clay	$-394.19 + 12.2(\text{ELEV})^{0.5} - 3.12\log(\text{STRI})$
Sand	$536.32 - 13.7(\text{ELEV})^{0.5} + 23.1(\text{SLOP})^{0.5}$
Silt	$100 - (\text{Sand} + \text{Clay})$
Na	$\text{EXP}(1.90 + 5.73*(\text{WETI})^{-0.5}$
Ca	$-951.69 + 30.69(\text{ELEV})^{0.5} - 104.69(\text{SLOP})^{0.5}$
P	$(-1.15 + 0.003(\text{Ca}) + 0.21\text{Log}(\text{Na}) + 0.13(\text{K})^{0.5} - 0.02(\text{Clay}))^{-2}$
Total P	$(0.03(\text{Silt})^{0.5} + 0.11(\text{OM})^{0.5} - 0.677(\text{N}) + 0.04)^{2}$

Source: Rhew *et al.* 2004

Notes: 1) All estimates based on a 30m grid size. 2) ELEV=elevation; SLOP=slope; WETI=wetness index; STRI=streampower index; OM=organic matter; Na=natrium; and Ca=Calcium.

In a first stage, the prediction was based on 285 soil samples and 910 GPS measurements from a single hillslope in the village of Magada collected in 2000 (Ruecker 2005). Predictive models were initially estimated from these data and scaling effects were explored to find robust estimators at different scales.

In a second stage, a new round of 120 soil samples and GPS measurements were collected in two villages in 2003. These data were used to validate the predictive models of the first stage, and to subsequently modify these (Rhew *et al.* 2004). In addition, these data were used to establish pedotransfer functions for phosphorus and total phosphorus in the soil.

The explanatory power of the estimated models varied (Rhew *et al.* 2004). Although the model fit for soil texture (clay and sand) and K was relatively good, it did not fit well for organic matter, acidity, and N. What this exercise showed was that terrain parameters alone are not enough to explain the variation in soil properties across the landscape. Land management and land-use history, for example, are likely to be more important in determining soil properties. Spatial data about these factors were not collected in the survey and would be a useful addition to a future exercise.

5.2.7 Soil physical properties (layers 11-12)

Soil physical properties were included to capture soil erosion. A relatively simple approach for modeling soil erosion was used based on the Universal Soil Loss Equation (USLE) (Wischmeier and Smith 1978). Two factors were included in the model: slope length factor and soil erodibility (**Chapter 6**).

Soil physical properties were derived from the DEM developed by Rhew *et al.* (2004) using the hydrologic and spatial analyst extension in AcrView GIS. Slope length was calculated following Moore and Burch (1986) as:

(9) $$\text{Slope Length} = ([\text{Flow Accumulation}] * [\text{Cell size}]/22.13)^{0.4}$$

Flow accumulation was expressed as a number of grid cells derived from the watershed delineation using the hydrologic and spatial analyst extension to ArcView GIS. The cell size was set to 71 meters. A layer of slope length was created using the map calculator tool in ArcView GIS.

The slope length factor was estimated from both slope and slope length following Wischmeier and Smith (1978). **Table 5.5** shows these factors for different combinations of these two variables. A constant factor of 0.40 was assumed for soil erodibility by lack of more precise data.

Table 5.5: Slope length factors

	Slope length (meters)			
Slope (%)	0 – 7.5	7.5 – 12.5	12.5 – 15	> 15
0 – 1.5	0.00	0.00	0.00	0.10
1.5 – 3.5	0.13	0.15	0.17	0.17
> 3.5	0.22	0.30	0.38	0.38

Source: Based on Wischmeier and Smith (1978)

5.3 The agents

A methodology to randomly generate agent populations from a random sample of farm households was developed based on Monte Carlo techniques. The objective was to generate a multitude of potential agent populations, with all agents being different both within a single population and between different populations.

5.3.1 Data Sources

Data came from a farm household survey conducted in the villages of Magada and Buyemba by Johannes Woelcke in 1999-2000 (Woelcke 2004, 2006). He used a stratified random sample to select 44 farm households participating in agricultural trials of the International Center for Tropical Agriculture (CIAT) and the Africa 2000 Network (A2N), and 62 randomly selected non-trial farm households. Probability weights were used in the analysis to correct for the over-representation of trial farm households in the sample. These probability weights were calculated as the inverse of the probability of a farm household being selected in the random sample.

5.3.2 Generating an agent population

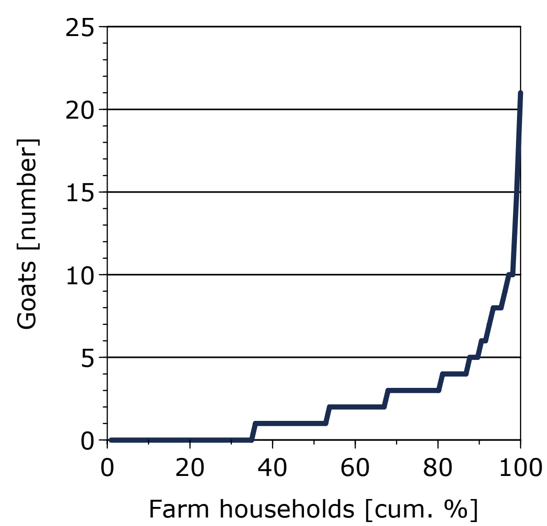
One of the challenges in generating an empirically based agent population was to represent each real-world farm household with a unique agent. Yet, data were available only for a sample of 17 percent of the farm households. The challenge hence became how to extrapolate from the sample population to generate the remaining 83 percent.

The most obvious route would be to multiply the sample farm households with their probability weights. Average values in this agent population would exactly match those of the sample survey. Yet, this copy-and-paste procedure is unsatisfactory for several reasons. First, it reduces the variability in the population, as a sampling fraction of 17 percent would imply about five identical agents, or clones. This might affect the simulated system dynamics, as these agents are likely to behave analogously. It then becomes difficult to interpret, for instance, a structural break in simulation outcomes; *e.g.*, is the structural break endogenous, caused by agents breaking with their path dependency, or is the break simply a computational artifact resulting from the fact that many agents are the same? This setback becomes more serious for smaller sampling fractions, because a larger share of the agents is then identical. Second, the random sample might not well represent the population. The sample size is small and the sampling error is unknown but can be large. When using the copy-and-paste procedure, only a single agent population can be created, while for sensitivity analyses a multitude of potential agent populations would be desired. For these reasons, the procedure for generating agent populations is automated using random seed numbers to generate a whole collection of possible agent populations.

5.3.3 Random data generation

Monte Carlo studies are generally used to test the properties of estimates based on small samples. It is suitable for this study where data about a relatively small sample of farm households is available but the interest goes to the properties of an entire population. The first stage in a Monte Carlo study is modeling the data generating process, and the second stage is the creation of artificial sets of data.

Figure 5.5: Empirical cumulative distribution of goats over all households in the sample



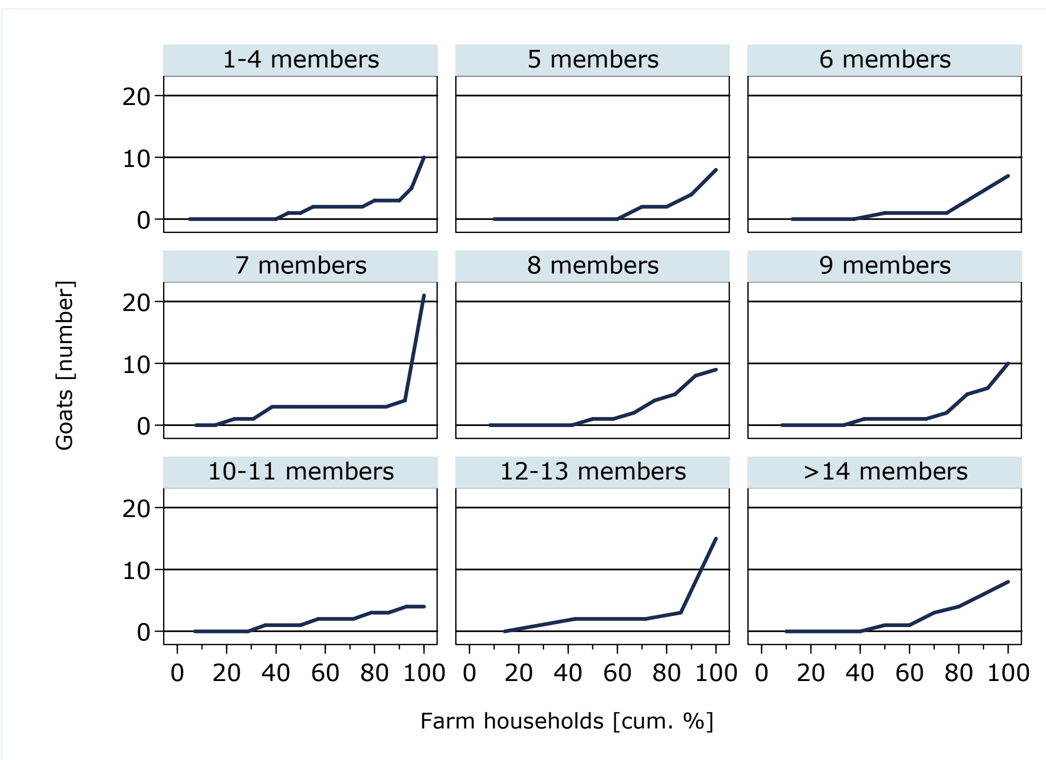
The methodology is based on the use of empirical cumulative distribution functions. **Figure 5.5** illustrates such a function for the distribution of goats. The figure shows that 35 percent of the farm households in the sample have no goats, the following 8 percent have one goat, etc. This function can be used to randomly generate the endowment of goats, and all other resources, in an agent population. For this, a random integer between 0 and 100 is drawn for each agent and the number of goats is then read from the y-axis. Repeating this procedure many times recreates the depicted empirical distribution function.

Each resource can be allocated using this procedure. Yet, each resource would then be allocated independently, excluding the event of possible correlations between resources. Actual resource endowments typically correlate, for example, larger households have more livestock as well as more land. To include these correlations in the agent populations, first the resources that most strongly correlate with all other resources are identified and used to organize the survey population into a number of clusters. Cumulative distribution functions are then calculated for each cluster of sample observations.

The sample was divided into clusters defined by household size because this was the variable most strongly correlated with all other variables. Cluster analysis can also be used for this purpose, but clusters produced under this procedure are more difficult to interpret, especially when using many variables. Nine clusters were chosen, as this number captured most of the different household sizes and allocated at least five observations to each cluster.

Each agent was allocated quantities of up to 80 different resources in the random procedure. These resources included 68 different categories of household members (34 age and two sex groups), 4 types of livestock (goats, young rams, cows, young bulls), an area under coffee plantation, female head, liquidity, leverage, and innovativeness. Agents were generated sequentially, that is, 80 random numbers in 80 different cumulative distribution functions were first drawn for agent No.1 before the same was done for agent No.2.

Figure 5.6: Cumulative distribution functions of goats over households per cluster



As most resources only came in discrete units, a piecewise linear segmentation was used to implement the distribution functions. Five segments were chosen, as this captured most resource levels; more segments would have been needed if the number of resource levels per cluster had been larger than five or if many resources had continuous distribution functions.

The above is illustrated in **Figure 5.6** and **Table 5.6** for the random allocation of goats to the agents. **Figure 5.6** plots the cumulative distribution functions for each of the nine clusters defined by household size. **Table 5.6** shows the piecewise linear segmentation of these functions in five segments for each of the nine clusters, the structure of this table also reflects how the data entered into the software.

Table 5.6: Piecewise segmented cumulative distribution function

Example of cumulative distribution of goats over nine clusters

	Segment									
	1		2		3		4		5	
cluster	cum.%	goats	cum.%	goats	cum.%	goats	cum.%	goats	cum.%	goats
1.	36.4	0	55.5	1	77.8	2	97.9	3	100.0	6
2.	70.5	0	95.1	2	97.6	3	100.0	6	0.0	0
3.	28.9	0	65.4	1	82.7	3	100.0	5	0.0	0
4.	3.1	0	29.1	1	98.6	3	100.0	15	0.0	0
5.	45.6	0	69.3	1	71.7	2	95.3	4	100.0	7
6.	34.4	0	75.4	1	93.9	2	95.9	4	100.0	6
7.	33.8	0	63.8	1	68.5	2	100.0	3	0.0	0
8.	0.6	0	24.1	1	94.8	2	97.4	3	100.0	11
9.	26.2	0	50.8	1	95.1	3	97.6	5	100.0	6

Source: Estimated from the 1999-2000 ZEF Survey

5.3.4 Consistency checks

In order to get realistic agents, four procedures were applied in addition to the cluster-specific cumulative distribution functions for each resource:

First, checks for inconsistencies at the agent-level. An agent with 20 household members is very unlikely to have only one plot of land. Yet, because of the purposeful randomness of the resource allocation, unrealistic settings can occur in the agent population. By defining a lower and/or upper bound for some combinations, this problem was overcome. If a resource combination lay outside such bound the generated agent was rejected, and the procedure was repeated. Two sets of bounds were included. The first set defined minimum land requirements for livestock and the second set defined demographic rules to ensure realistic family compositions (**Table 5.7**).

Table 5.7: Imposed rules on the agent populations

Nr.	Rule	Min.	Max.	Unit
1.	Cow	0.50	-	ha
2.	Young bull	0.33	-	ha
3.	Goat	0.10	-	ha
4.	Young ram	0.10	-	ha
5.	Household size	1	24	number
6.	Adults	1	13	number
7.	Children	0	16	number
8.	Male adults	0	6	number
9.	Female adults	0	7	number
10.	Male children	0	9	number
11.	Female children	0	10	number
12.	Children / adults	0	5	ratio
13.	Adults / children	0.2	4	ratio
14.	Females / males	0	8	ratio
15.	Males / females	0.125	5	ratio
16.	Female adults / male adults	0	4	ratio
17.	Male adults / female adults	0	3	ratio

Source: Estimated from the 1999-2000 ZEF Survey

Second, checks for inconsistencies at the cluster level. The generated mean resource endowments had to lie within the confidence interval of the estimated sample mean; and the correlation matrix of the agent population had to reflect the correlation matrix of the sample population. If not, the generated agent population was rejected and the procedure repeated.

Third, checks for inconsistencies at the population level. The mean resource endowments of the agent population had to lie within the confidence intervals of each estimated sample mean, and again, the procedure was automatically repeated if this was not the case.

Fourth, if individual agents, clusters of agents, or entire agent populations were continuously being rejected on one of the above criteria, then the cluster-specific distribution functions were fine-tuned.

5.4 Validation of results

To test the methodology, a large number of agent populations were generated by applying different random seed values. Their properties are analyzed at three levels: (1) the population level; (2) the cluster level; and (3) the level of the individual agents. Each of these is discussed in the following. It was not attempted to show the entire variation between and within agent populations. An attempt to do so would have the following pages covered with tables. Instead, the results are illustrated with a few examples and snap shots of alternative agent populations.

5.4.1 Population level

At the population level, it was checked whether the averages in the agent population resembled those of the survey population. For this, average resource allocations for hundred agent populations were calculated (**Table 5.8**). For all resources, average endowments in the agent population fell within the confidence interval of the survey average and the difference between the two averages was generally small. The random agent generator was hence well able to reproduce population averages.

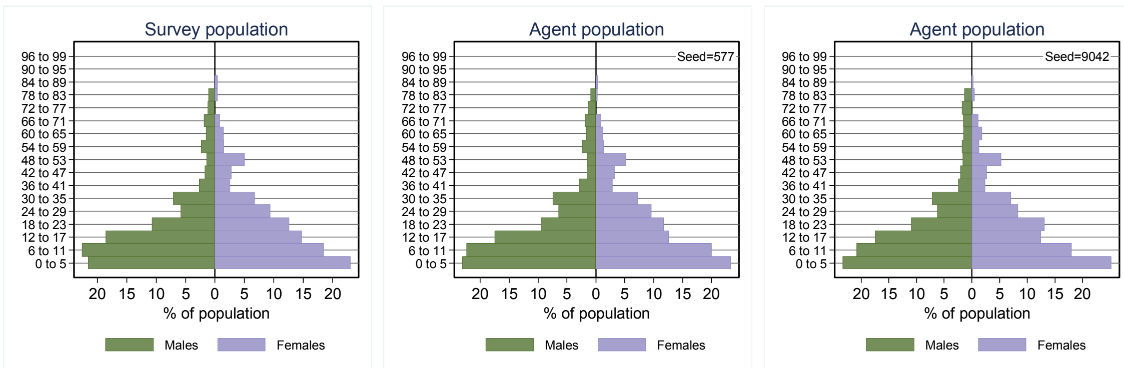
Table 5.8: Resource endowments of the survey population compared to meta-averages of the agent population

Resource	Population	Average	SE/ SD[1]	Confidence interval	
Household members	Survey	7.87	0.45	6.99	8.75
	Agent	7.89	0.11		
% children	Survey	55.06	2.47	50.22	59.91
	Agent	54.87	0.75		
Cows	Survey	0.81	0.18	0.45	1.17
	Agent	0.81	0.02		
Young bulls	Survey	0.08	0.04	0.01	0.16
	Agent	0.09	0.01		
Goats	Survey	1.29	0.16	0.98	1.61
	Agent	1.23	0.04		
Young rams	Survey	0.14	0.04	0.06	0.23
	Agent	0.14	0.02		
Coffee, ha	Survey	0.31	0.10	0.11	0.51
	Agent	0.31	0.02		
Plots, 0.5 ha	Survey	4.58	0.51	3.58	5.58
	Agent	4.34	0.00		
Innovativeness	Survey	3.88	0.17	2.35	3.03
	Agent	3.85	0.04		

Sources: Estimated from the 1999-2000 ZEF Survey and simulated with MAS
Notes: The shown agent population is an average over 100 alternative populations. [1]SE is the standard error of the average survey population; SD is the standard deviation of the average across 100 agent populations.

To get more details about the demographic structure of the population, a population pyramid for the survey population was calculated and compared with pyramids for two agent populations in **Figure 5.7**. This form of presentation may not be very suitable for comparing age groups exactly but does illustrate the similarity between survey and agent populations. The random agent generator hence creates demographic structures that are realistic.

Figure 5.7: Population pyramids comparing survey estimate with agent populations

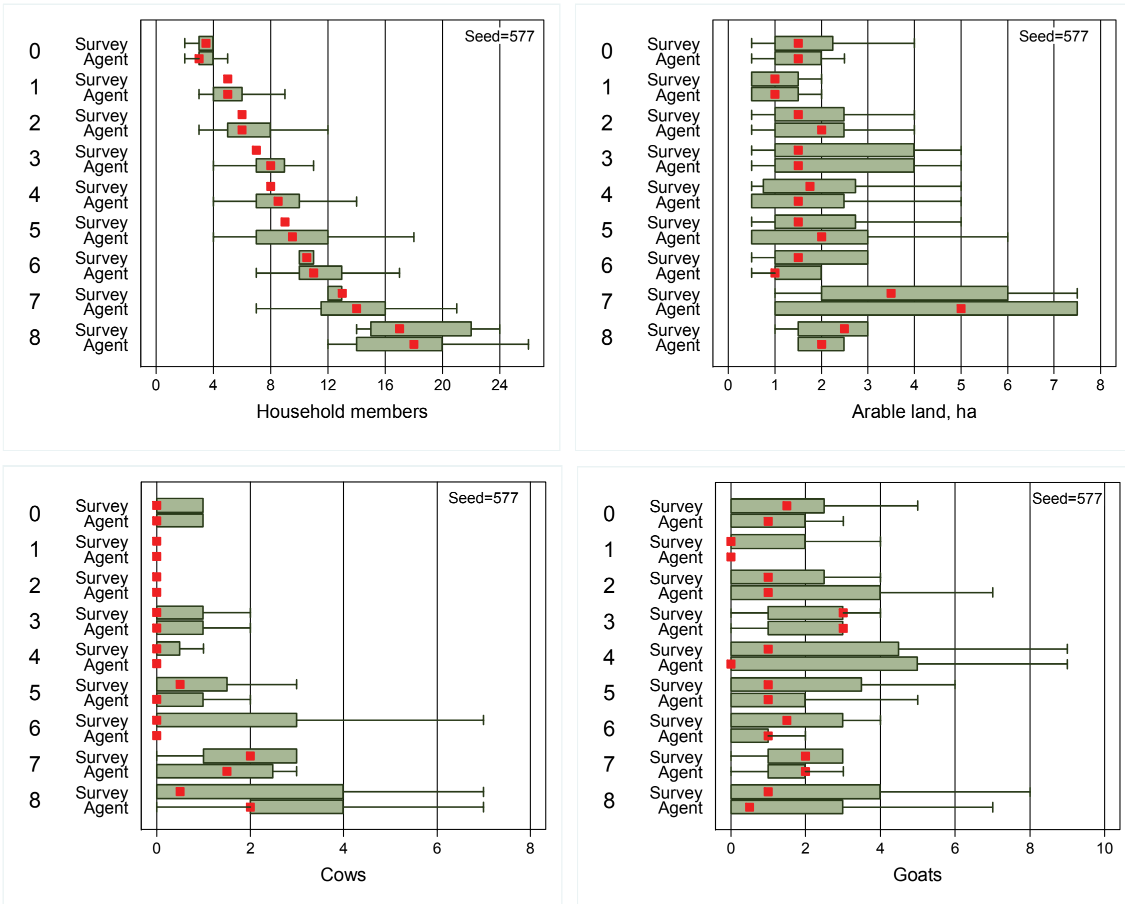


5.4.2 Cluster level

The above has shown that the survey population was well replicated at the aggregate level, which may not necessarily be so at lower levels of aggregation. The following graphs and figures thus look at the cluster level.

Figure 5.8 depicts four box plots comparing the distribution of household size, area under coffee, goats, and cows in the sample with an agent population with random seed value 577. Each box ranges from the 25th to the 75th percentile (the interquartile range) with the median also marked in it. Clusters are based on household size, which is why there is a strong correlation between the cluster number and the median household size in the left upper diagram. The figure shows that median values do not differ much between the survey population and the agent population. Most interquartile ranges are of comparable width.

Figure 5.8: Boxplots for the distribution of the four major resources over clusters

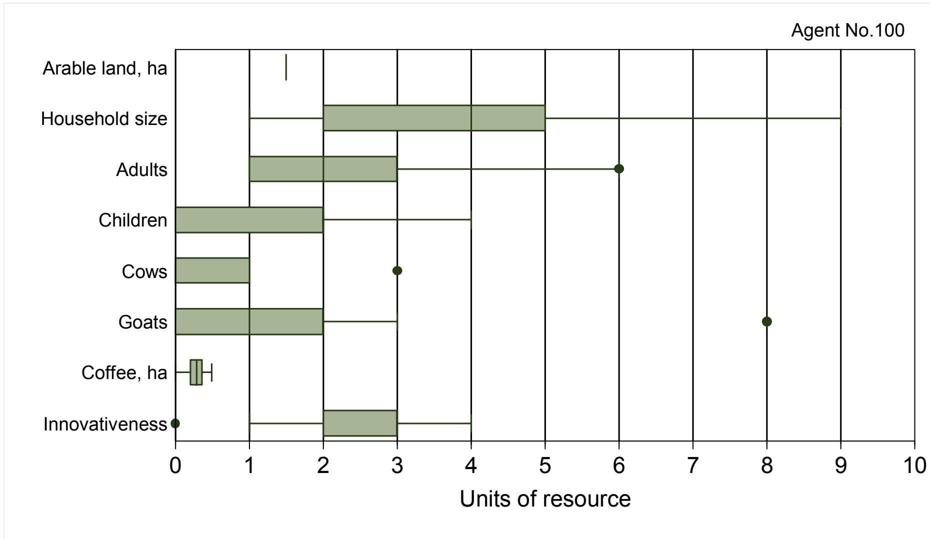


Note: Graphs show minimum, maximum, interquartile range (bar), and median value (squares).

5.4.3 Agent level

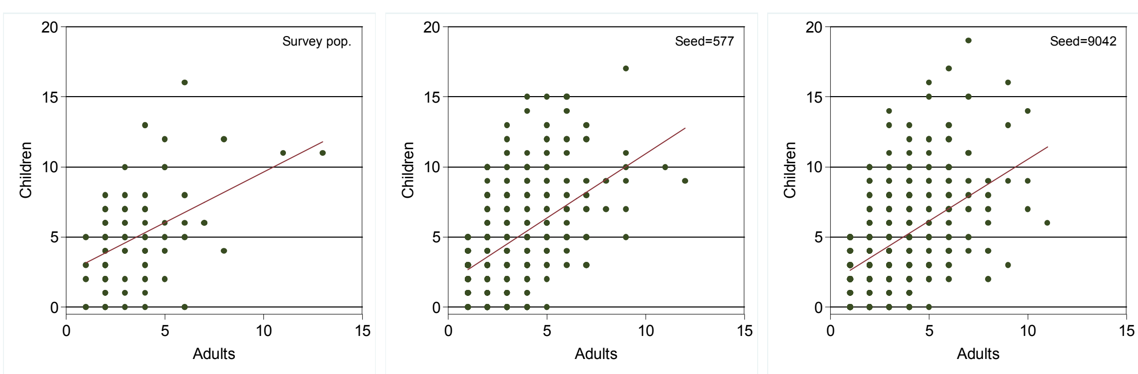
One objective of generating agents randomly was to endow each agent differently in alternative agent populations. The success of the approach is illustrated with **Figure 5.9**. This boxplot shows the variation in resource endowments for agent No.100 in hundred generations of alternative populations. Agent No.100 has a fixed location for farmstead and plots in the landscape as can be seen from the zero variance in the agent's land area of 1.5 ha. The variation in resource endowments is high with, for instance, the household size varying between 1 and 9 members and the number of goats varying between 0 and 3.

Figure 5.9: Boxplot illustrating the variation in agent endowments in alternative agent populations



The reproduction of correlations is the third objective in the random agent generation. The left diagram in **Figure 5.10** plots the number of adults against the number of children in the survey population, while the two right diagrams do the same for two generated agent populations. The figures show that correlation between adults and children within the household is well replicated in the agent populations, ensuring that the agents are demographically consistent in this respect.

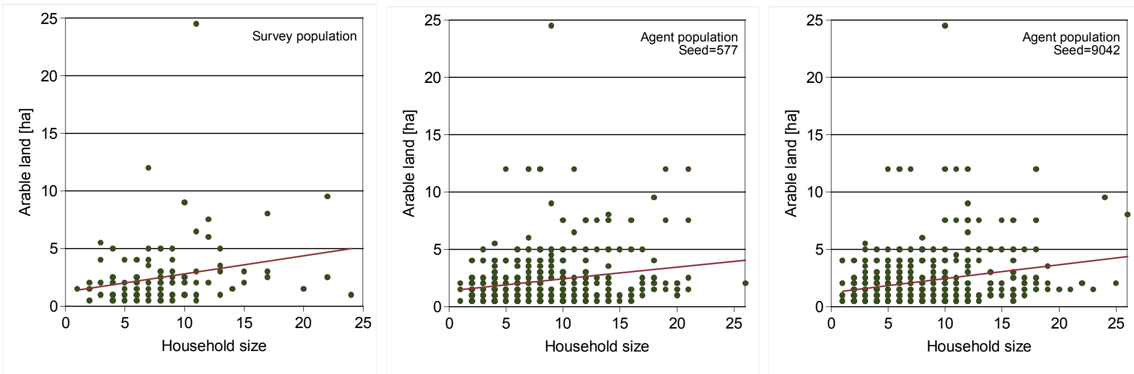
Figure 5.10: Scatter plots correlating the number of children and adults



Note: Scatter plots including a linear regression fit.

In addition, **Figure 5.11** plots household size against the number of plots per agent and shows that the observed positive correlation between household size and land size reappears in the agent populations.

Figure 5.11: Correlation between household size and amount of arable land



Note: Scatter plots including a linear regression fit.

5.5 Summary

A Monte Carlo technique based on cumulative distribution functions and random seed values can successfully generate sets of artificial agent populations from farm household survey data. The combination of Monte Carlo studies and predictive soil maps estimated from a Digital Elevation Model can be applied for generating agent populations as was shown for two village communities in southeast Uganda. The generated agents are statistically consistent with the survey population at aggregate, cluster, and agent levels.

6 Crop yield and soil property dynamics

6.1 Introduction‡

This chapter specifies and calibrates the model component simulating crop yields and soil property dynamics. For this, the Tropical Soil Productivity Calculator (TSPC) was calibrated to agriculture in the Lake Victoria region in southeast Uganda. The chapter begins in Section 6.2 with some relevant background information on the issue of soil fertility decline. It then continues in Section 6.3 with a specification of the model. Parameter values are given in Section 6.4 and the component is validated in Section 6.5. The chapter ends with a summary in Section 6.6.

6.2 Background

6.2.1 Problem background

Most of the soils in Uganda are old, highly weathered, and of relatively low fertility (Mubiru and Ssali 2002). In spite of this, farmers use only small quantities of manure, and mineral fertilizer use is less then 1 kg/ha on average (*ibid.*). In the absence of time series data on soil fertility, the nutrient dynamics of an agricultural system can be evaluated by studying current soil nutrient flows. In one of the first of such approaches in Uganda, Wortmann and Kaizzi (1998) found negative nutrient balances for nitrogen (N), phosphorus (P), and potassium (K) for all crops except banana in central and eastern Uganda. Ssali and Vlek (2002) confirmed this finding from a comparison of soil samples at selected sites in the Banana-Coffee Lakeshore farming system at two points in time. They found a general decline in crop nutrients (P, K, and calcium) and a substantial decline in soil pH and bases (Ca and K), despite little change in soil organic matter. Many studies in other parts of Uganda as well as

‡ This chapter is based on: SCHREINEMACHERS, P; J.B. AUNE and T. BERGER. 2004. An Application of the Tropical Soils Productivity Calculator to the Study of Soil Fertility Decline in Uganda. *ZEF Documentation of Research 3/2004*.

comparable studies in neighboring Kenya also show net nutrient outflows (**Table 6.1**).

Despite a fair amount of research on soil fertility decline, relatively little is known how this affects crop yields and even less is known how it affects farm household incomes and food security. Many studies have assumed, that soil fertility decline is behind the decline in crop yields and a worsening of farm household incomes and food security, yet few studies were able to show this empirically.

Table 6.1: Nutrient balances in Sub-Saharan Africa

Location	Balance (kg/ha/year)			Source
	N	P	K	
Uganda, Mukono district	-49.0	-13.3	-17.3	Aniku *et al.* 2001
Uganda, Arua district	-33.4	-6.0	-7.3	Aniku *et al.* 2001
Uganda, Palissa district	-21.2	-8.2	-43.0	Wortmann & Kaizzi 1998
Uganda, Kamuli, Iganga & Mpigi districts	-30.6	-4.0	-38.9	Wortmann & Kaizzi 1998
Magada village, Mayuge District	-108	-14	-94	Kaizzi *et al.* 2003
Eastern Uganda (average 8 villages)	-83	-10	-60	Kaizzi *et al.* 2003
Western Kenya	-76.0	-3.8	-	Shepherd *et al.* 1995, 1996
Kenya, Machakos District	-53.0	1.0	-9.0	De Jager *et al.* 2001
Sub-Saharan Africa (total)	-22.0	-2.5	-15.0	Stoorvogel *et al.* 1993

6.2.2 Theoretical background

There are two basic approaches to estimating crop yield. One is crop growth modeling, which involves simulating different phases of plant growth from germination to reproduction. This approach is data intensive as it considers many factors influencing plant growth, often at a daily time scale, and which are furthermore crop variety and location specific. The other approach is empirical yield modeling, which has lower data requirements as it uses only data on growth determining factors and the resulting level of crop yield to statistically estimate empirical yield response functions. This study uses this second approach for the practical reasons that crop yield functions are needed for all crops in the system and data to calibrate crop growth models are unavailable.

There are competing theories of how crop yields respond to nutrients and other soil properties. Two of the most accepted theories are associated with the names of Von Liebig and Mitscherlich. Both assume a diminishing growth of yield to increasing quantities of soil nutrients, which is constrained by a yield potential. The theories, however, differ in the functional form of the yield response. The theory initially proposed by Von Liebig holds that factors are generally complementary; each factor has a unique role in crop growth, which cannot be substituted. Increasing levels of one factor increases crop yield linearly until another factor becomes limited, at which point the yield level plateaus. The plateau can be raised only from an increase in the most limiting factor. Mitscherlich also accepted that factors act in a complementary way but allowed for a possible interaction between factors. This interaction implies that an increase in one factor positively affects the response to all other factors. In other words, additions of each factor have a positive impact on yield but this impact is moderated by the availability of all other factors.

6.2.3 The Tropical Soil Productivity Calculator (TSPC)

This study uses the Tropical Soil Productivity Calculator (TSPC) to estimate yields. The TSPC was developed by J.B Aune and R. Lal (Aune and Lal 1995, 1997). According to its developers, the reasoning behind the design of the TSPC was to create a model that is relatively "simple to use but comprehensive enough to account for main factors that influence crop growth and yield in the tropics" (Aune and Lal 1995: 502). The TSPC was specifically designed for tropical soils and includes a factor for the effect of acidity on yield. Acidity is often a limiting factor in highly weathered tropical soils, but is commonly omitted in crop models designed for temperate regions where acidity levels do not commonly constrain crop yield *(ibid.)*. For the purpose of this study, Jens Aune (Norwegian University of Life Sciences) specified and calibrated the TSPC to Ugandan agro-ecological conditions.

Theoretically, the TSPC is based on the theory of Mitscherlich as it assumes factors to be complementary and that yields plateau if one or more of these factors is in limited supply. In this study, the (non-linear) crop yield equation is specified as:

$$Y = P*(F_{LAB})*(F_{NAV})*(F_{PAV})*(F_{KAV})*(F_{SOC})*(F_{pH}) \quad (6\text{-}1)$$

with Y denoting crop yield and P the crop yield potential, the six variables within brackets are reduction factors: management (F_{LAB}), available N in the soil (F_{NAV}), available P (F_{PAV}), available K (F_{KAV}), soil organic carbon (F_{SOC}), and acidity (F_{pH}). All factors but one were specified as logarithmic functions while a quadratic function was used soil organic carbon (C). Factors range from zero (crop yield completely constrained) to unity (not constrained).

6.3 Four phases in soil property dynamics

Soil properties change as a result of harvesting, as well as various other endogenous and exogenous processes at the plot, farm, and catchment level. These soil property dynamics influence future yields. The processes underlying these dynamics are modeled with an extended and dynamic version of the TSPC (Aune and Massawe 1998).

In this section, the crop-soil processes are formulated in mathematical terms. These processes are divided into four phases:

- phase 1: calculation of yield determining factors;
- phase 2: calculation of crop and residue yields;
- phase 3: calculation of soil property dynamics; and
- phase 4: evaluation of these dynamics in balance equations.

In the following equations, the subscripts *p*, *t*, *c*, and *f* denote plot, time, crop type, and fertilizer type respectively.

Multiple crops can be grown simultaneously on a single plot. Soil properties affecting present crop yields not only depend on current input use but also on past crop choice and past levels of input use. Keeping track of the land use history at the subplot level would be overly complicated. The updating of soil property dynamics was therefore simplified by averaging all crop-specific calculations to the plot level at the end of each period. Acreage weights were used for this purpose, which were calculated as the proportion of the plot a crop occupied. For example, if the plot was cultivated with maize in the first season and beans in the second, then the acreage weight of both crops would be 0.5. This means that at the beginning of each period, each plot had homogeneous soil properties, which is only justifiable for small plot sizes; the plot size was therefore set to 0.5 ha in with farm households having four such plots on average.

6.3.1 Phase 1: yield determinants

Phase 1 computes the available amount of N (NAV_{pct}), carbon (SOC_{pct}), P (PAV_{pct}), and K (KAV_{pct}), and the level of acidity (APH_{pct}). Yield determinants were calculated at the crop level because input decisions are crop specific. The following subsections describe how each determinant was calculated.

6.3.1.1 Available nitrogen

The amount of available N in mineral form in year *t* and plot *p* (NAV_{pct}) is a function of four variables:

- atmospheric deposition of N in the current period (ATN_{pt});
- N mineralized from the total stock of soil organic N (MIN_{pt});
- mineralized N from manure, residues, and roots from previous periods (RNM_{pt});
- N from mineral fertilizer applied in the current period (FER_{pfct}), or:

(6-2) $$NAV_{pct} = 0.5(ATN_{pt} + MIN_{pt} + RNM_{pt}) + \sum_f (FER_{pfct} \frac{mf\_n_f}{100})$$

in which mf_n_f is the share of N in mineral fertilizer. The contribution of natural sources of N was divided in half to capture their reduced efficacy as compared to mineral fertilizer, which can be applied when plants need nutrients the most. The variable ATN_{pt} was calculated from the annual amount of rainfall in mm (rt_t):

(6-3) $$ATN_{pt} = 0.14\sqrt{rt_t}$$

MIN_{pt} was calculated from the soil's total stock of organic N in year *t* (SON_{pt}, in kg/ha) and a mineralization constant (dc_p):

(6-4) $$MIN_{pt} = SON_{pt}(1 - e^{-dc_p})$$

with SON_{pt} itself a balance of the soil organic N stock in the previous period plus changes in the stock resulting from remaining residues in the previous period, plus N stored in the crops' roots in the previous period (RTN_{pt-1}) minus what is lost from N erosion (NER_{pt-1}):

(6-5) $$SON_{pt} = (SON_{pt-1})\, e^{-dc_p} + 0.15\,(\frac{RES_{pt-1}}{100} * STN_{pt-1} + \frac{of\_n}{100} * MNR_{pt}) + 0.35(RTN_{t-1}) - NER_{pt-1}$$

In which RES_{pt} is the percentage of residues remaining after harvest, STN_{pct} is the total quantity of N (in kg) stored on one hectare of residues and the factor 0.15 is the proportion of residue N and manure entering into soil organic N. MNR stands for the quantity of manure and *of_n* for the percentage of N in manure. MNR was calculated as the weighted sum over three previous periods to adjust for mineralization over several years (Jenkinson and Ayanaba 1977):

(6-6) $$MNR_{pt} = 0.5\,MAN_{pt-1} + 0.3\,MAN_{pt-2} + 0.2\,MAN_{pt-3}$$

Unlike mineral fertilizer, the application of manure is not crop-specific (and therefore has no subscript *c*). This is because manure management is uncommon in the research area, though extension services promote the practice. A uniform distribution of manure over all plots and all crops was therefore assumed.

Total dry matter manure production (MAN_t) is endogenous in the model since the size of the animal herd is a management decision (**Chapter 7**). Manure production was specified as a function of feed intake by assuming that an average kilogram of dry matter feed contains 9 million joule of energy and has a digestibility of 50 percent (LEAD 1999):

$$(6\text{-}7) \qquad MAN_t = 0.5\sum_n \sum_m (IME_{nmt}/9)$$

In which subscripts *n* and *m* denote animal specie and age respectively and the variable IME_{nmt} is the total metabolizable energy need of the entire livestock at time *t* and will be specified in equation (6-22) below.

Residues and manure not only contribute to the stock of organic N but also mineralize directly and add to the available N in mineral from. This was captured by the factor RNM_{pt} in equation (6-2). RNM_{pt} was calculated from the remaining quantities of residues on the plot in three previous years:

$$(6\text{-}8) \qquad RNM_{pt} = \sum_{t=-1}^{t=-3}(0.75 * 0.33 * \frac{RES_{pt}}{100} * \sum_c STN_{pct}) + MNR_{pt} * \frac{of\_n}{100}$$

Livestock in the research area mostly graze on the farm with little off-farm grazing (Wortmann and Kaizzi 1998). The potential of livestock to transport nutrients from grazing areas into the farming area is therefore small. Yet, the metabolizable energy need of cattle was reduced by 25 percent to capture off-farm grazing.

6.3.1.2 Soil organic matter

Soil organic C was included in the model, even though it is not a plant nutrient. Yields in the tropics have been found to be correlated to soil organic C (Aune and Lal 1997) due to its effects on the physical properties of the soil, including its water holding capacity. The amount of soil organic C in the upper 20 cm of the soil was calculated as a function of bulk density (bd_p) and the total amount of C in the soil (KGC_{pt}):

$$(6\text{-}9) \qquad SOC_{pt} = 100(KGC_{pt}) * 10^{-6} * bd_p * 0.2$$

6.3.1.3 Available phosphorus

In forest soils of Uganda organic P is a major source of available P through mineralization (Udo and Ogunwale 1977) and significant correlations have been found between soil organic matter and response to P and between soil organic matter and extractable P (Foster 1978). Release of P was therefore calculated from the release from the organic P pool. The organic P pool has been found to represent 60 to 80 percent of the active P fraction. The C:P ratio was set to 100 (Shepherd and Soule 1998). The amount of available P (PAV_{pct}) was therefore calculated as a function of decomposition of soil organic matter plus any P added through mineral fertilizers (FER_{pcft}) and animal manure (MAN_{pt}) plus a carry-over effect of P from the previous period (PCO_{pt-1}):

(6-10)
$$PAV_{pct} = bd * 10^2 * 0.2 * SOC * dc + \sum_f (FER_{pcft} \frac{mf\_p_f}{100}) + MAN_{pt} \frac{of\_p}{100} + PCO_{pt}$$

in which *mf_p* and *of_p* are the percentage P in mineral and organic fertilizers. Soils bind P in a form that is unavailable for plant growth in subsequent periods. Only part of the P not taken up by the plants in one period is therefore available in the next period, and this part depends on the sorption capacity of the soil. A sorption capacity of 80 percent, for instance, means that only 20 percent of the P not taken up by the plants will be available for plants in the following period. Let *sp* be the sorption capacity of the soil, then the carry over of P (PCO_{pt}) is:

(6-11)
$$PCO_{pt} = PRM_{pt-1}(1 - sp)$$

in which PRM_{pt-1} is the P remaining on the plot in the previous period, which is a function of balance between available P minus P removal from grain and residues (GRP_{pct-1}, STP_{pct-1}) plus P additions form manure and fertilizer (MAN_{pt-1} and FER_{pcft-1}):

(6-12)
$$PRM_{pt-1} = \sum_c (PAV_{pct-1} - GRP_{pct-1} - STP_{pct-1}(1 - \frac{RES_{pct-1}}{100}) + \sum_f (FER_{pcft-1} \frac{mf\_p}{100}) + 0.5(MAN_{pt-1}) \frac{of\_p}{100})$$

6.3.1.4 Available potassium

The available amount of K (KAV_{pct}, in kg/ha) is a function of the total K in the soil (KIS_{pt} in percent) plus what is added from mineral fertilizers in the current year and the decomposition of manure from previous years:

$$KAV_{pct} = \frac{KIS_{pt}}{100} * 10^6 * bd_p * 0.2 * 0.5 + \sum_f (FER_{pcft} \frac{mf\_k_f}{100}) + MNR_{pt} \frac{of\_k}{100} \quad (6\text{-}13)$$

in which *mf_k* and *of_k* are the percentage K in mineral and organic fertilizers. Total K in the soil (KIS_{pt}) is assumed constant as no good model for the dynamics of the K stock was available, hence:

$$KIS_{pt} = KIS_{pt_0} \quad (6\text{-}14)$$

6.3.1.5 Acidity

The final equation in this phase calculates the level of acidity in the soil (APH_{pt}), which is a function of the acidity in the previous year minus an acidifying constant (-0.018) and the type and quantity of mineral fertilizer used. The level of acidity can be reduced by applying lime (in kg). Lime was therefore considered as a special type of mineral fertilizer with a neutralizing effect (*li_f*).

$$APH_{pct} = APH_{pt-1} - 0.018 - 0.00091 \sum_f (ph\_f_f * FER_{pcft} \frac{mf\_n_f}{100} + 0.27 * 10^{-3} * li\_f_{pcft} * FER_{pcft}) \quad (6\text{-}15)$$

6.3.2 Phase 2: crop yield

Phase 2 calculates crop yields from available amounts of soil nutrients (NAV_{pct}, PAV_{pct}, KAV_{pct}, SOC_{pct}), the level of acidity (APH_{pct}), and crop management (LAB_{pct}). The yield of crop *c* at time *t* on plot *p* was specified as:

$$\begin{aligned} YLD_{pct} = {} & P_c * (m1_c * e^{\ln(LAB_{pct}) + m2_c}) * (1 - n1_c * e^{-n2_c * NAV_{pct}}) \\ & *(1 - p1_c * e^{-p2_c * PAV_{pct}}) * (1 - k1_c * e^{-k2_c * KAV_{pct}}) \\ & *(c1_c + c2_c * SOC_{pct} - c3_c * SOC_{pct}^2) * (1 - a1_c * e^{-a2_c * APH_{pct}}) * fg_c \end{aligned} \quad (6\text{-}16)$$

with the last factor (fg_c) being a crop-specific adjustment factor used to calibrate the model. Crops also produce residues, which can be used for animal feeding. This residue production of crop *c* at time *t* on plot *p* was calculated as:

(6-17) $$YST_{pct} = \frac{(100 - h_c)\, YLD_{pct}}{h_c} + fs_c$$

with h_c denoting the harvest index of crop *c*, which is the percentage of grain to the grain plus residue biomass. The last factor (fs_c) is a crop-specific adjustment factor to cater for the production of residues by those plants that do not produce grain, such as grasses and other fallow crops.

6.3.3 Phase 3: soil property updating

Phase 3 updates the soil properties to form the starting conditions for next period's yield estimation. All the equations in this phase are crop, plot, and time specific and calculated on a per hectare basis. Soil properties were updated for the loss of nutrients (N, P, and K) with the removal of harvested products:

(6-18) $$GRN_{pct} = \frac{gn_c(YLD_{pct})}{1000}$$

(6-19) $$GRP_{pct} = \frac{gp_c(YLD_{pct})}{1000}$$

(6-20) $$GRK_{pct} = \frac{gk_c(YLD_{pct})}{1000}$$

in which gn_c, gp_c, and gk_c stand for the proportion of N, P, and K in the harvested product, expressed in mg/kg of grain yield. It was assumed that all edible crop products were completely removed from the plot. The equations imply that the loss of nutrients is more severe the greater the crop yields.

Feeding crop residues to livestock furthermore removes nutrients from the plot. To quantify this loss, every crop's residue yield (YST_{pct}) was converted to energy units (joule) by multiplying it with a crop specific metabolizable energy contents (in million joule/kg) (me_c):

(6-21) $$YSM_{pt} = \sum_{c} (YST_{pct} * me_c)$$

The total livestock feed requirement at times *t* was expressed in metabolizable energy units following Close and Menke (1986). The metabolizable energy requirement (IME_t) is the sum over all animal species *n* and ages *m* and is calculated as the sum of energy required for maintenance (MAW_{nmt}, which is the sum of daily metabolic body sizes), milk production (MLK_{nmt}), gain in animal weight (GRW_{nmt}), and weight gain during pregnancy (PRE_{nmt}):

$$IME_t = \sum_n \sum_m (a_n (MAW_{nmt})^{0.75} + b_n (MLK_{nmt}) + c_n (GRW_{nmt}) + d_n (PRE_{nmt})) \quad (6\text{-}22)$$

with subscripts *a*, *b*, *c* and *d* denoting the animal-specific parameters and all right-hand side variables being expressed in kilograms. This equation excludes the catabolism of body tissue, which is a factor capturing the reduction in energy needs if feeding is not optimal. Animal growth was assumed exogenous to food intake, as no data were available on catabolism.

Metabolizable energy requirements (IME_t) can be satisfied by feeding crop residues or grain to livestock. The economic component included separate decision variable for feeding different types of grain to the livestock (**Chapter 7**). Residues for feeding were, however, not included as separate decision variables but it was assumed that the most nutritious residues were fed to the livestock first. Let YRM_{pct} be the residues fed to livestock of crop *c* from plot *p*; then the percentage remaining residues on the plot is:

$$RES_{pct} = \frac{(YSM_{pct} - YRM_{pct})}{YSM_{pct}} * 100 \quad (6\text{-}23)$$

The total amount of nutrients in residues was calculated by multiplying the crop residue yield by its nutrient contents; as the nutrient contents was expressed in mg/kg, all values were divided by 1000 to get kilograms of nutrients:

$$STN_{pct} = \frac{sn_c (YST_{pct})}{1000} \quad (6\text{-}24)$$

$$STP_{pct} = \frac{sp_c (YST_{pct})}{1000} \quad (6\text{-}25)$$

$$STK_{pct} = \frac{sk_c (YST_{pct})}{1000} \quad (6\text{-}26)$$

Multiplying these values by (1-RES_{pct}/100) gave the nutrient loss through residue removal.

Processes other than harvest and residue removal affect soil properties as well. These processes include erosion, leaching, and the loss of N through the plant's roots. A simplified version of the Universal Soil Loss Equation was used to model erosion. Total soil loss is a function of a rainfall factor (rt_t), soil erodibility (se_p), a slope length factor (sl_p), and a yield-specific erosion factor (ERF_{pct}):

(6-27) $$ERO_{pct} = 1000 * rt_t * se_p * sl_p * ERF_{pct}$$

The slope length factor was spatially explicit and derived from a Digital Elevation Model. The erosion factor (ERF_{pct}) was specified as a function of crop yield for some but not all crops, as yield specific erosion data were not available for all crops:

(6-28) $$ERF_{pct} = \sum_c (e1_c + 1 - e2_c(YLD_{pct}))$$

in which $e2_c$ is a crop and yield-specific erosion parameter, and $e1_c$ a crop-specific constant level of erosion.

Some soil nutrients are more affected by soil erosion than others are. Nutrient-specific erosion equations were specified to cater for this variation. The amount of eroded nutrients (N, P, K, and C) was specified as:

(6-29) $$NER_{pct} = ERO_{pct} * ef_p * SON_{pct} / (bd_p * 10^6 * 0.2)$$

(6-30) $$PER_{pct} = ERO_{pct} * ef_p * SOC * 10^{-4}$$

(6-31) $$KER_{pct} = ERO_{pct} * ef_p * KIS_{pt} / 100$$

(6-32) $$CER_{pct} = ERO_{pct} * ef_p * SOC_{pt} / 100$$

in which ef_p is a plot-specific erosion factor.

Apart from erosion, soil N is lost through the roots (RON_{pct}) of a crop, which is a function of residue and crop yield, the shoot-to-root ratio ($s1_c$) and the N contents in the roots, which is assumed to equal the N contents of stover (sn_c):

(6-33) $$RON_{pct} = 0.65 * 10^{-3} * sn_c \frac{(YLD_{pct} + YST_{pct})}{s1_c}$$

Finally, the total amount of C in the soil (KGCpct, in kg/ha) was updated as:

(6-34)
$$KGC_{pct} = KGC_{pt\text{-}1}(1 - dc_p) + (YST_{pt\text{-}1}\frac{RES_{pct}}{100}0.5 * 0.11) + (RTS_{pt\text{-}1} * 0.5 * 0.3) - CER_{pct} + 0.3 * MAN_{pt}\frac{of\_c}{100}$$

in which the amount of C stored in the roots (RTS_{pct}) was calculated as:

(6-35)
$$RTS_{pct} = \frac{YLD_{pct} + YST_{pct}}{s1_c}$$

The updated soil properties served as the initial conditions for the subsequent period. For this, the crop-specific values were aggregated to the plot level using an acreage-weighting factor (w_{pct}), which is the proportion of plot *p* under crop *c* at period *t*.

6.3.4 Phase 4: soil property balances

In phase 4, all crop-soil processes were evaluated by means of balances for N ($BALN_{pt}$), P ($BALP_{pt}$), K ($BALK_{pt}$), organic C ($BALC_{pt}$), and acidity ($BAPH_{pt}$). These balances were calculated at a plot level by using the weighting factor (w_{pct}):

(6-36)
$$BALN_{pt} = MNR_{pt}\frac{of\_n}{100} + ATN_{pt} - \sum_c w_{pct} * (NER_{pct} - GRN_{pct} - (1 - \frac{RES_{pct}}{100})STN_{pct} + \sum_f FER_{pcft}\frac{mf\_n_f}{100})$$

(6-37)
$$BALP_{pt} = MNR_{pt}\frac{of\_p}{100} - \sum_c w_{pct} * (PER_{pct} - GRP_{pct} - (1 - \frac{RES_{pct}}{100})STP_{pct} + \sum_f FER_{pcft}\frac{mf\_p_f}{100})$$

(6-38)
$$BALK_{pt} = MNR_{pt}\frac{of\_k}{100}\sum_c w_{pct} * (KER_{pct} - GRK_{pct} - (1 - \frac{RES_{pct}}{100})STK_{pct} + \sum_f FER_{pcft}\frac{mf\_k_f}{100})$$

(6-39)
$$BALC_{pt} = \sum_c w_{pct} * (SOC_{pct} - SOC_{pt-1})$$

(6-40) $$BAPH_{pt} = \sum_{c} w_{pct} * (APH_{pct} - APH_{pt-1})$$

6.4 Model calibration

6.4.1 Crops included

Eleven crop species plus one fallow activity were included in the model and listed in **Table 6.2**. Most crops can be grown twice a year. Cassava and plantain were treated as annual crops although farm households sometimes postpone harvesting cassava as its tubers preserve well in the soil. Coffee was the only perennial crop included.

Table 6.2: Cropping activities

Seasonal crops		Annual and permanent crops
1. Sweet potato	5. Millet	9. Cassava (annual)
2. Maize, traditional	6. Sorghum	10. Plantain (annual)
3. Maize, improved	7. Groundnut	11. Coffee (permanent)
4. Maize, hybrid	8. Bean	

The time interval of the biophysical model is one year. This is a simplification of reality as most crops are seasonal; yet, the seasons are of unequal length, which would require two different decision models. The chosen time scale implies that at the start of the period the agent simultaneously decides for both seasons what crops to grow and how many inputs to use. Initial values for soil properties at the plot level were derived from the predictive soil maps as described in **Chapter 5**.

6.4.2 Crop physical characteristics

Crop physical characteristics are often variety specific and average values were used instead. Some values shown in **Table 6.3** were derived from literature; these include the ratio of above ground biomass to the roots, the ratio of roots to above ground biomass, the planting density, and the seed rate.

The yield potential was estimated from survey data rather than from on-station experiments. Farmers grow many different varieties, many of which have a lower yield potential than those grown at experiment stations. Using the experiment station potential as a reference would therefore overestimate the on-farm yield potential and the 90-percentile yield was used instead (**Table 6.4**). The table compares it to the mean and the median yield as well as the yield potential from literature on Ugandan agriculture. Values were comparable except for sweet potato and groundnut, which survey values were much lower than reported in literature.

Table 6.3: Crop physical characteristics and recommended planting densities

Crop	Shoot/root ratio	Harvest index[1]	Density	Seed rate[2]
		[%]	[plants/ha]	[kg/ha]
	s_1	h_1		
Sweet potato	3.00	50	15,000	-
Traditional maize	2.33	40	44,500	20
Improved maize	2.33	40	44,500	20
Hybrid maize	2.33	40	44,500	25
Millet	2.33	40	140,000	8
Sorghum	2.33	40	170,000	10
Bean	2.33	40	200,000	60
Groundnut	2.33	40	150,000	80
Cassava	4.00	50	10,000	-
Plantain	2.33	60	1,100	-
Coffee	5.00	10	1,100	-
Fallow	2.33	0	-	-

Sources: Abalo *et al.* 2003; NARO Maize growers guide; Sasakawa Farmer guide & fact sheet; Nyende *et al.* 2001; Tenywa *et al.* 1999: 582; Oryokot 2001: 36; Otim-Nape *et al.* 2001: 203; Musoli *et al.* 2001: 394; Ebiyau and Oryokot 2001: 47

Notes: 1) The percentage of grain to grain plus stover. 2) Sweet potato, coffee, and cassava are mostly propagated vegetatively from stem or vine cuttings, plantain is propagated from its suckers.

Table 6.4: Average yield estimates and yield potential

Crop	Average	P-50	P-90	P-95	Literature	Source
Sweet potato	2748	1975	5836	7407	20,000	1)
Traditional maize	973	556	2305	2305	3,000	6)
Improved maize	1133	617	1244	4741	5,000	2)
Hybrid maize	784	**	**	**	8,000	2)
Millet	764	790	2116	2305	2,500	4,5)
Sorghum	711	593	1556	1975	3,000	
Bean	325	79	658	1235	2,000	3)
Groundnut	260	123	470	702	3,000	3)
Cassava	2883	1235	11852	18107	15,000	6)
Plantain	3107	1738	7111	11852	40,000	6)
Coffee	367	281	856	1111	2,000	6)

Sources: Columns 2-5 estimated from 2000-2001 IFPRI Survey, column 6: 1) Abalo *et al.* 2003; 2) NARO. Maize growers' guide; 3) Sasakawa Farmer guide & fact sheet; 4) Nyende *et al.* 2001; 5) Tenywa *et al.* 1999: 582; 6) Oryokot 2001: 36; Otim-Nape *et al.* 2001: 203; Musoli *et al.* 2001: 394

Notes: Survey estimates based on the agroclimatic zone with a bimodal high rainfall. ** Number of observations for hybrid maize was too small (12); the yield potential was taken as 6000 kg/ha/season.

6.4.3 Crop chemical characteristics

Table 6.5 shows the crop chemical characteristics including the N, P, K, and C content of the consumable product and residues.

Table 6.5: Chemical composition of crop and non-crop products, in grams/kg

Products	nutrients / commodity			nutrients / stover			Carbon / grain+residues
	N	P	K	N	P	K	C
Crop products	*gn_c*	*gp*	*gk*	*sn*	*sp*	*sk*	*sc*
Sweet potato	6.0	1.0	7.0	10.0	2.0	5.0	50
Hybrid maize	11.2	1.0	6.2	6.0	1.0	15.0	50
Improved maize	11.2	1.0	6.2	6.0	1.0	15.0	50
Traditional maize	11.2	1.0	6.2	6.0	1.0	15.0	50
Millet	23.1	2.5	16.3	12.6	1.5	17.8	50
Sorghum	15.0	2.5	2.4	8.0	1.5	17.0	50
Bean	33.0	2.8	11.0	8.0	1.2	8.0	50
Groundnut	37.2	5.9	8.1	12.2	1.2	9.4	50
Cassava	4.2	0.5	4.3	4.6	0.9	1.4	50
Plantain	1.2	0.3	4.5	1.6	0.3	11.9	50
Coffee	15.0	2.0	24.0	0.0	0.0	0.0	50
Fallow	0.0	0.0	0.0	15.0	1.5	15.0	50
Manure	*of_n*	*of_p*	*of_k*				*of_c*
	5.0	2.0	7.0				500
Min. fertilizer	*mf_n*	*mf_p*	*mf_k*				
DAP	180	460	0				
Urea	460	0	0				

Sources: Aune; Wortmann & Kaizzi 1998: 117; Harmand *et al.* 2004

Notes: Parameter symbols refer to the specified model for soil nutrient dynamics in Section 6.3.3; DAP=diammonium phosphate.

6.4.4 Crop yield response functions

Crop yield response to soil nutrients was estimated from fertilizer experiments and based on literature values. **Table 6.6** shows the parameter values for the crop yield functions. Crop yield response to labor was derived from a frontier production analysis using survey data (see **Chapter 7**).

Table 6.6: Crop yield parameters

A. Nitrogen, phosphorus, and potassium

Crop	Nitrogen		Phosphorus		Potassium	
	n_1	n_2	p_1	p_2	k_1	k_2
Sweet potato	0.810	0.007	-	-	-	-
Hybrid maize	1.075	0.024	0.950	0.204	-	-
Improved maize	0.734	0.016	0.950	0.204	0.395	0.830
Traditional maize	0.392	0.007	0.950	0.204	0.790	1.660
Millet	0.370	0.011	0.950	0.204	-	-
Sorghum	0.392	0.007	-	-	-	-
Bean	0.700	0.023	-	-	-	-
Groundnut	-	-	0.233	0.170	-	-
Cassava	0.500	0.018	0.653	0.138	-	-
Plantain	0.700	0.072	-	-	-	-
Coffee	0.370	0.006	-	-	-	-
Fallow	-	-	-	-	-	-

B. Labor, organic matter, and acidity

Crop	Labor		Organic matter			Acidity	
	m1	*m2*	c_1	c_2	c_3	a_1	a_2
Sweet potato	0.062	0.340	-	-	-	-	-
Hybrid maize	0.021	0.493	0.27	0.67	0.16	82.5	1.2
Improved maize	0.021	0.493	0.27	0.67	0.16	82.5	1.2
Traditional maize	0.030	0.493	0.27	0.67	0.16	82.5	1.2
Millet	0.011	0.588	0.27	0.67	0.16	-	-
Sorghum	0.082	0.341	0.27	0.67	0.16	-	-
Bean	0.008	0.610	-	-	-	-	-
Groundnut	0.006	0.567	0.27	0.67	0.16	-	-
Cassava	0.168	0.240	0.27	0.67	0.16	-	-
Plantain	0.026	0.532	-	-	-	-	-
Coffee	0.130	0.284	0.27	0.67	0.16	-	-
Fallow	-	-	-	-	-	-	-

Sources: 2000-2001 IFPRI Survey (labor response, see **Chapter 7**), J. Aune (organic matter and acidity); Bourke (1985); Kaizzi and Wortmann (2001); Smithson *et al.* (2001); Harmand *et al.* (2004)

Note: Parameter symbols refer to the specified yield response function in Section 6.3.3.

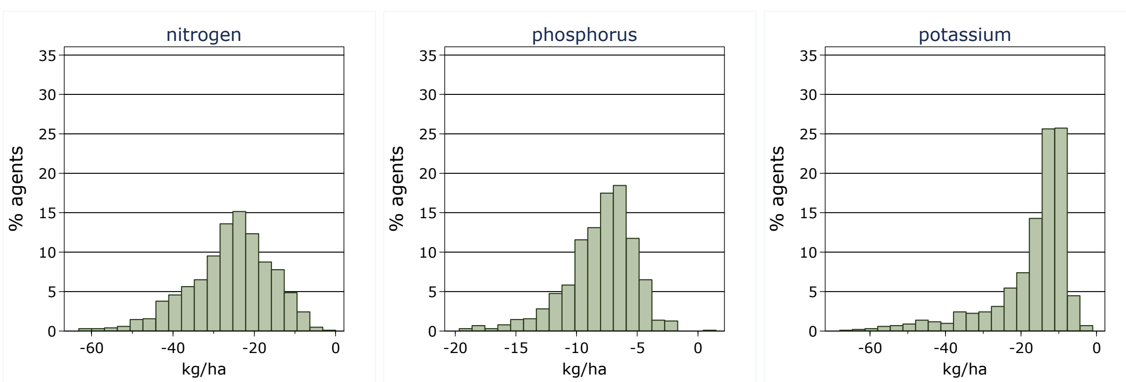
6.5 Validation of results

Model results were validated by comparing simulated crop yields and nutrient balances in the baseline year with crop yields estimated from the household survey and the nutrient balances from literature (**Table 6.1**). Because crop yield is a direct function of labor use, which will be discussed in the following chapter, the validation of the simulated crop yields is postponed until then.

Soil nutrient studies are usually based on in-depth information about a single or limited number of farm plots. Nutrient balances are widely different between plots and between farm households, mainly depending on the crops grown and the amount of inputs applied. The distribution of nutrient balances over farm households is, however, unknown for in nutrient balance studies (**Table 6.1**).

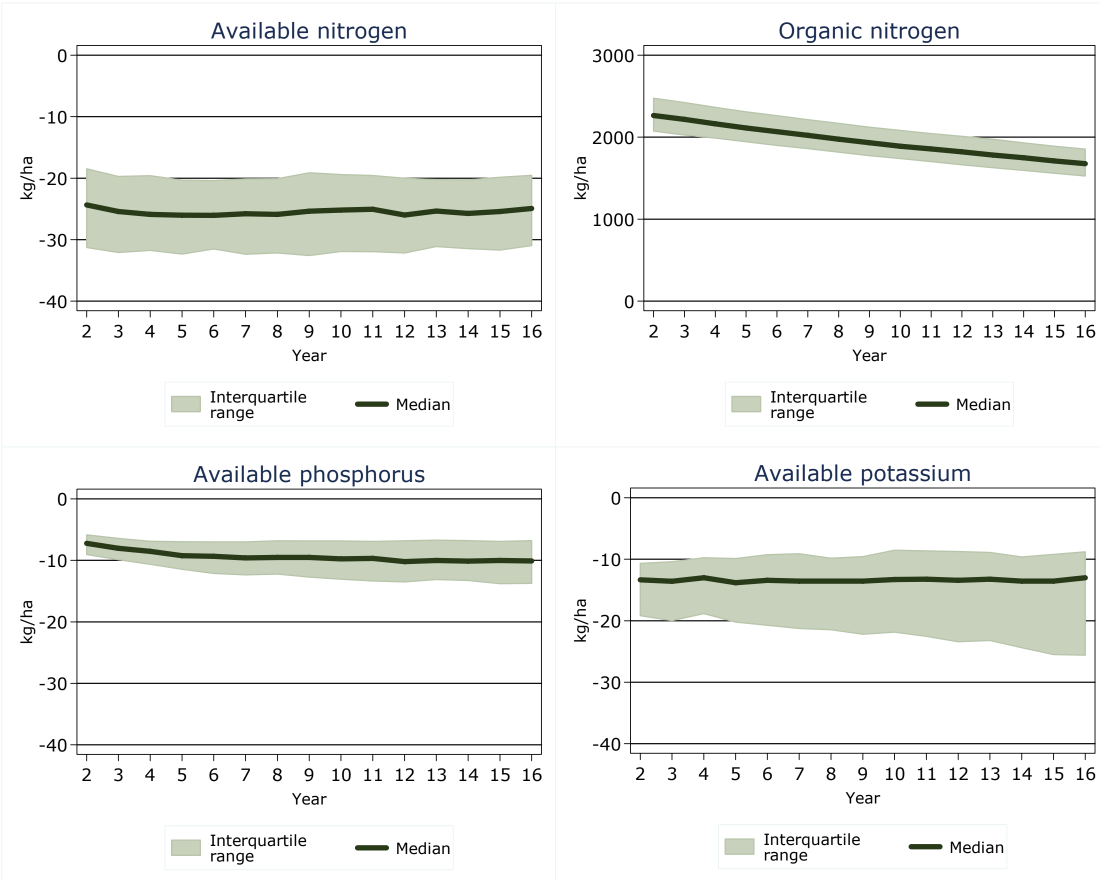
Figure 6.1 shows the distribution of average changes in soil nutrients (N, P, and K) over all 520 agents in the baseline scenario. These three diagrams clearly show a wide variation between agents in simulated nutrient balances. **Figure 6.2** shows the development of the nutrient balance over the 15 years of the simulation as median values while the variation around the median is shown as an interquartile range. It shows that the loss of soil nutrients is about constant over time, with the result that the stock of soil N decreases linearly over time (upper right diagram).

Figure 6.1: Simulated nutrient balances



Notes: Simulated changes in available nutrients in the second and third year of the simulated baseline scenario.

Figure 6.2: Simulated annual change in soil properties



Notes: Simulation results from baseline scenario. Interquartile range is the range between the 25th and 75th percentile.

6.6 Summary

Though it is generally accepted that levels of soil nutrients are declining in many parts of Sub-Saharan Africa, little is known about how this process affects crop yields in the short and long-term, and even less is known about the impact on poverty and household food security. The Tropical Soil Productivity Calculator (TSPC) is a suitable tool for closing this knowledge gap as it can be used to simulate both crop yields and soil property dynamics. When integrated into MAS, input decisions can be made endogenous and links between crop yield and soil dynamics on the one hand, and poverty and food security on the other hand can be analyzed.

7 Production behavior

7.1 Introduction

Agriculture is a managed ecosystem. Whereas the previous chapter focused on ecology, this chapter focuses on its management. The chapter is structured as follows. In Section 7.2 the crop yield response to labor use is modeled by adding a labor response factor to the crop yield model. Section 7.3 describes how the diffusion of innovations is simulated and Section 7.4 describes how agents form expectations on crop yields and prices. Production activities not captured in the biophysical model, such as livestock production are detailed in Section 7.5. The remaining constraints and opportunities to agent decision-making are specified in Section 7.6, while Section 7.7 validates the production component, and the final section summarizes.

7.2 Crop yield response to labor use

In this study, experimental data are mainly used for calibrating yield response functions. Experiment data were, however, not available for the effect of labor use on crop yield and survey data were used instead. The effect of labor use on crop yield was introduced in equation (6-1), where crop yield was specified as a function of yield potential and six reduction factors. This section describes how the reduction factor for labor was calibrated.

7.2.1 Frontier production function

Agents take a large number of decisions, including which crops to grow, on which plots to grow these, and how much input to apply. Assuming that the quantity of output varies with the quantity of input in a systematic way, this relation can be expressed as a production function, which is defined as the maximum quantity of output for each combination of a specified quantity of inputs (Colman and Young 1989).

There are two analytical approaches for estimating such input-output relationships. The first is the optimization of a production function subject to a profit, revenue, or cost function using duality theory. This approach is particularly useful to derive information on input elasticities, rates of substitution between inputs, and optimal resource allocations. The second approach is production frontier analysis, which compared to the first approach, is more pragmatic as it only returns parameter estimates of the production function together with measures of inefficiency. However, because the response of crop yield to labor use is the only interest for estimating a production function, the second approach is taken.

An empirical production frontier represents the best production bundles (input-output combinations) achievable in a sample (Coelli *et al.* 1998). Two methodologies are available for estimating the parameters for this function within the frontier approach. One is Data Envelopment Analysis (DEA), which is a non-parametric method based on mathematical programming. The other is Stochastic Frontier Analysis (SFA), which is a parametric method using econometrics. SFA is more suitable for empirical applications using survey data because this method is better able to separate technical inefficiency from statistical noise, and hence yields a better-shaped empirical production function.

The production function is specified as a Cobb-Douglas function with output of crop c (Y_c) and labor use per crop (LAB_c) on a per hectare basis. Non-labor inputs are rarely applied with less than 5 percent of the farm households using fertilizers or animal traction. These factors are therefore included as two dummy variables (DF_c for fertilizer use and DA_c for use of animal traction). To control for differences in agro-ecological conditions, six dummies for the seven agro-ecological zones (AEZ_c) are included. The production function in logarithmic form is specified as:

(7-1) $$\mathrm{Ln}(Y_c) = a_c \mathrm{Ln}(LAB_c) + b_c(DF_c) + c_c(DA_c) + \sum_{i}^{6} d_i AEZ_i + \mathrm{Ln}(f_c)$$

In which *a*, *b*, *c*, *d*, and *f* stand for the parameters to be estimated. Stochastic frontier analysis estimates the parameters of a linear regression model for crop *c*:

(7-2) $$Y_c = F(X_c, B_c) f^{\varepsilon_c}$$

in which Y_c is a vector of crop production, X_c is a vector of variable inputs, B_c is a vector of coefficients capturing the production technology, and ε_c is a vector of error components. These error components are assumed to comprise two separable parts:

(7-3) $$\varepsilon_c = v_c - u_c$$

The first part (v_c) is random disturbance capturing factors outside the agents' control as well as all sources of statistical noise. The second part (u_c) is technical inefficiency. These parts can be separated based on assumptions about their respective distributions. The random part (v_c) is assumed to come from a symmetric random probability distribution. The inefficiency part (u_c) is assumed to come from a normal distribution truncated above zero (to give only positive inefficiencies, *i.e.* a half-normal distribution. Both distributions are furthermore assumed independent from each other and to be independently and identically distributed.

7.2.2 Production data used

Production data come from a farm household survey conducted by IFPRI (2000-2001) within the research project *Policies for Improved Land Management in Uganda.* The survey included 451 farm households located in 32 of Uganda's 45 districts, excluding those districts, mostly in the north of the country, that are unsafe due to war. The IFPRI survey has detailed information at the village, household, and plot level.

The IFPRI survey is also the only large data set from which crop yields can be estimated. The 1999-2000 Uganda National Household Survey would be preferable for its larger sample size and sheer coverage of the country, but although it recorded production of each crop, it did not record the area under each crop (only the total area under crops) and can therefore not be used.

A similar problem was encountered in the IFPRI survey, but then for labor. Labor use was recorded on a plot level but not on a crop level. This is problematic since farmers in Uganda typically grow more than one crop on a plot. This issue was resolved by disaggregating labor use to the crop level assuming that a crop's share in the plot area equals its share in the plot's total labor use.

7.2.3 Model estimates

Model estimates are shown in **Table C1** in the appendix. The coefficient on labor is significant for all crops, though least significant for sweet potato and cassava. This might have to do with inaccurate production estimates for these crops, which is a typical problem for continuously harvested crops. The coefficients on mixed cropping are generally negative, which implies that intercropping reduces partial yields - though not necessarily the total land productivity (see Section 7.6.5). The five agroclimatic dummies compare different areas of Uganda against the bimodal high-rainfall area. The uni-modal area (with a single cropping season) has generally higher yields per season, though this effect is not pronounced for all crops. The parameters for animal traction and fertilizer use have significant negative signs. The use of these inputs is, however, extremely low, with less than 5 percent of the farmers using any of them at all. This means that the magnitude and sign of these coefficients are based on only a few observations and it is therefore better not to interpret them. The dummies for the two varieties of maize: the improved Longe 1 and hybrid maize are positive as expected, with hybrids giving a higher yield than Longe 1. The model for each crop is significant at a 95 percent confidence interval.

7.2.4 Labor response factor

The estimated labor response was added to the crop yield equation defined in the previous chapter. The addition of this factor to the Mitscherlich function assumes a complementary relationship between labor and all other factors, as well as interactions between all factors. The effect of increased labor use is more pronounced when soil nutrients are less limited then when they are more limited. In other words, the efficiency of labor is greater on fertile soils than on marginal soils.

The labor factor was derived from the estimated Cobb-Douglas yield function by multiplying it by the inverse of the yield potential to get values between 0 and 1. A value of 0 means that labor fully constrains crop yield, while a value of 1 means that labor does not limit yields. The labor reduction factor for crop c ($F_{labor,c}$) was calculated by dividing equation (7-1) by the yield potential while omitting the effect of fertilizer and animal traction:

(7-4) $$F_{labor,c} = g_c * e^{(a_c * Ln(LAB_c))} \qquad \text{with } g_c = \frac{1}{P_c} * e^{f_c}$$

Where LAB_c is labor supply to crop c, P_c is the yield potential of crop c, a_c is the labor coefficients of the estimated Cobb-Douglas function and f_c is the constant term of the same function.

The empirical Cobb-Douglas frontier does not necessarily plateau at the yield potential, so the reduction factor could exceed unity at high levels of labor use. To prevent this, labor use was constrained to a maximum, ($LAB_{max,c}$), which was obtained by setting equation (7-4) equal to 1:

(7-5) $$LAB_{max,c} = e^{a_c^{-1} * Ln(g_c^{-1})}$$

The labor response function was included using a piecewise linear segmentation with three levels of labor use (that is, two linear segments) (see **Table C2** of the appendix). The smallest level of labor use was set at one-tenth of the maximum, the second level was set at half the maximum, and the highest level of labor use was set equal to the maximum labor use at which yield is labor-unconstrained.

7.3 The diffusion of innovations

7.3.1 Theoretical background

Not all farm households adopt innovations at the same time. Literature suggests two alternative explanations for this (Stoneman 2002; Diederen *et al.* 2003): (1) the disequilibrium explanation states that market imperfections and lack of transparency make that the potential of an innovation is not immediately realized even though the innovation is profitable for all (*e.g.*, Griliches 1957); and (2) the equilibrium explanation states that because farm households are heterogeneous in terms of structural characteristics (*e.g.*, land, labor, and education), the innovation is not equally profitable to all farm households who as a result not all adopt, or not all adopt at the same time. The first explanation refers to a knowledge transfer between farm households, which can be analyzed using behavior models (*e.g.*, network models), while the second explanation refers to individually observable characteristics and is usually analyzed by econometric estimation of dichotomous choice models.

These two explanations do not compete but complement one another, as knowledge and profitability explain different aspects of the diffusion process. MAS are able to capture both explanations by simulating adoption as a two-stage process (Berger 2001; *cf.* Deffuant *et al.* 2001, 2005). In a first stage, an agent gains access to an innovation through an interpersonal network, while in a second stage, the agent adopts an innovation if expecting a positive contribution to the objective function. Details on each stage follow.

The diffusion of information that characterizes the first stage can be based on interpersonal networks of communication with pre-defined individual network-threshold values (Valente 1994; Berger 2001). These individual network-threshold values express the proportion of peers in a network who must already have adopted before the individual will consider adopting as well. Risk-taking agents adopt after minimal information has reached them and have low threshold values, while risk-averse agents need much information before considering to adopt and have high threshold values. The lower the threshold value, the earlier an agent adopts relative to its peers in the interpersonal network. For instance, in an interpersonal network of 10 agents, an agent with a threshold of 0.3 will only consider adoption after three of its peers have adopted.

Notice the word 'consider' because the actual decision to adopt depends on the expected contribution of the innovation to the agent's objectives, which is a function of the agent's structural characteristics. These characteristics are agent-specific as each agent is endowed individually with resources using empirically observed resource distributions (**Chapter 5**). Hence, land-saving technologies become most profitable for agents with the highest opportunity cost of land. Reaching individual network thresholds are therefore a necessary, but not a sufficient condition for adoption. The following sections describe how this was implemented.

7.3.2 Empirical application

Individual network-threshold values are ideally constructed from social network analyses in which farm households were asked about the year of adoption and the peers with whom they had communicated on a particular innovation. Such network studies were unavailable for the study villages, yet one adoption study on improved maize (a variety called Longe 1) by Ntege-Nanyeenya *et al.* (1997) was available for the Iganga district. A methodology was therefore developed to approximate network-thresholds from this study.

Adoption studies commonly group farm households into five adopter categories, from innovators to laggards, as shown in **Table 7.1**. These categories can be seen as individual network-threshold values by interpreting them as the minimum proportion of peers in the interpersonal network that needs to have adopted before the agent considers adopting as well. Instead of social network data, these thresholds can also be approximated from econometrically estimated adoption models.

Dichotomous choice models, such as logit or probit models are a standard approach to analyze the structural determinants of adoption, though sometimes criticized for not being dynamic, as they do not consider information on the time of adoption. These models explain the adoption decision from a range of variables such as farm size, family size, age, and education—some of which could also be proxies for behavioral characteristics. The parameter estimates of these models can be used to calculate a predicted probability of adoption for each farm household given its resources endowments. When ranking all households in ascending order by their predicted probability of adoption, they can be divided into threshold groups: the top 2.5 percent of farm households are the innovator segment, while the bottom 16 percent are the laggard segment. The right column in **Table 7.1** shows the estimated probabilities averaged by adoption category. As no similar adoption studies were available for other innovations, it had to be assumed that these threshold categories reflect a farm households' overall willingness to adopt innovations, not only its willingness to adopt improved maize. The evaluation of individual network thresholds is implemented in the MAS source code while the adoption decision is simulated in the mathematical programming model by constraining the access to innovations as explained in **Table C3** of the appendix.

Table 7.1: Threshold groups and their average probability of adopting Longe 1

Threshold	Characterization	Average probability of adoption[1]
< 2.5%	Innovators	0.99
2.5 – 16%	Early adopters	0.91
16 – 50 %	Early majority	0.67
50 – 84 %	Late majority	0.33
84 – 100 %	Laggards	0.08

Source: Ntege-Nanyeenya *et al.* 1997 (average prob. of adoption as calculated from logit estimate)

Note: [1]Based on significant variables in logit model. Let Y be the estimated dependent variable based on the estimated Logit equation. Y is the natural logarithm of the probability of adoption divided by the probability of non-adoption, *i.e.*, $\frac{P}{(1-P)}$. The probability of adoption (P) is then $\frac{Y}{(1+Y)}$.

7.4 Agent yield expectations

7.4.1 Theoretical background

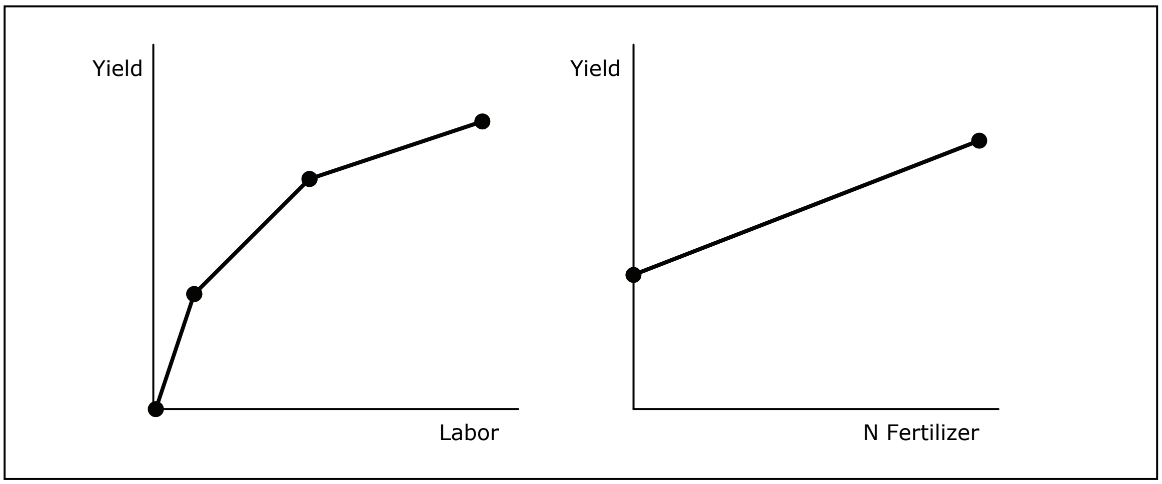
Decisions in agriculture and forestry are rather unique because of a relatively long time-span between implementation and outcomes and a high level of uncertainty in these outcomes due to the vagaries in weather, pests, and diseases. Nobody can predict the future perfectly, and farm household even less so (Brandes 1989). In the case of land-use decisions of farm households, the three main areas of uncertainty are price uncertainty, yield uncertainty, and uncertainties in the resource supply.

There are two main microeconomic theories about the formation of expectations: the theory of rational expectations and the theory of adaptive expectations (see Arrow 1987). First, rational expectations imply that decision-makers use all relevant pieces of information and make optimal land-use decisions based on stochastic foresights. Positive and negative unexpected events cancel out over time, so that expectations on average are accurate. Second, adaptive expectations imply that decision-makers base their views about the future only on past trends and experiences, ignoring newly available and potentially relevant information. The theoretical justification for this rather myopic form of behavior is that access to and processing of new information might be too costly, so that under these conditions adaptive expectations are rational too.

7.4.2 Empirical application

Like real-world farm households, agents do not have complete knowledge about the quality of their soils and the yield they will get from it. However, agents do have an idea about yield response to labor and fertilizer use for each soil quality based on the observed yield response in previous years. **Figure 7.1** illustrates the limited knowledge agents have about yield response, the left pane shows the response to labor hours as three linear pieces, and the right pane shows the response to fertilizer as one linear piece. This is a realistic way of modeling decision-making by real-world farm households.

Figure 7.1: Agents' limited knowledge of crop yield response

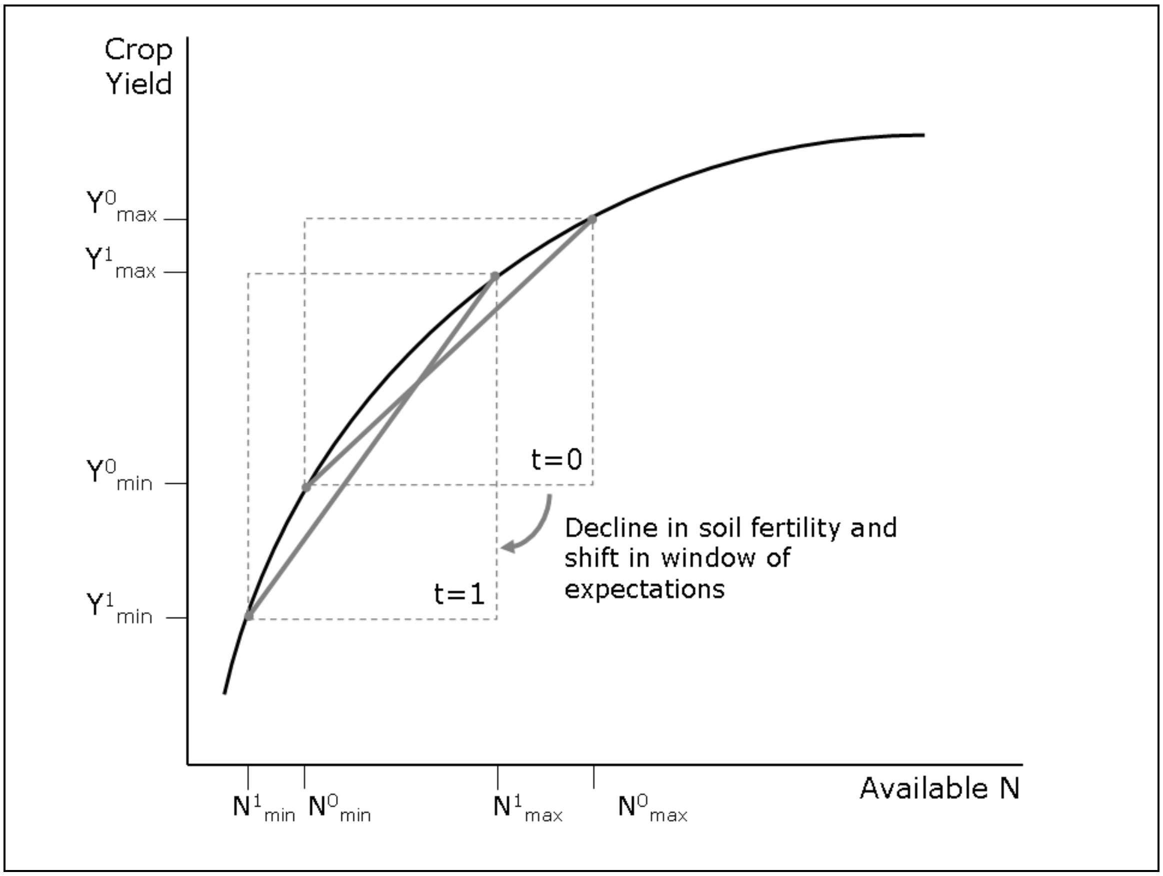


The linear segment that the agent has knowledge of does not represent the whole fertilizer response curve. Neither does the maximum level of fertilizer coincide with a yield level at the plateau. The agent's limited knowledge means that the agent can only see a snapshot of the entire yield response curve.

In comparison, **Figure 7.2** shows a complete yield response function to mineral fertilizer with the snapshot of **Figure 7.1** marked by a dotted rectangle with t=0. The difference is that the x-axis represents all available N for plant growth in the soil (mineral fertilizer plus all other sources of N) instead of only fertilizer N. At period t=0, the agent expects a yield of $Y_{0,min}$ if applying no fertilizer, and a yield of $Y_{0,max}$ if applying the maximum amount of fertilizer ($N_{0,max}$-$N_{0,min}$). The exact quantity the agent applies is determined by optimizing the farm plan, which uses the agent's expected yield response shown by the linear segment inside the dotted rectangle.

Imagine the agent applies no fertilizer. Imagine also that because of this, the plot has a net nutrient deficit at the end of period t=0, which gives a reduced soil fertility at the start of the following period ($N_{1,min}$). The agent's perspective on yield response has now changed, which is shown by a shift of the dotted rectangle towards the left lower corner and a new linear segment that is now steeper—indicating increased marginal returns to fertilizer use. At this point, the agent expects a maximum quantity of fertilizer ($N_{1,max}$) to yield a quantity $Y_{1,max}$. This linear expectation is again entered into the programming matrix to simulate the agent's decision-making.

Figure 7.2: Agent's shifting expectations of crop yield



This approach has certain advantages that are worth noting. First, it represents an agent's knowledge more realistically than giving the agent full knowledge of the entire response curve. Second, it allows the agent to learn over time by shifting its yield expectations. Third, soil fertility decline increases the marginal return to fertilizers, which could make this technology profitable in the future, even if it is not so now. Fourth, the approach more fully captures heterogeneity, as an agent's response curve is crop, plot, and time specific.

7.5 Production of livestock, coffee, vegetables and fruits

7.5.1 Livestock production

Labor, feed, pastureland, and cash are the inputs to livestock production. Goats and cows are the main types of animals kept (see **Figure 5.5** and **Table 5.8**). Few data are available on livestock production in Uganda and none of the available farm surveys was suitable for estimating a livestock production function, as these recorded neither labor use nor the age of animals. Literature values were therefore used to model growth in animal live weight, reproduction, milk production, and feed requirements.

Table 7.2 shows the basic production data used and the sources these were obtained from. The production cycle was divided into four periods of birth, puberty, maturity, and slaughter. Input requirements and production values were interpolated between these periods and averaged per year. **Table 7.3** furthermore gives metabolizable energy values of crop products and stover when used for animal feeding. These enter the model through the constraints in the mathematical programming component: when selecting livestock in production, then energy requirements have to be met by selecting combinations of stover and grain (**Chapter 6**).

Table 7.2: Livestock production data

		Cow	Young bull	Goat	Young ram
1	Growth[3,5]				
	Live weight at birth (kg)	20	20	3	3
	Live weight at puberty (kg)	180	200	30	30
	Live weight at maturity (kg)	220	250	40	40
	Live weight at slaughter (kg)	250	300	40	40
	Age at puberty (months)	30	30	8.5	8.5
	Age at maturity (months)	60	60	12	12
	Max. age at slaughter (months)	132	96	96	24
2	Milk production[4]				
	milk yield (kg/year)	700	-	200	-
	Milk to calves (proportion)	0.25	-	1.00	-
3	Reproduction[3]				
	Gestation period (months)	10	-	5	-
	Average progeny/ annum	0.64	-	1.2	-
	First pregnancy (months)	34	-	15	-
	Last calving (months)	144	-	96	-
4	Input requirements[1,2]				
	Bull service (.00 Ush)	25	-	-	-
	Maintenance cost (.00 Ush/TLU)	150	150	150	150
	Buying cost (.00 Ush)	1500	-	150	-
	Probability of loss	0.05	0.05	0.08	0.08
	Pasture requirement (ha/TLU)	0.28	0.28	0.28	0.28
	Labor (hours/TLU)	1180	1180	1180	1180

Sources: 1) Ngugi *et al.* 1990; 2) Estimated from the 1999-2000 ZEF Survey; 3) Farnworth 1997; 4) 2000-2001 IFPRI Survey; 5) personal communication with K.Zander (ZEF), October 2004

Notes: 1 TLU = to 250 kg live weight. Maintenance cost refers to median maintenance costs per TLU multiplied by the TLU of the particular livestock type and includes veterinary services (*e.g.* deworming, salt, and spraying).

Table 7.3: Nutritional contents of livestock feeds

Crop	Crop part	Dry matter (%)	Metabolizable energy (joule 10^6)	Source
Sweet potato	tuber	29.0	14.1	*1)*
	vines	25.5	10.0	*2)*
Plantain	whole plant	16.0	9.9	*1)*
	peels	15.0	10.0	*1)*
	peeled fruit	31.0	12.4	*1)*
	pseudo stem	5.1	9.0	*3)*
	leaves	25.3	9.0	*2)*
Cassava	tops	36.1	9.2	*2)*
	root	30.0	13.2	*1)*
	peel	22.7	9.8	*2)*
Maize	grain	88.0	14.6	*1)*
	stover	88.0	8.1	*1)*
Millet	grain	88.0	11.9	*1)*
	stover	88.0	8.1	*1)*
Sorghum	grain	88.0	12.7	*1)*
	stover	88.0	9.8	*1)*
Bean	grain	88.0	14.6	*1)*
	stover	88.0	6.4	*1)*
Groundnut	grain	88.0	20.9	*1)*
	stover	19.9	10.2	*2)*
Milk	-	12.0	22.1	*3)*
Grazing	-	28.0	7.7	*3)*

Sources: 1) Close and Menke 1986; 2) Bakrie *et al.*1996; 3) personal communication with Romney, ILRI, June 2004

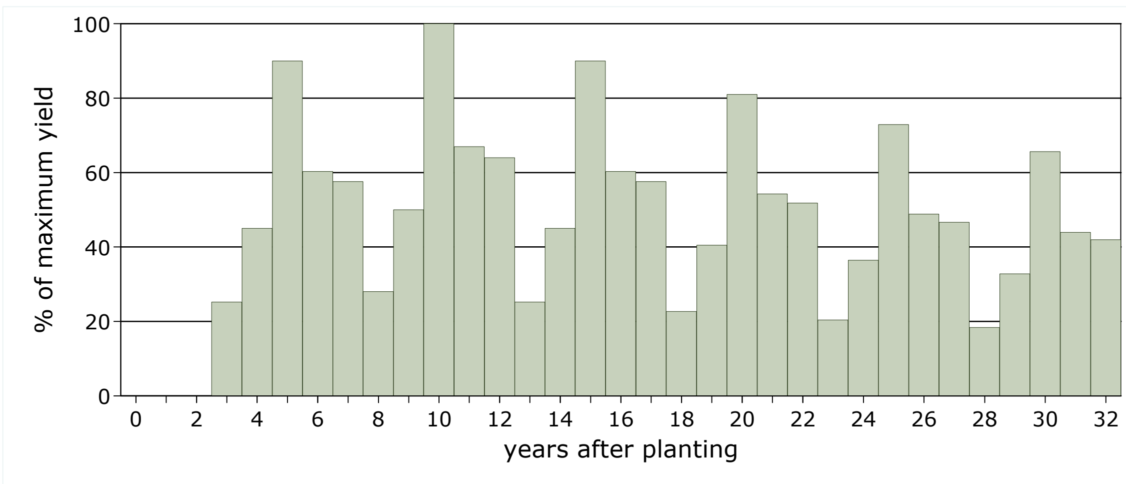
7.5.2 Coffee production

C. Canephora, or robusta coffee, is the main coffee variety grown in Uganda. The crop is indigenous to Uganda, though commercial production only began in the first half of the 20th century (Musoli *et al.* 2001). The Lake Victoria region is the main coffee producing area in Uganda with mostly smallholder farmers producing it (*ibid.*). Most coffee plantations in Uganda are relatively old, as farmers do not usually replace trees (*ibid.*). The average plantation is about 20 years old, as estimated from the 2000-2001 IFPRI Survey.

Coffee gives a first small yield after three years and a high yield after 5 to 6 years (Cambrony 1992). It was assumed that coffee production follows a five year pruning cycle, with three years of high yield and two years of low yield. The maximum age of a coffee plantation was set to 32 years (6 cycles). It was furthermore assumed that the crop bears a maximum amount of beans in the second cycle, and that the yield in each subsequent cycle declines by 10 percent. This cyclical trend is shown in **Figure 7.3** with the yield expressed as a percentage of the maximum yield.

Coffee, as livestock, is an investment. The decision to expand the plantation or herd is simulated in model's investment stage, while the decision whether or not to maintain it can be revisted in the production and consumption stages (see Section 4.6). **Table C4** and **Table C5** in the appendix explain its implementation in the mathematical programming model.

Figure 7.3: Relative coffee yield in an assumed five-year pruning cycle



7.5.3 Fruit and vegetable production

Fruit and vegetables form an important part of the local diet. Main vegetables grown include dodo (a spinach-like vegetable), onion, and cabbage; main fruits include mango, jackfruit, and tomato (UNHS 1999). Survey-based estimates of quantities of vegetables and fruits are notoriously unreliable as farm households grow and harvest these crops in a continuous fashion. Experiment data on these crops in Uganda were absent, as these are not considered major crops. It was therefore not possible to establish crop yield response functions for each of these crops. Yet, excluding them would create a bias, especially in the consumption model as fruits and vegetables are an important component of the local diet. Vegetables and fruits were therefore included at their median labor use and median yield and agents were allowed to consume but not to trade these crops. Since vegetables and fruits are commonly grown on only small areas around the compound and are fertilized with kitchen waste, they neither require land in the model nor affect the soil property dynamics.

Table 7.4: Median vegetable and fruit production and labor use

Crop	Median yield [kg/ha]	Median labor use [hours/ha]
Vegetables	740.7	564.2
Fruits	1514.0	224.5

Source: Estimated from the 2000-2001 IFPRI Survey

Note: Median values for the agroclimatic zone with a bimodal high rainfall.

7.6 Further constraints and incentives to production

7.6.1 Labor availability

Available labor hours were postulated as a function of age of the household head and the households' size, age, and sex composition. The coefficients of this function were estimated from the 2000-2001 IFPRI Survey, which was previously used to estimate labor response functions. Labor allocation data collected from household surveys are typically inaccurate. As the recall period is long, households tend to underestimate their labor use. It is therefore important that both the available labor hours and the yield response functions are estimated from the same data source so that possible data errors in the labor supply and demand cancel out.

The used data set did not specify labor supply for each household member but only total labor supply for male and female adults, and children. A functional relation between age and labor supply was established by regressing the average daily adult labor supply on the age of the household head while controlling for household size, the share of children, and district dummies (**Table 7.5**). Calibrating the resource endowments in the mathematical programming component, however, requires not the average labor supply but to the maximum labor supply. Labor demand in agriculture is unevenly distributed over the year and the average labor use therefore underestimates the available labor at peak seasons. To correct the function, the average labor use (3.15 hours/day) was compared with the labor use in the September peak season (3.92 hours/day), the difference between both (0.77 hours) was used to shift the average labor supply function upward to approximate the maximum labor supply.

Table 7.5: Regression estimates for adult labor supply

Explanatory variable	Coefficient	(SE)	R^2=0.36
Age head	8.44E-02	(3.72E-02)	**
Age head-squared	-9.10E-04	(3.36E-04)	**
Household size	-9.06E-02	(1.78E-02)	**
Share of children [%]	3.01E-02	(8.04E-03)	**
Iganga District [0/1]	1.03E+00	(5.13E-01)	**
Constant	-5.01E-01	(9.86E-01)	

Source: Estimated from the 2000-2001 IFPRI Survey

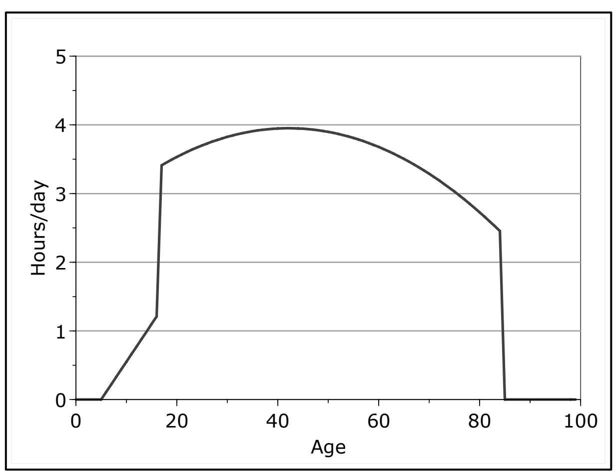
Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

These figures may seem low for semi-subsistence agriculture but are not. For example, Nyende *et al.* (2001: 509) estimated a person-day of 4 hours in Eastern Uganda, while Kidoido et al (2002: 113) equated it to 6 working hours. Yet again, available labor hours need to be consistent with the labor response function rather than with literature.

Children contribute about 12 percent to the total labor use in Ugandan agriculture (2000-2001 IFPRI Survey). Kidoido *et al.* (2002) noted that family labor use has recently been constrained by the introduction of a universal primal education program, as children are send to school by their parents instead of working in the field. The 2000-2001 IFPRI Survey was conducted after this program had been introduced and hence reflects the possibly reduced child labor supply.

It was not possible to estimate child labor supply in the same way as for adults. Yet, the survey showed that children generally start working at the age of six years old and that children work 0.61 hours/day on average in the peak season. Using this information, a linear function was drawn, so that the average working hours equalled the observed average in the peak season. To complete the age-specific labor function, it was assumed that labor supply drops to zero at the age of 85 years. **Figure 7.4** depicts the age-specific labor supply path by age of household members.

Figure 7.4: Age-specific male labor supply in average available working hours/day



According to FAO (2003), Uganda is loosing about 14 percent of its agricultural labor force due to HIV/Aids between 1985 and 2020. This implies an average annual labor loss of 0.43 percent over 35 years. As the survey was conducted in 2000, much of this predicted loss must already have occurred. Based on this FAO estimate, the agricultural labor force in the year 2000 was 6.3 percent lower than what it would have been without HIV/Aids.

A HIV/Aids related decrease in labor force can have three causes. One is the increase in mortality, the other is a reduction in labor hours of affected families, and the last is migration. Most of the reduction in labor force must come from reduced activity levels and migration: net population growth is still positive at 2.4 percent and based on the demographic composition in the study villages and the above age-specific labor supply the annual growth in agricultural labor supply is also about 2.4 percent. Literature also suggests that the relatively long period between infection and death (5-7 years on average) much reduces physical activity levels of the sick and puts a heavy burden on the other family members (FAO 2006).

The effects of reduced activity levels on total labor supply cannot be fully separated from the effect of migration, and this is also out of the scope of this thesis. Yet, to judge the impact of HIV/Aids on agricultural production, scenarios analysis was used for different levels of reduced activity levels. The baseline scenario was based on current physical activity levels estimated from the survey while two scenarios increased labor availability with 20 and 40 percent to represent the situation without HIV/Aids.

7.6.2 Labor time allocation

Seasonal variation in labor demand was captured by labor allocation requirements. Woelcke (2004) divided the year into five periods of unequal length. July, August, and September are peak periods including the harvesting of the first crop as well as the planting and weeding of the second crop.

Labor allocation per cropping period was estimated from the 2000-2001 IFPRI Survey as a proportion of total labor use per crop, as shown in **Table 7.6**. The coefficients in this table are equal for each level of input use. Although the total labor input per crop is a decision variable, the periodic distribution of this labor requirement is fixed. Agents can therefore not decide to use less labor for land preparation and compensate this with more labor for weeding. Multiplying the proportion in **Table 7.6** with the total labor requirement gives the labor requirement per period per crop.

Table 7.6: Relative time allocation to crop production by period

As a proportion of the total labor use per crop per season

Crop	January - February	March - June	July - August	September	October - December
First season	Land preparation	Weeding	Harvesting, processing		
Sweet potato	0.62	0.23	0.14	-	-
Traditional maize	0.58	0.27	0.15	-	-
Improved maize	0.49	0.34	0.17	-	-
Hybrid maize	0.38	0.48	0.14	-	-
Millet	0.38	0.42	0.20	-	-
Sorghum	0.49	0.33	0.18	-	-
Bean	0.52	0.31	0.17	-	-
Groundnut	0.49	0.27	0.24	-	-
Second season			Land preparation	Weeding	Harvesting, processing
Sweet potato	-	-	0.52	0.33	0.15
Traditional maize	-	-	0.46	0.35	0.18
Improved maize	-	-	0.42	0.42	0.16
Hybrid maize	-	-	0.52	0.33	0.15
Millet	-	-	0.43	0.31	0.26
Sorghum	-	-	0.39	0.39	0.22
Bean	-	-	0.45	0.39	0.16
Groundnut	-	-	0.56	0.24	0.20
Annual, and perennial crops, and livestock[1]					
Cassava	0.45	0.23	0.12	0.06	0.15
Plantain	0.09	0.43	0.09	0.04	0.35
Coffee	0.01	0.50	0.00	0.00	0.50
Cows	0.17	0.33	0.17	0.08	0.25
Goats	0.17	0.33	0.17	0.08	0.25
Chicken	0.17	0.33	0.17	0.08	0.25

Source: Estimated from the 2000-2001 IFPRI Survey

Note: [1)] Time allocation for livestock is assumed equal for each month by lack of more detailed data.

7.6.3 Labor allocation by gender

Gender relations affect agricultural production as male and female labor are no perfect substitutes. Men are usually more involved in land preparation while women are proportionally more involved in weeding and harvesting. The consequence is that while the total labor endowment might be abundant in some periods, male or female labor might still be a constraint.

To take gender into account, the average share of male and female labor was estimated for each crop and for each period (land preparation, weeding, and harvesting). For each average, the difference in means between male and female labor was statistically tested using a *t*-statistic; if not significant then they were taken as perfect substitutes, but if significant then they were taken as incomplete substitutes and a constraint was imposed that one of the sexes has to perform a minimum share of an activity. Summarizing, the following procedure was used:

- Estimate the average share of male and female labor per crop, per season, and for each of the three periods (land preparation, weeding, and harvesting)
- Test whether these averages are significantly different at a 95% confidence level
- If significant, then take the highest average labor share and constrain the activity as greater than or equal to this average labor share minus 1 standard deviation

For example, women did 61 percent (sd=27) of the weeding in local maize in the first season, which was significantly different (p=0.026) from men's share of 39 percent. A constraint was hence added requiring at least 34% (=61-27) of the weeding to be performed by women. This constraint was made only binding if the agent had both male and female adult labor; if an agent had only either of the two then it was allowed to substitute one for the other. **Table C6** in the appendix gives all averages, standard deviations, significance tests, and resulting constraints. **Table C7** furthermore explains the implementation in the mathematical programming model.

7.6.4 Rotational constraints

Rotational constraints refer to the sequence of crops grown over subsequent cropping cycles. Data about this come from the 2000-2001 IFPRI Survey, which for each plot recorded the three main crops grown in the past 20 cropping seasons. From these data, it appeared that farm households do not commonly practice crop rotations; especially maize was almost continuously grown on many plots. The main type of variation stems from crop combinations within the plot, especially the combination of maize with beans. Rather than alternating one season of maize with one season of beans, farm households intercrop maize with beans for many consecutive years (see next section).

This practice ensues that a plot is never fully covered by a single crop. From the last 10 cropping seasons (1996-2000), the frequency of cultivation was estimated for each crop. As rotation practices are location specific, only observations for the Iganga, Mayuge, and the neighboring Jinja Districts were included. **Table 7.7** shows the results. If for example, maize was grown 5 times in 10 seasons, then the rotational constraint was estimated to be 50 percent, implying that not more than 50 percent of a plot can be covered by maize in any season. A comparison of these estimates with those used by Woelcke (2004), who based his estimates on expert opinion, showed that estimated constraints were similar for maize, millet, bean, and sweet potato, but lower for sorghum and groundnut (**Table 7.7**). These rotations were included in the mathematical programming component by constraining the land allocation to either a maximum or a minimum proportion that could be grown under a specific crop; **Table C8** in the appendix illustrates this.

Table 7.7: Rotational constraints

Expressed as the maximum proportion of land under a crop

Crop	Proportion	Crop	Proportion	Crop	Proportion
Maize	0.531 (0.500)	Bean	0.393 (0.330)	Cassava	0.256 (0.500)
Millet	0.196 (0.250)	Groundnut	0.181 (0.333)	Plantain	1
Sorghum	0.125 (0.250)	Sweet potato	0.236 (0.330)	Coffee	1

Sources: Estimated from the 2000-2001 IFPRI Survey (Iganga and Jinja Districts); Woelcke (2004)

Note: Woelcke's estimates in brackets.

7.6.5 Intercropping

Intercropping can bring significant benefits to farm households, including increased protection from soil erosion, better pest and weed control, greater nutrient efficiency, and the spread of risk. Intercropping is common in southeast Uganda with mixed intercropping being the most common form. Crops are planted between each other, usually by broadcasting seed in a standing crop. The most common combinations of crops in the research communities are maize with bean, cassava with maize, maize with groundnut, and millet with sorghum (Woelcke 2004).

The average yield of an intercrop is commonly estimated based on the total area under all crops grown in a particular combination. For example, if maize and bean are intercropped on one hectare, then the yield is estimated by dividing both the maize and the bean output by one hectare rather than dividing each crop's output by half a hectare. Individual crop yields from intercropped fields are usually below yields from single cropped fields. Yet, the total land productivity of intercropped fields is usually greater than that in pure stand since not one but two crops are harvested.

None of the available surveys was suitable for estimating partial land equivalent ratios (PLER) for intercrop combinations. The surveys recorded whether or not a crop was grown in combination but not which crops were exactly combined. Only the yield difference between a crop grown in single stand and a crop grown in all types of mixed stands could therefore be estimated and this is shown in **Table 7.8**. However, these data might be unreliable as it is unclear whether the enumerators recorded the intercropped area in a consistent manner.

Table 7.8: Estimated yield differences for intercropping

Estimated average yield [kg/ha/season] and PLER

Crop	Single stand		Mixed stand		PLER
Maize	1,308	(255)	1,083	(207)	0.83
Groundnut	954	(627)	187	(57)	0.20
Bean	315	(212)	419	(73)	1.33
Millet	1,688	(1,199)	997	(180)	0.59
Sorghum	798	(145)	644	(135)	0.81

Source: Estimated from the 2000-2001 IFPRI Survey

Note: Standard error in parentheses. PLER=partial land equivalent ratio.

Literature values from Uganda and other countries in Africa were therefore used to verify the survey estimates. For this purpose, all publications were summarized in the *African Crop Science Journal* on the issue of intercropping that estimated LERs, as this journal has published relatively much on the issue. A number of other publications specifically on Uganda were added to these data. **Table 7.9** summarizes the results of this review.

With few exceptions, all of the published PLERs were above 0.5 and LERs above unity. This suggests definitive advantages in land productivity of intercrop combinations over crops in sole stand. **Table 7.10** shows the PLER values finally used in this study and specifies on what sources these values were based. **Table C9** in the appendix explains the implementation of intercropping in the mathematical programming model.

Although PLERs are a standard way of expressing yield premiums of intercropping compared to single cropping, it is noted that the measure is incomplete as it only considers the land saving aspect and not the labor saving aspect of intercropping.

Yet, crop science literature has not yet produced 'labor equivalent ratios' and neither did the available data allow estimation of such ratios. Intercropping was hence only included as a land-saving management decision while the labor requirement was averaged over both crops.

Table 7.9: Land Equivalent Ratios from literature

Intercrop combination	Crop 1 PLER	Crop 2 PLER	Total LER	Location	Source
Maize / bean, mixed intercropping	0.23	0.77	1.00	Ethiopia	1)
Maize / bean, row intercropping	0.38	0.92	1.30	Ethiopia	1)
Millet / bambara nut	0.62	0.46	1.08	Botswana	2)
Sorghum / bambara nut	0.90	0.77	1.68	Botswana	2)
Maize / bambara nut	0.79	0.48	1.27	Botswana	2)
Maize / groundnut	0.95	-	>1	Zimbabwe	3)
Maize / bean	0.99	-	>1	Zimbabwe	3)
Maize (low density) / bean	0.53	0.58	1.11	Zimbabwe	4)
Maize (high density) / bean	0.71	0.56	1.28	Zimbabwe	4)
Sorghum / cowpea	0.19	0.99	1.18	Uganda	5)
Maize / potato	0.44	0.58	1.02	Uganda	6)
Maize / potato (both high density)	0.60	0.98	1.58	Uganda	6)
Maize / soybean (sufficient rain)	0.63	0.33	0.95	Nebraska (USA)	7)
Maize / soybean (insufficient rain)	0.43	0.61	1.04	Nebraska (USA)	7)
Bean / potato	-	-	1.45	Uganda	8)
Maize / cassava	-	-	1.60	Uganda	9)
Cassava / cowpea	-	-	1.80	Uganda	9)

Sources:
1) Fininsa 1997
2) Karikari *et al.* 1999
3) Musambasi *et al.* 2002
4) Mutungamiri *et al.* 2001
5) Nampala *et al.* 2002
6) Ebwongu *et al.* 2001
7) Ennin *et al.* 2002
8) Opio *et al.* 2001
9) Otim-Nape *et al.* 2001

Notes: Some ratios are averaged over different experiments and locations. - means that the PLER was not specified.

Table 7.10: Used intercrop premiums

	Crop 1 PLER	Crop 2 PLER	Total LER	Note
Traditional maize / bean	0.62	0.57	1.19	Average Zimbabwe values in Table 7.9
Improved maize / bean	0.62	0.57	1.19	*Ibid.*
Traditional maize / groundnut	0.79	0.48	1.27	Botswana values Table 7.9
Improved maize / groundnut	0.79	0.48	1.27	*Ibid.*
Traditional maize / cassava	0.80	0.40	1.20	Uganda value Table 7.9, equally distributed over both crops
Improved maize / cassava	0.80	0.40	1.20	*Ibid.*
Millet / sorghum	0.59	0.81	1.40	Table 7.8

7.6.6 Crop pests and diseases

Crop pests and diseases constrain crop growth, though the severity of it varies strongly between crops, locations, seasons, and years. The magnitude of this effect was not explicitly included in the present model, and neither was the ameliorating use of pesticides, insecticides, and fungicides. Rather, the information about the effect of pests and diseases was used to fine-tune the model to observed data as explained in the next section. **Table 7.11** summarizes main pests and diseases, ordered by importance and compiled from various contributions to Mukiibi (2001) *Agriculture in Uganda*. The four crops most affected are cassava, plantain, maize, and sorghum. Least affected are millet, groundnut, coffee and bean.

7.6.7 Risk

Farm households not only value the level of yield but also the annual variation around the average and the variation in revenues obtained from it. Some crops, like cassava, have high average yields but are strongly affected by pests and diseases. Other crops, like plantain, are risky because they are perishable, which creates dependence on market prices. Raising chicken is profitable but the probability of a chicken surviving is low. Coffee is also profitable, but its price fluctuates much. Hence, risk strongly affects the production behavior of farm households. The model representation of farm household behavior considered risk in four ways:

- Empirically based network threshold levels partly captured the perceived risk associated with technology adoption.
- A difference between expected farm gate and expected market prices captured price risk.

- A premium to the future returns from investments, such as livestock and coffee, gave preference to future consumption over present consumption, which smoothed consumption and reduced income risk.
- Certainty equivalents captured the risk from yield variability by altering the yield expectations in production and investment mode. A crop, like cassava, which has a high yield risk, has a lower certainty equivalent, reducing the yield expectation below the average and hence making the crop less attractive. **Table C10** in the appendix illustrates how this was implemented in the programming matrix.

Table 7.11: Main crop pests and diseases and their average yield loss

Crops ordered by magnitude of yield loss

Crop	Main pests	Main diseases	Relative yield loss	Source
Cassava	Cassava mealybug, cassava green spider mite	Cassava brown streak virus, Cassava Mosaic Disease, bacterial blight and leaf spot	+++	Otim-Nape *et al.* 2001
Plantain	Banana weevil, nematodes	Banana leaf spot, black and yellow sigatoka, banana streak virus	+++	Tushemereirwe *et al.* 2001
Sweet potato	Rodents, sweet potato weevils	Alternaria blight, leaf spot, Fusarium wilt, viral diseases	+++	Muwanga *et al.* 2001
Maize	Stem borers, termites, weevils, grain borers	Maize streak virus, leaf blight, rust	++	Kikafunda-Twine *et al.* 2001
Sorghum	Stalk borers, sorhum midge, sorghum shoot fly, birds	Grain mold, leaf blight, Anthracnose, smuts, striga	++	Ebiyau and Oryokot 2001
Bean	Bean fly, black bean Aphid	Common bacterial blight, Angular leaf spot, rust	++	Opio *et al.* 2001
Coffee	Antestia bug, lacebug, berry borer	Coffee leaf rust, Coffee Berry Disease, coffee wilt disease	+	Musoli *et al.* 2001
Groundnut	-	Groundnut rosette disease. fungal leaf spots, groundnut blight	+	Busolo-Bulafu and Obong 2001
Millet	Birds	Millet blast	+	Oryokot 2001

Source: Compiled from Mukiibi (2001)

Note: Relative yield losses were roughly estimated from the literature source.

7.6.8 Input prices

Table 7.12 shows input prices used in the analyses. These were estimated from the 2000-2001 IFPRI Survey. The reference period of this survey was the first and second season of 2000. This only partly overlapped with the UNHS, which referred to the second season in 1999 and the first season in 2000. The consumer price inflation and the price inflation for goods and services both were about 2 percent in 2000 (UBOS 2002), which was used to adjust input prices downward to 1999 price level of the 1999 UNHS.

Table 7.12: Prices of variable inputs

	Input	Price [.00 Ush]		Input	Price [.00 Ush]
1	Seed (kg)		3	Land (ha)	
	Traditional maize	2.00		Renting in	1000.00
	Improved maize	12.00		Renting out	1000.00
	Hybrid maize	25.00	4	Labor (hour)	
	Bean	5.00		Hiring in, male	2.50
	Groundnut	13.00		Hiring in, female	1.88
	Millet	4.00		Hiring out, male	2.50
	Sorghum	3.00		Hiring out, female	1.88
2	Mineral fertilizer (kg)	6.00	5	Leasing in tractor (hour)	4.70

Source: Estimated from the 2000-2001 IFPRI Survey

7.7 Validation of results

Model results were validated in three steps. First, it was tested if all programming matrices for all agents gave feasible solutions. If this was not the case then the matrices were analyzed and changes made where necessary. Second, price experiments were conducted to test if an increase in farm gate price of a particular enterprise lead to this enterprise being increasingly selected in the solution. Third, model outcomes in the first five years of the baseline scenario (representing the observed situation and current trends, see **Chapter 9**) were compared with observed values from the survey. A few built in tools were available to bring the baseline closer to the observed values if needed. These included:

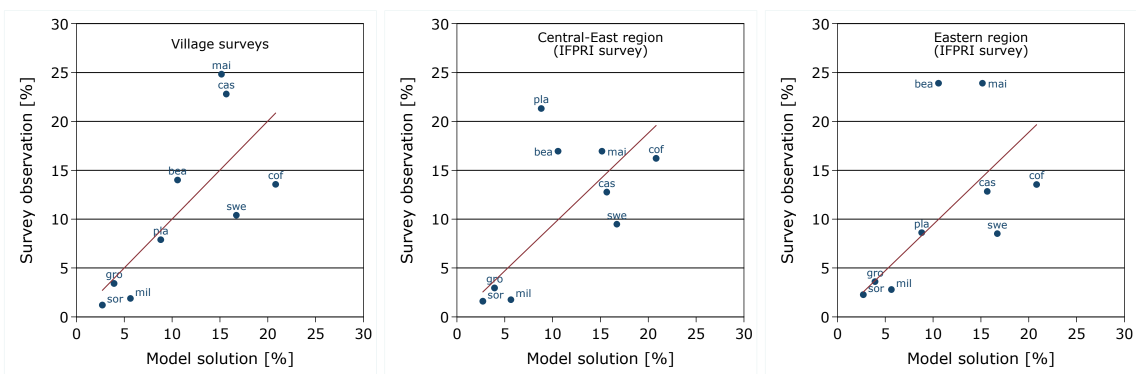
- Yield reduction factors (fg_c, in **Chapter 6**) to adjust the average estimated yield levels to the average observed yield levels.
- Certainty equivalents to adjust the area under crops in accordance to their level of perceived risk.
- Future price premiums for livestock and coffee to alter the preference of future income relative to current income.

Observed and predicted values were compared for four main production aspects of the farming system: (1) land-use patterns; (2) average and median crop yields; (3) the share of intercropping; and (4) the adoption of innovations.

Figure 7.5 compares the simulated land-use pattern as a percentage of the total cropped area with the observed land use pattern. Validation was complicated by a high variation in the observed land-use between separate surveys and areas within the surveys. The left diagram validates the predicted land-use against the 1999-2000 ZEF Survey. The regression fit falls exactly together with a 45°-line from the origin, which indicates a good fit (coeff.=1.00, SE=0.15. R^2=0.82). The second and third diagrams validate model results against the 2000-2001 IFPRI Survey, which was used to estimate crop yields, and shows a somewhat lower fit. Not so surprisingly, the fit was worst when taking the central and eastern region in the survey together as in the right-most diagram.

Figure 7.5: Validation of land-use with linear regression fit

Comparison between model results and three different survey estimates

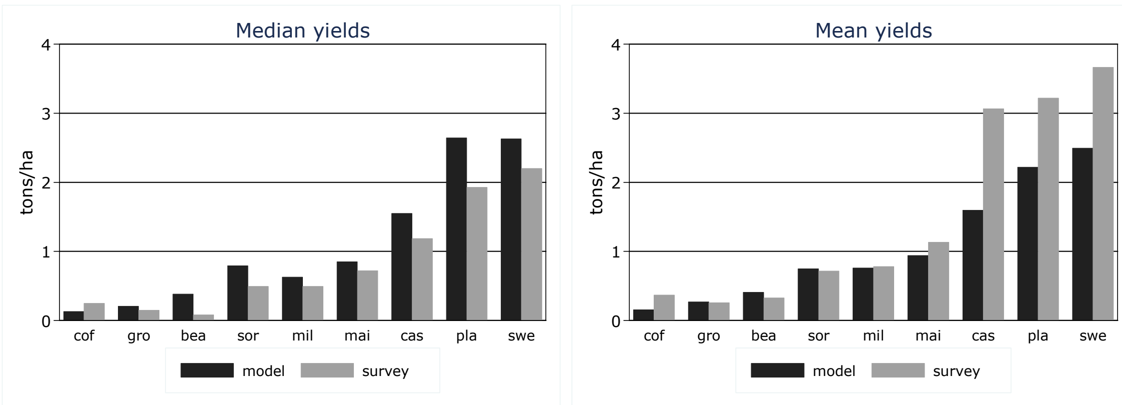


Notes: Simulated means over year 2-3 in the simulation. Line segments indicate the linear regression fit of survey vs. model estimates without a constant term. Model fit: Coeff=1.00; SE=0.15, R^2=0.82 (left); Coeff=0.94; SE=0.18, R^2=0.74 (middle); Coeff=0.94; SE=0.16, R^2=0.79 (right). Crop codes: cof=coffee; gro=groundnut; bea=bean; sor=sorghum; mil=millet; mai=maize; cas=cassava; pla=plantain; and swe=sweet potato.

Figure 7.6 compares simulated crop yields with observed crop yields as estimated from the 2000-2001 IFPRI Survey. As discussed in **Chapter 2**, mean and median estimates of crop yields differ widely for the distribution of yields is negatively skewed. The observed negative skew is strongest for sweet potato, plantain, and cassava. Yield estimates for these, continuously harvested crops, are notoriously difficult and statistical errors might therefore explain part of the problem. The negative skew is not well replicated in the model as means and medians are much closer.

Figure 7.6: Validation of crop yields

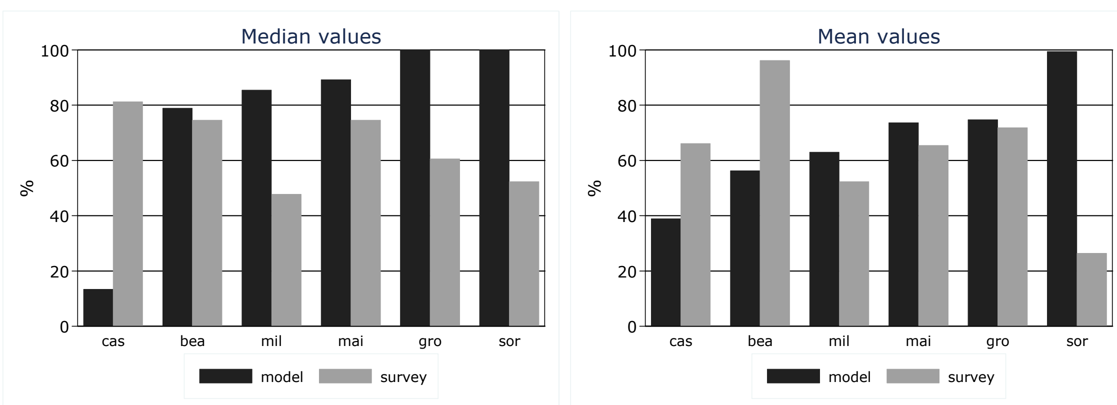
Comparison of medians (left) and means (right)



Notes: Simulated means and medians over year 2-3 in the simulation. See **Figure 7.5** for an explanation of crop codes.

Figure 7.7: Validation of intercropping

Comparison of median (left) and mean (right) percentage intercropping of crop area



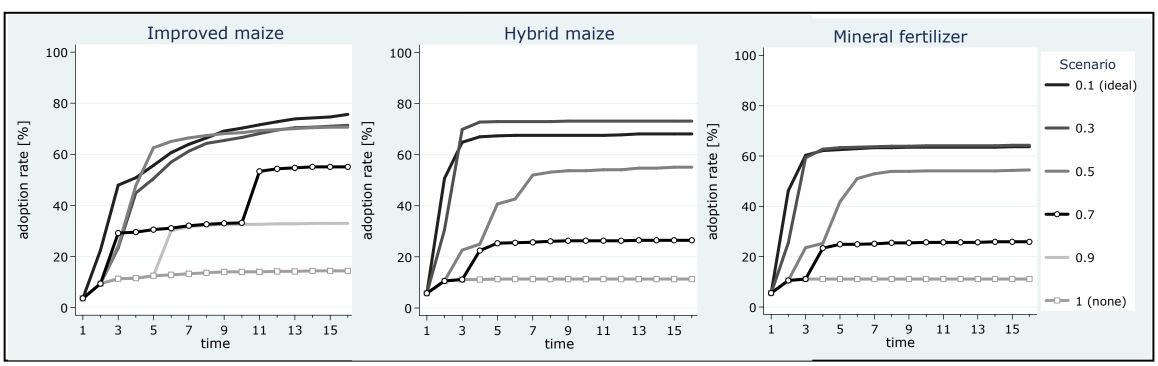
Notes: Simulated means and medians over year 2-3 in the simulation. See **Figure 7.5** for an explanation of crop codes.

Figure 7.7 compares the simulated and observed patterns of intercropping. The percentage intercropping of a crop was calculated as the area intercropped divided by the total area under a crop. Again, the diagrams show a wide difference between mean and median estimates from the survey. The model overestimates the intercropping of sorghum, which is a minor crop, but attains a good fit for all other crops.

The final issue is the diffusion of innovations. This diffusion is difficult to validate by lack of empirical time series data on the speed of diffusion. The challenge is therefore to select the most plausible diffusion pattern. To change the speed of diffusion, the model includes a parameter through which threshold values can be adjusted (see Section 7.3). **Figure 7.8** shows the simulated diffusion patterns for three innovations—improved maize, mineral fertilizer, and hybrid maize—and for six different values of this parameter ranging almost from ideal technical change (0.1) to no technical change (1). In the almost ideal scenario, all innovations become available to all agents in year 1 of the simulation, while in the 'no diffusion' scenario, the diffusion is frozen at current levels as observed from survey data.

The figure shows that the diffusion pattern is highly sensitive to the chosen parameter value. The diffusion of improved maize and hybrid maize levels off at about 80 percent diffusion, even in the ideal scenario; while the diffusion of mineral fertilizers levels off at an even lower level of about 60 percent. Two scenarios, with parameter values 0.5 and 0.7, give an S-shaped diffusion path and look plausible. The scenario with value 0.7 is more realistic, as the diffusion of improved maize is more gradual while the diffusion of hybrid maize and mineral fertilizer reach a plateau at about 25 percent diffusion. This scenario was chosen as the most plausible since hybrid maize and mineral fertilizers are no new technologies in Uganda, but their adoption has remained low, while the adoption of improved maize is still increasing.

Figure 7.8: Alternative diffusion patterns by adjustment of threshold values



Notes: All scenarios are based on the baseline (Chapter 9). The numbers in the legend refer to values of the threshold adjustment parameter.

7.8 Summary

The calibration of the present model involved the estimation of a large number of parameters. This chapter specified how the production parameters were estimated and explained how these parameters were incorporated in the mathematical programming component. The production side of the mathematical programming-based MAS was validated against observed data by comparing mean and median values for land-use, crop yields, and the percentage intercropping.

8 Consumption behavior

8.1 Introduction[§]

In the three-stage sequence of farm agent decision-making, the consumption of generated income constitutes the third and final stage. This chapter outlines a novel methodology for representing consumption decisions and simulating consumption poverty of farm households. The methodology is based on mathematical programming but adds three innovations: First, poverty levels are quantified by including a three-step budgeting system, including a savings model, a Working-Leser model, and an Almost Ideal Demand System. Second, the model is extended with a disinvestment model to simulate farm household coping strategies to food insecurity. Third, multi-agent systems are used to tailor each mathematical program to a real-world household and so to capture the heterogeneity of opportunities and constraints at the farm level as well as to quantify the distributional effects of change.

The chapter continuous in Section 8.2 with an outline of the three-step budgeting process. Section 8.3 shows the model estimation results and gives other relevant data used to model the consumption behavior of agents. The consumption component is validated in Section 8.4 and the chapter ends with a summary in Section 8.5.

8.2 A three-step budgeting process

8.2.1 Theoretical background

If consumption decisions systematically vary with the level of income, prices, and household characteristics then these decisions can be presented as a set of demand equations. The theory of consumer behavior offers the best guidelines as to how these consumption decisions are made (Sadoulet and de Janvry 1995). The theory

---

[§] Parts of this chapter were used for SCHREINEMACHERS, P. & T. BERGER 2006. Simulating farm household poverty: from passive victims to adaptive agents, *Selected paper for the Tri-annual meetings of the International Association of Agricultural Economics*, Gold Coast, Australia, 12-18 August 2006.

postulates that consumers maximize their utility subject to a budget constraint. The solution to this maximization problem yields a set of demand equations from which income and price elasticities can be derived.

Using duality theory, demand equations can also be derived from a cost function; costs are then minimized subject to a utility function (Deaton and Muellbauer 1980). In a cost function, unit values (expenditures divided by quantities) replace prices (Huang and Lin 2000). Cost functions are commonly applied in empirical studies since expenditures and quantities are usually recorded in surveys but prices are not (*e.g.*, Elsner 1999; Brosig 2000).

Weak separability of consumer decision-making implies that "commodities which act closely in yielding utility can be grouped together, while goods which interact only in a general way through the budget constraint are kept in separate groups" (Sadoulet and de Janvry 1995: 36). If assuming weak separability then the expenditure decisions of agents can be conceptualized as a stepwise budgeting-process.

8.2.2 Theoretical model

For the present study, the budgeting-process was divided into three steps as is shown in **Figure 8.1**. In the first step, agents choose what part of the disposable income to spend and what part to save. This is a decision between current and future expenditures as income set aside for savings is available for investment and hence contributes to future income and expenditures. Let SAV be the part of disposable income saved and EXP the part spent, which gives the following income identity:

(8-1) $$INC = SAV + EXP$$

The level of savings itself is a function of disposable income (INC), household size (HHS), and a constant term (C_a):

(8-2) $$SAV = f(INC, HHS, C_a)$$

Let LIQ be the available stock of liquid means and INV the amount of investment, for example in livestock, then investments cannot exceed savings plus liquid means:

(8-3) $$INV \leq SAV + LIQ$$

In the second step, consumers allocate income to food or non-food items. Let FEX be the part of income spent on food and NEX be the part of the income spent on non-food items, which gives the following equality condition:

(8-4) $EXP - FEX - NEX = 0$

In which FEX is a function of total expenditures (EXP), household size (HHS) and a constant term (C_b):

(8-5) $FEX = g(EXP, HHS, C_b)$

In the third and last step, consumers spend the allocated food budget on eight categories of food products, such as cereals, roots and tubers, and animal products. Let CAT_i be the expenditures on category i, then:

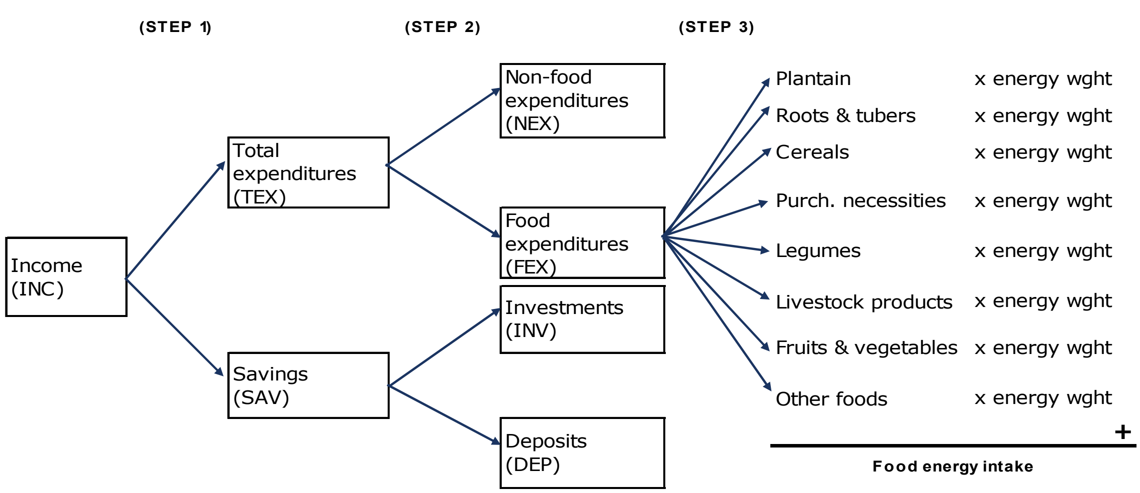
(8-6) $FEX = \sum_i CAT_i$

with expenditures to food categories being a function of the total food budget (FEX), household size (HHS), prices (c_i) and a constant (C_i):

(8-7) $CAT_i = h(FEX, HHS, c_i, C_i)$

This equation is parameterized using a linear version of the Almost Ideal Demand System (LA/AIDS). The three-step budgeting system was included in a mathematical programming model that optimized the expected cash and in-kind income and future income from investments. The is-equal-signs in equations 8-1, 8-4, and 8-7 ensure that the agents can only increase the in-kind income (*i.e.*, the consumption of own produced food) through increases in the disposable income from farm and off-farm activities.

Figure 8.1: Three-step budgeting process



The assumption of separability furthermore facilitates empirical estimation, as different models can be fitted at each step in the budgeting process. The following sections specify estimable models for each of the three steps in the budgeting process.

8.2.3 Savings and expenditures (Step 1)

In line with micro-economic theory, it was assumed that the proportion of income saved increases with the level of income. Let the variable SAV be the savings and INC the total income, HHS the household size, D the matrix of district dummies, and C_a a constant term. The amount of savings was specified as a quadratic function of income:

$$SAV = \alpha_1 INC + \alpha_2 INC^2 + \beta HHS + \sum_{i=1}^{n-1} D + C_a \qquad \text{with } \alpha_2 > 0 \tag{8-8}$$

As long as α_2 is positive, savings increase more than proportionally with income. Alternatively, a share-log functional form could have been chosen as this also allows for increasing savings with income. However, a share-log does not allow for negative net savings, hence the preference for a quadratic specification.

8.2.4 Food and non-food expenditures (Step 2)

It is assumed that the budget allocation to food and non-food items systematically varies with the level of income, prices, and household characteristics, and that this relation can be described by a modified version of the Working-Leser model (Working 1943; Leser 1963). The Working-Leser model has a share-log functional form in which the share of food and non-food items in the total budget is a linear function of the logarithm of total expenditures. Theoretically, this function is consistent with utility maximization and the adding up property can also be met. The modified version also includes a number of variables found to influence the budget share, such as the size and composition of the household, and regional dummies to capture spatial variation in consumption patterns.

The Working-Leser model has a flexible functional form that allows income elasticities to vary with income levels. This flexibility is desirable as high income households are expected to spend a lower share of their income on food expenditures. The Working-Leser demand function is expressed as:

(8-9) $$v_i = \beta_0 + \beta_1 \ln TEX + \beta_2 HHS + \sum_{n=1}^{40} \beta_3 D$$

where v_i is the expenditure share on food or non-food items, TEX is total expenditures (calculated as INC–SAV), HHS is household size, D is a matrix of district dummies, and the betas are the parameters to be estimated. The estimated parameters should satisfy the properties of adding up and symmetry. The first is satisfied if constant terms across equations add up to unity ($\sum_i \beta_{1,i} = 1$), the second is satisfied if the sum of parameters across the equations is zero ($\sum_i \beta_{1,i} = \sum_i \beta_{2,i} = \sum_i \beta_{3,i} = 0$). In this two-equation case, only one equation needs to be estimated, as the parameters of the other equation can be derived from the first using the above properties.

8.2.5 Almost Ideal Demand System (Step 3)

In the final step of the budgeting process, the consumer chooses which food products to consume. A linear approximation of the Almost Ideal Demand System (AIDS) was used for modeling the decision-making at this third step (Deaton and Muellbauer 1980). The AIDS is specified as:

(8-10) $$w_k = \delta_{0,k} + \sum_k \delta_{1,k,l} \ln p_l + \delta_{2,k} \ln(M/P^*) + \delta_{3,k} HHS + \sum_{n=1}^{40} \delta_{4,k} D$$

where the subscripts *k* and *l* denote individual food categories of a total of *n* categories (*k,l*=1,2,..,n) and the gammas denote the parameters to be estimated. The variable w_k is the share of category *k* in the total food budget, M is per capita food expenditures, and P^* is an index of prices, which in the original (non-linear) version of the model has a translog functional form. For practical estimation, the system can be made linear by replacing the non-linear price index with the logarithm of the Stone geometric price index (Deaton and Muellbauer 1980):

(8-11) $$\ln P^* = \sum_k w_k \ln p_k$$

The use of the Stone price index introduces a simultaneity problem as the budget share (w_k) now appears both as a dependent and independent variable (Eales and Unnevehr 1994). In panel data analysis, the lagged budget share is commonly used, but in cross sectional data this is not possible. The impact of the simultaneity problem is, however, likely to be small as it enters the equation as a logarithm of its inverse form and is multiplied by income.

A further complication occurs if some of the dependent variables have a large frequency of zero values, that is, many households do not consume a particular food category. If this is the case, estimation can give biased parameter estimates. A two-stage Heckman procedure can abate this problem (Heckman 1976, as in Kennedy 1998). First, probit models are estimated that include the variables of the demand equation (8-10) minus one variable and including the zero observations. From this estimate, inverse Mill's ratios are computed for each category, which is included as an additional explanatory variable in the demand equation. Only the non-zero values of the dependent variable are then used in the demand estimation.

To ensure that the estimated model is in harmony with the theory of consumer behavior, the estimated parameters must meet the property of adding up ($\sum_k \delta_{1,kl} = \sum_l \delta_{1,kl} = \sum_k \delta_{2,k} = 0$) to satisfy the budget constraint, as well as the properties of homogeneity ($\sum_k \delta_{0,k} = 1$) and symmetry ($\delta_{1,kl} = \delta_{1,lk}$) to allow utility maximization.

Error terms across demand equations correlate because the sum of budget shares is constrained to unity (Sadoulet and de Janvry 1995: 45). Zellner's seemingly unrelated regression model is used to overcome this. To ensure theoretical consistency, parameters need either to be tested or constrained across equations. To avoid singularity in the variance-covariance matrix, one system equation has to be omitted in the estimation but it can be recovered from the remaining equations.

Own and cross-price elasticities, compensated for the income effect, are derived from the estimated parameters of the demand functions as follows (Sadoulet and de Janvry 1995):

(8-12) Own price elasticity: $$e_{kk} = -1 + \frac{\delta_{1,kk}}{w_k} + w_k$$

(8-13) Cross price elasticity: $$e_{kl} = \frac{\delta_{1,kl}}{w_k} + w_k$$

(8-14) Income elasticity: $$\eta_k = 1 + \frac{\delta_{2,k}}{w_k}$$

These elasticities are estimated at the average budget share and are only used to visually check the plausibility of the estimates. Because the budget share is endogenous in the complete model, not the average elasticities but the original equations (8-1, 8-2, and 8-3) are included in the mathematical programming component, as is explained in the appendix in **Table D1**.

8.2.6 Quantifying poverty from food energy needs and intake levels

In a developing country context, poverty is best quantified in terms of food energy consumption (Coudouel *et al.* 2003). To quantify consumption poverty for each agent in the present study, the estimated energy intake from the three-step budgetting system was compared to the agent's food energy needs.

Food energy intake was estimated from the value of expenditures on the *k*-th food category, *i.e.*, as w_k*FEX. To express food consumption in joule, each value was divided by its unit value to get a physical quantity and then multiplied by an energy weight as shown in **Figure 8.1**.

Food energy needs were estimated from the age and sex composition of each agent household. James and Schofield (1990) suggested estimating age and sex-specific energy needs from basal metabolic requirements and physical activity related requirements. They divided human energy needs into two parts: basal metabolic requirements and physical activity related requirements. The basal metabolic rate (BMR) of a population was calculated from data on body weight, age, and sex. Multiplying this BMR by a value capturing the physical activity level (PAL) gave total energy needs. The food energy requirement of an agent *p* (E_p) was calculated as the sum of individual requirements of its members *q* (*q*=1,2,..,n):

(8-15) $$E_p = \sum_q (BMR_q * PAL_q + 0.418 * f_q)$$

in which the variable BMR is a function of an age and sex-specific constant (α_q) and the member's body weight multiplied by an age and sex-specific coefficient (β_q):

(8-16) $$BMR_q = \alpha_q + \beta_q * Weight_q$$

Equation (8-16) has a correction of 0.418 million joule/day for the increased food energy intake during pregnancy; this correction factor was multiplied with an age specific fertility rate (*f*) that takes a value of zero if the member is a male.

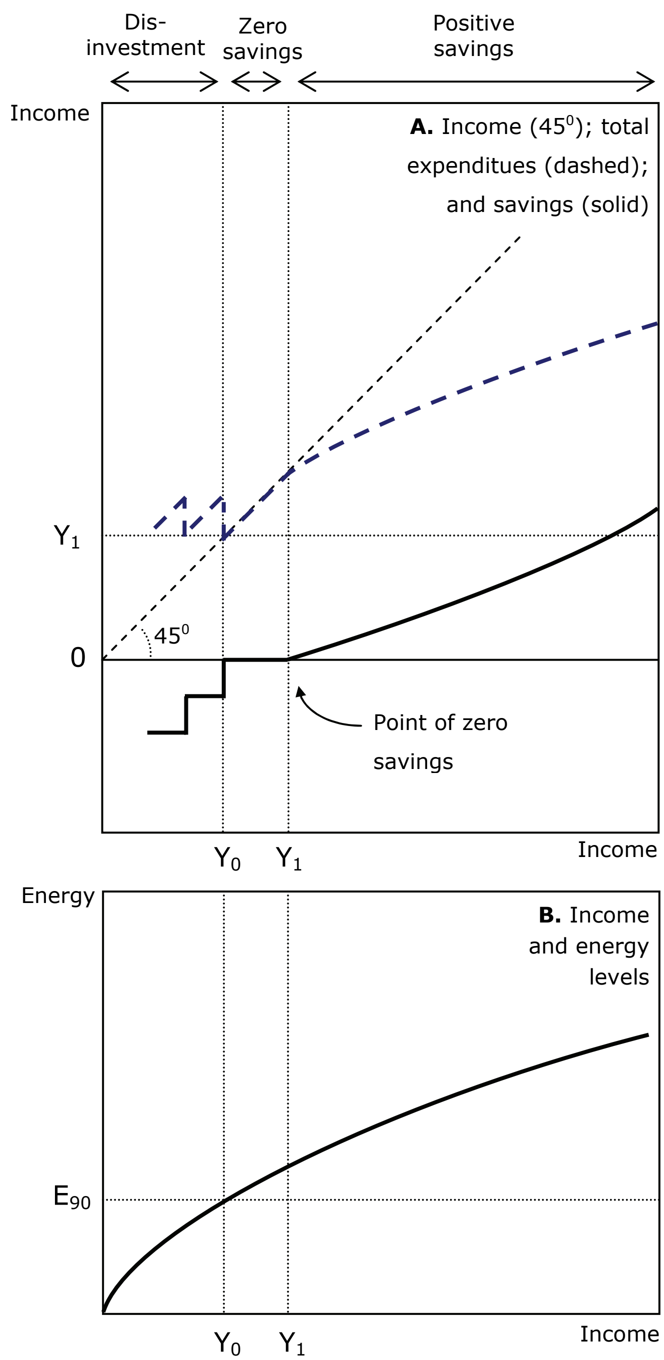
The estimated household size in joule was used throughout the study as a measure of household size. Dividing it by the average male energy requirement, 3.229 billion joule/year, expresses household size in male adult equivalents.

8.2.7 Coping strategies to food insecurity

Farm households are food secure when they have strategies to smooth consumption in the event of disaster. One such commonly observed strategy in African farming systems is keeping livestock (Kristjanson *et al.* 2004). Farm households buy and maintain animals in good years—when yields or prices are high, and sell them in bad years—when yields or prices are low. By investing and disinvesting in livestock, households smooth their income and consumption between years.

The difference between a bad and a good year can be conceptualized as a food energy balance. A bad year is, for instance, when income is not enough to satisfy 90 percent of the food energy needs. This level is arbitrary and should be ideally based on in-depth interviews. **Figure 8.2** shows the theoretical model for this disinvestment process. The upper diagram shows that savings increase (the solid line) and expenditures decrease (the dashed line) as a share of income. The lower diagram shows food energy as a function of income with the horizontal line depicting the food energy level at 90 percent (E_{90}) fulfillment of the (physical)

Figure 8.2: Theoretical model for coping strategies against food insecurity



needs; the income level where both functions intersect is denoted as Y_0.

Agents try to avoid poverty by keeping their income above level Y_0. Yet, if income would fall below Y_0, then an agent has two options: (1) add savings to the disposable income; or (2) sell livestock and add the returns to the disposable income. Both options entail the substitution of current consumption for future income. Agents continue their disinvestment until their food energy needs are satisfied for at least 90 percent, which is when the level of income equals Y_0.

If the agent is unable to sell assets or consume savings then it runs into an energy deficit and falls into poverty. **Figure 8.2** shows that the consumption and savings/dissavings functions are non-smooth below level Y_0, this is because selling livestock is a discrete rather than a continuous event; a smooth function would, however, represent the consumption of savings, which is not shown here.

At income levels between Y_0 and Y_1, no income is saved – the savings function is flat at the zero level, as all income is consumed. Point Y_1 represents the highest level of income at which agents do not save income and can be derived by equating the savings equation to zero (see **Table D1** of the appendix). Disinvestment decisions were included in the mathematical programming component, which is explained in **Table D2**.

8.2.8 Fertility and mortality

Uganda has one of the highest fertility rates in sub-Saharan Africa with a woman bearing about six children on average in a lifetime (The World Bank 2004). However, mortality rates are high too, with male infants having only a 25 percent chance of surviving to the age of 65 (*ibid.*).

There is a long lasting debate about whether the decision to have children is subject to economic rationale (*e.g.*, Nugent 1985). If economic rationale motivates the decision then family planning is an investment decision. Having more children increases the amount of available labor and offers economic security for aging parents. On the other hand, some scholars have argued, that households can choose to take more children than would be economically optimal if this has a social value (*e.g.*, rising social status). While others have argued, that the decision to have children is only partly controllable when contraceptives are either unavailable or inaccessible. For the sake of simplicity, family planning was not taken as endogenous. Instead, current fertility and mortality rates were uniformly imposed on all households.

8.3 Data and estimation

8.3.1 Budget data used

The consumption model was estimated from the 1999-2000 Uganda National Household Survey, as this was the only available data set suitable for this purpose. This nationwide survey recorded food expenditure by recall methods over the previous seven days, while non-food expenditures were recorded from a recall of one month for non-durables and one year for (semi-)durables.

The expenditure module was of low quality as some quantities and values had been filled out erroneously for at least some of the study areas. Quantities were recorded in 87 different units, ranging from pieces to heaps and sacks, the exact weight of which are often location specific, but were not recorded in the survey. Further, some enumerators are likely to have converted all quantities into kilograms but without adjusting the original units given by the respondents (*e.g.*, households consuming 10 bunches of plantain in a week). Values of food items were also messily recorded as some, but not all enumerators, denominated the currency by dividing values by a thousand.

All study areas with some suspiciously high level of quantities or unit values were deleted from the survey. Fortunately, data errors were found to be concentrated in some particular areas only, but unfortunately, many of these were in the eastern region where the research area was located. Severe outliers in quantities or unit values were identified as those values above the 75th percentile plus three times the interquartile range (p75+3(p75-p25)). Households with severe outliers in quantities were omitted from the analysis; severe outliers in unit values were replaced by regional level median values.

8.3.2 Savings and expenditures (Step 1)

The savings function was estimated in unrestricted form using ordinary least squares (OLS). A share-equation gave mostly insignificant results and the total savings and total income were regressed instead. **Table D3** in the appendix shows the estimates. The parameter signs were as expected with savings increasing more than proportionally with income. The squared-income term was, however, very small and insignificant, which indicated that a linear function would actually be a good representation of the relation between income and savings. The table also shows a regression for total expenditures (food and non-food) on income terms. The two equations do not add up because the budget constraint is not binding if disinvestment takes place.

8.3.3 Food and non-food expenditures (Step 2)

The food consumption model was estimated in the unrestricted form using OLS. **Table D4** in the appendix shows the results. The signs were as expected with a negative coefficient for income in the food equation and rural and farm households spending a larger share of their budget on food than urban and non-farm households. The coefficient on household size was negative in the food equation, indicating economies of scale in food consumption for larger households. The overall fit of the model was poor with an explained variance of only 13 percent.

8.3.4 Almost Ideal Demand System (Step 3)

Food items were aggregated into eight categories (**Table 8.1**) in accordance with the type of item, source (home produced vs. purchased), and nutritional content. Unit values for food categories were estimated as weighted averages using household specific expenditure shares as weights (Brosig 2000).

Table 8.1: Aggregate food items

Composition, average budget share and unit values

No.	Food category	Food items included	Budget share %	Unit values .00 Ush
1	Plantain	Plantain	11.5	1.181
2	Roots and tubers	Cassava, sweet potato, irish potato	22.0	1.175
3	Cereals	Maize (cobs, grains and flour), sorghum, millet, rice (mainly home-produced)	12.6	4.701
4	Legumes	Bean, groundnut, pea, sesame, bread (mainly purchased)	11.7	5.246
5	Animal products	Beef, pork, goat, other meat, fresh milk, fish, margarine / butter, egg, chicken	8.9	8.667
6	Purchased necessities	Sugar, salt, cooking oil/ghee	15.5	12.438
7	Fruits and vegetables	Spinach (‘dodo’), cabbage, onion, tomato, jackfruit, mango, passion fruit, etc.	7.0	3.189
8	Luxuries	Soda drinks, beer, tea and coffee, tobacco, restaurant food	10.6	10.433

Source: Estimated from the 1999-2000 UNHS

Note: All values are 10% trimmed means.

Demand equations were estimated using Zellner’s seemingly unrelated regression model. Parameter estimates were restricted for symmetry and homogeneity. The adding up restriction was satisfied by omitting one demand equation and calculating the coefficients of this equation from the imposed restrictions (*i.e.*, the sum of constant terms across equations is 1, the sum of price coefficients in each equation is zero, and the sum of income coefficients across equations is zero). **Table D5** in the appendix shows the results. Parameter values on income were positive and highly significant in all equations. The explained variance was relatively low and varied between 11 and 34 percent across the equations.

Average price and income elasticities were estimated as shown in **Table 8.2**. Price elasticities were as expected with all own price elasticities having a negative sign, and most cross price elasticities having a positive sign, except for category 8, which included mostly luxury goods. Income elasticities indicate whether goods are necessities ($0<\eta<1$) or luxuries ($\eta>1$). As expected, the income elasticities were above unity for animal products and luxury foods; plantain and cassava had relatively high income elasticities, slightly above that of cereals, which is somewhat surprising as cereals are commonly expected to substitute roots and tubers at higher income levels.

Table 8.2: Estimated elasticities

Income, own, and cross price elasticities

	Income elasticity[2]	Own and cross price elasticities[1]							
		1	2	3	4	5	6	7	8
1. Plantain	0.89	**-0.46**	0.18	0.12	0.03	-0.03	0.09	0.05	0.01
2. Roots & tubers	0.94	0.05	**-0.64**	0.15	0.07	0.04	0.12	0.08	0.01
3. Cereals	0.92	0.11	0.24	**-0.71**	0.05	0.02	0.10	0.05	0.11
4. Legumes	0.80	0.03	0.20	0.07	**-0.59**	0.08	0.11	0.11	0.04
5. Purchased Necces.	0.64	0.03	0.20	0.08	0.11	**-0.41**	0.08	0.08	-0.01
6. Animal products	1.26	0.03	0.20	0.06	0.06	-0.04	**-0.51**	-0.01	0.07
7. Fruits & vegetables	0.90	0.10	0.22	0.12	0.14	0.10	0.12	**-0.83**	0.08
8. Other	1.55	0.03	0.17	0.13	0.06	-0.05	0.13	0.04	**-0.47**

Source: Estimated from the 1999-2000 UNHS

Notes: [1] Compensated (Hicksian) price elasticities; [2]Income elasticity at average level of income.

8.3.5 Market prices

The 1999-2000 UNHS did not distinguish between farm gate (selling) prices and market (buying) prices of food products. To estimate these two different prices, sample household were divided for each food product into 'net buyers', when purchasing more than 50 percent of the consumed product, and 'home-consumers', when producing more than 50 percent of the consumed product. The unit values of net buyers were then assumed to reflect market prices, while the average unit values of both all households were assumed to reflect farm gate prices (as home consumers are not necessarily 'net sellers'). Prices were estimated for 46 food products and are shown in **Table 8.3**.

Table 8.3: Estimated market prices and farm gate prices [.00 Ush/kg]

Product	Cat.	Market	Farm gate	Product	Cat.	Market	Farm gate
Plantain	1	1.677	1.181	Beef	6	18.525	18.462
Sweet potato	2	1.663	1.290	Pork	6	14.241	14.222
Cassava, fresh	2	0.884	0.888	Goat meat	6	19.074	18.979
Irish potato	2	2.865	2.050	Other meat	6	14.147	14.121
Millet	3	4.627	3.807	Chicken	6	16.101	14.301
Sorghum	3	2.785	2.513	Fresh fish	6	12.665	12.616
Rice	3	8.969	8.838	Dry/smoked fish	6	10.802	10.786
Maize, grains	3	3.350	3.015	Eggs	6	18.360	17.927
Maize, cobs	3	3.059	2.387	Fresh milk	6	4.128	3.889
Maize, flour	3	4.941	4.635	Margarine, butter	6	32.605	-
Bean	4	4.418	3.600	Passion fruits	7	17.233	15.554
Sesame	4	5.188	4.888	Sweet bananas	7	2.365	1.829
Bread	4	12.276	12.229	Mangoes	7	1.651	1.630
Gnut, in shell	4	3.393	3.125	Oranges	7	3.500	3.083
Gnut, shelled	4	10.254	9.874	Onions	7	5.732	5.723
Gnuts, pounded	4	10.932	10.626	Tomatoes	7	2.435	2.426
Peas	4	5.466	4.363	Cabbages	7	1.400	1.358
Coffee	5	4.316	-	Dodo	7	1.624	1.675
Cooking oil/ghee	5	14.662	-	Other vegetables	7	2.291	2.206
Sugar	5	11.223	-	Infant foods	7	43.503	-
Tea	5	3.612	-	Soda/juice	8	15.207	-
Salt	5	4.416	-	Beer	8	19.586	-
				Restaurant food	8	17.879	-
				Other food	8	17.879	-

Source: Estimated from the 1999-2000 UNHS
Notes: Cat.=food category (1-8). Unit values for restaurant and other food expenditures are based on the weighted average of unit values in food category 8. All values are 10%-trimmed averages.

8.3.6 Food energy needs and intake levels

Data on age and sex-specific human weight levels and physical activity levels (PAL) for a developing country like Uganda were taken from James and Schofield (1990: 120) and are shown in **Table 8.4** and **Table D6**. Each food product was converted to energy equivalents. For this, consumption data from the 1999-2000 UNHS were combined with standard waste fractions and conversion factors as taken from Latham (1979, 1997).

Table 8.4: Human basal metabolic energy requirements by age and sex

In million joule/day

Age group	Male	Female
0 – 16	*Standard value*	*Standard values*
10 – 17	0.0732 * Weight + 2.72	0.0510 * Weight + 3.12
18 – 29	0.0640 * Weight + 2.84	0.0615 * Weight + 2.08
30 – 59	0.0485 * Weight + 3.67	0.0364 * Weight + 3.47
> 60	0.0565 * Weight + 2.04	0.0439 * Weight + 2.49

Source: James and Schofield 1990: 25

Note: The complete age and sex specific data are included in appendix **Table D6**

Food products were consumed in many different forms (*e.g.*, fresh, dried, or processed beans). Each food product was therefore first converted into energy units (joule) and the median energy content was calculated for each food product (*e.g.*, joule/kg maize). Nutrient equivalents for aggregate food categories were calculated using the relative contribution of each food product to the total value of the category as weights. **Table 8.5** shows the resulting median nutrient values for all food products.

Table 8.5: Median energy and protein values of food items

Separated by home produced and purchased (aggregated) items

	Purchased	Energy MJ/kg		Home-produced	Energy MJ/kg
1.	Plantain	3.95	1.	Plantain	3.95
2.	Roots & tubers	5.13	2.	Sweet potato	4.56
				Cassava	5.61
3.	Cereals	14.74	3.	Maize	14.98
				Millet	14.02
				Sorghum	14.52
4.	Legumes	15.28	4.	Bean	14.06
				Groundnut	18.45
5.	Necessities	23.53	5.	-	
6.	Animal products	7.71	6.	Chicken meat	5.82
				Eggs	5.27
				Fresh milk	3.31
7.	Fruits & vegetables	1.43	7.	Fruits	2.58
				Vegetables	1.22
8.	Other	4.61	8.	-	

Source: Estimated from the 1999-2000 UNHS and conversion factors from Latham (1979, 1997)

Note: 1 MJ=$1*10^6$ joule=23.9 kcal.

The household model optimizes consumption in value terms. The optimization procedure is, however, indifferent about the quality or physical quantity of consumption. For example, the optimization procedure sees no difference between 1000 shillings of chicken and 1000 shillings of beef if both are in the same category, even though the first would buy 1 kg while the second would buy only 500 grams. To circumvent this indifference, only one buying activity per food category (*e.g.*, 'animal products') was included, which was given an average nutrient content (**Table 8.5**).

8.3.7 Opportunity cost of farm labor and migration

If income from farming is low and the expected income outside agriculture is greater, then agents might decide to leave farming, perhaps to move to a town. The present model optionally allows for out-migration by specifying a threshold, called a pull factor (Berger 2001). If the income from farming divided by the agent's opportunity cost is below this pull factor then the agent will migrate; but if above, then the agent will stay. For example, a pull factor of 0.5 means that the agent will migrate if income from farming is only half the opportunity cost.

The opportunity cost of farm labor is difficult to estimate for it is a function of an unobserved perceived probability of finding a job and an expected wage rate. The opportunity cost for the potentially migrating farmers can be estimated as the product of the marginal unemployment rate and the wage rate for new arrivers.

The 1999 UNHS did not ask if (and when) a person had migrated from a rural area, so early arrivers could not be discerned from the sample. Instead, **Table 8.6** shows the median wage rates of four alternative employments as estimated from the 1999-2000 UNHS. The weighted average of these four wage rates is about 800 thousand shillings per year. This median wage rate probably overestimates the wage rate of new arrivers, so perhaps half the median would be more realistic. The (1997) urban unemployment rate was about 22 percent (UBOS 2000: 41), which can be interpreted as a probability of 78 percent of finding employment. Multiplying this probability by half the median wage rate gives an opportunity wage of 312 thousand shillings per year.

Problems arise validating the migration pattern against observed data. Farm household surveys typically ask questions only about the current household members and not about members who migrated. In absence of such data, it is impossible to validate out-migration in the model. In addition, opportunity costs are not equal for all farmers in reality. It is often observed that the best educated rather than the least educated farmers leave agriculture first. The use of a uniform opportunity costs would create the opposite effect by pulling out those farmers with the lowest opportunity cost. It was therefore decided to set the migration pull factor to zero and hence not to allow migration in the baseline scenario.

Table 8.6: Wage income of alternative employments

Employment	Median [.000 Ush/year]	Number of observations
Fishing	500	5
Mining and quarrying	720	2
Construction	800	30
Manufacturing	840	41
Total (average)	800	78

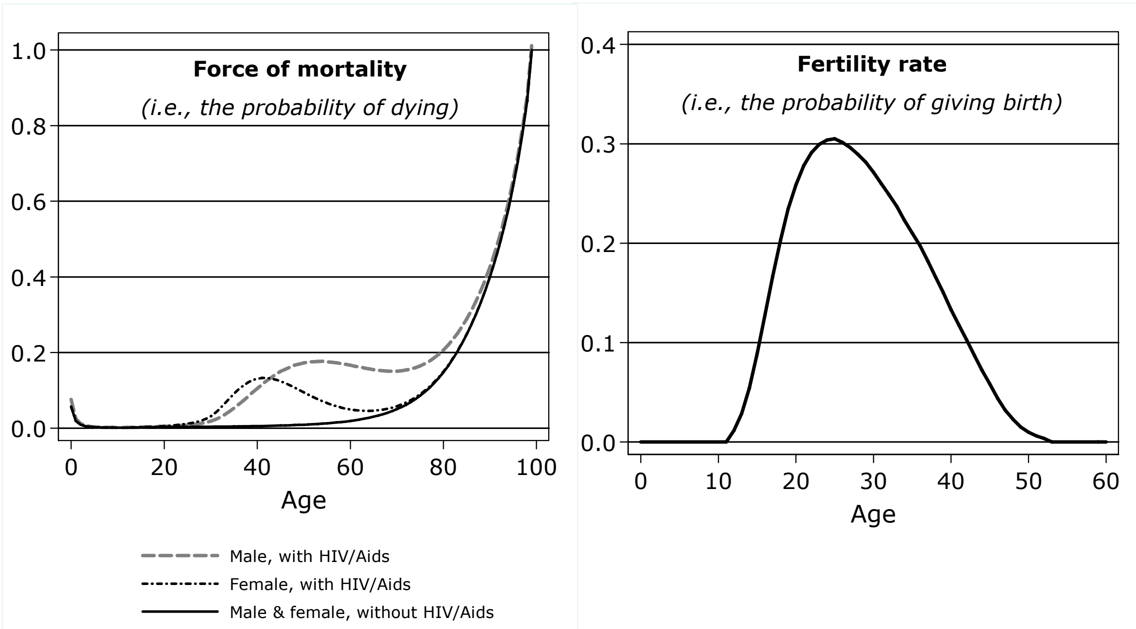
Source: Estimated from the 1999-2000 UNHS

8.3.8 Population growth and HIV/Aids

The effect of HIV/Aids on agricultural production was assessed with a random demographic simulation component. Demographic data were derived from Feeney and Zaba (2001) and Nunn *et al.* (1997) and included in **Table D6** of the appendix. Based on a survey of 15 villages in Southwestern Uganda, Nunn *et al.* (1997) estimated a life expectancy for people uninfected with HIV/Aids of 56.5 years for men and 60.5 years for women. Age-specific mortality rates for the uninfected were estimated using a Brass general standard lifetable with level parameters set to -0.55 for females and -0.40 for males using an Excel workbook from Feeney and Zaba (2001). Age specific fertility rates were also taken from Feeney and Zaba (2001).

The age-specific mortality rates for people infected with HIV/Aids was calculated from the incidence ratio and force of mortality as provided by Feeney and Zaba (2001). **Figure 8.3** shows the two demographic trends included in the model. The left diagram shows the force of mortality for males, which is the probability of dying at a certain age. The solid line represents mortality rates without HIV/Aids, which is about equal for males and females, and the dotted lines represent mortality rates with HIV/Aids, which are sex-specific. The right-hand diagram shows the fertility rate, which is the probability that a woman of a certain age gives birth. In the model used here, no adjustment is made for the change of infection rates over time; the HIV/Aids scenario hence represents the situation at the end of the 1990s.

Figure 8.3: Demographic model for Uganda



Source: Based on Feeney and Zaba (2001)

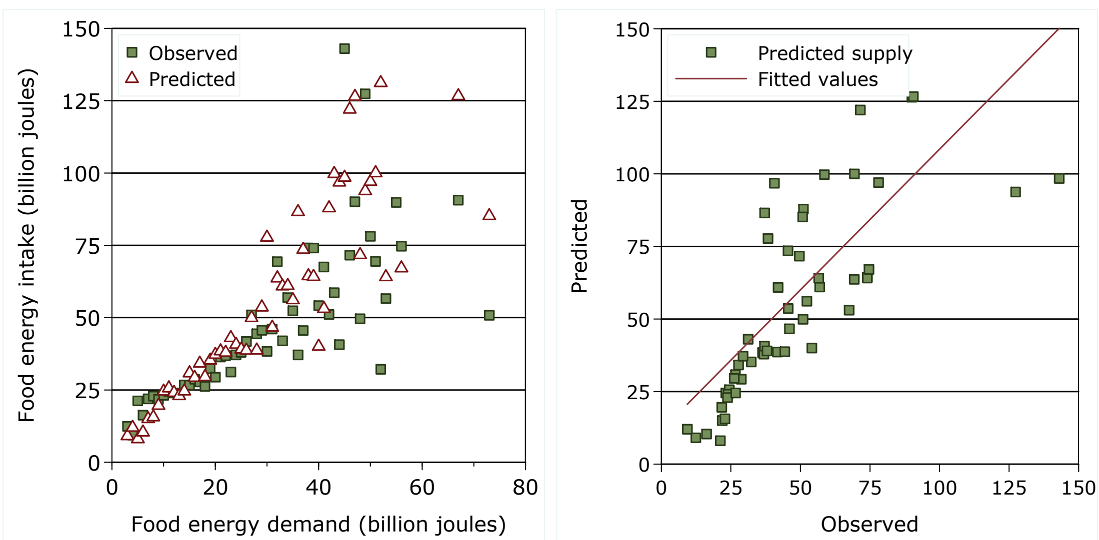
8.4 Validation of results

The consumption component was validated using three methods: (1) the three separate regression models were scrutinized for the signs and significance of parameter estimates and their explained variance as measured by R-squared; (2) the overall fit of the combined models in a three-step budgeting system was examined at given levels of disposable income; and (3) the overall fit of the three-step budgeting system when integrated into the MP-MAS (and hence endogenous levels of disposable income) was examined. Since the first method was already discussed (Section 8.3) this section focuses on the two other methods.

To compare the sample observations from the 1999-2000 UNHS with predicted observations based on the estimated models, the sample households were divided into categories of household size. Households with a size of 2.5–3.5 billion joule were taken into category 2, households with a size of 3.5-4.5 into category 3, etc. Average household size, income levels, food expenditures, and food energy consumption were calculated for each category from the survey.

The validation method tested how good the three combined regression models can predict the observed level of food energy consumption. For this, food energy consumption was predicted from the average household size and income levels per category of household size. It appeared that the predicted level of food energy consumption was lower than the observed level, which was mainly because the Working-Leser model underestimated total food expenditures. Adjusting the composite constant in the Working-Leser model from 0.70 to 0.95 and multiplying the food energy contents of all food products by 1.6 brought the predicted and observed level of food consumption closer together. The left diagram in **Figure 8.4** plots both the observed and the predicted values after adjustment against categories of household size; the figure shows that the model is well able to predict food energy intake for smaller households as the unexplained variance is small, while for very large households (larger than 35 billion joule) the unexplained variance is greater. It is, however, noted that 96 percent of the sample households were smaller than 35 billion joule and hence the model fit for almost all households is very good.

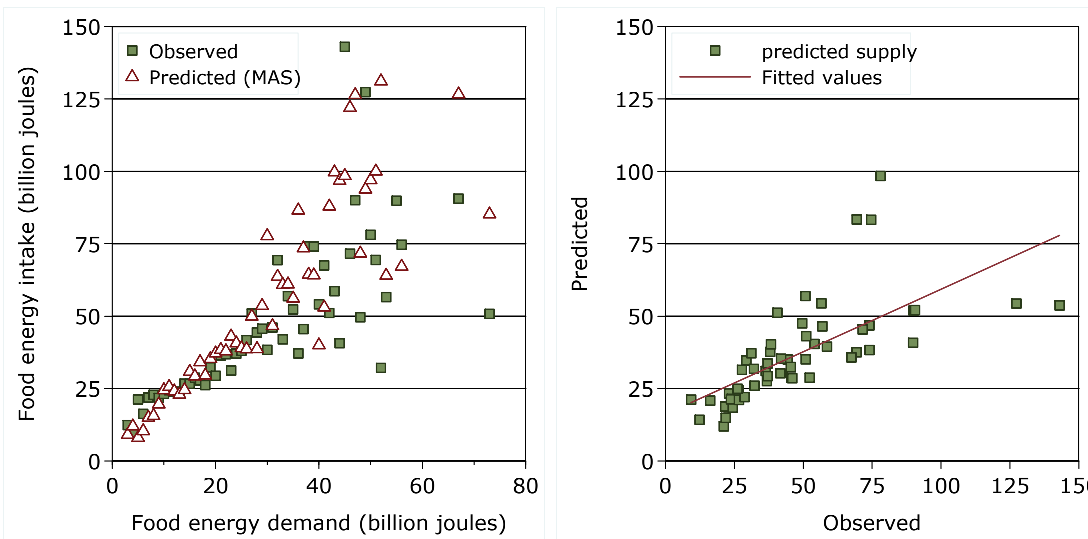
Figure 8.4: Validation of the three-step budgeting system at fixed income levels



Notes: Household sample observations are divided into 53 groups by size. Model fit for the right diagram: coeff=1.15; SE=0.06; R^2=0.86.

The last validation method scrutinized the consumption side of the complete MP-MAS. This differed from the previous method in that income was now endogenous and food energy consumption was simulated for all agents. Agents were categorized by household size in the same way as above. The left diagram of **Figure 8.5** plots these categories against food energy consumption for both the survey observations and simulation outcomes. The figure shows a relatively close match between model results and survey estimates, especially for smaller households, while the fit is worse for larger households as the model tends to underestimate disposable income levels for large households. Yet, the share of large households in the village survey (that was used to calibrate the agent populations) was larger with only 78 percent of the households smaller than 35 billion joule. As a result, the linear fit between predicted and observed values gave a coefficient less than unity (0.69). It is, however, noted that this linear fit was based on an unweighted regression of household categories which under-represents smaller household sizes for which the fit was much better.

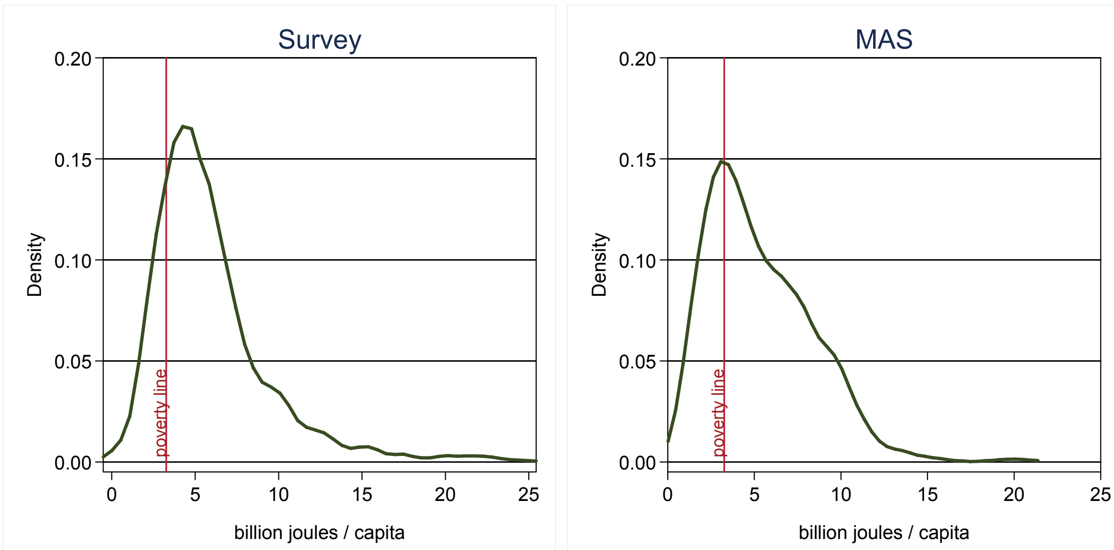
Figure 8.5: Validation of the three-step budgeting system in the MAS



Notes: Model values are simulated means for 520 agents over the first 5 years of the simulation assuming constant soil fertility levels. Note that unlike **Figure 8.4**, income levels are endogenously determined in the model. Model fit for right diagram: coeff=0.69; SE=0.04; R^2=0.87.

The above validation relied on average values per category of household size. A complementary method is to scrutinize the distribution of food energy consumption at the agent level. A kernel density function was used for this purpose. **Figure 8.6** plots the kernel distribution for both the survey (left diagram) and the MAS simulation (right diagram). Again, the kernel estimates were not fully comparable because the survey estimation was based on a much larger area and population and the average household sizes were different from the study villages. The figure shows that both distribution functions are similar in shape; yet, the simulated energy consumption is a little above the observed food energy consumption with the result that the share of poor agents (those left of the indicated vertical poverty line) is somewhat larger than the share of poor farm households in the survey.

Figure 8.6: Validation per capita food energy consumption



Notes: In male adult equivalents. Epanechnikov kernel used. The vertical line indicate the poverty line of 3.259 billion joule, at which agents' food energy demand equals supply. Survey estimate based on the farm households in southeast Uganda as recorded by the 1999-2000 UNHS. MAS based on the baseline scenario averaged over 15 years.

8.5 Summary

A three-step budgeting model was estimated for household expenditure decisions, including the decisions to save, to allocate income to food, and to spend the food budget on specific food categories. A disinvestment model was furthermore included to allow agents to substitute current income for future income in the face of a consumption shortfall. These three models were implemented in a mathematical programming framework and used to simulate poverty in terms of food energy consumption. The consumption model was validated in three ways from regression estimates to MAS simulation. The model fit was good on average but food energy consumption for large households might have been underestimated. Better quality expenditure data at the level of the study villages could improve the model.

9 Simulation results

9.1 Introduction

This chapter uses the empirically calibrated MAS to analyze crop yield gaps at the farm household level. The chapter has six sections. Section 9.2 defines the baseline scenario and analyzes it with respect to soil fertility decline and population growth. Section 9.3 decomposes the yield gap in proximate and underlying factors and assesses its relevance to food security. Section 9.4 analyzes the impact of crop breeding and Section 9.5 the effect of the HIV/Aids epidemic. The chapter ends with a summary of the major findings.

9.2 The baseline scenario

9.2.1 Defining the baseline scenario

The baseline scenario is the simulation run that reflects the present situation and the present sources of change. The baseline assumes that current trends in demography, soil processes, and the diffusion of innovations will continue and that there are no new external interventions. The model was calibrated in the previous chapters in such way that the baseline scenario reflects the observed conditions from the various surveys.

Because it was impossible to calibrate the monetary liquidity reserves of the agents to observed data (**Chapter 7**), all agents were given a fixed amount of liquidity in the first year. The model was run for a 16-year period, in which the first year was used to let agents adjust their liquidity reserves and omitted from the presentation.

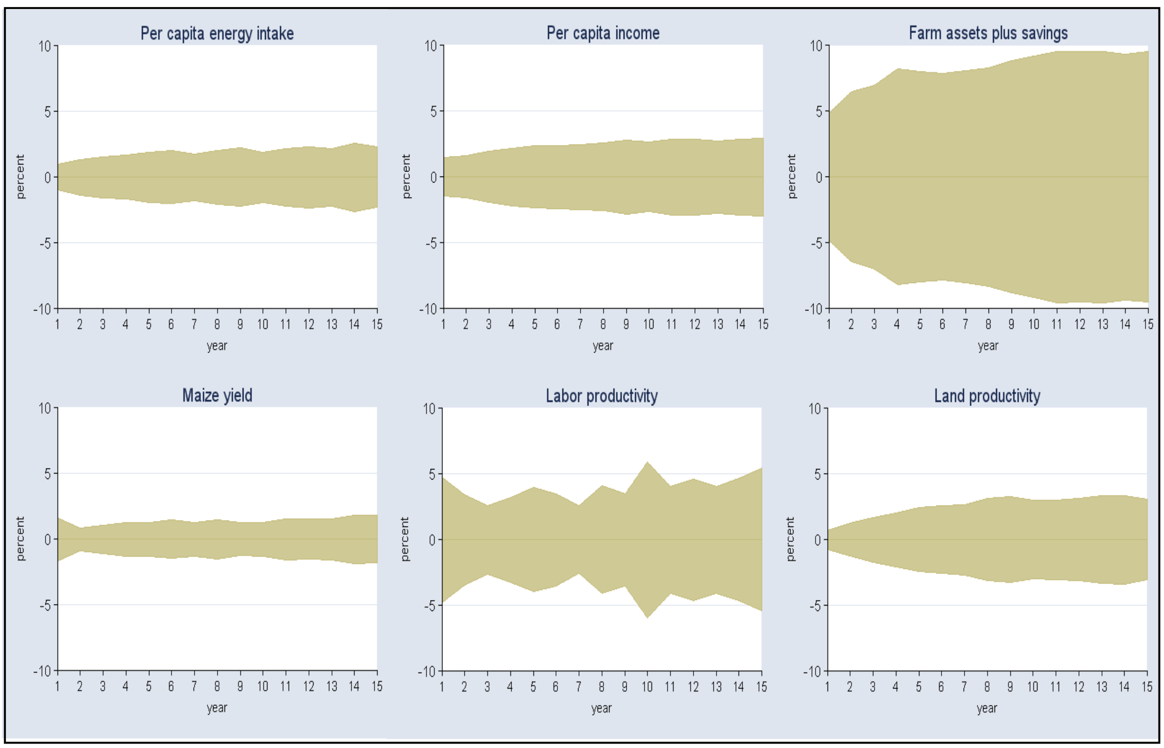
9.2.2 Sensitivity of the baseline to initial conditions

The Monte Carlo approach generated many possible and statistically consistent agent populations. These alternative agent populations can be used to check the robustness of simulation experiments to variations in initial conditions. The model was run for fifty different agent populations and average values for a selection of important variables were calculated for each. The variation in these averages indicates the 'noise' in simulation results. If this noise is substantial then simulation results are sensitive to initial conditions. Variables with little noise are therefore preferred over variables with much noise.

The noise was quantified as the standard deviation of the average over all fifty agent populations. The noise estimate was normalized by dividing it by the meta average and expressed as a percentage deviation to make it comparable between variables. Hence, a 5 percent noise for a variable implies that the average varies +/- 5 percent depending on the initial conditions of the agent population. Note that this noise estimate is constructed from averages and its accuracy hence depends on how close the distribution of variables comes to a standard normal distribution.

Figure 9.1: Variation of simulation results to differences in initial conditions

Standard deviations expressed as a percentage of the normalized mean



Notes: Graphs show variation in simulation outcomes due to variations in initial populations, and not the variation between agents within these populations. Number of agent populations is 50. Total number of mathematical programs solved for this is 1.17 million.

Figure 9.1 show the estimated variation to initial conditions. The diagrams have in common that the variation gradually increases over time. Model outcomes are hence more robust in the initial years of the simulation than in later years. Yet, the patterns of variation are different for each variable. The variation in per capita incomes and per capita energy consumption is relatively low at 1-3 percent; also, the variation in land productivity and especially maize yields is low. These variables are therefore robust to initial conditions. The picture is different for labor productivity and especially farm assets. The variation in farm assets is wide because assets are directly allocated through the Monte Carlo procedure, while the other variables measure outcomes, which are a function of all assets and agent characteristics together. Labor productivity varies more than land productivity because the composition of land sizes is constant between populations as only agent characteristics (including the age and sex composition) vary.

9.2.3 Baseline dynamics: soil fertility decline and population growth

This section analyzes how the baseline scenario is driven by the joint dynamics of soil fertility changes and population growth. To do this, each of these two dynamics was switched off in turns and compared to the baseline in which both dynamics were switched on. **Figure 9.2** illustrates this exercise. The left diagram shows changes in soil organic nitrogen while the right diagram shows changes in population.

Figure 9.2: Soil fertility decline and population growth.

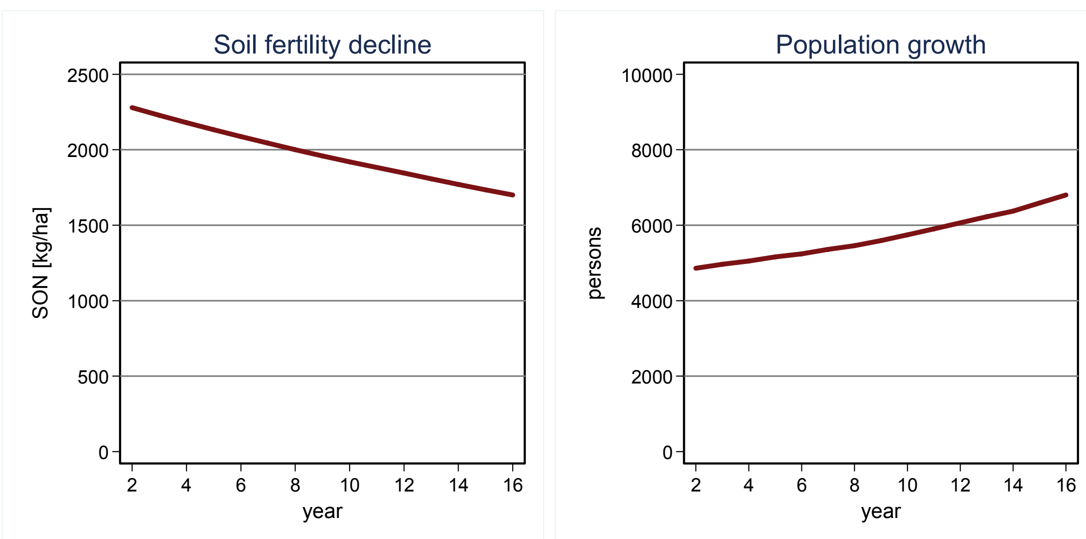
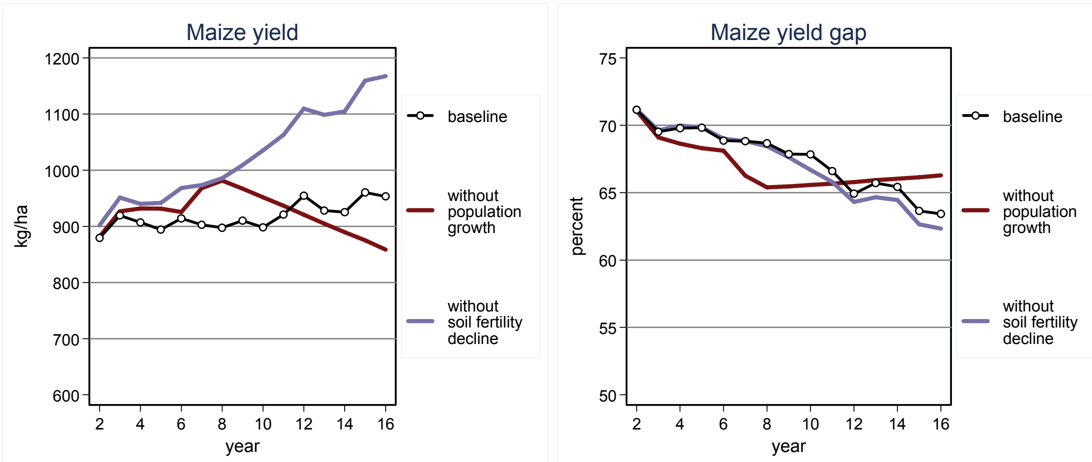


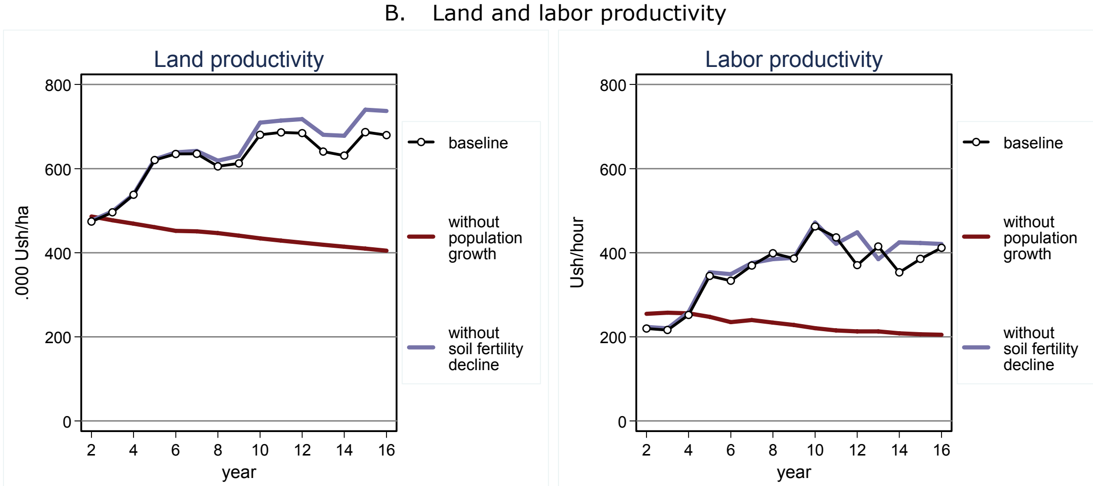
Figure 9.3 shows how these scenarios impact on a range of indicators that were introduced in **Chapter 3**. Indicators include maize yields, partial productivity indicators, and the fulfillment of private and social objectives.

Figure 9.3: The effect of population growth and soil fertility change on the baseline

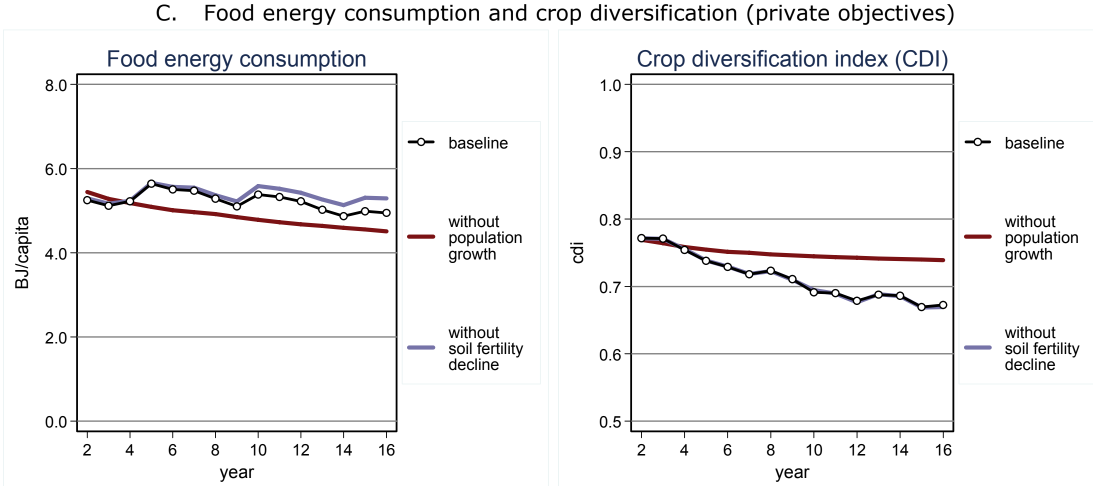
A. Maize yields and the width of the yield gap



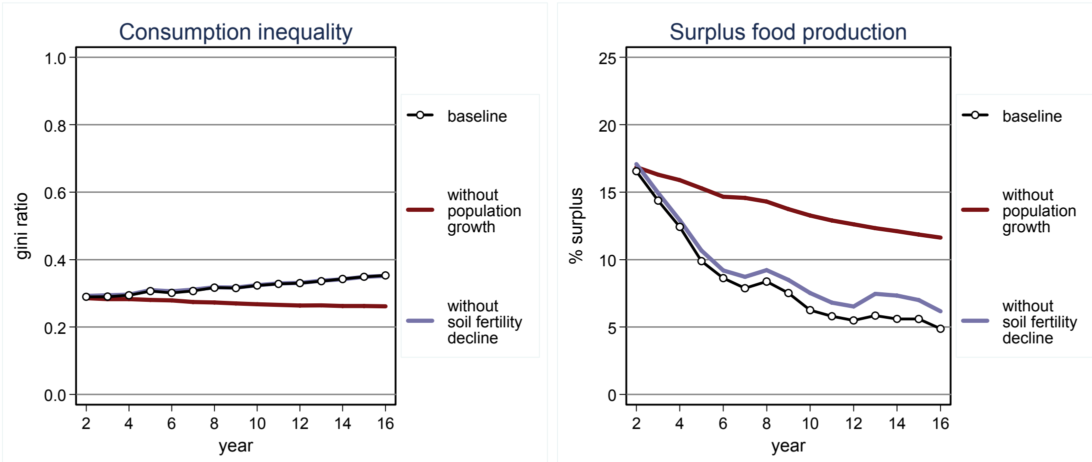
B. Land and labor productivity



C. Food energy consumption and crop diversification (private objectives)



D. Inequality and surplus food production (social objectives)



Notes: Surplus food production is calculated as the difference between food production and food consumption in energy units and expressed as a percentage of total production. Food energy intake is per male adult equivalent per year.

The most important thing these diagrams show is that population growth drives the baseline scenario while soil fertility decline has only a minor impact on the trend and variation in the baseline. Trends in about all variables closely follow the trend in the scenario without soil fertility decline, while the trend in the scenario without population growth is more distant.

Although soil fertility decline depresses land and labor productivity, population growth compensates for this decline and leads to an overall positive growth in the simulated land and labor productivity. The level of food energy consumption does not clearly decrease in spite of soil fertility decline and population growth. Most remarkable, per capita levels of food energy consumption are above the scenario with population growth than in the scenario without population growth. Without population growth, poverty would worsen as soils degrade, yet population growth seems to compensate for the loss in soil productivity.

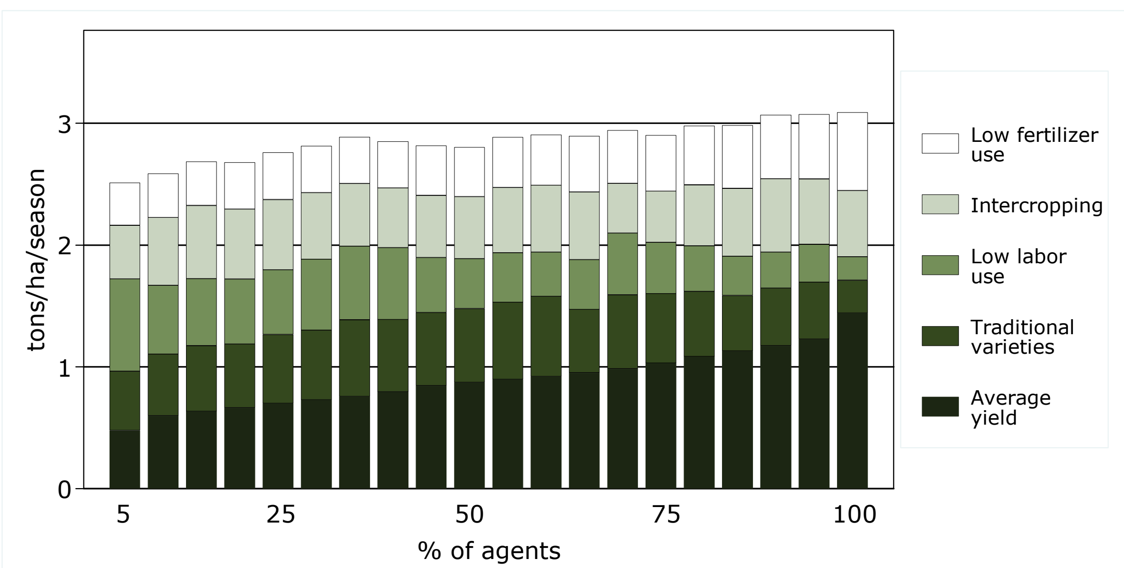
The last diagrams in the above figure show the strong effect of population growth on the social objectives of equality and surplus food production. Population growth increases the inequality between agents and reduces the amount of food sold on the market.

9.3 The maize yield gap

9.3.1 Decomposition in proximate factors

Figure 9.4 shows the decomposition of the maize yield gap in five proximate factors as was explained in **Chapter 3**. The dark bars on the bottom show average maize yields as simulated in the baseline scenario. Values for each agent were averaged over 15 years. Agents were subsequently ranked by their average maize yields and divided into 20 equal groups. The left most bar shows the group of agents with the lowest average yields, while the right most bar shows the group of agents with the highest average yields. The bars stacked on top of the average show the potential increase from four factors: using improved maize, using maximum amounts of labor, growing the crop in pure stand instead of intercropping, and using 100 kg of mineral fertilizer per hectare. The figure shows that each of these factors account for about a quarter of the yield gap. The total height of each bar shows the maize yield potential faced by the agents. Agents with higher average yields tend to have a higher yield potential.

Figure 9.4: Decomposition of the maize yield gap in proximate factors



Notes: Simulated average values per category of agents over the first 5 years in the simulation. Two varieties are included: traditional maize and improved maize. Optimal fertilizer level is 46 kg N. Yield potential under agent conditions is estimated for maximum levels of labor and fertilizer use, grown in pure stand and with the best variety. The effect of pests, diseases, and weather is excluded from the estimation and may account for remaining gap to the genetic yield potential, which is about 4.5 tons/ha/season.

The result is remarkable as intercropping and low labor use account for about half of the exploitable yield gap. Both these factors are not normally included in yield gap decompositions based on crop yield experiments; yet, **Figure 9.4** shows that these two factors are very important.

Figure 9.4 furthermore shows a wide variation in average simulated maize yields. Extension services could use this figure to estimate potential increases in maize yields from further adoption of improved varieties, or use it to promote mineral fertilizers. They would probably want to target those farm households with the lowest average yields, meaning those with the widest yield gap, as these have most to gain and are probably food insecure. Section 9.3.3 will address whether this hypothesis is correct.

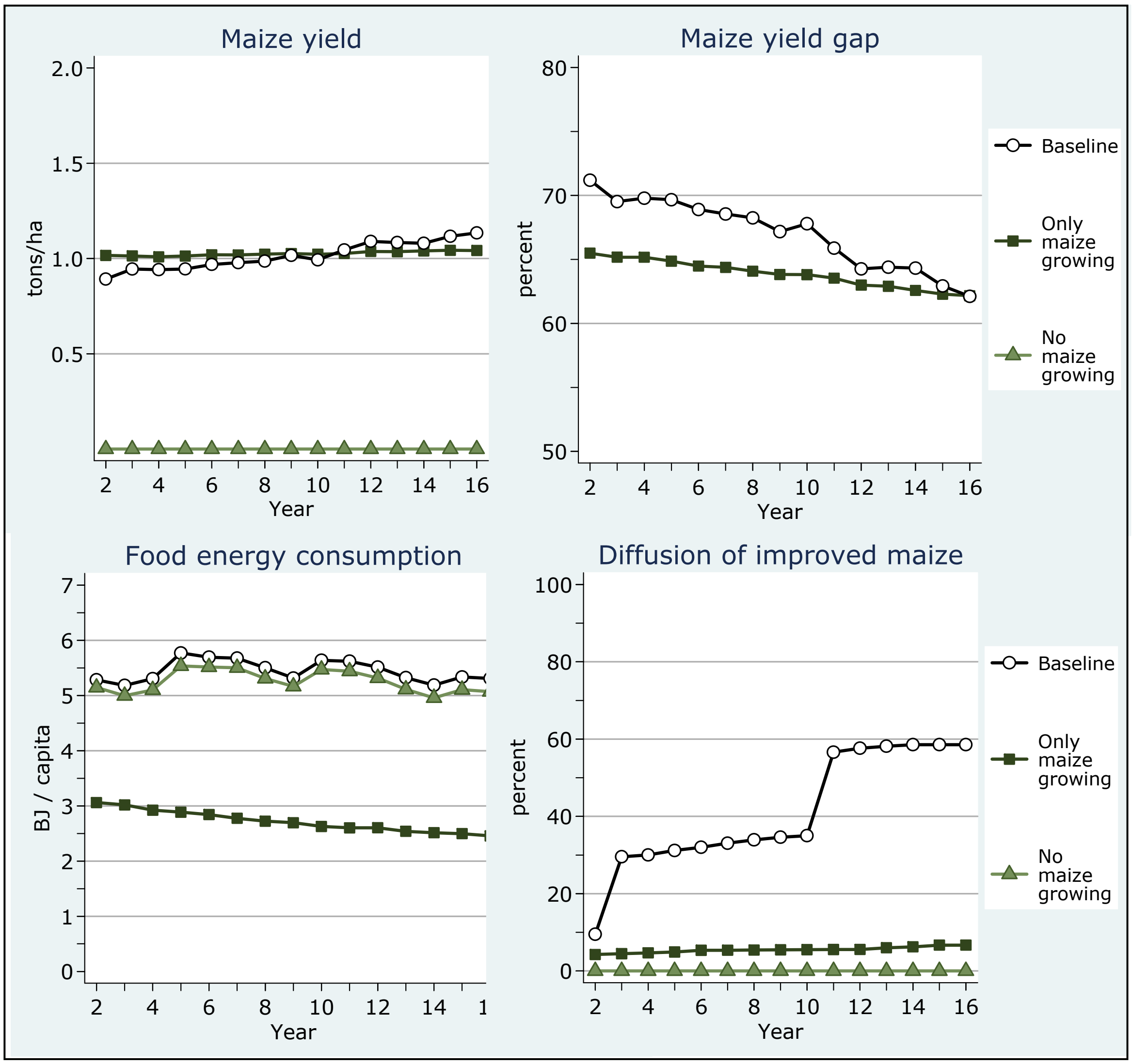
9.3.2 The maize yield gap and farm performance

The existence of maize yield gaps does not necessarily signal inefficiencies. Inefficiency refers to the performance of a farm holding, not to the yield of a single crop. Three scenarios were used to gauge the effect of maize yields on farm performance. The first scenario is the baseline as described above. In the second scenario, the opportunity cost of all farm activities other than maize growing was set to zero; hence, maize growing was the only means for an agent to satisfy its food energy needs. For this, the rotation constraints on maize were relieved so that agents could plant all their land to maize. In the third scenario, which is the opposite of the second scenario, the agents could grow any crop but maize. **Figure 9.5** compares the three scenarios with respect to average maize yield, the yield gap, the per capita food energy intake, and the adoption of improved maize.

If agents would only grow maize then maize yields would be about 10 percent above those in the baseline scenario during the first years of the simulation. Yet, in the 'only maize scenario', the average agent would fall below the poverty line of 3.3 billion joule/capita as shown in the lower left diagram. As agents are hence food insecure, livestock is sold and savings are consumed, which impedes on investments, including the purchase of improved seeds. The diagram on the lower right hence shows a stagnant diffusion of improved maize varieties in the scenario with only maize growing.

In comparison, if agents would not grow maize at all, the impact on poverty would be much less severe (-4 percent). The reason is that agents substitute other crop and livestock activities for maize growing, *i.e.*, the opportunity costs of maize growing is relatively high as the cash and in-kind income generated from these activities is only slightly less than what was generated by growing maize.

Figure 9.5: The baseline with, without, and with only maize cultivation



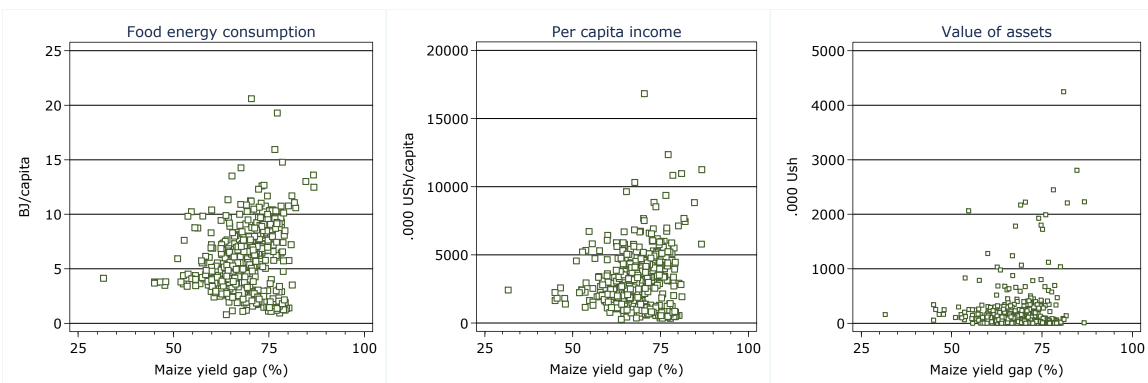
Notes: The maize yield gap is the average yield as a percentage of the maximum yield of the improved maize variety. Per capita food energy consumption is expressed in male adult equivalents.

These scenarios are rather synthetic experiments, yet they point to three important observations with empirical relevance. First, a diversified farm operation is important and contributes to food security; hence, the promotion of the maize area expansion beyond a certain point might induce inefficiencies. Second, a narrow yield gap does not mean a better performance. Third, the importance of maize for food energy consumption cannot be judged from the area it occupies or the yield it generates.

9.3.3 The maize yield gap vs. economic well-being and food security

The (ir)relevance of the maize yield gap to food security was further scrutinized in **Figure 9.6**. This figure shows three scatter diagrams correlating the size of the maize yield gap (on the x-axis) with three indicators of economic well-being and food security (on the y-axis). Values were averaged over 15 years and each dot in the diagrams represents a single agent. The diagrams show no clear relation between the variables, but it appears that agents with a low yield gap are more likely to be poor than rich and have lower values of farm assets. This result shows that the yield gap is not a good indicator of economic well-being or food security: agents with wide yield gaps are not necessarily poor and food insecure.

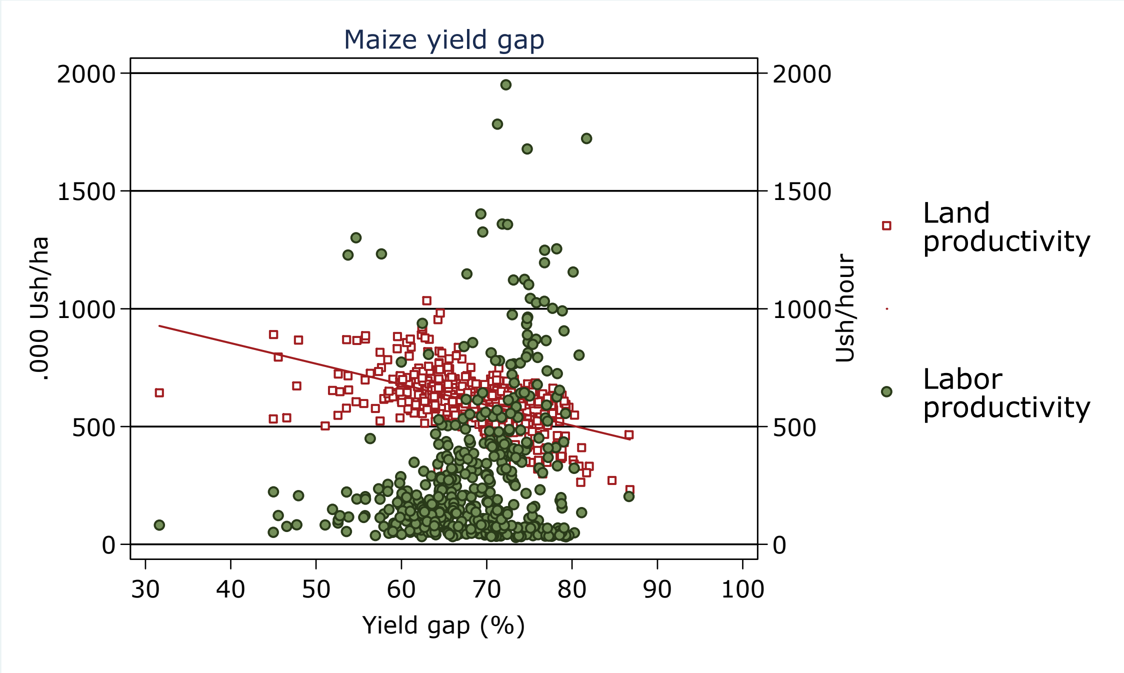
Figure 9.6: Scatter plot of yield gap vs. economic well-being and food security



Notes: Simulated average values over the 15 years using the baseline scenario. Food energy intake and per capita income in male adult equivalents. Assets include livestock (valued in terms of live weight at farm gate prices) plus monetary savings.

Maize yields affect the economic well-being of agents through the returns to land and labor use. **Figure 9.7** focuses on this relationship by plotting land and labor productivity on two different y-axes against the maize yield gap on the x-axis. The circles stand for labor productivity and the squares stand for land productivity. The figure shows that agents with wider yield gaps have lower average land productivity but also tend to have greater average labor productivity.

Figure 9.7: Land and labor productivity correlated with the maize yield gap

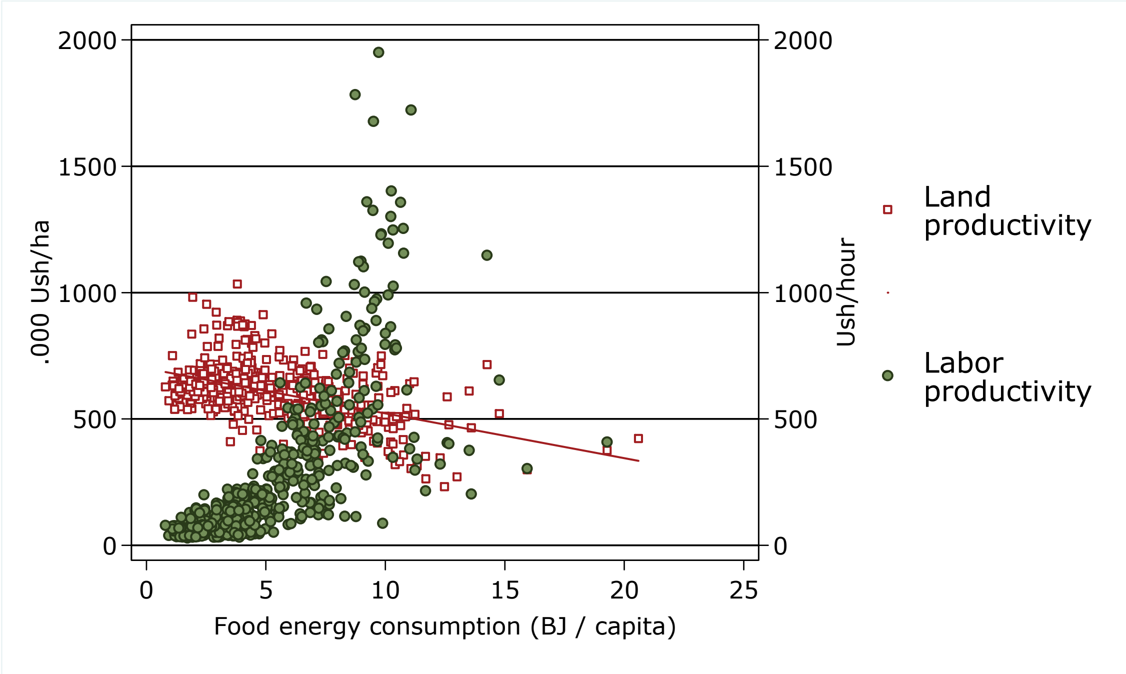


Notes: Simulated average values over the 15 years. Partial productivity indicators including crops and livestock production valued at farm gate prices. Maize yield gap is the average yield as a percentage of the farm level yield potential using maximum levels of input use (labor, fertilizer, variety and pure stand).

Figure 9.8 shows how this relates to poverty levels. It clearly shows that economic well-being is positively correlated with labor productivity, yet negatively correlated with land productivity. Hence, agents with high land productivity—*i.e.*, a low yield gap—are more likely to be poor, while agents with high labor productivity are more likely to be rich.

It appears that the high reliance on manual labor—agents use no mechanization or animal traction—makes labor a relatively scarce factor in production. The returns to labor are greater when distributing the available labor more equally over the land than when concentrating it on a few plots to get a high crop yield. In other words, there is a trade-off between high crop yields and high labor productivity, and agents choose for the second, and thereby lower their crop yields deliberately. The simulated crop yield gap due to low labor use in **Figure 9.4** does hence not signal inefficiencies.

Figure 9.8: Land and labor productivity correlated with poverty

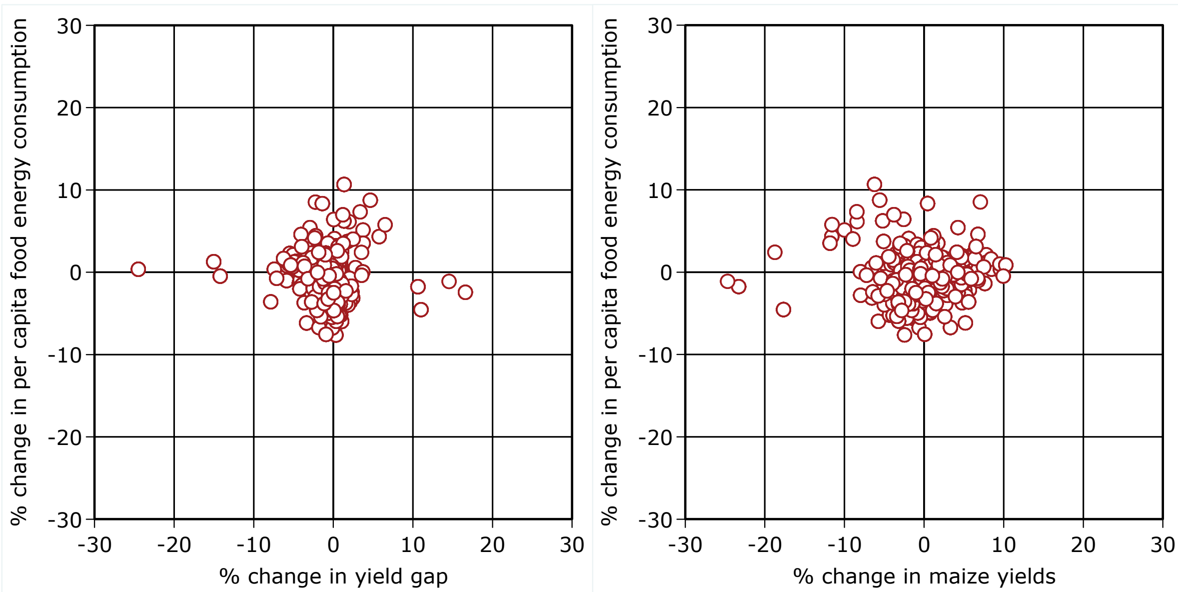


9.3.4 Maize yield gap dynamics

Though the width of the maize yield gap might be not be an indicator of economic well-being and food security, this does not answer the question whether a change in yield gap is a good indicator for change in economic well-being and food security. It could be hypothesized that as population grows, soil fertility will further decline, land will become scarcer in production, and the yield gap would hence gain in relevance.

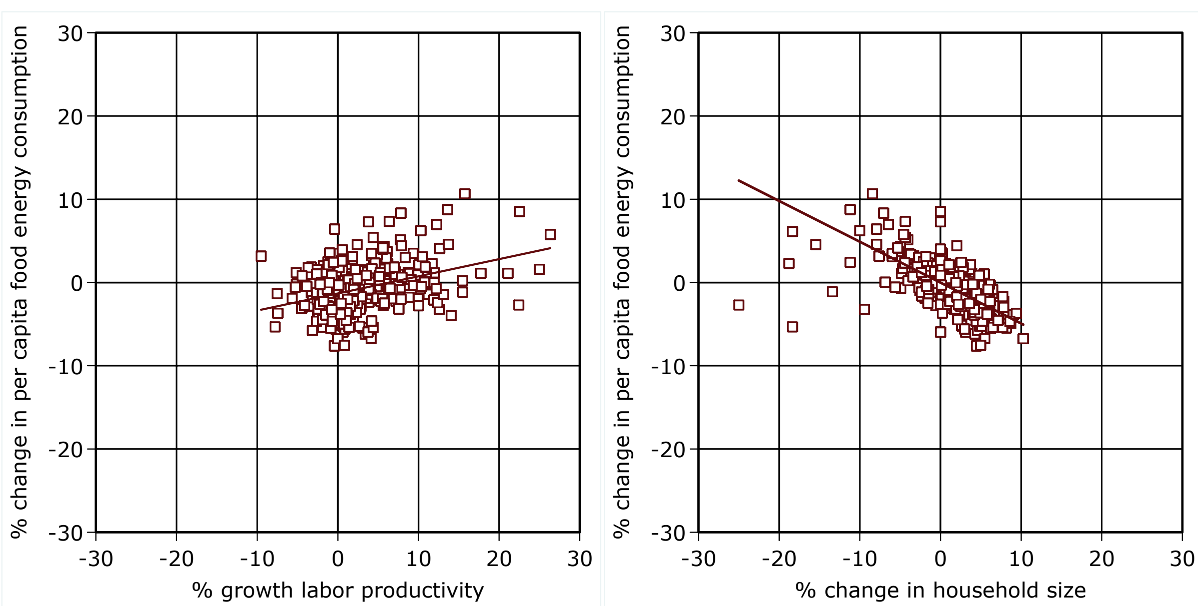
To test this hypothesis, the annual growth rate in the maize yield gap was estimated for each agent from a linear regression of the time variable on the logarithm of the yield gap. **Figure 9.9** plots the percentage annual growth in maize yield gap and maize yield against the percentage annual growth in per capita food energy consumption. The figure shows that there is no correlation between the growth in yield gap or maize yields on the one hand and the growth in food energy consumption on the other hand. A change in yield gaps is hence no good indicator for a change in food consumption levels.

Figure 9.9: Correlation between growth in maize yield (gap) and poverty



Notes: The average annual growth rate (g) calculated from the linear regression estimates ln(% yield gap) = a + b * time, with g = (exp(b) – 1) * 100. Confidence interval is 95%. Significance level is only shown for the variable on the y-axis.

Figure 9.10: Correlating between growth in labor productivity / household size and growth in poverty

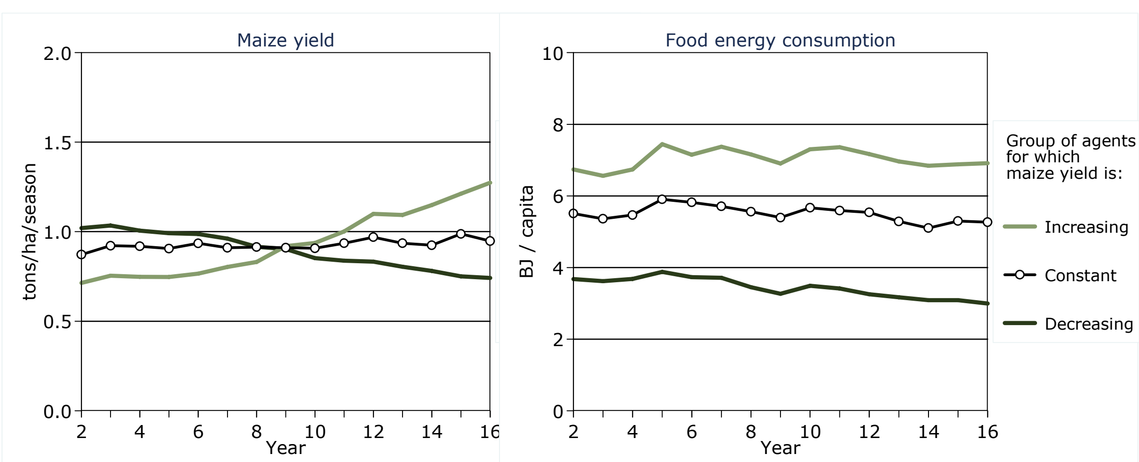


Note: see notes under **Figure 9.9**

The question emerges, what does correlate with changes in food energy consumption? Two possible candidates emerged from the above: labor productivity and population growth. The growth rates of these two variables are plotted against food energy consumption in **Figure 9.10**. The left diagram shows that food energy consumption positively and significantly correlates with the growth in labor productivity, while the right diagram shows a significant and negative correlation with growth in household size. Hence, growth in labor productivity and growth in household size are better proxies for growth in economic well-being than growth in maize yield and growth in maize yield gaps.

That is not to say that maize yields are unrelated to the food security of agents. Based on the above estimates of average growth in maize yields, all agents were divided into three groups: agents with a significant decline in maize yields (14 percent of the total); agents without a significant change in maize yields (59 percent), and agents with a significant increase in maize yields (26 percent). **Figure 9.11** plots the growth in maize yields for each group in the left diagram and combines it with the average level of food energy consumption for the same groups. The figure shows two things. First, those agents with an increase in maize yields do not experience an increase their food energy consumption (the line is flat). Second, agents with an increase in maize yields had a significantly greater level of food energy consumption already in the initial years of the simulation. This implies that wealthier agents are more likely than poorer agents to increase their maize yields.

Figure 9.11: Change in maize yields and change in per capita energy consumption

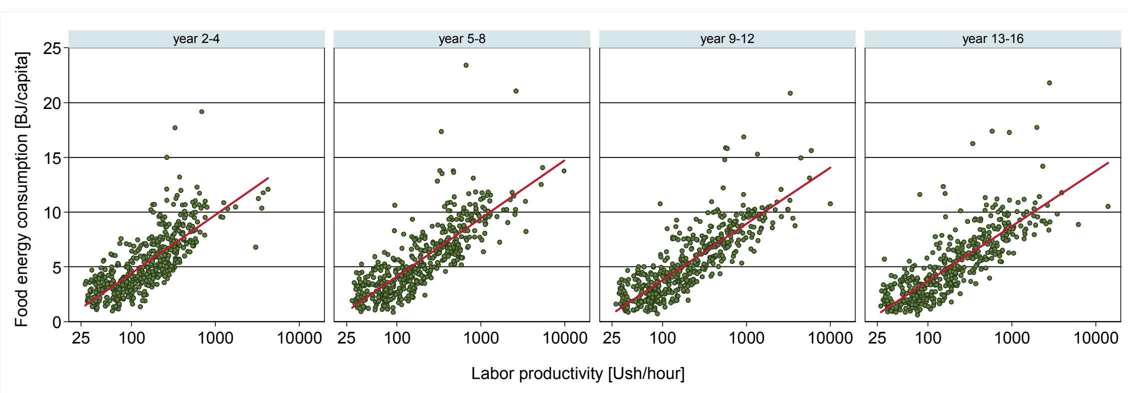


Because of soil fertility decline and strong population growth, one could hypothesize that land will increasingly become a scarce production factor. The importance of land productivity could therefore increase relative to that of labor productivity. This would imply that as land becomes scarcer, the negative correlation between land productivity and economic well-being would eventually turn into a positive correlation. **Figure 9.12** contains three rows of diagrams plotting land productivity, maize yield and labor productivity against per capita food energy intake. The 15-year simulation period was subdivided into four periods to reduce annual variation. The independent variables were furthermore put on a logarithmic scale so that their distributions come closer to a standard normal distribution, which allows a linear regression fit.

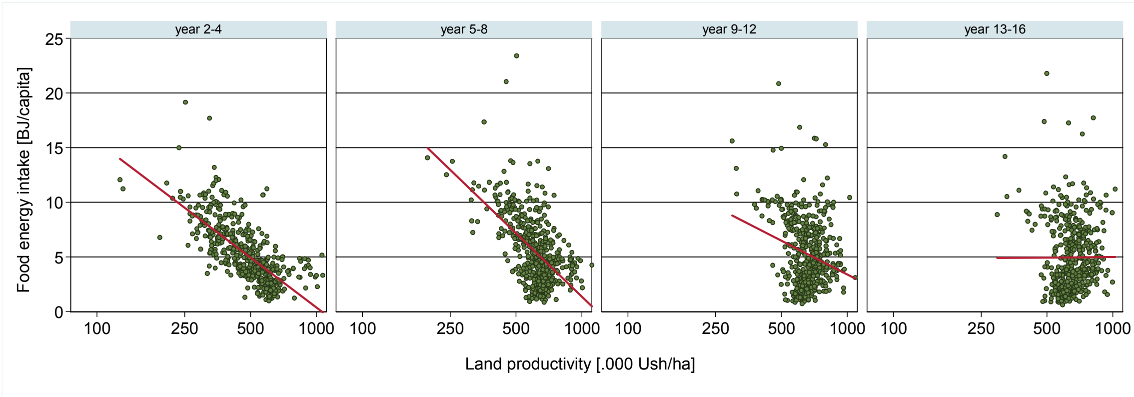
Moving from the left to the right, the diagrams show that the strong negative correlation between land productivity and economic well-being becomes cloudier over time and the linear fit flattens out. The negative correlation hence gradually disappears, though this effect is clearer total land productivity than for maize, for which no linear fit is drawn, as the correlation is insignificant. Yet, what does not disappear is the strong positive correlation between labor productivity and economic well-being, which is equally strong for each period. It is therefore concluded that in order to increase the economic well-being of agents, improvements in labor productivity are crucial.

Figure 9.12: Correlations between partial productivities and poverty for four periods

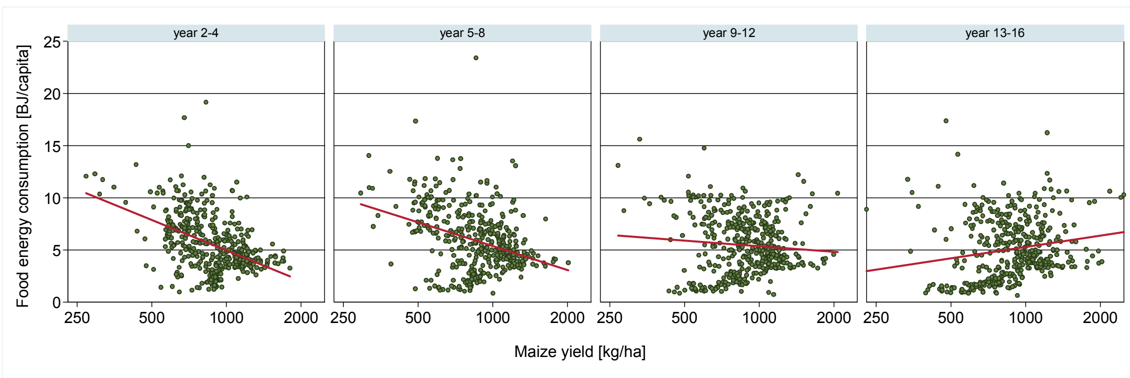
A. Labor productivity



B. Land productivity



C. Maize yields

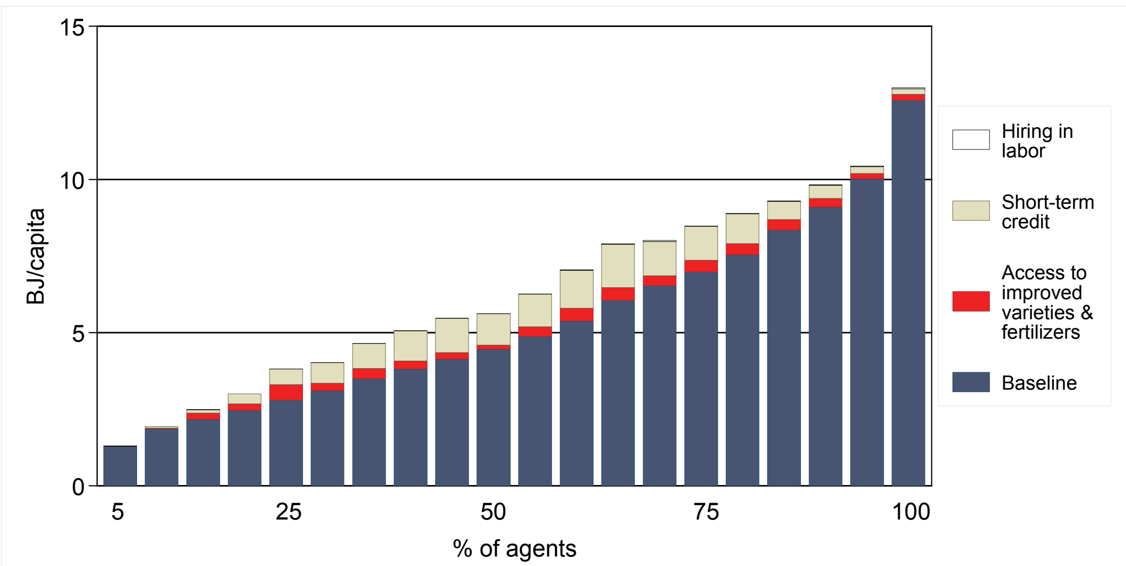


9.3.5 Decomposition in underlying factors

This section shifts focus from the maize yield gap to the broader issue of performance gaps. Performance gaps were decomposed in underlying factors using a factorial design of eight simulation experiments as described in **Chapter 3**. The three underlying factors included: (1) access to mineral fertilizers and improved maize varieties; (2) missing markets for credit; and (2) labor market imperfections. Per capita food energy consumption is the performance indicator used.

Figure 9.13 shows the results. The calculated performance gaps were averaged over 15 years for each agent. Agents were ranked by per capita food energy consumption and divided into 20 equal groups. Dark vertical bars on the bottom of the figure show the average level of food energy consumption for each group as simulated in the baseline. The left most bar shows the group of agents with lowest food energy consumption, while the right most right bar shows the group of agents with highest food energy consumption. The diagram shows that about 30 percent of the agents in the baseline fall below the poverty line of 3.3 BJ/capita. The stacked bars on top show the additional food energy consumption that can be attained when constraints on each underlying factor are relieved through a policy intervention. For example, the median group of agents has an energy intake of about 4.5 BJ/capita in the baseline. Giving this group full access to innovations would increase this groups' food energy intake by about 0.5 BJ to 5 BJ/capita and with additional access to short-term credit, its energy intake would even rise to 5.6 BJ/capita.

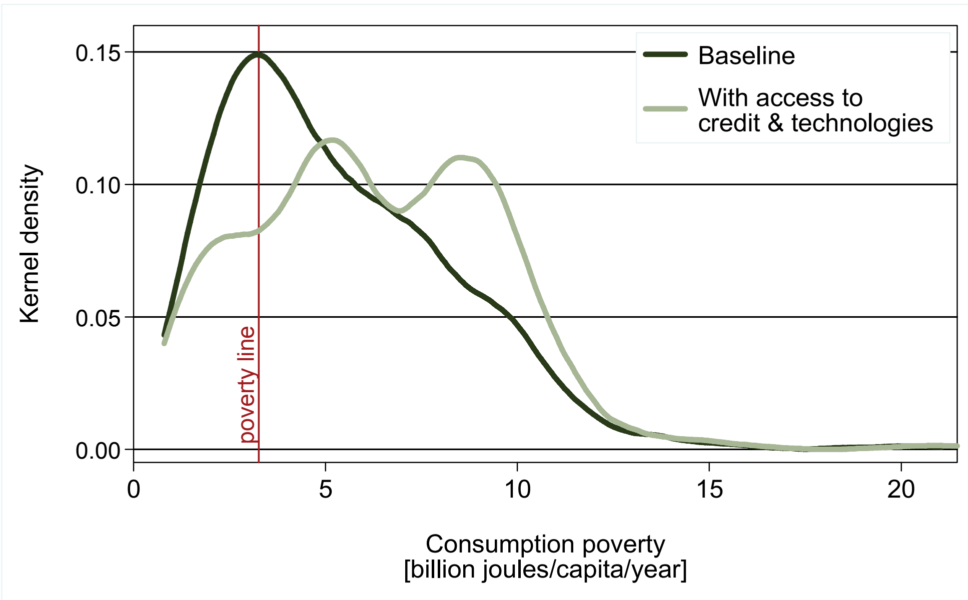
Figure 9.13: Decomposing the gap in per capita food energy consumption



Notes: Simulated average values over 15 years. Per capita income in male adult equivalents. Knowledge refers to the full access to innovations, including mineral fertilizers and improved seeds. Interest rate on short-term credit is 34 percent. Hiring includes both leasing a tractor for land preparation and hiring of labor up to 200 hours per farm per year.

The performance gaps in **Figure 9.13** are narrow relative to the maize yield gaps in **Figure 9.4**, but different from the maize yield gaps, the performance gaps directly relate to the well-being of the agents. In fact, the simulated change in poverty levels is enormous, as is shown in **Figure 9.14**. This figure overlays two kernel density distribution of poverty; one is the average baseline over the 15-year simulation period and the other is the scenario with full access to credit and innovations over the same period. The figure shows that the policy intervention would reduce poverty substantially, as the bulge of the poor have crossed the poverty line. The graph also shows the distributional effect of the policy intervention. It shows that neither the poorest nor the richest agents would benefit from the policy intervention, as the tails of the distribution have not moved. This policy intervention would hence not be suitable if the aim were to reach the poorest of the poor.

Figure 9.14: Kernel density graph showing the change in poverty distribution

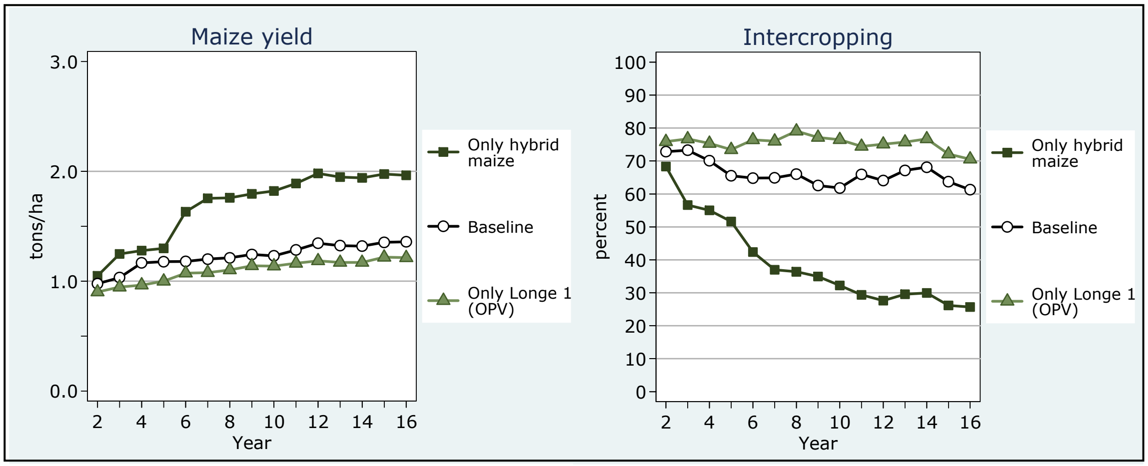


9.4 The impact of crop breeding

The baseline scenario included the diffusion of two maize varieties: first, an improved open pollinated variety, called Longe 1, which is an open pollinated variety for which agents need to purchase new seeds only once every five years. The variety is also suitable for intercropping with bean, groundnut, and cassava (**Chapter 7**). Second, a hybrid maize variety, which has a higher yield potential than Longe 1 but also has needs a more intense labor use and is not suitable for intercropping. To assess the impact of each variety, two scenarios were analyzed and compared to the baseline scenario. The first scenario simulated the effect in the case that there would be no Longe 1 variety but only traditional varieties and hybrid maize; the second scenario simulated the opposite case in which there would be no hybrid maize but only Longe 1 and traditional varieties.

Figure 9.15 shows the trends in average yields and intercropping for the three scenarios. Maize yields in the scenario with only hybrid maize increased rapidly with as the crop diffused, while maize yields in the scenario with only Longe 1 remained much lower. As agents cannot intercrop hybrid maize, the practice of intercropping maize rapidly diminished as is shown in the right diagram.

Figure 9.15: The impact of two improved maize varieties on maize yields and the share of land intercropped



It appears that the coexistence of both hybrid maize and Longe 1 impedes on each other's diffusion process: if either of the two varieties is excluded then the adoption of the remaining variety is more complete than in the baseline (**Table 9.1**). So, what variety is superior? Looking at the yield graph, one would say hybrid maize is superior; yet, **Table 9.1** shows that in fact, neither of the varieties is superior, as the average simulated food energy consumption does not significantly different between the three scenarios. It is therefore concluded that hybrid maize and Longe 1 act as substitutes rather than complements, with neither of them being superior to the other in terms of farm household performance. The results of this simulation experiment can explain the observed lack of adoption of hybrid maize varieties in Uganda.

Table 9.1: The impact of two improved maize varieties

Scenario	Maize yield (kg/ha)	Maize production (kg)	Adoption Longe 1 (%)	Adoption hybrid maize (%)	Inter-cropping (%)	Food energy cons. (BJ/capita)
Baseline	1,204	736.7	0.36	0.22	66.79	5.47
Hybrid maize only	1,638	751.8	0.00	0.55	41.25	5.45
Longe 1 only	1,082	680.6	0.53	0.00	75.51	5.50

Note: Average values over the 15-year simulation period.

9.5 The effect of HIV/Aids

The HIV/Aids epidemic has had a large impact on the labor supply through an increased mortality and a reduction in the available working hours of the surviving. To assess the impact of HIV/Aids on the farming system, three scenarios were designed that reflect the possible situation without the disease, which are then compared to the baseline that reflects the present situation with the disease. The three scenarios assumed age and sex-specific mortality rates as before the onset of the HIV/Aids epidemic and increased the total available labor supply by respectively 0, 20, and 40 percent in the first year of the simulation as compared to the baseline scenario. **Table 9.2** shows the results.

Table 9.2: Simulated effect of the HIV/Aids epidemic

	Scenario	Maize yield (kg/ha)	Maize yield gap (%)	Adoption of improved maize (%)	Value of livestock assets (.00 Ush)	Food energy consumption (BJ/capita)
1	Baseline (with HIV/Aids)	981	67.96	0.32	90.4	5.56
2	Pre-HIV/Aids mortality levels	1,026	66.65	0.43	108.0	5.34
3	Scenario 2 + 20% labor supply	1,089	64.76	0.52	132.1	5.45
4	Scenario 2 + 40% labor supply	1,140	63.28	0.57	151.5	5.59

Note: Average values over the 15-year simulation period.

The table shows that the impact of the HIV/Aids epidemic much depends on the how much the available labor supply was reduced, which is largely unknown. The epidemic has depressed maize yields, increased the maize yield gap, and impeded on the diffusion of improved maize varieties. The most dramatic effect is on the value of livestock assets: when assuming that the epidemic reduced labor supply by 20 percent, than the current livestock value is about 50 percent lower than in the scenario without the disease.

9.6 Summary

Trends in maize yields, land productivity, and labor productivity are most strongly determined by population dynamics rather than soil fertility decline. The maize yield gap is no good indicator of economic well-being or food security. Economic well-being strongly relates to high labor productivity while agents with high land productivity are more likely to be poor. Different from yield gaps, performance gaps directly relate underlying factors to the well-being of agents. The analysis shows that improved access to innovations in combination with credit can significantly contribute to food security. Although hybrid maize increases average maize yields significantly, the analysis suggested that its sluggish adoption in Uganda is because the variety does not have an advantage over Longe 1 with respect to the level of food energy consumption.

10 Discussion

10.1 Introduction

The final chapter puts the study in a wider perspective again. The limitations of the used model are made explicit and the possible impact on the findings is discussed in Section 10.1. Section 10.2 compares the used methodology to other approaches of modeling farm households. Section 10.3 gives recommendations for research.

10.2 Limitations of the study

Though the model is probably the most encompassing every built for a Ugandan farming system, like any model, it is incomplete and has limitations. These limitations are highlighted in the following and their impact on the results is discussed.

10.2.1 Low data quality

The data quality was low. The expenditure data from the 1999-2000 UNHS had recording errors in units of quantity and prices and income levels are likely underestimated. Crop production data from the 2000-2001 IFPRI survey did not allow reliable estimates of crop yields because of unclear recording of intercropped areas. Data on labor use per crop could only be approximated from data on labor use per plot times the area proportion of each crop. The effect of the low data quality on the simulation results is difficult to assess. Yet, the estimated income and price elasticities were plausible. Only when better data become available, can the model estimates be compared.

10.2.2 Migration

Rural to urban migration was not captured in the present model. Although growth in the non-farm sector is one of the most important determinants of agricultural change, it was outside the scope of the present model. Rural to urban migration is complex; especially in developing countries were the young and educated tend to migrate but also the poor with insufficient land. Migration data are, however, notoriously hard to obtain as household surveys generally only collect information on the present number of household members, and of households that have not migrated. Because the present model did not capture migration, it is very likely to have overestimated the growth in population density. This could mean that the observed negative correlation between land productivity and food energy consumption would disappear more gradually than in the predicted 15-year period.

10.2.3 Sources of heterogeneity

Heterogeneity in the agent population only refered to resource endowments and innovativeness. Sources of heterogeneity not included in the model are differences in price expectations, seasonal price variations, educational levels, skills, and health. It would be possible to also include these types of heterogeneity in the present framework, given data availability and quantitative models that relate them to either production or consumption decisions of the farm household.

10.2.4 Unknown crop yield response functions

Crop yield response functions were almost completely based on studies done in countries other than Uganda, as local fertilizer response data were not available. The yield response to some important factors was not explicitly considered due to limitations in both data availability and time. These factors are rainfall, crop pests and diseases, planting density, and the timing of management decisions.

10.2.5 Absence of local factor and output markets

Local factor markets were not included in the model. The quantity of land was constant over the life of the agents. This is problematic in combination with population dynamics as the land of agents that pass away was not redistributed but became idle. The model is therefore not suited for simulation long-term dynamics. In the 16-year period, the number of agents passing away was about 8 percent, which was still considered acceptable. Reallocation of land and labor across farms is, however, important and would be a welcome addition to the present model. Yet, additional data collection on how the process of this reallocation would be required.

10.3 An ex-post comparison of approaches

The present study combined farm household modeling with a MAS approach. When comparing the present study with the literature on farm household modeling and MAS, it fits better to the first than to the second. There are two reasons for this. First, most MAS have been developed for hypothetical and experimental settings with only few empirical applications to agricultural land-use in developing countries (*e.g.*, Becu *et al.* 2003; Huigen 2004; Manson 2005). Second, most of these MAS applied to agriculture in developing countries, have had little in common with agricultural economics as MAS modelers have frequently resorted to qualitative approaches to model farm decision-making based on behavioral heuristics (Schreinemachers and Berger 2006). These approaches have sought to understand farm household behavior using participatory methods or group interviews, observation, and role-playing games (*e.g.*, Castella 2004). These MAS are used for different purposes, such as to support stakeholder interaction rather than to simulate agricultural dynamics.

The remainder of this section will therefore describe how the present study relates to other farm household models rather than how it compares to other MAS applications. One way of describing the existing variety of modeling approaches is taking a historical perspective and putting them on a time line as shown in **Table 10.1**.

Approaches to modeling farm household behavior have changed over time, but mostly by adding additional components rather than discarding old ones. Farm households were mostly seen as production units in the 1960s and hence farm behavior was represented by agricultural production functions and mathematical programming models of farm production (Heady and Dillon 1960). This approach proofed unsatisfactory for developing country agriculture in which semi-subsistence households had a dual role of producers and consumers. Researchers tried to explain why an increase in the price of a staple did not significantly increase its marketed surplus in rural Japan and this lead to the development of the integrated farm household models in late 1970s (Taylor and Adelman 2003). Barnum and Squire (1979) developed an integrated farm household model that combined econometrically estimated agricultural production and expenditure functions to capture the farm household’s dual role of producer and consumer. Integrated farm household models have, however, a drawback that the number of endogenous variables is small and that they can be used for static projection only.

To do dynamic simulation, these econometric models were increasingly turned into whole farm programming models (Hazell and Norton 1986). Mathematical programming was a powerful tool to represent trade-offs in farm decision-making and its flexibility to include econometrically estimated production and consumption functions as well as other information and expert opinion was useful for the *ex ante* evaluation of policy interventions and improved technologies.

In the 1990s, when concerns for ecological sustainability came to the front, these whole farm programming models were integrated with biophysical process models. These so-called bio-economic models combined the advantages of dynamic whole farm programming models with biophysical processes such as soil fertility dynamics (Barbier 1998; Shiferaw & Holden 2000; Holden and Shiferaw 2004; Kruseman 2004).

From this point onwards, there are two observable trends. Some researchers have chosen to scale up farm household models by integrating them in vertical way with market and village models (Kuiper 2005). These models are able to capture market interaction and have endogenous prices. The MP-MAS approach goes, however, into the other direction by scaling out farm household models by horizontally integrating a very large number of farm households with a landscape. The MP-MAS approach does not rely on a few 'representative' farm households but every farm household in an area. These models are thereby able to capture the heterogeneity in socioeconomic and biophysical conditions and household interaction (Balmann 1997; Berger 2001; Happe 2004). It is worth noting that all subsequent approaches have had the same building blocks of production functions with each approach merely adding components to them. The present study is no exception.

Table 10.1: Comparison of farm household modeling approaches on a time line

	Modeling approach		**Advancement**	**Reference**
1	Agricultural production functions		Supply response to production factors	Heady and Dillon 1960
2	Integrated farm household models		Integrating production and consumption decisions	Barnum and Squire 1979; Strauss *et al.* 1986
3	Whole farm programming		Dynamic simulation of farm decision-making	Hazell and Norton 1986; Dillon and Hardaker 1993
4	Bioeconomic modeling of farm households		Integrating biophysical and economic processes in a dynamic simulation	Barbier 1998; Shiferaw and Holden 2000; Kruseman 2004
5	*Scaling up:*	Integrated household-village models	Markets endogenous; market interaction	Kuiper 2005
	Scaling out:	Mathematical programming-based multi-agent systems	Heterogeneity and interaction	Balmann 1997; Berger 2001; Happe 2004; this thesis

The present study falls clearly falls into the latter category but has more in common with earlier approaches than the other models in its category by Balmann (1997), Berger (2001), and Happe (2004). First, it has in common with the integrated farm household models that it has a strong representation of the consumption side of farm households by including a full and flexible expenditure system (Working-Leser model and Almost Ideal Demand System). It is also the first application of MP-MAS to semi-subsistence farm households as Balmann and Happe developed their model for a region in Germany and Berger for a region in Chile with commercial agriculture. Second, it has in common with the bioeconomic models that it has detailed biophysical dynamics, especially soil property dynamics. It is also the first application of MP-MAS to the study of soil fertility decline.

The present MP-MAS approach hence builds on the long tradition from production functions to bioeconomic modeling but adds two new components of heterogeneity of landscapes and farm households and interaction between households. The addition of these components seems especially relevant with respect to the current development agenda on poverty, inequality, and sustainability, which all relate to heterogeneity. Though Berger (2001) and Happe (2004) showed that inequality can be captured in MAS, the present study showed that also sustainability and poverty can very well be addressed using MP-MAS.

10.4 Recommendations for research

First, given the importance of labor productivity, research efforts need urgently to be diverted from an exclusive focus on crop yields and soil fertility. Labor remains the primary force of Ugandan agriculture and its availability is crucial, perhaps more crucial than the fertility of soils. On-farm trials should not only be selected based on variation in soil fertility but also on variation available labor. Crop varieties will be successful if they increase labor productivity, that is, if they give more output per hour of management. Labor use in on farm trials should be recorded along with crop yield to better estimate the potential of a crop under farm conditions.

Second, intercropping receives far less attention from researchers than it should. The majority of the arable land in Uganda is intercropped. Yet, none of the available data sets (UNHS, IFPRI, or ZEF), was suitable to analyze intercropping. Intercropping reduces maize yields, as well as yields of most other crops, but this is no reason to discourage the practice.

Research on intercropping should furthermore divert attention from crop yields and analyze the effect on labor productivity in addition. Agronomic literature has convincingly shown major benefits of intercropping on crop yields in Africa as indicated by land equivalent ratios exceeding unity. To assess the contribution of intercropping to the well-being of farm households, the estimation of labor equivalent ratios would be most useful. Yet, the first estimate has yet to appear in the *African Crop Science Journal*.

The effect of intercropping is better assessed through farm household surveys than experiments, for two reasons. First, the advantages of intercropping are likely to be most apparent under suboptimal biophysical and socioeconomic conditions, which are usually not well represented in experiments. Second, farm household surveys are more regular than experiments and questions on intercropping can be incorporated.

Third, basic empirical data on crop nutrient response is essential in the study of soil fertility decline. This study shows that negative nutrient balances do not immediately reduce crop yields and that dynamics in the farming system are more strongly influenced by population growth than soil fertility decline. Yet, there is a large uncertainty in the used response functions. This is, however, not a call is not for multi-million dollar experiments, like the fertilizer use response projects (FURP) as for example conducted in Kenya. The call is rather for a limited number of fertilizer experiments on average soil conditions but for all crops. This could serve as a baseline for further studies on soil fertility decline in Uganda and make the crucial linkage between soil fertility decline, changing crop yields, and adjustments at the farm level.

Fourth, the actual importance of maize is likely to be overstated. Though farm households grow maize on about 25 percent of their area, this does not reflect its economic importance as farm households grow a large number of crops, most of which are substitutes. A more diversified approach to agricultural development is hence required, in which the focus should be on raising the labor productivity of poor farm households.

References

Abalo, G., J.J.Hakiza, R.El-Bedewy & E.Adipala 2003. Genotype x environment interaction studies on yields of selected potato genotypes in Uganda. *African Crop Science Journal*, 11(1): 9-15.

Adesina, A.A. & J.Baidu-Forson.1995. Farmers' perceptions and adoption of new agricultural technology: evidence from analysis in Burkina Faso and Guinea, West Africa. *Agricultural Economics*, 13(1): 1-9.

Aggarwal, P.K., K.K. Talukdar & R.K. Mall. 2000. Potential yields of rice-wheat system in the Indo-Gangetic plains of India. *Rice-Wheat Consortium Paper Series*, 10.

Ali, M. 1999. Evaluation of green manure technology in tropical lowland rice systems. *Field Crops Research*, 61: 61-78.

Aniku, J., D. Kataama, P. Nkedi-Kizza & S. Ssessanga. 2001. PDCO and Soil Fertility Management: Uganda Results and Experience. Aniku J., D. Kataama, P. Nkedi-Kizza and Ssesanga S. PDCO and soil fertility management: Uganda results and experience. In R.N. Roy and H. Nabhan. Soil nutrient management in sub-Saharan Africa in support of the Soil Fertility Initiative. Proceedings of the conference held in Lusaka, Zambia 6-8 December, 1999. Food and Agriculture Organization of the United Nations, Rome, 2001. pp. 65-75.

Arrow, K.J. 1987. Economic theory and the hypothesis of rational expectations. *The New Palgrave*, 2: 69-74.

ASARECA. 2004. http://www.asareca.org/ecamaw/maize/maize-res.htm

Aune J.B. and R. Lal. 1997. Agricultural productivity in the tropics and critical limits of properties of oxisols, ultisols and alfisols. *Tropical Agriculture*, 74:96-103.

Aune, J.B. & A. Massawe. 1998. Effects on Soil Management on Economic Return and Indicators of Soil Degradation in Tanzania: a Modelling Approach. *Advances in GeoEcology*, 31: 37-43.

Aune, J.B. & R. Lal. 1995. The Tropical Soil Productivity Calculator—A Model for Assessing Effects of Soil Management on Productivity. In Lal, R., Stewart, B.A. (eds.): Soil Management. *Experimental Basis for Sustainability and Environmental Quality*. Lewis Publishers, London.

Austin, R.B. 1999. Yield of wheat in the United Kingdom: recent advances and prospects, *Crop Science*, 39: 1604-1610.

Bakrie, B., J. Hogan, J.B. Liang, A.M.M. Tareque & R.C. Upadhyay. 1996. Ruminant nutrition and production in the tropics and subtropics. ACIAR Monograph, 36.

Balmann, A. 1997. Farm-based Modelling of Regional Structural Change: A Cellular Automata Approach. *European Review of Agricultural Economics*, 24: 85-108.

Barbier, B. 1998. Induced innovation and land degradation: results from a bioeconomic model of a village in West Africa, *Agricultural Economics*, 19: 15-25.

Barnum, H.N. & L. Squire. 1979. *A Model of an Agricultural Household. Theory and Evidence*. Johns Hopkins University Press, Baltimore

Becu, N., P. Perez, B. Walker, O. Barreteau, C. Le Page. 2003. Agent-based simulation of a small catchment water management in northern Thailand. Description of the Catchscape model. *Ecological Modelling*, 170, pp.319–331.

Bellon, M. R. & J. Risopoulos. 2001. Small-scale farmers expand the benefits of improved maize germplasm: a case study from Chiapas, Mexico, *World Development*, 29(5): 799-811.

Berger, T. & P. Schreinemachers. 2006. Creating agents and landscapes for multi-agent systems using random sample data. *Ecology and Society*. Forthcoming.

Berger, T. 2001. Agent-based 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., P. Schreinemachers & J. Woelcke, 2006. Multi-agent simulation for the targeting of development policies in less-favored areas. *Agricultural Systems*, 88: 28-43.

Bourke, R.M. 1985. Influence of nitrogen and potassium fertilizer on growth of sweet potato (Ipomoea batatas) in Papua New Guinea. *Field Crops Research*, 12: 363-375.

Brandes, W. 1989. On the limitations of armchair economics: some views of an armchair agricultural economist. *European Review of Agricultural Economics*, 16: 319-343.

von Braun, J. 1988. Effects of technological change in agriculture on food consumption and nutrition: rice in a West African setting. *World Development*, 16(9): 1083-1098.

von Braun, J. 2003. Agricultural economics and distributional effects. *Agricultural*

Economics, 32(s1): 1-20.

Brosig, S. 2000. A model of household type specific food demand behaviour in Hungary, *Discussion Paper*, 30, Institute of Agricultural Development in Central and Eastern Europe, Halle (Saale), Germany. Available online at: http://www.iamo.de/dok/dp30.pdf (accessed 18 April 2006).

Brunner, A.C. 2003. *Sensitivity of the Water Erosion Prediction Project (WEPP)-model in assessing the spatial erosion susceptibility of soils on a hillslope in Uganda*. Master Thesis, University of Bonn, Germany.

Busolo-Bulafu, C.M. & Y. Obong. 2001. Groundnuts. In: Mukiibi, J.K. (ed.) *Agriculture in Uganda*. Volume II: Crops, pp.87-102.

Byerlee, D. & A. Siddiq. 1994. Has the green revolution been sustained? The quantitative impact of the seed-fertilizer revolution in Pakistan revisited, *World Development*, 22(9): 1345-1361.

Cambrony, H.R. 1992. *Coffee Growing.* The Tropical Agriculturalist. CTA / MacMillan.

Cameron, L.A. 1999. The importance of learning in the adoption of high-yielding variety seeds. *American Journal of Agricultural Economics*, 81: 83-94.

Canevara, M.G., M. Romani, M. Corbellini, M. Perenzin & B. Borghi. 1994. Evolutionary trends in morphological, physiological, agronomical and qualitative traits of Triticum aestivum L. cultivars bred in Italy since 1900. *European Journal of Agronomy*, 3: 175–185.

Cassanova, D., J. Goudriaan, J. Bouma, G.F. Epema. 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice. *Geoderma*, 91: 191-216.

Cassman, K.G. & Pingali, P.L. 1995. Extrapolating trends from long-term experiments to farmers' fields: the case of irrigated rice systems in Asia. In: D. Barnett and R. Steiner (eds.) *Agricultural sustainability in economic environmental and statistical terms*. London: John Wiley & Sons.

Cassman, K.G. (ed.) 1994. *Breaking the yield barrier,* proceedings of a workshop on rice yield potential in favorable environments, IRRI, 29 November - 4 December 1993, IRRI: Los Baños, Philippines.

Cassmann, K.G. 1999. Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *Proceedings of the national Academy of Science*, U.S.A., 96: 5952-5959.

Castella, J.C., T.N. Trung, & S. Boissau. 2005. Participatory simulation of land-use changes in the northern mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. *Ecology and Society*, 10(1):27. Available online at: http://www.ecologyandsociety.org/vol10/iss1/art27/ (accessed 7 November 2005).

CIMMYT Bangladesh Regional Office. 2004. Agricultural statistics. Available online at: http://www.cimmyt.org/bangladesh/ (accessed 10 February 2004).

CIMMYT. 2002. CIMMYT Maize Program, International Maize Testing Unit. Available online at: http://www.cimmyt.org/Research/Maize/Index.htm (accessed 12 March 2002).

Clampett, W.S. 2001. Integrated crop management and technology transfer in the New South Wales rice industry: the Australian Ricecheck model. *International Rice Commission Newsletter*, 50, 7-16.

Close, W. & K.H. Menke. 1986. Selected topics in animal nutrition. *A manual for the 3rd Hohenheim course on animal nutrition in the tropics and semi-tropics*, 2nd edition. Deutsche Stiftung für Internationale Entwicklung & Universität Hohenheim.

Cochrane, W.W. 1979. *The development of American agriculture: a historical analysis*. Minneapolis: University of Minnesota Press.

Cochrane, Willard W. 1958. *Farm prices. myth and reality.* Minneapolis: University of Minnesota Press.

Coelli, T., D.S. Rao, Prasada & G. Battese. 1998. *An introduction to efficiency and productivity analysis*, Boston: Kluwer.

Colman, D. & T. Young. 1989. *Principles of agricultural economics: markets and prices in less developed countries.* Wye studies in agricultural and rural development. Cambridge: Cambridge University Press.

Cooper, M. 1999. Concepts and strategies for plant adaptation research in rainfed lowland rice. *Field Crops Research*, 64: 13-34.

Cornia, G.A., 1985. Farm size, land yields and the agricultural production function: an analysis for fifteen countries. *World Development*, 13(4): 513–534.

Coudouel, A., J.S. Hentschel & Q.T. Wodon. 2002. Poverty Measurement and Analysis. In: Klugman, J. *A Sourcebook for Poverty Reduction Strategies.* Washington, D.C.: The World Bank.

Dawe, D. 2002. The changing structure of the world rice market, 1950-2000. *Food Policy*, 27(4): 355-370.

Dawe, D., A. Dobermann, P. Moya, S.Abdulrachman, Bijay Singh, P. Lal, S.Y. Li, B. Lin, G. Panaullah, O. Sariam, Y. Singh, A. Swarup, P.S. Tan & Q.-X. Zhen. 2000. How widespread are yield declines in long-term rice experiments in Asia? *Field Crops Research*, 66: 175-193.

De Datta, S.K. 1981. *Principles and practices of rice production*. New York: Wiley-Interscience Publications.

Deaton, A. & J. Muelbauer 1980. An Almost Ideal Demand System. *The American Economic Review*, 70(3): 312-326.

Deaton, A. 2000. *The analysis of household surveys: a microeconometric approach to development policy*. Johns Hopkins University Press, Baltimore and London.

Deffuant, G., Huet, S. & Amblard, F. 2005. An individual-based model of innovation diffusion mixing social value and individual benefit, *American Journal of Sociology*, 110: 1041-1069.

Deffuant, G., D. Neau, F. Amblard & G. Weisbuch. 2001. Mixing beliefs among interacting agents, *Advances in Complex Systems*, 3: 87-98.

Deininger, K. & J. Okidi. 2001. Rural households: incomes, productivity, and nonfarm enterprises. In: R. Reinikka & P. Collier (eds.), *Uganda's recovery: the role of farms, firms, and government.* The World Bank, Washington, D.C.

Diederen, P., H. van Meijl, A. Wolters & K. Bijak. 2003. Innovation adoption in agriculture: innovators, early adopters and laggards, *Cahiers d'économie et sociologie rurales*, 67.

Dillon, J.L. & J.B. Hardaker. 1993. Farm management research for small farmer development, *FAO Farm Systems Management Series*, 6, Rome: Food and Agriculture Organization of the United Nations.

Dorward, A., 1999, Farm size and productivity in Malawian smallholder agriculture. *Journal of Development Studies*. 35(5): 141-161.

Duvick, D.N. & K.G. Cassman. 1999. Post-green revolution trends in yield potential of temperate maize in the north-central United States, *Crop Science*, 39: 1622-1630.

Duvick, D.N. 1992. Genetic contributions to advances in yield of U.S. maize, *Maydica*, 37: 69-79.

Duvick, D.N. 1996. Plant breeding, an evolutionary concept, *Crop Science*, 36: 539-548.

Duwayri, M., D.V. Tran & V.N. Nguyen. 1999. Reflection on yield gaps in rice production. *International Rice Commission Newsletter*, 48: 13-26.

Duwayri, M., D.V. Tran & V.N. Nguyen. 1999. Reflections on yield gaps in rice production, *International Rice Commission Newsletter*, 50: 13-31.

Duwayri, M., D.V. Tran & V.N. Nguyen. 2000 Reflections on yield gaps in rice production: how to narrow the gaps. In: M.K. Papademetriou, F.J. Dent & E.M. Herath (eds.) *Bridging the yield gap in the Asia-Pacific region*, FAO Regional Office for Asia and the Pacific, Bangkok, Thailand, October 2000. Available online at: http://www.fao.org/DOCREP/003/x6905e/x6905e05.htm (accessed 18 April 2006).

Eales, J.S. & L.J. Unnevehr. 1994. The inverse almost ideal demand system. *European Economic Review*, 38: 101-115.

Ebiyau, J. & J.O.E. Oryokot, 2001. Sorghum. In: Mukiibi, J.K. (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/CTA/NARO, pp. 42-54.

Ebwongu, M., E. Adipala, C.K. Ssekabembe, S. Kyamanywa & A.S. Bhagsari. 2001. Effect of intercropping maize and solanum potato on yield of the component crops in central Uganda. *African Crop Science Journal*, 9(1): 83-96.

Elsner, K. 1999. Analyzing Russian food expenditures using micro-data. *Discussion Paper*, 23, Institute of Agricultural Development in Central and Eastern Europe, Halle (Saale), Germany.

Ennin, S.A., M.D. Clegg & C.A. Francis. 2002. Resource utilization in soybean/maize intercrops. *African Crop Science Journal*, 10(3): 251-261.

ESRI. 2003. *ArcView GIS 3.2*, Environmental Systems Research Institute, Inc.

Estudillo, J.A., M. Fujimura & M. Hossain. 1999. New rice technology and comparative advantage of rice production in the Philippines. Journal of Development Studies, 35(5): 162-184.

Evans, L.T. & R.A. Fischer. 1999. Yield potential: it's definition, measurement, and significance, *Crop Science*, 39: 1544-1551.

Evans, L.T. 1996. *Crop evolution, adaptation and yield*. Cambridge: Cambridge University Press.

Evenson, R.E. & D. Gollin. 2003. Assessing the impact of the green revolution, 1960 to 2000. *Science*, 300: 758-762.

Eyhérabide, G. H., A. L. Damilano & J. C. Colazo. 1994. Genetic gain for grain yield of maize in Argentina. *Maydica,* 39: 207-211.

Fan, S. & P. Hazell. 1999. Are returns to public investment lower in less-favored rural areas? An empirical analysis of India. *EPTD Discussion Paper*, 43, Environment and Production Technology Division, IFPRI, Washington, D.C.

FAO. 1999. *Report of the expert consultation on bridging the rice yield gap in the Asia-Pacific region*. FAO-RAP publication. FAO-RAP, Bangkok, Thailand.

FAO. 2006. *FAOSTAT Database*. Rome: Food and Agricultural Organization of the United Nations. Available online at: http://www.faostat.fao.org (accessed 3 April 2006).

FAO. 2006. HIV/Aids, food security and rural livelihoods. Fact sheet. Available online at: http://www.fao.org/worldfoodsummit/english/fsheets/aids.pdf (accessed 18 April 2006).

Farnworth, J. 1997. *Agri Info: guidelines for world crop and livestock production.* Chichester and New York: John Wiley & Sons

Feeney, G. & B. Zaba. 2001. A minimalist model for projecting HIV/Aids. Geneva: UNAIDS. Available online at: http://www.gfeeney.com/ (accessed 18 April 2006).

Fininsa, C. 1997. Effects of planting pattern, relative planting date and intra-row spacing on a Haricot bean/maize intercrop. *African Crop Science Journal*, 5(1): 15-22.

Fischer, G., H. van Velthuizen, M. Shah & F. Nachtergaele. 2002. *Global agro-ecological assessment for agriculture in the 21st century: methodology and results*, RR-02-02, January, International Institute for Applied Systems Analysis, Laxenburg, Austria & Food and Agriculture Organization of the United Nations, Rome, Italy.

Flinn, J.C. & S.K. De Datta. 1984. Trends in irrigated-rice yields under intensive cropping at Philippine research stations, *Field Crops Research*, 9: 1-15.

Flinn, J.C., S.K. De Datta & E. Labadan. 1982. An analysis of longterm rice yields in a wetland soil. *Field Crops Research*, 5: 201–216.

Gini, C. 1912. Variabilita e Mutabilita, Bologna.

Goldman, I.L. 1998. From out of old fields comes all this new corn: a historical perspective on heterosis in plant improvement, in: Kendall R. Lamkey & Jack E. Staub (eds.) Concepts and breeding of heterosis in crop plants, CSSA Special Publication Nr. 25, Madison, Wisconsin: Crop Science Society of America, pp.1-12.

Gomez, K.A. 1977. On farm assessment of yield constraints: methodological problems, in: *IRRI constraints to high yields on Asian rice farms: an interim report*, Los Baños, Philippines: IRRI, pp.1-16.

Griliches, Z. 1957. Hybrid corn: an exploration in the economics of technological change. *Econometrica*, 25(4): 501-522.

Happe, K., 2004, Agricultural policies and farm structures - agent-based modelling and application to EU-policy reform. *Studies on the agricultural and food sector in central and Eastern Europe*, 30, IAMO. Available online at: http://www.iamo.de/dok/sr_vol30.pdf>http://www.iamo.de/dok/sr_vol30.pdf (accessed 7 November 2005).

Harmand, J.M., Njiti, C.F., Bernhard-Reversat, F. & Puig, H. 2004. Aboveground and belowground biomass, productivity and nutrient accumulation in tree improved fallows in the dry tropics of Cameroon. *Forest Ecology and Management,* 188: 249-265.

Hayami, Y. & V.W. Ruttan. 1985. *Agricultural development; an international perspective*. Baltimore and London: The Johns Hopkins University Press.

Hazell, P. & L. Haddad. 2001. Agricultural research and poverty reduction, *Food, Agriculture, and the Environment Discussion Paper*, 34, International Food Policy Research Institute/Technical Advisory Group (TAC), Consultative Groups on International Agricultural Research, Washington, D.C.

Hazell, P. & R. Norton. 1986. *Mathematical programming for economic analysis in agriculture*. New York: Macmillan.

Hazell, P.B.R. & Ramasamy, C. 1991. *The green revolution reconsidered; the impact of high-yielding rice varieties in South India*. The John Hopkins University Press: Baltimore and London.

Heckman, J.J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5, 475-92.

Heisey, P.W., M.A. Lantican & H.J. Dubin. 2002. *Impacts of international wheat breeding research in developing countries, 1966-97*. Mexico, D.F.: CIMMYT.

Heltberg, R., 1998, Rural market imperfections and the farm size-productivity relationship: evidence from Pakistan. *World Development*, 26(10): 1807-1826.

Herdt, R.W. & Mandac, A.M. 1981. Modern technology and economic efficiency of Philippine rice farmers. *Economic Development and Cultural Change*. 29(2): 375-399.

Herdt, R.W. 1987. A retrospective view of technological and other changes in Philippines rice farming, 1965-1982. *Economic Development and Cultural Change,* 35(2): 329-349.

Hildebrand, P.E. 1976. Multiple cropping systems are dollars and "sense" agronomy. In: *ASA Special Publication*, 27: Multiple Cropping: 347-371.

Holden, S. & B. Shiferaw, 2004. Land degradation, drought and food security in a less-favored area in the Ethiopian highlands: a bio-economic model with market imperfections, *Agricultural Economics*, 30(1): 31:49.

Holden, S., B. Shiferaw & J. Pender 2004. Non-farm income, household welfare, and sustainable land management in a less-favoured area in the Ethiopian highlands. *Food Policy*, 19(4): 369-392.

Huang, K.S. & B.H. Lin. 2000. Estimation of food demand and nutrient elasticities from household survey data. *ERS Technical Bulletin*, 1887, Washington, D.C: Economic Research Service, U.S. Department of Agriculture.

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

Illinois Experiment Station. 1960-2001. *Performance of Commercial Corn Hybrids in Illinois, Bulletin Series*, various editions 1960-2001, Urbana, Illinois: Experiment Station.

IndiaStat. 2004. Statistical Database. Available online at: http://www.indiastat.com/ (accessed February 2004).

Instituto Brasileiro De Geograpia e Estadística (IBGE) 2004. *Sistema IBGE de Recupera Ção Automatica (SIDRA)*. Available online at: http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=1612&z=t&o=1 (accessed February 2004).

International Rice Research Institute (IRRI). 1978. *Interpretive analysis of selected papers form changes in rice farming in selected areas of Asia*. Los Baños, Philippines.

International Rice Research Institute (IRRI). 2000. *Annual Report 1999-2000*. Available online at: http://www.irri.org/publications/annual/ar2000.asp (accessed 18 April 2006).

International Rice Research Institute (IRRI). 2004. *World rice statistics*. Available online at: http://www.irri.org/science/ricestat/index.asp (accessed February 2004).

James, W.P.T. & E.C. Schofield. 1990. *Human energy requirements: a manual for planners and nutritionists*, Food and Agricultural Organization of the United

Nations, New York: Oxford University Press.

Jenkinson, D.S. & A. Ayanaba. 1977. Decomposition of carbon-14 labelled plant material under tropical conditions. *Soil Science Society of America Journal*, 41: 912-915.

Just, R.E. & O. Weninger. 1999. Are crop yields normally distributed? *American Journal of Agricultural Economics*, 81: 287-304.

Kaizzi C.K. 2002. The potential benefit of green manures and inorganic fertilizers in cereal production on contrasting soils in eastern Uganda. University of Bonn Dissertation. *Ecology and Development Series*, 4. Cuvillier Verlag, Göttingen.

Kaizzi, C.K, H. Ssali, A. Nansamba, G. Ruecker & P.L.G. Vlek 2003. Estimated farm-level nutrient balances for the farming systems of eastern Uganda. Draft. Submitted to *African Crop Science Journal*.

Kaizzi, C.K. & C.S. Wortmann. 2001. Plant materials for soil fertility management in subhumid tropical areas. *Agronomy Journal*. 93: 929-935.

Karikari, S.K., O. Chaba & B. Molosiwa. 1999. Effects of intercropping bambara groundnut on pearl millet, sorghum and maize in Botswana. *African Crop Science Journal*, 7(2): 143-152.

Kebbeh, M. & K.M. Miezam. 2003. *Ex ante* evaluation of integrated crop management options for irrigated rice production in the Senegal River Valley. *Field Crops Research*, 81: 87-94.

Kennedy, P. 1998. *A guide to econometrics,* Fourth edition. The MIT Press: Cambridge, MA.

Khush, G.S. & S. Peng. 1996. Breaking the yield frontier of rice. In: Reynolds, M.P., S. Rajaram & A. McNab (eds.) Increasing the yield potential in wheat: breaking the barriers. Proceedings of a workshop held in Ciudad Obregón, Sonora, Mexico, CIMMYT, March 1996, pp.38-51.

Khush, G.S. 1995a. Modern varieties - their real contribution to food supply and equity, *GeoJournal*, 35(2): 275-284.

Khush, G.S. 1995b. Breaking the yield frontier of rice. *GeoJournal*, 35(3): 329-332.

Kidoido, M.M., V. Kasenge, S. Mbowa, J.S. Tenywa & P. Nyende. 2002. Socioeconomic factors associated with finger millet production in eastern Uganda, *African Crop Science Journal*, 19(1): 111-120.

Kikafunda-Twine, J., Kyetere, D.T., Bigirwa, G., Kalule, T. & Wamaniala, M. 2001. Maize. In: Mukiibi, J.K. (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers

/CTA/NARO. pp.55-69.

Kristjanson, P., A. Krishna, M. Radeny & W. Nindo. 2004. Pathways Out of Poverty in Western Kenya and the Role of Livestock. ILRI, *PPLPI Working Paper*, 14.

Kruseman, G. 2000. *Bio-economic household modelling for agricultural intensification*. Ph.D. Thesis, Wageningen: Wageningen University.

Kuyvenhoven, A. 2004. Creating an enabling environment: policy conditions for less-favored areas. *Food Policy*, 19(4): 407-429.

Kuyvenhoven, A., J. Pender & R. Ruben. 2004. Development strategies for less-favored areas, *Food Policy*, 19(4): 295-302.

Kuyvenhoven, A., R. Ruben & G. Kruseman. 1998. Technology, market policies and institutional reform for sustainable land use in southern Mali. *Agricultural Economics*, 19: 53-62.

Lacy, J., W. Clampett & J. Nagy. 2000. Bridging the rice yield gap in Australia. In: M.K. Papademetriou, F.J. Dent & E.M. Herath (eds.) *Bridging the yield gap in the Asia-Pacific region*, FAO Regional Office for Asia and the Pacific, Bangkok, Thailand, October 2000. Available online at: http://www.fao.org/DOCREP/003/x6905e/x6905e05.htm (accessed 18 April 2006).

Lampe, K. 1995. Rice research: food for 4 billion people, *GeoJournal*, 35(2): 253-259.

Lanjouw, P. & M. Ravallion. 1995. Poverty and household size. *Economic Journal*, 105:1415–1434.

Lantican, M.A., P.L. Pingali & S. Rajaram. 2003. Is research on marginal lands catching up? The case of unfavorable wheat growing environments. *Agricultural Economics*, 29: 353-361.

Latham, M.C. 1979. *Human nutrition in tropical Africa. A textbook for health workers with special reference to community health problems in East Africa*. Rome: Food and Agriculture Organization of the United Nations.

Latham, M.C. 1997. Human nutrition in the developing world. *FAO Food and Nutrition Series*, 29. Rome.

Leser, C.E.V. 1963. Forms of Engel functions, *Econometrica*, 31(4): 694-703.

Litchfield, J.A. 1999. *Inequality: methods and tools*. The World Bank PovertyNet, Available online at: http://www.worldbank.org/poverty (accessed 18 April 2006).

Livestock, Environment and Development (LEAD) Initiative. 1999. *Livestock and environment toolbox*. Food and Agriculture Organisation of the United Nations.

http://lead.virtualcenter.org/en/dec/toolbox/Index.htm (accessed 6 July 2004).

Manson, S.M., 2005, Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, Ecosystems and Environment*, 111(1): 47-62.

Martin, W.J. & P.G. Warr. 1994. Determinants of agriculture's relative decline: Thailand, *Agricultural Economics*, 11: 219-235.

Ministry of Finance, Planning and Economic Development. 2000. *Uganda's poverty eradication action plan, Poverty Reduction Strategy Paper, Summary and main objectives*, Kampala.

Moore, I. & G. Burch 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal*, 50: 1294-1298.

Moser, C.M. & C. Barrett 2003. The disappointing adoption dynamics of a yield-increasing, low external-input technology: the case of SRI in Madagascar. *Agricultural Systems*, 76(3): 1085-1100.

Mubiru, D. & H. Ssali (eds.) 2002. *Soils and soil fertility management programme*, Annual report 2001-2002, Kawanda Agricultural Research Institute, NARO.

Mukiibi, J.K. (ed.). 2001. *Agriculture in Uganda.* Volume II: Crops. Fountain Publishers / CTA/NARO.

Musambasi, D., O.A. Chivinge & I.K. Mariga. 2002. Intercropping maize with grain legumes for Striga control in Zimbabwe. *African Crop Science Journal*, 10(2): 163-171.

Musoli, P.C., G.J. Hakiza, J.B. Birikunzira, I. Kibirige-Sebunya & P. Kucel. 2001. Coffee. In: J.K. Mukiibi. (ed.) *Agriculture in Uganda*, Volume II: Crops. Fountain Publishers/CTA/NARO, pp.376-436.

Mutungamiri, A., I.K. Mariga & O.A. Chivinge. 2001. Effect of maize density, bean cultivar and bean spatial arrangement on intercrop performance. *African Crop Science Journal*, 9(3): 487-497.

Muwanga, R.O.M., C.N. Ocitti p' Obwoya, B. Odongo & G.M. Turyamureeba. 2001. Sweet potatoes. In: J.K. Mukiibi. (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/CTA/NARO, pp.233-251.

Nampala, P., M.W.Ogenga-Latigo, S.Kyamanywa, E.Adipala, N. Oyobo & L.E.N. Jackai. 2002. Potential impact of intercropping on major cowpea field pests in Uganda, *African Crop Science Journal*, 10(4): 335-344.

Ngugi, D., P.K. Karau & W. Nguyo. 1990. *East African Agriculture. A textbook for secondary schools*. 3rd edition, MacMillan Education Ltd, London and Oxford.

Nknonya, E., J. Pender, P. Jagger, D. Sserunkuuma & C.K. Kaizzi. 2002. Strategies for Sustainable Livelihoods and Land Management in Uganda, Research report manuscript submitted to the IFPRI Publications Review Committee, December 2002.

Nkonya, E., C. Kaizzi & J. Pender 2005. Determinants of nutrient balances in a maize farming system in eastern Uganda, *Agricultural Systems*, 85(2): 155-182.

Ntege-Nanyeenya, W., M. Mugisa-Mutetikka, W. Mwangi & H. Verkuijl. 1997. *An assessment of factors affecting the adoption of maize production technologies in Iganga District, Uganda*. Addis Ababa, Ethiopia: National Agricultural Research Organization (NARO) and International Maize and Wheat Improvement Center (CIMMYT).

Nugent, J.B. 1985. The old-age security motive for fertility. *Population and Development Review*, 11(1): 75-97.

Nunn, A.J., D.W.Mulder, A. Kamali, A. Ruberantwari, J.F Kengeya-Kayondo & J. Whitworth. 1997. Mortality associated with HIV-1 infection over five years in a rural Ugandan population: cohort study, *British Medical Journal*, 315: 767-771.

Nyende, P., J.S. Tenywa, J. Oryokot & M. Kidoido. 2001. Weed profiles and management assessment for increased finger millet production in Uganda. *African Crop Science Journal*, 9(3): 507-516.

Opio, F., Ugen, M.A., Kyamanywa, S., David, S. & Mugisa-Mutetikka, M. 2001. Beans. In: J.K. Mukiibi (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/ CTA/NARO, pp.162-191.

Oryokot, J.O.E. 2001. Finger Millet. In: J.K. Mukiibi (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/CTA/NARO. pp.29-41.

Otim-nape, G.W., Semakula, G.N., Bua, A., Baguma, Y.K. & Ogwal, S. 2001. Cassava. In: J.K. Mukiibi (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/CTA/NARO, pp.194-228.

Paarlberg, R. 2000. The weak link between world food markets and world food security. *Food Policy*, 25: 317-335.

Papademetriou, M.K. 2000. Rice production in the Asia-Pacific region: issues and perspectives. In: M.K. Papademetriou, F.J. Dent & E.M. Herath (eds.) *Bridging the* *yield gap in the Asia-Pacific region*, FAO Regional Office for Asia and the Pacific, Bangkok, Thailand, October 2000. Available online at: http://www.fao.org/DOCREP/003/x6905e/x6905e04.htm (accessed 18 April 2006).

Papademetriou. 2001. Issues and perspectives of rice production in Asia and the Pacific region. In: *FAO. Yield gap and productivity decline in rice production*, proceedings of the expert consultation held in Rome, 5-7 September 2000, Food and Agriculture Organization of the United Nations, Rome.

Parker, D.C., T. Berger & S. Manson. (eds.). 2002. Agent-based models of land use/land cover change. *LUCC Report Series*, 6, International Human Dimensions Programme on Global Environmental Change (IHDP).

Payne, T.S. 2002. *Grain yield and occurrence data extraction for 22 consecutive CIMMYT Elite Selection Wheat Yield Trials, 1980-2001, from the International Wheat Information System (IWIS-NT)*. 28 August 2002. Mexico, D.F.: CIMMYT.

Pender, J., P. Jagger, E. Nkonya & D. Sserunkuuma. 2001. Development pathways and land management in Uganda: causes and implications. *EPTD Discussion Paper*, 85, Washington, D.C.: International Food Policy Research Institute (IFPRI).

Pender, J., P. Jagger, E. Nkonya & D. Sserunkuuma. 2004. Development pathways and land management in Uganda: causes and implications, *World Development*, 32(5): 767-792.

Peng, S., K.G. Cassman, S.S. Virmani, J. Sheehy & G.S. Khush. 1999. Yield Potential Trends of Tropical Rice since the Release of IR8 and the Challenge of Increasing Rice Yield Potential, Crop Science, 39: 1552-1559.

Pingali, P.L. & M. Rosegrant. 1995. Agricultural commercialization and diversification: processes and policies. *Food Policy*, 20(3): 171-185.

Pingali, P.L. & M. Rosegrant. 1998. Supplying wheat for Asia's increasingly westernized diet, *American Journal of Agricultural Economics*. 80(5): 954-959.

Pingali, P.L. & P.W. Heisey. 1999. Cereal crop productivity in developing countries: past trends and future prospects, *CIMMYT Economics Paper*, 99-03. Mexico, D.F.: CIMMYT.

Pingali, P.L. & S. Pandey. 2001. Meeting world maize needs: technological opportunities and priorities for the public sector. In: P.L. Pingali (ed.) *CIMMYT 1999-2000 World Maize Facts and Trends. Meeting World Maize Needs: Technological Opportunities and Priorities for the Public Sector*. Mexico, D.F.: CIMMYT, pp.1-24.

Pingali, P.L. & S. Rajaram. 1999. Global wheat research in a changing world: options for sustaining growth in wheat productivity. In: P.L. Pingali (ed.) *CIMMYT 1998-99 world wheat facts and trends. Global Wheat Research in a Changing World: Challenges and Achievements.* Mexico, D.F.: CIMMYT: 1-18.

Pingali, P.L. 1997. From subsistence to commercial production systems: the transformation of Asian agriculture, *American Journal of Agricultural Economics*, 79: 628-634.

Pingali, P.L., M. Hossain, & R.V. Gerpacio 1997. *Asian rice bowls: the returning crisis?,* CAB International in association with the International Rice Research Institute.

Pinstrup-Anderson, P. 1985. Food prices and the poor in developing countries. *European Review of Agricultural Economics*, 65: 69-85.

Pinstrup-Andersen, P. (ed.) 1988. *Food subsidies in developing countries: costs, benefits, and policy options.* Baltimore, Md., USA: Johns Hopkins University Press for IFPRI.

Poehlman, J.M. & J.S. Quick. 1980. Crop breeding in a hungry world. In: D.R. Wood, K.M. Rawal & M.N. Wood. *Crop breeding.* American Society of Agronomy and Crop Science Society of America, pp.1-19.

Rejesus, R.M., P.W. Heisey & M. Smale. 1999. Sources of productivity growth in wheat: a review of recent performance and medium- to long-term prospects, *Economics Working Paper,* 99-05, Mexico, D.F.: CIMMYT.

Renkow, M. 1994. Technology, production environment, and household income: assessing the regional impacts of technological change. *Agricultural Economics*, 10: 219-231.

Reynolds, M.P., S. Rajaram & A. McNab (eds.). 1996. *Increasing the yield potential in wheat: breaking the barriers*. Proceedings of a workshop held in Ciudad Obregón, Sonora, Mexico, CIMMYT, March 1996.

Reynolds, M.P., S. Rajaram & K.D. Sayre. 1999. Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand, *Crop Science*, 39: 1611-1621.

Rhew, H., S. Park & G. Ruecker 2004. *Predictive soil mapping at regional scale in Iganga District, Uganda. A final report.* University of Seoul, South Korea.

Ricardo, D. 1957. *The Principles of political economy and taxation*, Everyman's Library: J.M. Dent & Sons Ltd.: London and New York.

Rockström, J. & M. Falkenland. 2000. Semiarid crop production from a hydrological perspective: gap between potential and actual yields. *Critical Reviews in Plant Sciences*, 19(4): 319-346.

Rogers, E.M., 1995. *Diffusion of innovations*, fourth edition, Free Press, New York.

de Rouw, A. 2004. Improving yields and reducing risk in peal millet farming in the African Sahel, *Agricultural Systems*, 81(1): 73-93.

Ruben, R. & J. Pender. 2004. Rural diversity and heterogeneity in less-favored areas: the quest for policy targeting. *Food Policy*, 19(4): 303-320.

Ruben, R., G. Kruseman & H. Hengsdijk. 1994. Farm household modelling for estimating the effectiveness of price instruments on sustainable land use in the atlantic zone of Costa Rica, *DLV Report*, 4, Wageningen, December.

Ruecker, G.R. 2005. Spatial variability of soils on national and hillslope scale in Uganda. *Ecology and Development Series*, 24, Cuvillier Verlag Göttingen.

Sadoulet, E. & A. de Janvry. 1995. *Quantitative development policy analysis*. The Johns Hopkins University Press, Baltimore and London.

Sall, S.; N. David & A.M. Featherstone. 1998. Adaptability of improved rice varieties in Senegal, *Agricultural Systems*, 57(1): 101-114.

Sanders, J.H. & J.K. Lynam. 1982. Evaluation of new technology on farms: methodology and some results from two crop programmes at CIAT, *Agricultural Systems*, 9(2): 97-112.

Sasakawa Global 2000. Nalongo. Quality protein maize seed. A farmer guide & fact sheet.

Schreinemachers, P & T. Berger. 2006. *Simulating farm household poverty: from passive victims to adaptive agents*. Selected paper at the tri-annual conference of the International Association of Agricultural Economists, Brisbane, Australia, 12-18 August 2006.

Schreinemachers, P. & T. Berger. 2006. Land-use decisions in developing countries and their representation in multi-agent systems, *Journal of Land Use Science*, 1(1): 1-16.

Schreinemachers, P, J.B. Aune & T. Berger. 2004. An application of the Tropical Soils Productivity Calculator to the study of soil fertility decline in Uganda. *ZEF Documentation of Research*, 3/2004.

Schultz, T.W. 1945. *Agriculture in an unstable economy*, Committee for Economic

Development Research Study, McGraw-Hill Book Company: New York and London.

Servicio de Información y Estadística Agroalimentaria y Pesquera (SIAP). 2004. *Sistema de Información Agropecuaria de Consulta (SIACON)* Available online at: http://www.siea.sagarpa.gob.mx//sistemas/siacon/SIACON.html (accessed February 2004).

Shepherd, K.D. & M.J. Soule. 1998. Soil fertility management in west Kenya: dynamic simulation of productivity, profitability and sustainability at different resource endowment levels, *Agriculture, Ecosystems & Environment*, 71: 131-145.

Shiferaw, B. & S. Holden. 2000. Policy instruments for sustainable land management: the case of highland smallholders in Ethiopia, *Agricultural Economics*, 22(3): 217-232.

Siddiq, E.A., K.V. Rao & A.S.R. Prasad. 2001. Yield and factor productivity trends in intensive rice production systems in India: a case study, *International Rice Commission Newsletter*, 50: 17-35.

Slafer, G.A., J.L. Araus & R.A. Richards, 1999. Physiological traits to increase the yield potential of wheat. In: E.H. Satorre & G.A. Slafer (eds.) *Wheat: ecology and physiology of yield determination*, Food Product Press, New York, pp. 379–415.

Smith, A. 1937. *An inquiry into the nature and causes of the wealth of nations*, New York: The Modern Library.

Smithson, P.C., B.D. McIntyre, C.S. Gold, H. Ssali, & I.N. Kashaija. 2001. Nitrogen and potassium fertilizer vs. nematode and weevil effects on yield and foliar nutrient status of banana in Uganda. *Nutrient Cycling in Agroecosystems*, 59: 239-250.

Snapp, S.S., M.J. Blackie & C. Donovan. 2003. Realigning research and extension to focus on farmers' constraints and opportunities. *Food Policy*, 28: 349-363.

Ssali, H. & Vlek, P.L.G. 2002. Soil organic matter and its relationship to soil fertility changes in Uganda. NARO(KARI): Kampala, pp.16-33.

Sserunkuuma, D., Pender, J., and Nkonya, E. 2001. *Land Management in Uganda: characterization of problems and hypotheses about causes and strategies for improvement*. International Food Policy Research Institute. Environment and Production Technology Division, Washington, D.C. Mimeo.

Stoneman, P. 2002. *The economics of technological diffusion*, Malden, MA: Blackwell Publishers.

Stoorvogel, J.J., E.M.A. Smaling, and B.H. Janssen. 1993. Calculating soil nutrient balances in Africa at different scales I. supra-natural scale, *Fertilizer Research*, 35: 227-235.

Tenywa, J.S., P. Nyende, M. Kidoido, V. Kasenge, J. Oryokot & S. Mbowa. 1999. Prospects and constraints of finger millet production in Eastern Uganda. *African Crop Science Journal*, 7(4): 569-583.

The World Bank. 2004. *Making services work for poor people. World Development Report 2004*. Oxford.

Theil, H. 1967. *Economics and information theory*, North Holland, Amsterdam.

Timsina, J. & D.J. Connor. 2001. Productivity and management of rice-wheat cropping systems: issues and challenges. *Field Crops Research*, 69: 93-132.

Tiongco, M. & D. Dawe. 2002. Long-term evolution of productivity in a sample of Philippine rice farms: implications for sustainability and future research, *World Development*, 30(5): 891-898.

Tran, D.V. & V.N. Nguyen. 2001. Understanding yield gap and productivity decline under intensive rice-based cropping systems. In: *Yield gap and productivity decline in rice production*, Proceedings of the Expert Consultation held in Rome, 5-7 September 2000. Rome: International Rice Commission, Crop and Grassland Service, FAO Plant Production and Protection Division, Food and Agricultural Organization of the United Nations, pp.13-37.

Tushemereirwe, W.K., D. Karamura, H. Ssali, D. Bwamiki, I. Kashaija, C. Nankinga, F. Bagamba, A. Kangire & R. Ssebuliba. 2001. Bananas. In: J.K. Mukiibi (ed.) *Agriculture in Uganda*. Volume II: Crops. Fountain Publishers/CTA/NARO, pp.281-319.

Udo, E.J. & J.A. Ogunwale. 1977. Phophorus fractions in selected nigerian soils, *Soil Science Society of America Journal*, 41: 1141-1146.

Uganda Bureau of Statistics. 2002. *Statistical abstracts 2002*. Uganda Bureau of Statistics (UBOS): Entebbe.

United States Department of Agriculture. 2002. Agricultural statistics data base, U.S. and state level data, Washington: Economic Research Service, available online at: http://www.ers.usda.gov/Data/ and http://www.nass.usda.gov:81/ipedb/ (accessed October 2002).

Valente, T.W. 1994. Network models of the diffusion of innovations, The Hampton Press communication series, Cresskill, NJ: Hampton Press, Inc.

Wade, L.J., S. Fukai, B.K. Samson, A. Ali & M.A. Mazid. 1999. Rainfed lowland rice: physical environment and cultivar requirements. *Field Crops Research*, 64: 3-12.

Wiebe, K. 2003. Linking land quality, agricultural productivity, and food security. *Agricultural Economic Report*, 823. Resource Economics Division, Economic Research Service, U.S. Department of Agriculture.

Wischmeier, W.H. & D.D. Smith. 1978. Predicting rainfall erosion losses – a guide to conservation planning. *Agriculture Handbook*, 537, U.S. Department of Agriculture.

de Wit, C.T. 1958. Transpiration and crop yields. *Agricultural Research Report*, 64(6), Pudoc, Wageningen.

Woelcke, J. 2004. Bio-Economics of sustainable land management in Uganda. *Development Economics and Policy Series*, 37. Peter Lang: Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, and Wien.

Woelcke, J. 2006. Technological and policy options for sustainable agricultural intensification in eastern Uganda, *Agricultural Economics*, 34: 129-139.

Working, H. 1943. Statistical laws of family expenditure, *Journal of the American Statistical Association*, 38: 43-56.

Wortmann, C.S. & C.K. Kaizzi. 1998. Nutrient balances and expected effects of alternative practices in farming systems of Uganda, *Agriculture, Ecosystems and Environment*, 71: 115-129.

Yuan, L.P. 1998. Hybrid rice development and use: innovative approach and challenges. *International Rice Commission Newsletters*, 47: 7-15.

Appendix

A. Survey data used

Table A: Socio-economic data used

	ZEF[1,2]	**IFPRI**[2]	**UNHS**[3]
Parts of survey used:	Household survey	Household survey All plot s survey Single plot survey	Household survey Crop survey
Data collection:	October 1999 - September 2000	January 2000 - December 2001	August 1999 - July 2000
Reference period:[5]	2nd season 1999 1st season 2000	1st season 2000 2nd season 2000	1st season 1999 2nd season 1999
Locations included:	2 villages (Magada and Buyembe) in the Iganga and Mayuge Districts	32 districts comprising 107 communities[4]	All districts except Kitgum, Gulu, Kasese, and Bundibugyo
Type of sample:	Stratified random sample	Stratified random sample	Stratified random sample
Sample weights used:	Yes	Yes	Yes
Sample size:	106 farm households	451 farm households; 1,681 plots	10,696 rural and urban households of which 8,452 farm households
Used to estimate:	Household resource endowments Location of farms and plots	Production functions Crop yields Input prices Crop rotations Labor availability Gender-specific labor use	Expenditure models Market prices Opportunity wage rate

Notes:

[1] Conducted by Johannes Woelcke (see Woelcke 2004, 2006)

[2] The *Project on Policies for Improved Land Management in Uganda* was a joint project by the International Food Policy Institute, Makerere University Faculty of Agriculture, National Agricultural Research Organization, Agricultural Policy Secretariat, and the Center for Development Research (ZEF-Bonn).

[3] Uganda National Household Survey 1999/2000. Uganda Bureau of Statistics, Ministry of Finance, Planning and Economic Development.

[4] Districts included are Kabale, Kisoro, Rukungiri, Bushenyi, Ntungamo, Mbarara, Rakai, Masaka, Sembabule, Kasese, Kabarole, Kibale, Mubende, Kiboga, Luwero, Mpigi, Nakasongola, Mukono, Kamuli, Jinja, Iganga, Bugiri, Busia, Tororo, Pallisa, Kumi, Soroti, Katakwi, Lira, Apac, Mbale, and Kapchorwa.

[5] The 1st cropping season is from March to August and the 2nd season is from September to February in the following year.

B. MAS implementation

Table B: Differences between the three mathematical programming models simulating investment, production, and consumption decisions

	Variable	Investment decisions	Production decisions	Consumption decisions
1	Resources			
	Expected future availability of labor	X		
	Current availability of labor		X	X
	Current availability of land	X	X	X
	Current availability of liquid means before investment	X		
	Current availability of liquid means after investment		X	X
	Current availability of assets before investment	X		
	Current availability of assets after investment		X	X
	Capital requirement in the current year	X	X	X
	Average capital requirement in future years	X	X	X
2	Crop and livestock yields			
	Expected crop and livestock yield in current year	X	X	
	Expected yield in future years for investment goods		X	X
	Simulated actual crop and livestock yield in current year			X
	Expected average future returns to new investments	X		
	Expected average future returns to existing investments	X	X	X
3	Prices			
	Expected farm gate (selling) and market (buying) prices	X	X	
	Simulate actual prices farm gate and market prices			X
4	Income and expenditure coefficients			
	Savings function based on future exp. household size	X		
	Savings function based on current household size		X	X
	Food/non-food function based on future exp. hh. size	X		
	Food/non-food function based on current hh. size		X	X
	Consumption function based on future exp. hh. size	X		
	Consumption function based on current hh. size		X	X
	Energy needs based on current household size			X
5	Constraints			
	New investment activities can not be selected		X	X
	Production activities are fixed (to production mode)			X
	Inclusion of food energy constraint			X

C. Production behavior

Table C1: Frontier production function estimates

	(1) Sweet potato	(2) Maize	(3) Millet	(4) Sorghum	(5) Groundnut	(6) Bean	(7) Cassava	(8) Plantain	(9) Coffee	(10) Groundnut + Bean
Labor use (tot. hours, Ln)	0.354	0.535	0.587	0.373	0.566	0.609	0.241	0.456	0.284	0.667
	(0.079)***	(0.062)***	(0.089)***	(0.145)**	(0.140)***	(0.080)***	(0.133)*	(0.056)***	(0.122)**	(0.076)***
Intercropping (1/0)	-0.052	-0.053	0.360	-0.226	-0.942	-0.236	-1.220	-0.170	-0.514	-0.329
	(0.220)	(0.123)	(0.180)**	(0.232)	(0.212)***	(0.194)	(0.469)***	(0.140)	(0.188)***	(0.183)*
1st season (1/0)	-0.151	-0.023	0.663	0.296	0.200	-0.023	-	-	-	0.055
	(0.179)	(0.133)	(0.225)***	(0.147)**	(0.248)	(0.199)				(0.199)
Bimodal low rainfall (1/0)	-0.221	0.001	0.314	0.065	0.290	0.904	-0.729	0.500	-0.077	0.843
	(0.234)	(0.143)	(0.321)	(0.199)	(0.356)	(0.196)***	(0.351)**	(0.163)***	(0.308)	(0.191)***
Bimodal med. rain. (1/0)	0.020	0.283	-0.338	-0.308	0.366	0.126	-0.272	-0.032	0.367	0.362
	(0.202)	(0.210)	(0.270)	(0.366)	(0.199)*	(0.255)	(0.313)	(0.184)	(0.274)	(0.222)
Eastern highlands (1/0)	0.848	0.183	-0.192	-	-	0.776	-	0.045	0.076	0.890
	(0.374)**	(0.343)	(0.453)			(0.317)**		(0.310)	(0.377)	(0.285)***
Southwest (1/0)	-0.379	-1.025	0.006	-0.249	-0.342	0.031	-2.074	0.136	-0.020	-0.088
	(0.341)	(0.218)***	(0.289)	(0.190)	(0.419)	(0.229)	(0.490)***	(0.205)	(0.236)	(0.222)
Unimodel rainfall (1/0)	0.140	0.006	-0.553	0.329	-	1.163	-0.576	0.723	-0.049	1.099
	(0.366)	(0.272)	(0.322)*	(0.245)		(0.361)***	(0.736)	(0.204)***	(0.503)	(0.322)***
Animal traction (1/0)	-0.497	0.008	-	-	-0.444	-0.204	-0.699	-	-	-0.581
	(0.292)*	(0.203)			(0.275)	(0.339)	(0.345)**			(0.261)**
Fertilizer use (1/0)	-	0.347	-	-	-	-1.340	-	-	0.047	-0.964
		(0.347)				(0.440)***			(0.338)	(0.440)**
Hybrid maize (1/0)	-	0.488	-	-	-	-	-	-	-	-
		(0.179)***								
Longe1 (1/0)	-	0.289	-	-	-	-	-	-	-	-
		(0.129)**								
Groundnut (1/0)	-	-	-	-	-	-	-	-	-	-0.590
										(0.207)***
Age of plantation	-	-	-	-	-	-	-	0.108	-	-
								(0.017)***		
Age of plantation sq.	-	-	-	-	-	-	-	-0.002	-	-
								(0.000)***		
Constant	6.434	4.003	3.166	4.926	1.836	2.390	8.014	5.176	4.969	2.366
	(0.604)***	(0.451)***	(0.683)***	(0.833)***	(0.956)*	(0.627)***	(1.060)***	(0.421)***	(0.787)***	(0.582)***
Observations	234	443	136	138	137	306	117	400	231	392
Wald chi-squared	35.65	147.00	138.00	17.90	40.36	111.45	55.24	121.02	18.42	130.48
Prob > chi-squared	0.0000	0.0000	0.0000	0.0124	0.000	0.0000	0.0000	0.0000	0.0183	0.0000

Source: IFPRI 1999-2000 household survey

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Effect of location dummies estimated against the location with bimodal high rainfall, which is omitted from the estimation. - indicates that variable is omitted as all observations have zero value

Table C2: Crop production (example)

	Grow maize, season 1									Sell maize	Buy fertilizer	Buy fertilizer		
	Nutrient response unit 1													
	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3					
	Fer0	Fer0	Fer0	FerA	FerA	FerA	FerB	FerB	FerB					
Objective (MAX)										C	-C	-C		
1. Land, NRU 1	1	1	1	1	1	1	1	1	1				≤	LND
2. Labor	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3				≤	LAB
3. Fertilizer A				F_1	F_1	F_1					-1		≤	0
4. Fertilizer B							F_2	F_2	F_2			-1	≤	0
7. Balance maize	$-Y_1$	$-Y_2$	$-Y_3$	$-Y_4$	$-Y_5$	$-Y_6$	$-Y_7$	$-Y_8$	$-Y_9$	1			≤	0

Notes: A_i=labor requirements; F_i=fertilizer requirements; Y_i=yield expectations; C=objective function coefficients. The table shows the segmentation into levels of input use for the example of maize growing in the first cropping season. The complete matrix had 2320 activities and 556 constraints. The programming matrix was relatively large as it had to be generic for all agents; meaning that all possible choices of all agents are included in a single matrix. Crop production, with 1990 activities, accounted for most of the matrix size. Crop production was divided into 5 nutrient response units (NRUs). A NRU is an amount of land of homogenous quality that is agent specific. Each NRU was subdivided into 2 seasons, 12 crops, and 7 intercrop combinations. For each crop, there were 3 alternative levels of labor use, and two types of mineral fertilizer. Labor levels were set to 100, 50, and 10 percent of the yield maximizing level; fertilizer levels were set to either zero or to a maximum level of fertilizer use (100 kg for all crops).

Table C3: Adoption of innovations (example)

	1. Liquid means	2. Short term deposits	3. Purchase fertilizer from own means	4. Purchase fertilizer on credit	5. Access to credit	6. Access to fertilizer	7. Grow maize with fertilizer		
Objective (MAX)		*d*	-C	-C**i*			C		
1. Land							1	≤	LND
1. Liquid means	1							≤	LIQ
2. Access to credit				1	**-1**			≤	0
3. Access to fertilizer			1	1		**-1**		≤	0
4. Capital use	-1	1	C	C**i*				≤	0
5. Balance fertilizer			-1	-1			A	≤	0
5. Balance maize							-Y	≤	0

Notes: C=objective function coefficients; Y=expected maize yield; A=fertilizer requirement/ha; *i*=the interest rate+1 (*e.g.,* if the interest rate is 34 percent then *i* is 1.34); *d*=the interest rate on short-term deposits; LND=available land (ha); LIQ=the amount of available liquid means (ush), which is carried over from the previous period. Innovations such as improved seeds and mineral fertilizers can either be purchased from own liquid means or on credit if the agent has access to it. If purchasing on credit then the annuity of investment cost is augmented by an interest rate *i* as shown in this example (in constraint 4, activity 4 and in the objective function). Access to innovations is controlled through access activities (activities 5 and 6 in the table); a value of -1 is entered if an agent has access while it was set to zero otherwise. In this example, the agent has both access to fertilizer and credit. Besides using liquid means to purchase fertilizer, it can be deposited at a bank, *d* is the interest rate on short-term deposits, which can be set very low in the absence of financial intermediation.

Table C4: Investment in new coffee plantation (example of investment decision)

	1. Liquid means transfer	2. Invest in new coffee plants	3. Maintain existing coffee plants	4. Sell coffee	5. Sell coffee in future	6. Short-term deposits	7. Short-term credit		
Objective (MAX)		-C	-C	C	C	C	-C		
1. Land		1	1					≤	LND
2. Labor		L_i	L_c					≤	LAB_i
3. Capital	-1	A_i	A_c			1	-1	≤	0
4. Liquid means	1							≤	LIQ_i
5. Current coffee plantation			1					≤	COF_i
6. Max. coffee investment		1						≤	MXC
7. Short-term credit limit							1	≤	0
8. Coffee balance			$-Y_2$	1				=	0
9. Coffee balance (premium)		$-Y_1$			1			=	0

Notes: C=objective function coefficients; L=labor requirement (hrs) for maintenance (*c*) and investment (*i*); A=capital requirement (ush) for maintenance (*c*) and investment (*i*); LND=land endowment (ha); LAB_i=long-term labor availability (hrs) based on given fertility and mortality levels; LIQ_i=available liquid means (ush); COF_i=current area under coffee plantation (ha); MXC=upper bound (in ha) on investments in perennial crops per annum (optional);$-Y_1$=average expected yield over the lifespan of the coffee plantation as an annuity; $-Y_2$=the average expected yield over the remaining years of the coffee plantation expressed as an annuity. It was assumed that the liquid means available to the agent in the first year of the investment (LIQ) equals the agent's expected future availability of liquid means. Credit and deposits are explained in **Table C3**.

Table C5: Current production of coffee (example of production decision)

	1. Liquid means transfer	2. Invest in new coffee plants	3. Maintain existing coffee plants	4. Sell coffee	5. Sell coffee in future	6. Short-term deposits	7. Short-term credit		
Objective (MAX)		-C	-C	C	C	C	-C		
1. Land		1	1					≤	LND
2. Labor		L_i	L_c					≤	**LAB_c**
3. Capital	-1	A_i	A_c			1	-1	≤	0
4. Liquid means	1							≤	**LIQ_c**
5. Current coffee plantation			1					≤	**COF_c**
6. Max. coffee investment		1						≤	0
7. Short-term credit limit							1	≤	0
8. Coffee balance			**$-Y_3$**	1				=	0
9. Coffee balance (premium)			$-Y_2$		1			=	0

Notes: Same as **Table C4** with four differences: (1) the current amount of available labor replaces the long-term labor expectation; (2) the liquid means are reduced by the investment cost of additional plantation; (3) the current coffee plantation (COF_c) is incremented by the additional invested area; a current yield expectation ($-Y_3$) enters the matrix; (4) investments in additional coffee plantation are impossible (MXC=0) and the the yield level (Y_1) is set to zero. Current (Y_3) and average future expected yields (Y_2) are distinguished as the first add to the disposable income (and hence current food expenditures) but the latter do not. See also **Table B** for differences between investment, production, and consumption decisions.

Table C6: Relative time allocation to crop production by sex

A. Seasonal crops, first season

Crop	Period[1]	Female labor (%)		Male labor (%)		t-test	Constraint[2]	
		Mean	SD	Mean	SD	p-value	Minimum % female	Minimum % male
Sweet potato	1	51.33	30.84	48.67	30.84	0.584	-	-
	2	70.59	28.26	29.41	28.26	0.000	42.3	-
	3	74.05	28.75	25.95	28.75	0.000	45.3	-
Traditional maize	1	49.27	28.26	50.73	28.26	0.881	-	-
	2	57.04	23.82	42.96	23.82	0.099	-	-
	3	67.27	24.18	32.73	24.18	0.000	43.1	-
Improved maize	1	43.12	28.40	56.88	28.40	0.028	-	28.5
	2	51.83	27.80	48.17	27.80	0.554	-	-
	3	59.86	28.72	40.14	28.72	0.003	31.1	-
Hybrid maize	1	26.30	26.59	73.70	26.59	0.000	-	47.1
	2	50.19	27.18	49.81	27.18	0.972	-	-
	3	65.55	23.35	34.45	23.35	0.001	42.2	-
Millet	1	37.22	27.85	62.78	27.85	0.000	-	34.9
	2	60.21	27.17	39.79	27.17	0.000	33.0	-
	3	65.99	27.68	34.01	27.68	0.000	38.3	-
Sorghum	1	45.75	27.86	54.25	27.86	0.182	-	-
	2	63.35	33.18	36.65	33.18	0.001	30.2	-
	3	71.28	26.18	28.72	26.18	0.000	45.1	-
Bean	1	49.30	29.57	50.70	29.57	0.706	-	-
	2	58.85	29.80	41.15	29.80	0.000	29.1	-
	3	67.29	27.11	32.71	27.11	0.000	40.2	-
Groundnut	1	50.69	27.73	49.31	27.73	0.789	-	-
	2	61.43	31.60	38.57	31.60	0.000	29.8	-
	3	67.12	29.06	32.88	29.06	0.000	38.1	-

[continued from **Table C6** on the previous page]

B. Seasonal crops, second season

Crop	Period[1]	Female labor (%)		Male labor (%)		t-test	Constraint[2]	
		Mean	SD	Mean	SD	p-value	Minimum % female	Minimum % male
Sweet potato	1	58.00	29.69	42.00	29.69	0.000	28.3	-
	2	68.17	30.51	31.83	30.51	0.000	37.7	-
	3	70.90	30.14	29.10	30.14	0.000	40.8	-
Trad. maize	1	49.03	27.92	50.97	27.92	0.845	-	-
	2	61.33	27.03	38.67	27.03	0.026	34.3	-
	3	62.29	26.82	37.71	26.82	0.022	35.5	-
Improved maize	1	44.25	30.14	55.75	30.14	0.088	-	-
	2	55.91	30.31	44.09	30.31	0.081	-	-
	3	58.42	28.15	41.58	28.15	0.011	30.3	-
Hybrid maize	1	59.92	23.87	40.08	23.87	0.221	-	-
	2	77.06	31.18	22.94	31.18	0.023	45.9	-
	3	63.72	31.54	36.28	31.54	0.160	-	-
Millet	1	63.93	28.69	36.07	28.69	0.000	35.2	-
	2	80.81	25.89	19.19	25.89	0.000	54.9	-
	3	80.69	24.55	19.31	24.55	0.000	56.1	-
Sorghum	1	42.99	34.56	57.01	34.56	0.046	-	22.5
	2	67.82	29.91	32.18	29.91	0.000	37.9	-
	3	69.90	27.64	30.10	27.64	0.000	42.3	-
Bean	1	56.16	28.84	43.84	28.84	0.002	27.3	-
	2	63.48	29.87	36.52	29.87	0.000	33.6	-
	3	68.05	28.62	31.95	28.62	0.000	39.4	-
Groundnut	1	56.57	27.47	43.43	27.47	0.094	-	-
	2	68.27	28.52	31.73	28.52	0.000	39.8	-
	3	69.35	27.86	30.65	27.86	0.000	41.5	-

C. Annual and permanent crops

Crop	Period[1]	Female labor (%)		Male labor (%)		t-test	Constraint[2]	
		Mean	SD	Mean	SD	p-value	Minimum % female	Minimum % male
Cassava	1	48.12	28.87	51.88	28.87	0.182	-	-
	2	57.24	28.95	42.76	28.95	0.000	28.3	-
	3	62.20	30.06	37.80	30.06	0.000	32.1	-
Plantain	1	48.44	34.87	51.56	34.87	0.443	-	-
	2	50.22	31.52	49.78	31.52	0.878	-	-
	3	60.56	36.53	39.44	36.53	0.000	24.0	-
Coffee	1	47.75	36.24	52.25	36.24	0.410	-	-
	2	46.78	32.08	53.22	32.08	0.078	-	-
	3	54.22	33.16	45.78	33.16	0.031	21.1	-

Source: Estimated from IFPRI 2000-2001 survey
Notes: Period codes: 1=land preparation and planting; 2=weeding; 3=harvesting and processing. SD=standard deviation of the average. The constraints are calculated as the mean minus 1 SD and only if the t-test is significant and for the sex contributing most to the labor supply per crop. See text for details.

Table C7: Labor use (example)

Including 1 crop, male, female and child labor, 1 period

	Female labor 18+	Female labor 7-17	Female transfer	Female to mixed	Male labor 18+	Male labor 7-17	Male transfer	Male to mixed	Hiring out male 18+	Hiring in male 18+	Female transfer	Female 7-17 transfer	Male transfer	Male 7-17 transfer	Grow maize	Sell maize		
Objective (MAX)									C	-C						C		
Female 18+											1						≤	LAB_{fa}
Female 7-17												1					≤	LAB_{fc}
Male 18+													1				≤	LAB_{ma}
Male 7-17														1			≤	LAB_{mc}
Female total			-1	1											A_t		=	0
Female 18+	1										-1						≤	0
Female 7-17		1										-1					≤	0
Female transfer	-1	-1	1														=	0
Male total							-1	1							A_m		=	0
Male 18+					1				1	-1			-1				≤	0
Male 7-17						1								-1			≤	0
Male transfer					-1	-1	1										=	0
Mixed labor				-1				-1							A_i		≤	0
Balance crop 1															-Y	1	≤	0

Notes: C=objective function coefficients; Y=yield expectation; A_t, A_m and A_f=total, male, and female labor requirement in maize cultivation; LAB=age and sex specific labor availability. The model used age and sex specific labor supply and requirements. The table shows the implementation of labor use for male and female adults and male and female children. The coefficient A_t is positive if female labor is required, alternatively A_m is positive if male labor is required. If the activity is independent of sex then only a positive value for A_i appears in the matrix.

Table C8: Crop rotation and fallow requirements (example)

	1. Land	2. Grow maize	3. Grow bean	4. Fallow	5. Sell maize	6. Sell bean		
Objective (MAX)					C	C		
1. Labor		A	A				≤	LAB
2. Land	1						=	LND
3. Land transfer	-1	1	1	1			≤	0
4. Balance maize		$-Y_1$			1		≤	0
5. Balance bean			$-Y_2$			1	≤	0
6. Rotation maize	-0.7	1					≤	0
7. Rotation bean	-0.5		1				≤	0
8. Fallow	0.05			-1			≤	0

Notes: C=objective function coefficients; Y=yield expectation; LAB=available labor; LND= available land. The table shows how crop rotation constraints were implemented. In this example, the cultivated area under plot 1 can be used to a maximum of 70% for maize and a maximum of 50% for bean, while at least 5% of the plot should be set to fallow.

Table C9: Intercropping (example)

	1. Maize	2. Bean	3. Maize / bean	4. Bean / maize	5. Sell maize	6. Sell bean		
Objective (MAX)					C	C		
1. Labor	A	A	A	A			≤	LAB
2. Land	1	1	1	1			≤	LND
3. Maize yield	-Y		-Y		1		≤	0
4. Bean yield		-Y		-Y		1	≤	0
5. Maize / bean			1	-1			=	0

Notes: A=labor requirement; Y=expected yield; LAB=available labor; LND= available land. Intercrop combinations were included as separate activities as shown in the table. A combination of maize and bean was was treated as two separate activities with one additional constraint requiring that both activities must be selected in equal amounts. Input requirements were equalized by taking the average over both crops. Maize grown under intensive management can therefore not be intercropped with bean under extensive managment. Though intercropping could be treated as one activity that gives both bean and maize yield, it was separated in two activities as this was more straightforward in terms of computer coding.

Table C10: Certainty equivalents to adjust expectations (example)

	1. Grow maize	2. Adjustment	3. Sell crop	4. Consume crop		
Objective (MAX)	-C		C	C		
1. Labor	A				≤	LAB
2. Land	1				≤	LND
3. Maize yield balance	-Y	1			≤	0
4. Maize yield balance, adjusted		-Q	1	1	≤	0

Notes: C=objective function coefficients; Y=maize yield expectation; LAB=available labor; LND= available land. Certainty equivalents were used to adjust yield expectations in the face of yield risks. The table shows how this works. In column 1, a crop is produced, which enters an amount of yield into row 3. This yield is then transferred to row 4 through column 2 where an adjustment is made by factor Q. Setting Q below 1 would lower the yield expectations, setting it above 1 would increase the yield expectations. This adjustment is only made in production mode to consider the risk in the crop mix; in consumption mode, where actual yields are inserted in the matrix, Q is set back to 1, so that all produced crop can be sold or consumed.

Table D1: Implementation of the three-step budgeting system in a mathematical programming model

	1. Sell [kg]	2. **INC**	3. Savings	4. *Segment 1*	5. *Segment 2*	6. *Binary*	7. **TEX**	8. Constant	9. Household size	10. *Segment 1*	11. *Segment 2*	12. *Segment 3*	13. *Binary 1*	14. *Binary 2*	15. **FEX**	16. Constant	17. Household size	18. Price	19. *Segment 1*	20. *Segment 2*	21. *Segment 3*	22. *Binary 1*	23. *Binary 2*	24. Cons. food cat. 1	25. Cons. food cat. 2		
1. Maximize	c_3																							c_1	c_2		
2. Income function	c_3	-1																						c_1	c_2	=	0
3. Savings transfer		1		-1	-1																					=	0
4. Savings function			-1	0	Φ																					=	0
5. INC=SAV+TEX		1	-1				-1																			=	0
6. lower bound 1				-1		S_1																				≤	0
7. upper bound 1					1	$-S_2$																				≤	0
8. Food function								β_5	χ	λ_1	λ_2	λ_3			-1											=	0
9. Constant							1	-1																		=	0
10. Household size							1		-1																	=	0
11. TEX							1			-1	-1	-1														=	0
12. lower bound 1										-1			E_1													≤	0
13. lower bound 2											-1			E_2												≤	0
14. upper bound 2											1		$-E_2$													≤	0
15. upper bound 3												1		$-E_3$												≤	0
16. Food category 1																ψ_1	η_1	φ_1	$\zeta_{1,1}$	$\zeta_{1,2}$	$\zeta_{1,3}$			$-c_1$		≤	0
17. Food category 2																ψ_2	η_2	φ_2	$\zeta_{2,1}$	$\zeta_{2,2}$	$\zeta_{2,3}$				$-c_2$	≤	0
18. Constant															1	-1										=	0
19. Household size															1		-1									=	0
20. FEX															1				-1	-1	-1					=	0
21. lower bound 1																			-1			F_1				≤	0
22. lower bound 2																				-1			F_2			≤	0
23. upper bound 2																				1		$-F_2$				≤	0
24. upper bound 3																					1		$-F_3$			≤	0
25. upper bound obj. function				S_1	S_2	1				E_1	E_2	E_3	1	1					F_1	F_2	F_3	1	1			≤	0

Explanation of coefficients in Table D1

Step 1: savings model

S_1 The size of the first piecewise linear income segment at which savings are zero and calculated as: $(-\alpha_1 + ((\alpha_1^2) -4 * (\alpha_3 * H + \alpha_5) * \alpha_2)^{\wedge}0.5) / (2 * \alpha_2)$ in which α_5 is a composite constant of all relevant dummy variables in the model.

S_2 The size of the following segment and should be a sufficiently large number

Φ The average savings coefficient for a piecewise linear segment; calculated as:

$\alpha_1 + \alpha_2 * (S_1 + S_2 / 2)$

Step 2: Working-Leser model

E_i The width of the *i*-th piecewise linear segment of TEX

β_5 A composite constant of the relevant dummy variables in the model

χ The effect of household size (in joule) on food consumption, which is agent-specific and calculated as: $H * \beta_2$

λ_i The average food expenditure coefficient for the *i*-th piecewise linear segment, calculated as: $\beta_1 * \ln((E_{i+1}-E_i) / 2 * 100\% / H)$

Step 3: LA/AIDS

F_j The width of the *j*-th piecewise linear segment of FEX

ψ_k A composite constant of the relevant dummy variables in the model for the *k*-th food category

η_k The effect of household size (in joule) on the consumption of the *k*-th food category, which is agent-specific and calculated as: $H * \delta_3$

φ_k The price effect on the consumption of the *k*-the food category, calculated as: $\varphi_k = \sum_{l=1}^{7} \ln(\delta_{1,k,l} * \frac{p_l}{\bar{p}_l})$ in which p_l stands for the price of food category *l*, which is a weighted function of individual commodity prices.

$\zeta_{k,j}$ The average food item expenditure coefficient for the *k*-th food category and the *j*-th piecewise linear segment, calculated as: $\delta_{2,k} *(\ln (F_j+(F_{j+1}-F_j)/2)-\ln(H)-\ln(P^*))$

Table D2: Disinvestment decisions

	1. Deposit savings	2. Consume savings	3. Keep livestock	4. Sell livestock	5. Income transfer	6. Purchase food	7. Energy supply	8. Energy demand	9. Binary		
1. Objective (MAX)	d		c_1	c_2		$-c_3$					
2. Savings	1	1								≤	LIQ
3. Livestock			1	1						≤	LVS
4. Household size								1		=	H
5. Income transfer	d	1	c_1	c_2	-1	$-c_3$				=	0
6. Food consumption					α	$-c_3$				=	0
7. Energy supply						σ	-1			=	0
8. Energy supply=dem.							-1	0.90	-10^3	≤	0
9. Sell livestock				-1					LVS	≤	0
10. Consume savings		-1							LIQ	≤	0

Notes: C=objective function values (prices); *d*=the interest rate on short-term deposits; LIQ=liquid means (savings); LVS=livestock herd size; H=size of the household in billion joule; α=coefficient on income in a demand equation; σ=the energy equivalent of one unit of consumption.

Table D3: Regression estimates for the saving and expenditure models

	Savings	Total expenditures
Income (.00 Ush)	0.445 (0.047)***	0.555 (0.047)***
Income-squared (.00 Ush)	0.000 (0.000)	-0.000 (0.000)
Household size (10^9 Joule)	-325.875 (12.435)***	325.875 (12.435)***
Rural	4,799.418 (511.663)***	-4,799.418 (511.663)***
Eastern Highlands	-278.156 (340.958)	278.156 (340.958)
Karamoja Drylands	3,589.500 (336.063)***	-3,589.500 (336.063)***
Lake Albert Crescent	157.577 (350.795)	-157.577 (350.795)
Lake Victoria Crescent	-2,457.691 (313.641)***	2,457.691 (313.641)***
Mid Northern	822.138 (313.125)***	-822.138 (313.125)***
South East	-216.991 (284.341)	216.991 (284.341)
Southern Drylands	-1,947.670 (370.201)***	1,947.670 (370.201)***
Southern Highlands	-302.261 (380.082)	302.261 (380.082)
West Nile	2,358.659 (295.387)***	-2,358.659 (295.387)***
Western Highlands	-1,709.942 (387.551)***	1,709.942 (387.551)***
Constant	-5,575.383 (684.920)***	5,575.383 (684.920)***
Observations	9016	9016
R-squared	0.34	0.62

Source: Estimated from UNHS 1999/2000
Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%
Estimation method: unrestricted estimation using ordinary least squares (OLS) with survey estimation commands. The dependent variables are total savings and total expenditures.

Table D4: Regression estimates for the food and non-food expenditure models

Modified Working-Leser model

Variable	Food	Non-food
Ln(expenditures / household size)	-2.210e-02	2.210e-02
	(4.340e-03)***	(4.340e-03)***
Household size (bjoule)	-4.638e-04	4.638e-04
	(2.413e-04)*	(2.413e-04)*
Rural (0/1)	6.357e-02	-6.357e-02
	(9.368e-03)***	(9.368e-03)***
Farming household member (0/1)	3.072e-02	-3.072e-02
	(8.022e-03)***	(8.022e-03)***
Eastern	-1.614e-02	1.614e-02
	(8.747e-03)*	(8.747e-03)*
Eastern Highlands	2.632e-02	-2.632e-02
	(1.512e-02)*	(1.512e-02)*
Karamoja Drylands	-1.968e-02	1.968e-02
	(1.052e-02)*	(1.052e-02)*
Lake Albert Crescent	-5.105e-02	5.105e-02
	(7.368e-03)***	(7.368e-03)***
Lake Victoria Crescent	-4.638e-02	4.638e-02
	(9.020e-03)***	(9.020e-03)***
Mid Northern	-9.024e-03	9.024e-03
	(8.887e-03)	(8.887e-03)
South East	-2.783e-03	2.783e-03
	(9.025e-03)	(9.025e-03)
Southern Drylands	1.113e-02	-1.113e-02
	(9.308e-03)	(9.308e-03)
Southern Highlands	-3.580e-03	3.580e-03
	(9.809e-03)	(9.809e-03)
West Nile	-2.963e-02	2.963e-02
	(8.460e-03)***	(8.460e-03)***
Constant	6.376e-01	3.624e-01
	(3.414e-02)***	(3.414e-02)***
Observations	9078	9078
R-squared	0.13	0.13

Source: Estimated from UNHS 1999/2000

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All period dummies are omitted from the table. Agro-ecological dummy for the Western Highlands was omitted. Estimation method: unrestricted estimation using ordinary least squares (OLS) with survey estimation commands. Dependent variables are the proportions of total expenditures spent on food and non-food.

Table D5: Regression estimates for the LA/AIDS model

A. Food categories 1- 4

	Plantain (1)	Cassava and potato (2)	Maize, sorghum, and millet (3)	Bean, pea, and groundnut (4)
Unit value, ln (1)	4.952e-02	-7.949e-03	-4.379e-04	-9.706e-03
	(5.126e-03)***	(3.335e-03)**	(2.882e-03)	(2.151e-03)***
Unit value, ln (2)	-7.949e-03	2.972e-02	3.502e-03	-5.407e-03
	(3.335e-03)**	(4.110e-03)***	(2.833e-03)	(2.075e-03)***
Unit value, ln (3)	-4.379e-04	3.502e-03	2.078e-02	-7.503e-03
	(2.882e-03)	(2.833e-03)	(4.333e-03)***	(1.984e-03)***
Unit value, ln (4)	-9.706e-03	-5.407e-03	-7.503e-03	3.397e-02
	(2.151e-03)***	(2.075e-03)***	(1.984e-03)***	(2.180e-03)***
Unit value, ln (5)	-1.027e-02	-4.214e-03	-6.144e-03	-8.985e-04
	(2.075e-03)***	(1.973e-03)**	(2.086e-03)***	(1.577e-03)
Unit value, ln (6)	-9.608e-03	-5.438e-03	-8.876e-03	-6.741e-03
	(2.965e-03)***	(2.965e-03)*	(2.258e-03)***	(1.914e-03)***
Unit value, ln (7)	-1.705e-03	4.508e-04	-1.302e-03	2.955e-03
	(1.683e-03)	(1.582e-03)	(1.601e-03)	(1.326e-03)**
Unit value, ln (8)	-9.844e-03	-1.066e-02	-1.834e-05	-6.674e-03
	(2.161e-03)***	(2.002e-03)***	(1.583e-03)	(1.275e-03)***
Expenditure/stone1	-1.239e-02	-1.243e-02	-1.039e-02	-2.353e-02
	(6.092e-03)**	(5.142e-03)**	(3.893e-03)***	(3.184e-03)***
Eastern	9.956e-02	-7.918e-02	5.341e-03	1.195e-03
	(1.594e-02)***	(1.478e-02)***	(1.063e-02)	(8.849e-03)
Eastern Highlands	2.882e-02	-2.093e-02	-9.668e-03	-9.732e-03
	(6.307e-02)	(5.780e-02)	(4.179e-02)	(3.499e-02)
Karamoja Drylands	9.402e-02	-2.878e-04	-2.990e-02	2.332e-02
	(1.911e-02)***	(1.612e-02)	(1.162e-02)**	(1.189e-02)**
Lake Albert Cresc.	1.321e-01	-1.944e-02	-6.577e-02	1.141e-02
	(1.605e-02)***	(1.375e-02)	(1.012e-02)***	(8.587e-03)
Lake Victoria Cresc.	-3.407e-02	-5.877e-02	-6.181e-02	1.564e-02
	(2.825e-02)	(2.407e-02)**	(1.776e-02)***	(1.552e-02)
Mid Northern	3.171e-02	5.480e-03	-8.325e-03	-1.366e-02
	(1.651e-02)*	(1.529e-02)	(1.101e-02)	(9.158e-03)
South East	1.982e-01	-4.329e-02	-2.357e-02	2.355e-02
	(1.678e-02)***	(1.415e-02)***	(1.033e-02)**	(1.029e-02)**
Southern Drylands	1.731e-01	1.081e-02	-8.265e-03	3.540e-02
	(1.783e-02)***	(1.511e-02)	(1.093e-02)	(1.121e-02)***
Southern Highlands	-5.460e-02	7.916e-02	-5.339e-02	1.538e-02
	(2.566e-02)**	(2.107e-02)***	(1.584e-02)***	(1.369e-02)
West Nile	1.988e-01	-3.763e-02	-2.003e-02	2.644e-02
	(1.686e-02)***	(1.426e-02)***	(1.035e-02)*	(1.091e-02)**
Household size (mj)	-5.219e-04	4.224e-04	1.140e-03	-7.437e-05
	(2.852e-04)*	(3.734e-04)	(2.405e-04)***	(2.093e-04)
Farming (0/1)	4.092e-02	3.429e-02	-4.940e-03	9.617e-03
	(6.984e-03)***	(1.249e-02)***	(5.051e-03)	(3.880e-03)**
Rural (0/1)	2.436e-02	4.466e-02	-2.388e-03	4.495e-04
	(7.348e-03)***	(6.735e-03)***	(4.933e-03)	(4.067e-03)
Nonselection hazard	6.417e-02	-5.161e-02	6.984e-02	-1.559e-03
	(1.411e-02)***	(5.047e-02)	(1.743e-02)***	(2.512e-02)
Constant	7.028e-02	1.643e-01	1.878e-01	2.312e-01
	(5.319e-02)	(5.236e-02)***	(3.134e-02)***	(2.882e-02)***
Observations	1965	1965	1965	1965
'R-squared'	0.2801	0.2375	0.1452	0.1736

[continued on the next page]

B. Food categories 5 - 8

	Purchased necessities (5)	Animal products (6)	Fruits and vegetables (7)	Food luxuries[2] (8)
Unit value, ln (1)	-1.027e-02	-9.608e-03	-1.705e-03	-0.00984
	(2.075e-03)***	(2.965e-03)***	(1.683e-03)	
Unit value, ln (2)	-4.214e-03	-5.438e-03	4.508e-04	-0.01066
	(1.973e-03)**	(2.965e-03)*	(1.582e-03)	
Unit value, ln (3)	-6.144e-03	-8.876e-03	-1.302e-03	-1.8E-05
	(2.086e-03)***	(2.258e-03)***	(1.601e-03)	
Unit value, ln (4)	-8.985e-04	-6.741e-03	2.955e-03	-0.00667
	(1.577e-03)	(1.914e-03)***	(1.326e-03)**	
Unit value, ln (5)	4.513e-02	-1.180e-02	5.722e-04	-0.01237
	(2.942e-03)***	(1.678e-03)***	(1.487e-03)	
Unit value, ln (6)	-1.180e-02	5.135e-02	-5.517e-03	-0.00336
	(1.678e-03)***	(4.223e-03)***	(1.403e-03)***	
Unit value, ln (7)	5.722e-04	-5.517e-03	6.994e-03	-0.00245
	(1.487e-03)	(1.403e-03)***	(1.512e-03)***	
Unit value, ln (8)	-1.237e-02	-3.363e-03	-2.448e-03	4.54E-02
	(1.135e-03)***	(2.182e-03)	(9.581e-04)**	
Expenditure/stone1)	-3.199e-02	3.993e-02	-7.179e-03	0.057979
	(2.544e-03)***	(5.968e-03)***	(2.304e-03)***	
Eastern	3.372e-02	-8.135e-02	2.087e-02	-0.00016
	(7.349e-03)***	(1.571e-02)***	(6.337e-03)***	
Eastern Highlands	-2.565e-02	-7.308e-02	4.435e-02	0.06589
	(2.874e-02)	(6.165e-02)	(2.481e-02)*	
Karamoja Drylands	1.770e-02	-2.699e-02	-1.688e-02	-0.06098
	(8.006e-03)**	(1.734e-02)	(7.134e-03)**	
Lake Albert Cresc.	1.933e-02	-5.363e-02	-1.060e-02	-0.0134
	(6.534e-03)***	(1.387e-02)***	(5.797e-03)*	
Lake Victoria Cresc.	1.567e-02	3.840e-02	-1.670e-02	0.10164
	(1.201e-02)	(2.560e-02)	(1.034e-02)	
Mid Northern	3.678e-02	-4.098e-02	5.738e-03	-0.01674
	(7.601e-03)***	(1.630e-02)**	(6.558e-03)	
South East	1.641e-03	-5.159e-02	-1.161e-02	-0.09333
	(6.997e-03)	(1.510e-02)***	(6.348e-03)*	
Southern Drylands	-1.392e-02	-8.046e-02	-7.663e-03	-0.109
	(7.507e-03)*	(1.632e-02)***	(6.813e-03)	
Southern Highlands	8.541e-03	-8.015e-02	-6.499e-03	0.091558
	(1.047e-02)	(2.243e-02)***	(9.079e-03)	
West Nile	-2.605e-03	-4.511e-02	-1.316e-02	-0.10671
	(7.098e-03)	(1.550e-02)***	(6.535e-03)**	
Household size (mj)	-5.991e-04	2.696e-04	-4.181e-04	-0.00022
	(1.483e-04)***	(2.969e-04)	(1.473e-04)***	
Farming (0/1)	-2.188e-02	-2.448e-02	-7.699e-03	-0.02583
	(3.408e-03)***	(6.946e-03)***	(2.800e-03)***	
Rural (0/1)	-1.764e-02	-1.458e-02	-9.622e-03	-0.02524
	(3.371e-03)***	(7.214e-03)**	(2.955e-03)***	
Nonselection hazard	4.130e-02	-6.351e-02	2.828e-02	-0.08691
	(2.688e-02)	(1.128e-02)***	(1.424e-02)**	
Constant	3.130e-01	3.998e-02	1.316e-01	-0.13816
	(2.014e-02)***	(4.639e-02)	(1.848e-02)***	-0.00984
Observations	1965	1965	1965	-
'R-squared'	0.3439	0.1676	0.1075	-

Source: UNHS 1999/2000

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. 1) Logarithm of (food expenditures / household size) / Stone price index. 2) Eq.8 constructed from eq.1-7. All monthly dummies are omitted from the table. Agroclimatic dummy for the Western Highlands was omitted. Estimation method: Seemingly Unrelated Regression with 28 parameter restrictions.

Table D6: Demographic data

Age	Agricultural labor	Female					Male			
		Energy	Protein	HIV incidence	Mortality	Fertility	Energy	Protein	HIV incidence	Mortality
years	*hours /day*	*BJ /year*	*kg /year*	*per 1000*	*per 1000*	*per 1000*	*BJ /year*	*kg /year*	*per 1000*	*per 1000*
1	0.00	0.99	5.11	0.0	57.1	0.0	1.07	5.11	0.0	76.4
2	0.00	1.45	8.03	0.0	19.5	0.0	1.51	8.03	0.0	25.7
3	0.00	1.64	8.03	0.0	9.4	0.0	1.73	8.03	0.0	12.3
4	0.00	1.76	9.49	0.0	5.7	0.0	1.87	9.49	0.0	7.4
5	0.00	1.94	9.49	0.0	3.6	0.0	2.02	9.49	0.0	4.8
6	0.00	2.14	10.95	0.0	2.5	0.0	2.25	10.95	0.0	3.3
7	0.11	2.34	10.95	0.0	1.9	0.0	2.38	10.95	0.0	2.5
8	0.22	2.40	12.41	0.0	1.6	0.0	2.74	12.41	0.0	2.1
9	0.33	2.46	12.41	0.0	1.5	0.0	2.62	12.41	0.0	2.0
10	0.44	2.30	12.41	0.1	1.5	0.0	2.67	12.41	0.0	1.9
11	0.55	2.84	17.89	0.7	1.3	0.0	2.90	17.52	0.2	1.7
12	0.66	2.84	17.89	1.9	1.3	0.0	2.97	17.52	0.5	1.7
13	0.77	3.06	21.59	3.5	1.4	11.6	3.19	21.54	1.0	1.8
14	0.88	3.19	21.68	5.4	1.6	28.5	3.23	21.54	1.7	2.1
15	0.99	3.32	23.63	7.4	2.0	54.1	3.42	25.55	2.6	2.5
16	1.10	3.52	23.80	9.3	2.4	88.8	3.58	25.55	3.5	2.9
17	1.21	3.55	23.63	11.1	2.8	128.1	3.86	29.57	4.5	3.3
18	3.41	3.68	23.82	12.6	3.3	166.8	4.00	29.57	5.6	3.7
19	3.45	3.28	18.89	13.8	3.9	202.9	4.40	20.08	6.7	4.2
20	3.49	3.28	19.04	14.7	4.6	234.3	4.40	20.08	7.8	4.6
21	3.53	3.28	19.16	15.3	5.4	258.5	4.40	20.08	8.8	5.1
22	3.57	3.29	19.25	15.6	6.3	278.0	4.40	20.08	9.7	5.6
23	3.61	3.29	19.32	15.7	7.4	291.2	4.40	20.08	10.6	6.2
24	3.64	3.29	19.36	15.5	8.6	299.5	4.40	20.08	11.4	6.8
25	3.67	3.29	19.38	15.1	10.0	303.8	4.40	20.08	12.2	7.5
26	3.70	3.29	19.39	14.6	11.8	305.0	4.40	20.08	12.8	8.4
27	3.73	3.29	19.37	13.9	13.9	301.4	4.40	20.08	13.3	9.6
28	3.76	3.29	19.35	13.2	16.6	296.3	4.40	20.08	13.7	11.0
29	3.78	3.29	19.31	12.4	20.1	289.4	4.40	20.08	14.1	12.8
30	3.80	3.29	19.27	11.5	24.8	281.7	4.40	20.08	14.3	15.2
31	3.83	3.33	19.22	10.7	31.7	271.4	4.35	20.08	14.4	18.6
32	3.85	3.33	19.17	9.9	40.8	260.0	4.35	20.08	14.5	23.0
33	3.86	3.32	19.11	9.0	52.2	249.0	4.35	20.08	14.5	28.7
34	3.88	3.32	19.05	8.2	65.3	237.1	4.35	20.08	14.4	35.8
35	3.89	3.32	18.98	7.5	77.9	223.0	4.35	20.08	14.2	43.7
36	3.91	3.32	18.92	6.7	90.9	210.2	4.35	20.08	14.0	52.9
37	3.92	3.32	18.86	6.1	103.3	197.3	4.35	20.08	13.8	63.1
38	3.93	3.31	18.78	5.4	112.4	182.2	4.35	20.08	13.5	72.9
39	3.94	3.31	18.70	4.9	120.8	166.3	4.35	20.08	13.1	83.8
40	3.94	3.31	18.63	4.3	127.6	150.6	4.35	20.08	12.8	95.0
41	3.95	3.31	18.54	3.8	130.6	133.2	4.35	20.08	12.4	104.8
42	3.95	3.30	18.47	3.4	132.6	117.8	4.35	20.08	12.0	115.1
43	3.95	3.30	18.39	3.0	133.4	102.8	4.35	20.08	11.5	125.4
44	3.95	3.30	18.31	2.6	131.0	87.2	4.35	20.08	11.1	133.6
45	3.95	3.30	18.24	2.3	128.1	72.2	4.35	20.08	10.6	142.0
46	3.94	3.29	18.17	2.0	124.4	58.5	4.35	20.08	10.2	150.1
47	3.94	3.29	18.10	1.8	118.9	44.1	4.35	20.08	9.7	155.7
48	3.93	3.29	18.04	1.6	113.3	32.0	4.35	20.08	9.3	161.4
49	3.92	3.29	18.00	1.4	107.6	22.7	4.35	20.08	8.8	166.7
50	3.91	3.29	17.96	1.2	100.9	15.0	4.35	20.08	8.4	169.6

[continued on the next page]

[continuation of **Table D6**]

Age	Agricultural labor	Female Energy	Female Protein	Female HIV incid-ence	Female Mor-tality	Female Fertility	Male Energy	Male Protein	Male HIV incid-ence	Male Mor-tality
years	*hours /day*	*BJ /year*	*kg /year*	*per 1000*	*per 1000*	*per 1000*	*BJ /year*	*kg /year*	*per 1000*	*per 1000*
51	3.90	3.29	17.93	1.0	94.6	9.8	4.35	20.08	7.9	172.6
52	3.88	3.29	17.92	0.9	88.6	6.2	4.35	20.08	7.5	175.2
53	3.87	3.29	17.90	0.8	82.4	3.7	4.35	20.08	7.1	175.6
54	3.85	3.29	17.89	0.7	76.7	0.0	4.35	20.08	6.7	176.3
55	3.83	3.29	17.89	0.6	71.5	0.0	4.35	20.08	6.3	176.7
56	3.81	3.29	17.89	0.5	66.4	0.0	4.35	20.08	6.0	175.3
57	3.79	3.29	17.89	0.4	62.1	0.0	4.35	20.08	5.6	174.3
58	3.76	3.29	17.89	0.4	58.2	0.0	4.35	20.08	5.3	173.2
59	3.74	3.29	17.89	0.3	54.7	0.0	4.35	20.08	4.9	170.7
60	3.71	3.29	17.89	0.3	51.9	0.0	4.35	20.08	4.6	168.6
61	3.68	2.74	17.89	0.2	49.6	0.0	2.96	20.08	4.3	166.7
62	3.65	2.74	17.89	0.2	47.8	0.0	2.96	20.08	4.0	163.7
63	3.61	2.74	17.89	0.2	46.6	0.0	2.96	20.08	3.8	161.4
64	3.58	2.74	17.89	0.1	46.0	0.0	2.96	20.08	3.5	159.3
65	3.54	2.74	17.89	0.1	45.9	0.0	2.96	20.08	3.3	156.8
66	3.50	2.74	17.89	0.1	46.3	0.0	2.96	20.08	3.1	154.8
67	3.46	2.74	17.89	0.1	47.4	0.0	2.96	20.08	2.9	153.4
68	3.42	2.74	17.89	0.1	49.1	0.0	2.96	20.08	2.7	151.8
69	3.38	2.74	17.89	0.1	51.4	0.0	2.96	20.08	2.5	150.9
70	3.33	2.74	17.89	0.1	54.4	0.0	2.96	20.08	2.3	150.8
71	3.29	2.74	17.89	0.0	58.1	0.0	2.96	20.08	2.1	150.9
72	3.24	2.74	17.89	0.0	62.6	0.0	2.96	20.08	2.0	151.9
73	3.19	2.74	17.89	0.0	67.9	0.0	2.96	20.08	1.8	153.9
74	3.14	2.74	17.89	0.0	74.1	0.0	2.96	20.08	1.7	156.3
75	3.08	2.74	17.89	0.0	81.3	0.0	2.96	20.08	1.6	159.9
76	3.03	2.74	17.89	0.0	89.6	0.0	2.96	20.08	1.4	164.7
77	2.97	2.74	17.89	0.0	98.9	0.0	2.96	20.08	1.3	170.4
78	2.91	2.74	17.89	0.0	109.5	0.0	2.96	20.08	1.2	177.3
79	2.85	2.74	17.89	0.0	121.4	0.0	2.96	20.08	1.1	185.8
80	2.79	2.74	17.89	0.0	134.6	0.0	2.96	20.08	1.0	195.4
81	2.73	2.74	17.89	0.0	149.4	0.0	2.96	20.08	1.0	206.6
82	2.66	2.74	17.89	0.0	165.7	0.0	2.96	20.08	0.9	219.5
83	2.59	2.74	17.89	0.0	183.7	0.0	2.96	20.08	0.8	233.8
84	2.53	2.74	17.89	0.0	203.4	0.0	2.96	20.08	0.8	250.0
85	2.45	2.74	17.89	0.0	224.9	0.0	2.96	20.08	0.7	268.2
86	0.00	2.74	17.89	0.0	248.3	0.0	2.96	20.08	0.6	288.4
87	0.00	2.74	17.89	0.0	273.8	0.0	2.96	20.08	0.6	310.7
88	0.00	2.74	17.89	0.0	301.5	0.0	2.96	20.08	0.5	335.2
89	0.00	2.74	17.89	0.0	331.5	0.0	2.96	20.08	0.5	362.4
90	0.00	2.74	17.89	0.0	364.2	0.0	2.96	20.08	0.5	392.4
91	0.00	2.74	17.89	0.0	399.8	0.0	2.96	20.08	0.4	425.5
92	0.00	2.74	17.89	0.0	438.8	0.0	2.96	20.08	0.4	462.2
93	0.00	2.74	17.89	0.0	481.5	0.0	2.96	20.08	0.3	502.9
94	0.00	2.74	17.89	0.0	528.7	0.0	2.96	20.08	0.3	548.3
95	0.00	2.74	17.89	0.0	581.1	0.0	2.96	20.08	0.3	599.0
96	0.00	2.74	17.89	0.0	639.6	0.0	2.96	20.08	0.3	656.0
97	0.00	2.74	17.89	0.0	705.4	0.0	2.96	20.08	0.2	720.4
98	0.00	2.74	17.89	0.0	780.0	0.0	2.96	20.08	0.2	793.8
99	0.00	2.74	17.89	0.0	865.4	0.0	2.96	20.08	0.2	878.0
100	0.00	2.74	17.89	0.0	1000.0	0.0	2.96	20.08	0.2	1000.0

Sources: Estimated from IFPRI 2000-2001 survey (labor supply); Feeney and Zaba 2001 (mortality, fertility and incidence rates); James and Schofield 1990 (energy needs); Latham 1997 (protein needs)
Notes: BJ=billion joule. Agricultural labor is equal for males and females in the baseline. HIV incidence is the probability of becoming infected with the HIV virus multiplied by 1000.

Table of authorities